

New Dimensions in Fuzzy Logic and Related Technologies

Proceedings
of the 5th EUSFLAT Conference,
Ostrava, Czech Republic,
September 11–14, 2007

Volume I





Edited by
Martin Štěpnička
Vilém Novák
Ulrich Bodenhofer





Title: **New Dimensions in Fuzzy Logic and Related Technologies. Volume I**
Proceedings of the 5th EUSFLAT Conference, Ostrava, Czech Republic, September 11-14, 2007

Editors:	Mgr. Martin Štěpnička, Prof. Ing. Vilém Novák, DrSc. Univ. Doz. Dipl.-Ing. Dr. Ulrich Bodenhofer
Publisher:	University of Ostrava
Number of Copies:	170
Pages:	480
Edition:	First, 2007
Printing:	Tribun

Typesetting: by authors using LaTeX or MS Word
completion by Viktor Pavliska

Cover design: Robert Rytina, VR Ateliér

© University of Ostrava
© Mgr. Martin Štěpnička, Prof. Ing. Vilém Novák, DrSc., Univ. Doz. Dipl.-Ing. Dr. Ulrich Bodenhofer

ISBN 978-80-7368-386-3

Committees

Honorary President:
L. A. Zadeh (University of California, USA)

Chairpersons:
Vilém Novák (University of Ostrava)
Ulrich Bodenhofer (Johannes Kepler University)

Chairpersons of the affiliated conferences:
P. Hájek (Czech Republic) - 6th Workshop of the ERCIM Working Group on Soft Computing
R. Hampel (Germany) - 14th Zittau Fuzzy Colloquium

Local Organizing Committee:
M. Štěpnička (General Co-Chair)
M. Daňková
A. Dvořák
L. Hájková
L. Nosková
V. Pavliska
I. Perfilieva
D. Plšková
M. Pokorný
J. Ramík

International Programme Committee:

J. Aczel (Canada)

R. Babuška (Netherlands)
J. Baldwin (UK)
I. Batyrshin (Russia)
J. Bezdek (USA)
R. Bělohlávek (Czech Republic)
P. Bonissone (USA)
B. Bouchon-Meunier (France)
D. Butnariu (Israel)

G. Chen (China)

B. De Baets (Belgium)
G. De Cooman (Belgium)

M. Delgado (Spain)
A. Di Nola (Italy)
D. Dubois (France)

F. Esteva (Spain)

M. Fedrizzi (Italy)
J. Fodor (Hungary)

L. Godo (Spain)
F. Gomide (Brazil)
S. Gottwald (Germany)
M. Grabisch (France)

P. Hájek (Czech Republic)

F. Herrera (Spain)
K. Hirota (Japan)
E. Hüllermeier (Germany)

J. Kacprzyk (Poland)
N. Kasabov (New Zealand)
E. Kerre (Belgium)
E. P. Klement (Austria)
G. J. Klir (USA)
L. Kóczy (Hungary)
V. Kreinovich (USA)
R. Kruse (Germany)

D. Lakov (Bulgaria)
H. Larsen (Denmark)

L. Magdalena (Spain)
M. Mareš (Czech Republic)
T. Martin (UK)
J. Mattila (Finland)
R. Mesiar (Slovakia)
K. C. Min (South Korea)
F. Montagna (Italy)
M. Mukaidono (Japan)
D. Mundici (Italy)

M. Navara (Czech Republic)
H. T. Nguyen (USA)

M. Ojeda-Aciego (Spain)
H. Ono (Japan)

E. Pap (Serbia and Montenegro)
G. Pasi (Italy)
W. Pedrycz (Canada)
H. Prade (France)

A. Ralescu (USA)
D. Ralescu (USA)
R. Ribeiro (Portugal)
D. Ruan (Belgium)
E. Ruspini (USA)

E. Sanchez (France)
R. Scozzafava (Italy)
P. Sobrevilla (Spain)
A. Sokolov (Ukraine)
A. Šostak (Latvia)

T. Takagi (Japan)
E. Trillas (Spain)
E. Turunen (Finland)

J. Verdegay (Spain)
M.-A. Vila (Spain)

M. Wagenknecht (Germany)
G. J. Wang (China)
J. Watada (Japan)

R. Yager (USA)

A Message from the EUSFLAT President



Dear friends, dear colleagues,
it is already the fifth time that our society celebrates its biannual conference which has become a well-known and well-established event in the international soft computing community. We have the pleasure to meet in the Czech Republic for the first time, in the gorgeous city of Ostrava in the center of the Moravian-Silesian region. For fuzzy logic scientists, Ostrava has been a well-known name for long, as it is the place where one of the best-known and most influential groups in our community is located – the Institute for Research and Applications of Fuzzy Modeling (IRAFM) headed by Prof. Vilém Novák. I consider it more than appropriate to have our conferences at places where highly important advances of our research area have been made. Hereby, I want to thank Vilém and his numerous co-workers for a very smooth and highly professional organization of our conference, which will surely be another highlight in the history of EUSFLAT. In particular, I want to thank Martin Štěpnička who has been my main contact person during the preparations for the conference and who spent a tremendous effort to make EUSFLAT 2007 an unforgettable event.
An impressive number of papers have been submitted to the conference, out of which 129 have been accepted and included in the proceedings volume you are holding in your hands right now. I want to thank all participants for sharing their research results with us and for considering EUSFLAT 2007 as the right forum to present them to the community. I thank those who are already members of EUSFLAT for their continued faithfulness and support and those who have become new members by registering to the conference for their interest in EUSFLAT and its conference. I hope that – for both groups – EUSFLAT will continue to provide a useful exchange and support platform for researchers in the area of fuzzy logic and technologies.
I wish all of you a most successful event, interesting discussions, new ideas and insights, and pleasant days in Ostrava and I hope to welcome you at future events organized by our society!

Ulrich Bodenhofer
President of EUSFLAT

Foreword

It is a great pleasure for me to welcome you in Ostrava. The 5^{th} Conference of the European Society for Fuzzy Logic and Technology is an important opportunity to meet people especially from Europe but also from non-European countries. Recall that EUSFLAT Conferences are organized every two years with the goal to bring together scientists working on methods for fuzzy modeling, foundations of soft computing, computational intelligence and related areas. This conference provides a medium for exchange of ideas between theoreticians and practitioners in the above areas.

I am very happy that the local organization is under the auspices of the rector of the University of Ostrava, Prof. Jiří Močkoř and the conference is supported by the Town Council of Ostrava. It is also specific that there are two conferences affiliated with EUSFLAT 2007, namely 6th Workshop of the ERCIM Working Group on Soft Computing and also 14^{th} Zittau Fuzzy Colloquium.

Organizing EUSFLAT 2007 Conference in Ostrava is not occasional. We have a long tradition in the research of fuzzy logic, fuzzy modeling as well as of some other methods of soft computing, such as neural networks and evolutionary algorithms. The results of people from this region, part of them now working in the Institute for Research and Applications of Fuzzy Modeling of the University of Ostrava are internationally renowned. Let us also remind that we have also significantly participated in organizing of the International Fuzzy Systems Association World Congress, which took place in 1997 in Prague.

Ostrava is a large industrial town that has about 300 thousand inhabitants. The original coal mining, iron and steel works industry by which Ostrava was famous in the past, has now been significantly reduced (the coal mining has been completely abolished). After partial depression, Ostrava is now successfully gaining its prosperity.

Ostrava and its surroundings is a place where famous people have been born or lived. Among them, let us mention especially one of the best Czech music composers, Leoš Janáček, who was born in Hukvaldy (a small village about 25 km from Ostrava) and died in Ostrava, and Sigmund Freud who was born in Příbor (a small town about 30 km from Ostrava). In the distance of about 35 km, Ostrava is surrounded by Beskydy mountains (average height 1000 m; the highest mountain Lysá hora has 1300 m). It is also notable that a beautiful castle Hradec nad Moravicí (about 35 km from Ostrava) has been visited in 1802 by Ludwig van Beethoven who has finished there his piano sonata "Appasionata". Other notable place of interest is Landek — a small hill with and open-air mining museum with some archeological excavations and reconstruction of mammoth-huntsmen village. When climbing the Town Council Tower, a typical landmark of Ostrava, you will realize that this town is full of trees and parks. We have 3 theaters where operas, light operas and ballets are played and world renowned Janáček symphony orchestra giving regularly its concerts. There are also many small restaurants and bars, most of them located in famous street Stodolní.

As a chair of EUSFLAT 2007, I would like to express my thanks to all who participated at its success, namely to all members of the local committee, to all the organizers of invited sessions, and to all other colleagues who helped us with its organization. Our special thanks belong to the Town council of Ostrava for its support.

Vilém Novák

Contents

Invited Lectures **15**

Fuzzy Logic and Theories of Vagueness

Fermüller, Ch. G. 17

Algebraic Semantics for t-norm Based Fuzzy Logic

Gispert, J. 19

Mathematical Fuzzy Logic – a Survey and Some News

Hájek, P. 21

Recent Advances in the Field of Left-continuous t-norms

Jenei, S. 23

Semirings, Dioids and Their Links to Fuzzy Sets and Other Applications

Minoux, M. 25

Lessons Learned from Fuzzy Logic Applications in Complex Systems

Ruan, D. 27

Fuzzy Logic – a New Direction. The Concept of f-Validity

Zadeh, L. A. 29

Special Sessions **31**

1 Aggregation Operators – R. Mesiar and A. Kolesárová **33**

Dominance of Ordinal Sums of T_L and T_P

Saminger, S., Sarkoci, P. 35

Domination and Information Boundedness Principle for Aggregation

Šabo, M. 41

Triple Rotation: Gymnastics for t-norms

Maes, K. C., De Baets, B. 47

Representations of Archimedean t-norms in Interval-valued Fuzzy Set Theory

Deschrijver, G. 53

Construction of Aggregation Operators that Preserve Ordering of the Data

Beliakov, G., Calvo, T. 61

Fitting ST-OWA Operators to Empirical Data

Beliakov, G., Troiano, L. 67

Ordinal Means

Kolesárová, A., Mayor, G., Mesiar, R. 75

A New Look on Discrete Quasi-copulas

Aguiló, I., Suner, J., Torrens, J. 83

Introduction to Asymptotically Idempotent Aggregation Operators

Ghiselli Ricci, R. 91

MP and MT-implications on a Finite Scale

Mas, M., Monserrat, M., Torrens, J. 97

(U, N)-Implications and Their Characterizations

Jayaram, B., Baczyński, M. 105

Intersections Between Basic Families of Fuzzy Implications: (S, N)-, R- and QL-Implications

Jayaram, B., Baczyński, M. 111

Logical Aggregation Based on Interpolative Realization of Boolean Algebra

Radojević, D. G. 119

E_T-Lipschitzian Aggregation Operators

Jacas, J., Recasens, J. 127

2 Axiomatic Fuzzy Mathematics – P. Cintula and L. Běhounek 135

Fuzzy Class Theory: Some Advanced Topics

Cintula, P., Horčík, R. 137

Interior-Based Topology in Fuzzy Class Theory

Běhounek, L., Kroupa, T. 145

Valverde-Style Representation Results in a Graded Framework

Běhounek, L., Bodenhofer, U., Cintula, P. 153

Enriched Generic Algebras of Fuzzy Relations

Kohout, L. J. 161

Uncertainty as a Modality over t-norm Based Logics

Marchioni, E. 169

Modal Systems Based on Many-valued Logics

Bou, F., Esteva, F., Godo, L. 177

3 Copulas – E. P. Klement and F. Durante 183

Different Types of Convexity and Concavity for Copulas

Alvoni, E., Durante, F., Papini, P. L., Sempi, C. 185

A Method for Constructing Multivariate Copulas

Durante, F., Quesada-Molina, J. J., Úbeda-Flores, M. 191

Semigroups of Semicopulas and Evolution of Dependence at Increase of Age

Foschi, R., Spizzichino, F. 197

An Aggregate Claims Model Between Independence and Comonotone Dependence

Hürlimann, W. 205

Bounds for Value at Risk for Asymptotically Dependent Assets – the Copula Approach

Jaworski, P. 213

4 Fuzzy Measures and Integrals: Theory – E. Pap and R. Mesiar 219

Pseudo-integral of Set-valued Functions

Grbić, T., Štajner-Papuga, I., Nedović, L. 221

M-probability Theory on IF-events

Riečan, B. 227

Almost Everywhere Convergence in Family of IF-events With Product

Lendelová, K. 231

Outer Measure on MV-algebras

Michalíková, A. 237

A Fixed Point Theorem in Probabilistic Metric Spaces with a Convex Structure

Žikić-Došenović, T. 241

Measuring the Similarity of Different Types of Fuzzy Sets in FRDB

Takači, A., Škrbić, S. 247

A Universal Integral

Klement, E. P., Mesiar, R., Pap, E. 253

On the k-additive Core of Capacities

Grabisch, M., Miranda, P. 257

Decomposable Signed Fuzzy Measures

Mihailović, B., Pap, E. 265

5 Fuzzy Measures and Integrals: Applications in Decision Making and Game Theory – M. Grabisch 271

Extending the Choquet integral

Rossi, G. 273

Shortfall-dependant Risk Measures (and Previsions)

Baroni, P., Pelessoni, R., Vicig, P. 281

Representation of Two Bipolar Decision Strategies with Generalized Choquet Integrals

Labreuche, Ch., Grabisch, M. 289

Various Representations and Algebraic Structure of Linear Imprecision Indices

Lepskiy, A. E., Bronevich, A. G. 297

On the Moments and the Distribution of the Choquet Integral

Kojadinovic, I., Marichal, J. L. 305

The Core of Games on Distributive Lattices

Xie, L., Grabisch, M. 313

6 Fuzzy Methods in Learning and Data Mining – E. Hüllermeier and F. Klawonn 319

Towards a Robust Rank Correlation Measure for Numerical Observations on the Basis of Fuzzy Orderings

Bodenhofer, U., Klawonn, F. 321

SPoID: Do Not Throw Meaningful Incomplete Sequences Away!

Fiot, C., Laurent, A., Teisseire, M. 329

Measuring Variation Strength in Gradual Dependencies

Molina, C., Serrano, J. M., Sánchez, D., Vila, M. A. 337

Forest of Fuzzy Decision Trees and Their Application in Video Mining

Detyniecki, M., Marsala, C. 345

Fuzzy-Relational Classification: Combining Pairwise Decomposition Techniques with Fuzzy Preference Modeling

Hüllermeier, E., Brinker, K. 353

Selecting the Optimal Rule Set Using a Bacterial Evolutionary Algorithm

Drobics, M., Botzheim, J., Adlassnig, K. P. 361

7 Fuzzy Sets in Distributive Artificial Intelligence – A. Averkin 367

Fuzzy Multiagent Distributed Assembly Chart Planning in Agriculture

Tikhonov, E. 369

Synthesis of Distributed Fuzzy Hierarchical Model in Decision Support Systems in Fuzzy Environment

Averkin, A. N., Agrafonova, T., Titova, N. 377

Fuzzy Calculating and Fuzzy Control in Wireless Sensor Network

Kalganova, I. 381

Soft Computing in Wireless Sensors Networks

Averkin, A. N., Belenki, A. G. 387

Evaluation of Alternatives for the Disposition of Surplus Weapons-Usable Plutonium on the Basis of Fuzzy Sets

Averkin, A. N., Kosterev, V. V., Stankevich, K. V. 391

Towards to Automatic Tactics' Analysis in Soccer

Averkin, A. N., Gusev, A., Shestakov, M. 395

8 Fuzzy Sets – Philosphy and Criticism – R. Seising 399

Scientific Theories and the Computational Theory of Perceptions – A Structuralist View Including Fuzzy Sets
Seising, R. 401

Fuzzy Set Theory and Philosophical Foundations of Medicine
Limberg, J., Seising, R. 409

Fuzzy Logic as a Theory of Vagueness: 15 Conceptual Questions
Bradley, J. 417

Postmodernism and Control Engineering
Balas, V. E., Balas, M. M. 425

Power Sets and Implication Operators Revisited: A Retrospective Look at the Foundational and Conceptual Issues in Bandler and Kohout's Paper After 29 Years
Kohout, L. J. 429

Exploring Dialogue Games as Foundation of Fuzzy Logic
Fermüller, Ch. G. 437

Representing Groups with Imprecise Opinions
Nurmi, H., Kacprzyk, J. 445

Fuzziness – Representation of Dynamic Changes?
Kosiński, W., Prokopowicz, P. 449

9 The Two Decades of Fuzzy Research at the IPM (Zittau) – Past and Recent Developments – M. Wagenknecht and R. Hampel 457

Soft Computing at the Zittau IPM – an Overview
Kästner, W., Hampel, R. 459

Experiences in Soft Computing and Dynamical Simulation
Kästner, W., Hampel, R. 463

Modelling the Differential Pressure at Sieves with Artificial Neural Networks (Multilayer Perceptron) – a Contribution to Reactor Safety Research
Kratzsch, A., Kästner, W., Hampel, R. 469

On the Reliability of Multistate Systems with Imprecise Probabilities
Wagenknecht, M., Gocht, U. 473

Invited Lectures

Fuzzy Logic and Theories of Vagueness

Christian G. Fermüller
Theory and Logic Group,
University of Technology
Vienna

Abstract

Fuzzy Logic has been successfully applied to all kinds of scenarios, where degrees of membership and truth can be systematically discerned. However it is highly contentious whether this fact renders Fuzzy Logic an ideal tool for modelling reasoning with vague concepts and propositions. There is a lively and prolific debate in analytic philosophy about so-called theories of vagueness. Interestingly, degree based approaches don't fare well among most contemporary philosophers. In this talk we will survey this discourse on vagueness from a logician's perspective and try to explain why various alternatives to many valued logics are deemed necessary to model human reasoning under vagueness. We maintain that deductive Fuzzy Logic in itself is hardly a full-blown theory of vagueness, but may play an important role in formalizing and quantifying different aspects of handling logically complex vague information.

Algebraic semantics for t-norm based fuzzy logic

Joan Gispert
Facultat de Matemàtiques
Universitat de Barcelona
Barcelona

Abstract

In this talk we present from an algebraic point of view the general framework of core and Δ-core fuzzy logics. We consider three types of completeness with respect to any semantics of linearly orderd algebras and we give useful algebraic characterizations of these completeness. Moreover we distinguish some special semantics for these logics and we survey the known completeness methods and results for prominent logics.

Mathematical fuzzy logic – a survey and some news

Petr Hájek
Institute of Computer Science
Academy of Sciences of the Czech Republic
Prague
Czech Republic

Abstract

Mathematical fuzzy logic (or fuzzy logic in a narrow sense) is a many-valued symbolic logic with a comparative notion of truth. The present state of development of t-norm based fuzzy propositional and predicate logic will be surveyed and some recent results will be described (on witnessed models and others).

Recent Advances in the Field of Left-continuous T-norms

Sándor Jenei
Institute of Mathematics and Informatics
University of Pécs
Ifjúság u. 6, H-7624 Pécs, Hungary
jenei@ttk.pte.hu

Abstract

The recent advances in the research of left-continuous t-norms is summarized in this talk. The main focus is on construction methods, geometric description, and structural characterization. We point out further research directions and open problems.

1 Summary

The structure of continuous t-norms has been known since ages [19]. In many mathematical theories $left$-continuous t-norms have been used for a few decades without having a single (non-continuous) example at hand. After this long period – when no left-continuous t-norm[1] was known – the first example, the nilpotent minimum, was found by Fodor [2]. Even after finding this example many researchers believed that this is the only one, and a conjecture was published saying, roughly speaking, that the nilpotent minimum is the only left-continuous t-norm.

Next, a series of papers have appeared [11, 14, 15, 8, 7, 3, 1, 16, 5, 20, 6, 22], in which new construction methods provided the interested community with a huge number of new left-continuous t-norms. A state of the art is in [10].

On one hand, the complexity of different construction methods has reached a high level, therefore this line of research has become difficult to follow. On the other hand, the huge number of examples of left-continuous t-norms has called for a kind of systematization. One reply for that need is a comprehensive geometric characterization of residuated structures [9, 12], in particular, of left-continuous t-norms. This geometric description provides a kind of geometric view of algebraic phenomena, thus it makes the field of left-continuous t-norms, and its construction methods much more understandable. A recent result along this direction is [17], which resulted in another construction method in [18].

Concerning structural characterization, little is known as of today. As it will be explained, it seems to be hopeless to give classification for the whole class of left-continuous t-norms. However this does not exclude the possibility of characterizing certain subclasses of them. Two recent results concerns the cancellative, and the regular cases, respectively [4, 21].

These are the focuses of the talk.

Acknowledgement
Supported by the Hungarian-American Enterprise Scholarship Fund (HAESF).

References

[1] R. Cignoli, F. Esteva, L. Godo, F. Montagna, On a class of left-continuous t-norms, Fuzzy Sets and Systems 131 (2002), 283-296.

[1] From now under left-continuous t-norms we mean left-continuous but not continuous t-norms.

[2] J.C. Fodor, Contrapositive symmetry of fuzzy implications, Fuzzy Sets and Systems 69 (1995) 141-156.

[3] P. Hájek, Observations on the monoidal t-norm logic, Fuzzy Sets and Systems 132 (2002), 107-112.

[4] R. Horcik, Algebraic Properties of Fuzzy Logics, PhD Thesis, Czech Technical University, Prague 2004.

[5] S. Jenei, A characterization theorem on the rotation construction for triangular norms, Fuzzy Sets and Systems, Fuzzy Sets and Systems, 136 (2003), 283-289.

[6] S. Jenei, F. Montagna, A general method for constructing left-continuous t-norms, Fuzzy Sets and Systems, 136 (2003), 263-282.

[7] S. Jenei, A note on the ordinal sum theorem and its consequence for the construction of triangular norms, Fuzzy Sets and Systems, 126 (2002), 199-205.

[8] S. Jenei, F. Montagna, A proof of standard completeness of Esteva and Godo's logic MTL, Studia Logica, 70 (2002), 184-192.

[9] S. Jenei, Geometry of Associativity Theory and Application, Proceedings of the 25th Linz Seminar on Fuzzy Set Theory (Fuzzy Logics and Related Structures), (Linz, Austria), 2005, 63-66.

[10] S. Jenei, How to construct left-continuous triangular norms state of the art Fuzzy Sets and Systems, 143 (2004), 27-45.

[11] S. Jenei, New family of triangular norms via contrapositive symmetrization of residuated implications, Fuzzy Sets and Systems 110 (1999), 157-174.

[12] S. Jenei, On the geometry of associativity, Semigroup Forum (in press)

[13] S. Jenei, On the structure of rotation-invariant semigroups, Archive for Mathematical Logic, 42 (2003), 489-514.

[14] S. Jenei, Structure of left-continuous triangular norms with strong associated negations. (I) Rotation construction, Journal of Applied Non-Classical Logics 10 (2000), 83-92.

[15] S. Jenei, Structure of left-continuous t-norms with strong associated negations. (II) Rotation-annihilation construction, Journal of Applied Non-Classical Logics, 11 (2001) 351-366.

[16] S. Jenei, Structure of Girard monoids on $[0;1]$, in: E.P. Klement, S. E. Rodabaugh, (Eds.), *Topological and Algebraic Structures in Fuzzy Sets*, Kluwer Academic Publishers, Dordrecht, (A Handbook of Recent Developments in the Mathematics of Fuzzy Sets, Trends in Logic, vol 20.) 2003, pp. 277-308.

[17] K.C. Maes and B. De Baets, On the structure of left-continuous t-norms that have a continuous contour line, Fuzzy Sets and Systems 158 (2007), 843-860.

[18] K.C. Maes and B. De Baets, The triple rotation method for constructing rotation-invariant t-norms, Fuzzy Sets and Systems (in press)

[19] P.S. Mostert, A.L. Shields, *On the structure of semigroups on a compact manifold with boundary.* Ann. Math., 65 (1957), 117-143.

[20] D. Smutná, *On a peculiar t-norm*, Busefal 75 (1998), 60-67.

[21] T. Vetterlein, Regular left-continuous t-norms (submitted)

[22] Y. Ouyanga, J. Fanga, J. Lic, A conditionally cancellative left-continuous t-norm is not necessarily continuous, Fuzzy Sets and Systems 157 (2006), 2328-2332.

Semirings, Dioids and their links to Fuzzy Sets and other Applications

Michel Minoux
Professor, University Paris-6
France

Abstract

Algebraic structures such as Bottleneck Algebras (R, Max, Min), Fuzzy Algebras $((0,1), Max, Min)$ or more generally $((0,1), Max, T)$ where T is a t-norm have been extensively used as relevant tools for modeling and solving problems related to Fuzzy Sets, Fuzzy relations and systems. Many of these algebraic structures may be viewed as special instances of canonically ordered Semirings, (i.e. semirings in which the preorder relation induced by addition is an order), these being frequently referred to as Dioids.

Though Semirings or Dioids do not enjoy all the classical properties of rings or fields in ordinary algebra, many classical results can be shown to be still valid in those structures. The talk will provide an overview of some of the most important properties of Semirings and Dioids in particular those related to solving linear systems, computing eigenvalues and eigenvectors, testing linear dependence or independence. Special emphasis will be put on the subclasses of Dioids more closely related to Fuzzy Set theory and applications.

Lessons learned from fuzzy logic applications in complex systems

Da Ruan
Belgian Nuclear Research Centre (SCK•CEN)
Boeretang 200, B-2400 Mol, Belgium
phone: +32-14-332272
fax: +32-14-321529
e-mail: druan@sckcen.be
www.sckcen.be/people/daruan

Abstract

In recent years there has been a growing interest in the need for designing intelligent systems to address complex engineering problems. One of the most challenging issues for the intelligent system is to effectively handle real-world uncertainties that cannot be eliminated. These uncertainties include sensor imprecision, instrumentation and process noise and disturbances, unpredictable environmental factors, to name a few. These uncertainties result in a lack of the full and precise knowledge of the system including its state, dynamics, and interaction with the environment.

Fuzzy logic (FL) and soft computing (SC) techniques, as complimentary to the existing traditional techniques, have shown great potential to solve these demanding, real-world problems that exist in uncertain and unpredictable environments. These technologies have formed the foundation for intelligent systems. An overview on FL and SC in control and decision making for complex systems will be given over the last four decades.

Some real-world cases on power plant operation, informationdriven safeguards, cost estimation under uncertainty for a large engineering project, and decision support for long-term options of energy policy will be illustrated for the potential use of FL related techniques in complex systems. Essential steps on implementing FL related techniques in industry will be presented via R & D, demonstration, and commercialization. Challenges and future research directions will be concluded in this talk.

Da Ruan's bio

Da Ruan (PhD in Math, Ghent U, Belgium 1990) is a scientific staff member at the Belgian Nuclear Research Centre (SCK•CEN). He was a Post-Doctoral Researcher from 1991-93 and since 1994 has been a senior researcher and FLINS Project Leader at SCK•CEN.

He is the principal investigator for the research project on intelligent control for nuclear reactors, cost-estimation for large nuclear projects under uncertainty, and computerized decision making systems for society and policy support at SCK•CEN. He was a guest research scientist at the OECD Halden Reactor Project (HRP), Norway from April 2001 to September 2002 as a principal investigator for the research project on computational intelligent systems for feedwater flow measurements at HRP.

His major research interests lie in the areas of mathematical modelling, computational intelligence methods, uncertainty analysis and information/sensor fusion, decision support systems to information management, cost/benefit analysis, and safety and security related fields.

Dr Ruan currently serves as Scientific advisor at the National Institute for Nuclear Research of Mexico for the project *"Adaptive fuzzy control and its applications in nuclear systems"* (Mexico), Regional editor for Europe of *Int. J. of Intelligent Automation and Soft Computing* (TSI Press, Albuquerque, NM), co-editor-in-chief of *Int. J. of Nuclear Knowledge Management* (Interscience Publishers, Geneva), and Guest Professor at the Dept. of Applied Math. and CS in Ghent University and Adjunct Professor in the Faculty of Information Technology at University of Technology, Sydney, Australia.

Fuzzy Logic - A New Direction. The Concept of f-Validity

Lotfi A. Zadeh
Berkeley Initiative in Soft Computing
University of California
Berkeley
CA 94720-1776

Abstract

Science deals not with reality but with models of reality. In this perspective, fuzzy logic may be viewed as a system of concepts and techniques aimed at construction of models of reality which are better than those which can be constructed through the use of methods based on bivalent logic and bivalent-logic-based probability theory.

What lies beyond the boundaries of fuzzy logic? An uncharted territory which is explored involves settings in which precision is not an attainable objective. In such settings, it is necessary to retreat from precision and accept what may be called f-validity. The concept of f-validity has wide-ranging ramifications.

Special Sessions

Session 1

Aggregation Operators – R. Mesiar and A. Kolesárová

Dominance of Ordinal Sums of $T_{\mathbf{L}}$ and $T_{\mathbf{P}}$

Susanne Saminger
Dept. of Knowledge-Based Math. Systems
Johannes Kepler University
A-4040 Linz, Austria,
and
Dipartimento di Matematica "Ennio De Giorgi"
Università del Salento
I-73100 Lecce, Italy
susanne.saminger@jku.at

Peter Sarkoci
Dept. of Knowledge-Based Math. Systems
Johannes Kepler University
A-4040 Linz, Austria
peter.sarkoci@{jku.at|gmail.com}

Abstract

Dominance is a relation on operations which are defined on a common poset. We treat the dominance relation on the set of ordinal sum t-norms which involve either exclusively the Łukasiewicz t-norm or exclusively the product t-norm as summand operations. We show that in both cases, the question of dominance can be reduced to a simple property of the idempotent elements of the dominating t-norm. We finally discuss the obtained results and possibilities of their generalization.

Keywords: Triangular norm, Ordinal sum, Idempotent element, Dominance.

1 Introduction

The concept of dominance has been introduced within the framework of probabilistic metric spaces as a property relevant for building Cartesian products of such spaces [14]. Later on the dominance of t-norms was studied in connection with the construction of fuzzy equivalence relations [3, 4, 15] and the construction of fuzzy orderings [2]. Further, the concept of dominance was extended to the more general class of aggregation operators [7, 9]. The dominance of aggregation operators emerges when investigating which aggregation procedures applied to the system of T-transitive fuzzy relations yield again a T-transitive fuzzy relation [7] or when seeking aggregation operators which preserve the extensionality of fuzzy sets with respect to given T-equivalence relations [8]. The definition of dominance we use was given in [14].

Definition 1 *Let $(S, \leq)$ be a partially ordered set and let f and g be associative binary operations on S with common identity $e \in S$. Then f* dominates *g, and we write $f \gg g$, if the inequality*

$$f(g(x,y),g(u,v)) \geq g(f(x,u),f(y,v)). \qquad \text{(D)}$$

holds for all $x, y, u, v \in S$.

Recall that a *triangular norm* (t-norm for short) [5,14] is an associative, commutative, binary operation on the unit interval $[0, 1]$ which is non-decreasing in each argument and has neutral element 1. Most of our attention is confined to three prototypical t-norms – the minimum $T_{\mathbf{M}}(x, y) = \min\{x, y\}$, the product $T_{\mathbf{P}}(x, y) = xy$ and the Łukasiewicz t-norm $T_{\mathbf{L}}(x, y) = \max\{0, x+y-1\}$. Note that these t-norms are continuous. It is well known that $T_{\mathbf{M}}$ is the strongest t-norm, i.e., for any t-norm T we have $T_{\mathbf{M}} \geq T$. Moreover, for any t-norm T we have $T_{\mathbf{M}} \gg T$. Thanks to associativity and commutativity, each t-norm dominates itself; therefore dominance of t-norms is a *reflexive* relation. From its commutativity together with the fact that all t-norms have the common neutral element 1 it follows that dominance is a refinement of the standard point-wise order of t-norms, i.e., $T_1 \gg T_2$ implies that $T_1 \geq T_2$. Since *antisymmetry* of a binary relation is inherited by any refinement, the dominance of t-norms is an *antisymmetric* relation. Anyhow, the dominance of t-norms is *not transitive* [12] and this even if considered on continuous t-norms only. Plenty of counterexamples to the transitivity of dominance of continuous t-norms can be constructed by means of the results of the present paper. Before turning to the concrete contents we summarize facts and notions which will be essential later.

2 Ordinal Sums and Dominance

By an *order isomorphism* from $[a, b]$ to $[c, d]$ we mean any increasing bijection from the first interval to the second one. With the symbol $\psi_{[a,b]}$ we denote the unique *affine* order isomorphism from $[a, b]$ to the unit interval $[0, 1]$. If O is a binary operation on the interval $[c, d]$ and φ is an order isomorphism from $[a, b]$ to

$[c,d]$, the *φ-transform of O* is the binary operation

$$O_\varphi\colon [a,b]^2 \to [a,b] : (x,y) \mapsto \varphi^{-1}\big(O(\varphi(x),\varphi(y))\big).$$

All φ-transforms preserve dominance, i.e., $T_1 \gg T_2$ if and only if $(T_1)_\varphi \gg (T_2)_\varphi$ for any φ.

Recall that a continuous t-norm is said to be *Archimedean* if it is a φ-transform of $T_{\mathbf{L}}$ or of $T_{\mathbf{P}}$ for some order isomorphism $\varphi\colon [0,1] \to [0,1]$. For the brevity we write Archimedean t-norm instead of continuous Archimedean t-norm in the sequel.

Let I be an at most countable index set, let $\{T_i\}_{i\in I}$ be a system of arbitrary t-norms and let $\{[a_i,b_i]\}_{i\in I}$ be a system of subintervals of $[0,1]$ with pairwise disjoint interiors. The *ordinal sum* determined by these two systems is a binary operation on the unit interval given as

$$T(x,y) = \begin{cases} (T_i)_{\psi_{[a_i,b_i]}}(x,y) & \text{if } x,y \in [a_i,b_i]\,, \\ T_{\mathbf{M}}(x,y) & \text{otherwise.} \end{cases}$$

This operation is again a t-norm [5]; we denote it by $(\langle a_i,b_i,T_i\rangle)_{i\in I}$. To t-norms T_i we refer as the *summands of T* and to intervals $[a_i,b_i]_{i\in I}$ as the *summand carriers*. In the sequel we will always assume that I is an at most coutable index set and that the corresponding families of intervals have pairwisely disjoint interiors.

Observe that if for some t-norms T_1, T_2, with T_2 being the ordinal sum $(\langle a_i,b_i,T_{2,i}\rangle)_{i\in I}$ it holds that $T_1 \geq T_2$, then also T_1 is an ordinal sum with the same underlying summand carriers but with possibly different summands, i.e., $T_1 = (\langle a_i,b_i,T_{1,i}\rangle)_{i\in I}$. Since comparability is a necessarry condition for dominance, the same holds if assumed $T_1 \gg T_2$.

The dominance of ordinal sum t-norms can be examined summand by summand [10]:

Theorem 2 Let $T_1 = (\langle a_i,b_i,T_{1,i}\rangle)_{i\in I}$ and $T_2 = (\langle a_i,b_i,T_{2,i}\rangle)_{i\in I}$ be two t-norms with the same structure of summand carriers. Then $T_1 \gg T_2$ if and only if $T_{1,i} \gg T_{2,i}$ holds for all $i \in I$.

Recall that an element $x \in [0,1]$ is called an *idempotent element* of T if $T(x,x) = x$. The set of all idempotent elements of T will be denoted $\mathfrak{I}_T$. As 0 and 1 are idempotent elements of every t-norm we refer to them as *trivial* idempotent elements. The structure of $\mathfrak{I}_T$ is strongly tied to the dominance properties of T [10]:

Theorem 3 If $T_1 \gg T_2$ then $\mathfrak{I}_{T_1}$ is closed with respect to T_2, i.e., if $x,y \in \mathfrak{I}_{T_1}$ then also $T_2(x,y) \in \mathfrak{I}_{T_1}$.

Let us concentrate now on the class $\mathcal{OS}_{T_{\mathrm{s}}}$ of ordinal sums of type $(\langle a_i,b_i,T_{\mathrm{s}}\rangle)_{i\in I}$ where all summands are equal to a fixed t-norm T_{s}. The elucidation of dominance on such a class can be simplified substantially making use of Theorem 2. Let $T_1, T_2 \in \mathcal{OS}_{T_{\mathrm{s}}}$ with $T_1 \geq T_2$ which is a necessary condition for dominance. Observe that both these t-norms can be constructed as ordinal sum t-norms with the same summand carriers but different summands:

$$\begin{aligned} T_1 &= (\langle a_i,b_i,T_{1,i}\rangle)_{i\in I} \\ T_2 &= (\langle a_i,b_i,T_{\mathrm{s}}\rangle)_{i\in I} \end{aligned}$$

where $T_{1,i} \in \mathcal{OS}_{T_{\mathrm{s}}}$ for all $i \in I$. By Theorem 2 we have $T_1 \gg T_2$ if and only if $T_{1,i} \gg T_{\mathrm{s}}$ for all $i \in I$. Thus, in order to solve the dominance on $\mathcal{OS}_{T_{\mathrm{s}}}$ completely, it is sufficient to describe all t-norms $T \in \mathcal{OS}_{T_{\mathrm{s}}}$ that dominate T_{s}. In the following two chapters we solve this question for the case $T_{\mathrm{s}} = T_{\mathbf{L}}$ and $T_{\mathrm{s}} = T_{\mathbf{P}}$.

3 Dominance on $\mathcal{OS}_{T_{\mathbf{L}}}$

The principal result of this section is a generalization and strenghtening of methods developed in [12] (see also [11] for a nice survey). The main message is, that Theorem 3 can be strenghthened provided all t-norms are from $\mathcal{OS}_{T_{\mathbf{L}}}$.

Theorem 4 Consider a triangular norm $T \in \mathcal{OS}_{T_{\mathbf{L}}}$. Such T dominates $T_{\mathbf{L}}$ if and only if $\mathfrak{I}_T$ is closed with respect to $T_{\mathbf{L}}$.

Thus the question of dominance between two t-norms from the class $\mathcal{OS}_{T_{\mathbf{L}}}$ can be reduced to the verification of a very simple algebraic property of some subset of the unit interval. To demostrate the power of this characterisation, let us first consider some simple examples.

Let us define $T_\lambda = (\langle \lambda, 1, T_{\mathbf{L}}\rangle)$ with $\lambda \in [0,1]$; such t-norms form a one-parametrical family. Clearly, $T_{\lambda_1} \geq T_{\lambda_2}$ if and only if $\lambda_1 \geq \lambda_2$. In this situation T_1 can be expressed as the ordinal sum $(\langle \lambda_2, 1, T_{\lambda^*}\rangle)$ where $\lambda^* = (\lambda_1 - \lambda_2)/(1-\lambda_2)$. Clearly $\lambda^* \in [0,1]$. By Theorem 2 t-norm T_{λ_1} dominates T_{λ_2} if and only if $T_{\lambda^*} \gg T_{\mathbf{L}}$. Further, by Theorem 4 the latter holds if and only if the set $\mathfrak{I}_{T_{\lambda^*}} = [0,\lambda^*] \cup \{1\}$ is closed with respect to $T_{\mathbf{L}}$. Since this is the case for any $\lambda^* \in [0,1]$ we have that within this family $T_{\lambda_1} \gg T_{\lambda_2}$ if and only if $\lambda_1 \geq \lambda_2$. As a consequence, this family is completely ordered by the dominance relation (see [11] for the detailed treatement).

Now, let us redefine $T_\lambda = (\langle 0, \lambda, T_{\mathbf{L}}\rangle)$ with $\lambda \in [0,1]$; this is a one-parametrical family often refered to as the Mayor-Torrens family [6]. We have $T_{\lambda_1} \geq T_{\lambda_2}$ if and only if $\lambda_1 \leq \lambda_2$. In this situation T_1 can be expressed

as the ordinal sum $(\langle 0, \lambda_2, T_{\lambda^*}\rangle)$ where $\lambda^* = \lambda_1/\lambda_2$. Again, we have $\lambda^* \in [0,1]$. By Theorem 2 the t-norm T_{λ_1} dominates T_{λ_2} if and only if $T_{\lambda^*} \gg T_{\mathbf{L}}$. And by Theorem 4 the latter holds if and only if the set $\mathcal{I}_{T_{\lambda^*}} = \{0\} \cup [\lambda^*, 1]$ is closed with respect to $T_{\mathbf{L}}$. A very simple analysis reveals that this property is never satisfied unles $\lambda^* = 0$ or $\lambda^* = 1$. These two cases correspond to the situations $\lambda_1 = 0$ or $\lambda_1 = \lambda_2$ respectively. Observe that both these situations encode the trivial cases of dominance only; while the first one corresponds to the fact that $T_{\mathbf{M}}$ dominates any member of the family, the second one indicates that the dominance is a reflexive relation. As a consequence there are no other cases of dominance within this family. Therefore this family is ordered by the dominance relation although, in contrast to the previous example, the order is not linear (again, for a detailed treatment see [11]).

Finally, let us put $T_1 = (\langle \frac{1}{2}, 1, T_{\mathbf{L}}\rangle)$, $T_2 = (\langle 0, \frac{1}{2}, T_{\mathbf{L}}\rangle, \langle \frac{1}{2}, 1, T_{\mathbf{L}}\rangle)$ and $T_3 = T_{\mathbf{L}}$. Making use of Theorem 2 and Theorem 4 one can show easily that $T_1 \gg T_2$, $T_2 \gg T_3$ while $T_1 \not\gg T_3$ [12]. It follows in turn, that dominance is not transitive on the class $\mathcal{OS}_{T_{\mathbf{L}}}$. Plenty of other counterexamples to the transitivity of dominance of continous triangular norms can be constructed with the aid of the just presented method [11].

4 Dominance on $\mathcal{OS}_{T_{\mathbf{P}}}$

We say that a t-norm T is *ordinally irreducible* if the only way how to represent it as an ordinal sum is $(\langle 0, 1, T\rangle)$. We consider some ordinal sum t-norm $T = (\langle a_1, b_i, T_i\rangle)_{i\in I}$ and assume that all its summands are ordinally irreducible, i.e., the actual representation of T as an ordinal sum is the finest possible. Then we define the so called *axis* of T as a set

$$\mathcal{AX}_T = \{(x,x) \,|\, x \in [0,1]\} \cup \left(\bigcup_{i\in I} [a_i, b_i]^2\right).$$

Note that due to the chosen representation of T, its axis $\mathcal{AX}_T$ is uniquely defined.

The function $f\colon [0,1] \to [0,1]$ is said to be *superhomogenous* [1] if the mapping $x \mapsto f(x)/x$ is nonincreasing on the interval $]0,1]$. A t-norm is said to be superhomogenous if all its horizontal sections are superhomogenous. Note that $T_{\mathbf{P}}$ and $T_{\mathbf{M}}$ are examples of superhomogenous t-norms while $T_{\mathbf{L}}$ is not.

Theorem 5 Let T be a superhomogeneous ordinal sum t-norm. If

$$T(T_{\mathbf{P}}(x,y), T_{\mathbf{P}}(u,v)) \geq T_{\mathbf{P}}(T(x,u), T(y,v))$$

holds for any $(y,v) \in \mathcal{AX}_T$, then $T \gg T_{\mathbf{P}}$.

In particular, if all summands are superhomogeneous, the resulting ordinal sum is so; therefore, Theorem 5 applies to all $T \in \mathcal{OS}_{T_{\mathbf{P}}}$. Moreover in the case of $\mathcal{OS}_{T_{\mathbf{P}}}$ one can characterize the dominance in a style analogical to Theorem 4.

Theorem 6 Consider a triangular norm $T \in \mathcal{OS}_{T_{\mathbf{P}}}$. Such T dominates $T_{\mathbf{P}}$ if and only if $\mathcal{I}_T$ is closed with respect to $T_{\mathbf{P}}$.

Thus the question of dominance within the class $\mathcal{OS}_{T_{\mathbf{P}}}$ can be reduced to a verification of a very simple algebraic property of the set of idempotent elements of one of the t-norms involved. Again, we list a few of examples.

Similarly as in the previous section, we can define the one-parametric family $T_\lambda = (\langle \lambda, 1, T_{\mathbf{P}}\rangle)$ for $\lambda \in [0,1]$. Again, $T_{\lambda_1} \geq T_{\lambda_2}$ if and only if $\lambda_1 \geq \lambda_2$. In this case we can also write $T_1 = (\langle \lambda_2, 1, T_{\lambda^*}\rangle)$ where $\lambda^* = (\lambda_1 - \lambda_2)/(1 - \lambda_2)$. By Theorem 2 $T_{\lambda_1} \gg T_{\lambda_2}$ if and only if $T_{\lambda^*} \gg T_{\mathbf{P}}$. Obviously $\lambda^* \in [0,1]$, so $\mathcal{I}_{T_{\lambda^*}} = [0, \lambda^*] \cup \{1\}$; this set is closed with respect to multiplication regardless of λ^*. By Theorem 6 $T_{\lambda_1} \gg T_{\lambda_2}$ if and only if $\lambda_1 \geq \lambda_2$. Observe that the treatment of this one-parametric family was formally the same as the treatment of the analogic family in the previous section. One can also mimic the investigation of the second family from the previous section, thus obtaining an analogical result for dominance of triangular norms $T_\lambda = (\langle 0, \lambda, T_{\mathbf{P}}\rangle)$.

In order to construct less trivial t-norms $T \in \mathcal{OS}_{T_{\mathbf{P}}}$ dominating $T_{\mathbf{P}}$ let us fix arbitrary $p, q \in]0,1[$ and define the set $M = \{p^n q^m \,|\, n, m \in \mathbb{N}\} \cup \{1\}$. Obviously, M is a countable set closed with respect to multiplication. Moreover, for each $a \in M$ there exists a unique element $a' \in M$ with $a' < a$ and such that there exist no $c \in M$ with $a' < c < a$. Now define $T = (\langle a', a, T_{\mathbf{P}}\rangle)_{a\in M}$. For this t-norm we have $\mathcal{I}_T = M \cup \{0\}$. Evidently, also $\mathcal{I}_T$ is closed with respect to multiplication; from Theorem 6 follows immediatelly that $T \gg T_{\mathbf{P}}$.

5 Possible generalizations

It is interesting that both Theorems 4 and 6 are formally of the same structure captured by the following schema.

Schema 7 Consider a triangular norm $T \in \mathcal{OS}_{T_{\mathrm{s}}}$. Such T dominates T_{s} if and only if $\mathcal{I}_T$ is closed with respect to T_{s}.

The main message of the previous two chapters could be summarized as that Schema 7 holds if $T_{\mathrm{s}} = T_{\mathbf{L}}$

or $T_{\mathrm{s}} = T_{\mathbf{P}}$. Immediatelly a natural question arises — which other t-norms satisfy this schema ? In the sequel we provide an example of an Archimedean t-norm which violates Schema 7.

The requirement that a subset of the unit interval has to be closed with respect to $T_{\mathbf{L}}$ or $T_{\mathbf{P}}$ is rather restrictive. Among different consequences of this condition, the one summarized in the following theorem is of a particular importance for our following considerations.

Theorem 8 If some $M \subseteq [0,1]$ fulfills $\{0,1\} \subseteq M$ and is closed with respect to some Archimedean t-norm, then either $M = [0,1]$ or the point 1 is isolated in M, i.e., there exists some $a < 1$ such that $]a,1[\cap M$ is the empty set.

This observation taken into account together with Theorem 3 yields interesting consequences for dominance of ordinal sum t-norms. In particular, if the t-norm $T = (\langle a_i, b_i, T_i\rangle)_{i\in I}$ dominates some Archimedean t-norm, then either T is simply the trivial ordinal sum equal to $T_{\mathbf{M}}$ ($\mathcal{J}_T = [0,1]$) or the right endpoint of some summand carrier coincides with 1 (meaning that $b_i = 1$ for some index i). If we restrict to nontrivial cases only, then one summand carrier of T is of the type $[a,1]$. We will refer to this summand carrier as the *top summand carrier* and to the corresponding summand as the *top summand*. In the special case when the dominated t-norm is $T_{\mathbf{L}}$ we can say even more.

Theorem 9 If the ordinal sum t-norm T different from $T_{\mathbf{M}}$ dominates $T_{\mathbf{L}}$, then so does its top summand.

Now, let $T_{\mathrm{s}} = (T_{\mathbf{L}})_{\varphi}$ for some order isomorphism $\varphi\colon [0,1] \to [0,1]$. Clearly, T_{s} is Archimedean. Assume that the t-norm $T \in \mathcal{OS}_{T_{\mathrm{s}}}$ different from $T_{\mathbf{M}}$ dominates T_{s}. In other words T satisfies the conclusion of Schema 7. Denote by $[a,1]$ the top summand carrier of T. Since φ-transforms preserve dominance, we have $T_{\varphi^{-1}} \gg T_{\mathbf{L}}$. By Theorem 9 the top summand of $T_{\varphi^{-1}}$ has to dominate $T_{\mathbf{L}}$ as well. Simple computation reveals that this top summand is $(T_{\mathbf{L}})_{\phi_a}$ where

$$\phi_a = \varphi \circ \psi_{[a,1]} \circ \varphi^{-1} \circ \psi^{-1}_{[\varphi(a),1]}.$$

So we have the following necessary condition.

Theorem 10 Let $\varphi\colon [0,1] \to [0,1]$ be an order isomorphism. If the t-norm $T = (T_{\mathbf{L}})_{\varphi}$ satisfies Schema 7 then

$$(T_{\mathbf{L}})_{\phi_a} \gg T_{\mathbf{L}}$$

for all $a \in]0,1[$.

Put now $\varphi\colon x \mapsto x^2$. Corresponding $T_{\mathrm{s}} = (T_{\mathbf{L}})_{\varphi}$ is the so called Schweizer-Sklar t-norm [13]. Now for the choice $a = \frac{1}{2}$ one can show easily that the corresponding T_{ϕ_a} does not dominate $T_{\mathbf{L}}$. Thus T_{s} is an Archimedean t-norm violating Schema 7.

Acknowledgement

Susanne Saminger gratefully acknowledges the support by the Austrian Science Fund (FWF) in the framework of the Erwin Schrödinger Fellowship J2636 "Dominance from Probabilistic Metric Spaces to (Quasi-) Copulas and Applications". Peter Sarkoci gratefully acknowledges the support by the grants VEGA 1/3012/06 and VEGA 1/2005/05.

References

[1] P. Burai, A. Száz, Homogeneity Properties of Subadditive Functions, Ann. Math. Inf. 32 (2005) 189–201.

[2] U. Bodenhofer, A Similarity-Based Generalization of Fuzzy Orderings, Volume C 26, Schriftereihe der Johannes-Kepler-Universität Linz, Universitätsverlag Rudolf Trauner, Linz (1999).

[3] B. De Baets, R. Mesiar, Pseudo-metrics and T-equivalences, J. Fuzzy Math. 5 (1997) 471–481.

[4] B. De Baets, R. Mesiar, T-partitions, Fuzzy Sets and Systems 97 (1998) 211–223.

[5] E. P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.

[6] G. Mayor and J. Torrens. On a family of t-norms. Fuzzy Sets and Systems, 41 (1991) 161–166.

[7] S. Saminger, R. Mesiar, U. Bodenhofer, Domination of Aggregation Operators and Preservation of Transitivity, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 10 (2002) 11–35.

[8] S. Saminger, Aggregation of Extensional Hulls and Domination, in: E. P. Klement, R. Mesiar, E. Drobná, F. Chovanec (Eds.), *Abstracts of FSTA 2004 – The Seventh International Conference on Fuzzy Sets Theory and its Applications*, Liptovský Ján. Liptovský Mikuláš, 2004, pp. 92–93.

[9] S. Saminger, U. Bodenhofer, E. P. Klement, R. Mesiar, Aggregation of Fuzzy Relations and Preservation of Transitivity, Lecture Notes in Artificial Intelligence 4342 (2006) 184–206.

[10] S. Saminger, B. De Baets, H. De Meyer, On the Dominance Relation between Ordinal Sums of Conjunctors, Kybernetika 42 (2006) 337–350.

[11] S. Saminger, P. Sarkoci, B. De Baets, The Dominance Relation on the Class of Continuous T-Norms from an Ordinal Sum Point of View, Lecture Notes in Artificial Intelligence 4342 (2006) 334–354.

[12] P. Sarkoci, Dominance is not Transitive on Continuous t-norms, Aequationes Mathematicae, Springer, accepted.

[13] B. Schweizer, A. Sklar, Associative Functions and Statistical Triangle Inequalities. Publ. Math. Debrecen. 8 (1961) 169–186.

[14] B. Schweizer, A. Sklar, Probabilistic Metric Spaces, North-Holland, Amsterdam 1983.

[15] L. Valverde, On the structure of F-indistinguishability operators, Fuzzy Sets and Systems 17 (1985) 313–328.

Domination and Information Boundedness Principle for Aggregation

Michal Šabo
STU Bratislava, Radlinského 9,
81237 Bratislava, Slovakia
e-mail: michal.sabo@stuba.sk

Abstract

The information boundedness principle for rule based inference process requires that the knowledge obtained as a result of a rule should not have more information than that contained in the consequent of this rule. We formulate the information boundedness principle for aggregation and show that it can be expressed using the notion of dominance. We also investigate conditions under which this principle holds and give some open problems.

Keywords: Information boundedness principle, Aggregation, Specificity measure, Domination.

1 Introduction

The main aim of rule based fuzzy modelling inference process is to establish a value of the output for given input value. Such system usually involves the use of a rule base consisting of several fuzzy statements. The most used types of fuzzy statements are if-then rules with fuzzy predicates. Overall output can be obtained by aggregation of individual rule outputs and by a possible defuzzification. The individual rule output is often based on the use of fuzzy implications or fuzzy conjunctions. The generalization of these operators is a relevancy transformation operator (RET operator) [1, 4, 5]. So, the individual output is obtained from the relevancy of the used rule and the rule consequent. To have this process meaningful, we require: Knowledge obtained in the individual rule output should not have more information than that contained in the consequent of rule [4]. This principle is usually called the information boundedness principle (IBP) for a fuzzy rule. The IBP for RET operators is investigated in [1]. For purpose of this paper we need to define some measure of information. As we consider fuzzy subsets of a finite universe we shall use a specificity measure which is maximal for singletons. Of course, a pointwise aggregation of fuzzy sets with high specificity measures need not to be a fuzzy set with high specificity again; particularly the aggregation of singletons need not to be a singleton. Thus we can express IBP for aggregation as follows: *The specificity measure of aggregation of individual outputs should not exceed the aggregation of specificities of individual outputs*.

It is clear that the mentioned principle needs two aggregation operators in general, one for aggregation of fuzzy sets (individual outputs) and another one for aggregation of specifities. For simplicity, in this paper, we shall use the same aggregator in both cases.

The main goal of this contribution is to investigate mathematical aspects of the mentioned IBP for aggregation. We will not pay attention to philosophical background of this problem.

This contribution is organised as follows. We recall necessary basic notions in Section 2. We introduce IBP for aggregation operators in Section 3. Section 4 brings some new results and some open problems.

2 Preliminaries

In this paper we shall consider fuzzy subset of a finite universe X with the cardinality n. Then each fuzzy set can be represented by an n-tuple $(a_1, a_2, \ldots, a_n) \in [0, 1]^n$. Many information measures have been attached to finite fuzzy sets, e.g., entropy of Shannon, measure of imprecision etc. Motivated by Yager [4, 5, 6] we define a specificity measure as a mapping from

the system of all fuzzy sets on X to the unit interval [1].

Definition 1 A mapping Sp: $[0, 1]^n \longrightarrow [0, 1]$ is called a specificity measure (specificity for short) if the following axioms are satisfied:

(S1) For any permutation $(p1, p2, \ldots, pn)$ of $(1, 2, ..., n)$ is

$$\mathrm{Sp}(a_{p1}, a_{p2}, \ldots, a_{pn}) = Sp(a_1, a_2, \ldots, a_n).$$

(S2) $Sp(a_1, a_2, \ldots, a_n) = 1$ if and only if there exists the unique i such that $a_i = 1$ and $a_j = 0$ for $j \neq i$.

(S3) If $1 \geq b_1 > b_2 \geq a_2 \geq a_3 \geq \ldots \geq a_n$ then

$$Sp(b_1, a_2, \ldots, a_n) > Sp(b_2, a_2, \ldots, a_n).$$

(S4) If $1 \geq a_1 \geq a_2 \geq \ldots \geq a_n$ and $1 \geq a_1 \geq b_2 \geq \ldots \geq b_n$ and $a_i \geq b_i$ for $i = 2, 3, \ldots, n$, then

$$Sp(a_1, a_2, \ldots, a_n) \leq Sp(a_1, b_2, \ldots, b_n).$$

(S5) $Sp(0, 0, \ldots, 0) = 0$

Axioms (S3) and (S4) say that specificity is increasing in the greatest membership value and non increasing in the others. Axiom (S2) says that only singletons have maximal specificity.

Moreover we say that a specificity Sp is grounded if

$$Sp(b, b, \ldots, b) = 0$$

for all $b \in [0, 1]$ and we say that Sp is shift invariant if

$$Sp(a_1, a_2, \ldots, a_n) = Sp(a_1+b, a_2+b, \ldots, a_n+b)$$

for all $b \in [-\min_i(a_i), 1 - \max_i(a_i)]$.

Example 1 Consider fuzzy sets over universe with two elements only, so $n = 2$. Put

(i) $Sp_1(x, y) = \max\{x, y\} - \min\{x, y\}$

(ii) $Sp_2(x, y) = \max\{x, y\} - 0.5 \min\{x, y\}$.

The specificity Sp_1 is grounded and shift invariant, but Sp_2 is also specificity but it is neither grounded nor shift invariant.

The specificity Sp_1 belongs to the family of so called linear specifities [1, 4].

Definition 2 Let $1 \geq v_2 \geq v_3 \geq \ldots \geq v_n$ be given constants such that $v_2 + v_3 + \ldots + v_n = 1$. Then a linear specificity Sp: $[0, 1]^n \longrightarrow [0, 1]$ is defined by:

$$Sp(a_1, a_2, \ldots, a_n) = a_{k1} - v_2 a_{k2} - v_3 a_{k3} - \ldots - v_n a_{kn}$$

where $(k1, k2, \ldots, kn)$ is a permutation of $(1, 2, ..., n)$ is such that $a_{k1} \geq a_{k2} \geq a_{k3} \geq \ldots \geq a_{kn}$.

Note that each linear specificity can be expressed using an OWA operator [4, 5] by

$$Sp(a_1, a_2, \ldots, a_n) = a_i -$$

$$\mathrm{OWA}\ v_n, \ldots, v_3, v_2\ (a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$$

with $a_i = \max(a_1, a_2, \ldots, a_n)$. For $n = 1$ we put $Sp(a_1) = a_1$.

Every linear specificity is grounded and shift invariant. Yager [4] showed that any linear specificity Sp satisfies

$$Sp_{l\min} \leq Sp \leq Sp_{l\max}$$

where

$$Sp_{l\min}(a_1, a_2, \ldots, a_n) = a_1 - a_2$$

and

$$Sp_{l\max}(a_1, a_2, ..., a_n) = a_1 - \frac{a_2 + a_3 + ... + a_n}{n-1}$$

for $1 \geq a_1 \geq a_2 \geq a_3 \geq \ldots \geq a_n \geq 0$.

The notion of aggregation operator (aggregator for short) plays an important role in fuzzy inference processes, especially in the information fusion.

Definition 3 A mapping

$$Agg : \bigcup_{n \in N} [0,1]^n \rightarrow [0,1]$$

is called aggregator if for all naturals n

(*i*) $Agg(0, 0, \ldots, 0) = 0$, $Agg(1, 1, \ldots, 1) = 1$.

(*ii*) $Agg(x_1, x_2, \ldots x_n) \geq Agg(y_1, y_2, \ldots y_n)$ when

$x_i \geq y_i$ for all $i = 1, 2, \ldots, n$.

(*iii*) $Agg(x) = x$ for all $x \in [0, 1]$.

The arithmetic and weighted means, t-norms, t-conorms are examples of aggregators. In this paper we shall aggregate also fuzzy sets over the same universe pointwisely i. e.,

$$A = Agg\ (A_1, A_2, \ldots, A_k)$$

if for all $x \in X$

$$A(x) = Agg\ (A_1(x), A_2(x), \ldots, A_k(x)).$$

Remark that for given natural m the aggregator can be considered as m-ary operation

$$Agg: [0, 1]^m \longrightarrow [0, 1].$$

The notion of dominance [2, 3] plays an important role in many fields of mathematics. It can be defined for n-ary operations on posets. We restrict our definition on the unit interval.

Definition 4 Let U: $[0, 1]^m \longrightarrow [0, 1]$, V: $[0, 1]^n \longrightarrow [0, 1]$ be m-ary and n-ary operations respectively. We say that an operation U dominates an operation V or V is dominated by U, shortly $U \gg V$ or $V \ll U$, if for any matrix$(x_{i,j})$ of type $m \times n$ with elements from [0, 1] we have

$$U(V(x_{1,1}, \ldots, x_{1,n}),\ldots, V(x_{m,1}, \ldots, x_{m,,n})) \geq$$
$$V(U(x_{1,1}, \ldots, x_{m,1}),\ldots, U(x_{1,n}, \ldots, x_{m,n})).$$

If U and V are binary operations, $U \gg V$ is equivalent to

$$U(V(x, y), V(u, v)) \geq V(U(x, u), U(y, v))$$

for all $x, y, u, v \in [0, 1]$.

3 IBP for aggregators

Let m, n be naturals and B_1 , B_2, …, B_m be fuzzy sets (individual fuzzy outputs) over the same finite universe X with cardinality n. Thus B_1, B_2, …, $B_m \in [0, 1]^n$. Let Sp: $[0, 1]^n \longrightarrow [0, 1]$ be a specificity measure. Then the aggregator Agg: $[0, 1]^m \longrightarrow [0, 1]$ fulfills IBP with respect to a given specificity Sp on the universe X if

$$Sp(B) \leq Agg(Sp(B_1), Sp(B_2),..., Sp(B_m))$$

for any B_1 , B_2, …, $B_m \in [0, 1]^n$ and B (overall output) is given by

$$B= Agg\ (B_1, B_2, \ldots, B_m)$$

Recall that we shall use the same aggregation operator for the aggregation of fuzzy sets and also for specifities. The next proposition allows verifying IBP by means of dominance.

Proposition 1 An aggregator Agg:$[0, 1]^m \longrightarrow [0, 1]$ fulfills IBP with respect to a given specificity Sp: $[0, 1]^n \longrightarrow [0, 1]$ on a finite universe X with cardinality n if and only if

$$Agg \gg Sp.$$

Proof: Agg:$[0, 1]^m \longrightarrow [0, 1]$fulfills IBP with respect to a given specificity Sp: $[0, 1]^n \longrightarrow [0, 1]$ if and only if for any B_1 , B_2, …, $B_m \in [0, 1]^n$ and $B= Agg\ (B_1, B_2, \ldots, B_m)$ we have

$$Sp(B) \leq Agg(Sp(B_1), Sp(B_2),..., Sp(B_m)).$$

Let

$$B_1 = (b_{1,1}, b_{1,2}, \ldots, b_{1,n})$$
$$B_2 = (b_{2,1}, b_{2,2}, \ldots, b_{2,n})$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$B_m = (b_{m,1}, b_{m,2}, \ldots, b_{m,n})$$

Then

$$B= Agg\ (B_1, B_2, \ldots, B_m) =$$

$(Agg(b_{1,1}, b_{2,1}, \ldots, b_{m,1}), Agg(b_{1,2}, b_{2,2}, \ldots, b_{m,2}), Agg(b_{1,n}, b_{2,n}, \ldots, b_{m,n}))$.

The last inequality is equivalent to

$$Sp\ (Agg(b_{1,1}, b_{2,1}, \ldots, b_{m,1}), Agg(b_{1,2}, b_{2,2}, \ldots, b_{m,2}), \mathrm{Agg}(b_{1,n}, b_{2,n}, \ldots, b_{m,n})) \leq$$
$$Agg(Sp(b_{1,1}, b_{1,2}, \ldots, b_{1,n}), Sp(b_{2,1}, b_{2,2}, \ldots, b_{2,n}),..., Sp(b_{m,1}, b_{m,2}, \ldots, b_{m,n}))$$

which means that $Agg \gg Sp$.

4 Results and open problems

We have introduced IBP for an aggregator (considered as the an m-ary operation) with respect to a given specificity measure for fuzzy sets on given n-membered universe. We shall try to characterize m-ary aggregators fulfilling IBP with respect to given specificity and formulate some open problems.

Proposition 2 Consider a linear specificity Sp: $[0, 1]^2 \longrightarrow [0, 1]$ and the arithmetic mean $Amean$: $[0, 1]^m \longrightarrow [0, 1]$ as the aggregator. Then for any natural m

$$Amean \gg Sp$$

Proof. Evidently for all $(x, y) \in [0, 1]^2$

$$Sp(x,y) = |x - y|$$

and for all $A_1= (a_{11}, a_{12})$, $A_2 = (a_{2,1}, a_{2,2})$, …, $A_m = (a_{m,1}, a_{m,2})$ we have

$$Amean(Sp(a_{11}, a_{12}), Sp(a_{2,1}, a_{2,2}), Sp(\ a_{m,1}, a_{m,2}) = \frac{|a_{1,1} - a_{1,2}| + |a_{2,1} - a_{2,2}| + \ldots + |a_{m,1} - a_{m,2}|}{m} \geq$$

$$\frac{|a_{1,1} + a_{2,1} + \ldots + a_{m,1} - (a_{1,2} + a_{2,2} + \ldots + a_{m,2})|}{m} =$$

$$Sp(Amean((a_{1,1}, \ldots, a_{m,1}), Amean((a_{1,2} \ldots, a_{m,2})$$

which proves that $Amean \gg Sp$.

In the next example we shall show that the modification of Proposition 2 is not more valid for $n = 3$; i.e., for any linear specificity Sp: $[0, 1]^3 \longrightarrow [0, 1]$.

Example 2 Consider fuzzy sets over a universe with three elements and the minimal linear specificity Sp: $[0, 1]^3 \longrightarrow [0, 1]$ given by

$$Sp\ (a_1, a_2, a_3) = a_1 - a_2$$

where $1 \geq a_1 \geq a_2 \geq a_3 \geq 0$ and the arithmetic mean *Amean* as an aggregator. Put

$$A = (1, 0, 0.8),\ B = (0, 1, 0.8)$$

Then $Sp(A) = Sp(B) = 0.2$, $Amean(A, B) = (0.5, 0.5, 0.8)$ and $0.2 = Amean(Sp(A),\ Sp(B)) < Sp(Amean(A, B) = 0.3$. The aggregator *Amean* does not dominate the specificity *Sp*; IBP is not fulfilled.

Proposition 2 can be generalized for a wider class of specifities.

Proposition 3 Consider a specificity Sp: $[0, 1]^2 \longrightarrow [0, 1]$ given by

$$Sp(x, y) = \max\{x, y\} - \alpha \min\{x, y\}$$

for all $(x, y) \in [0, 1]^2$ and for a fixed $\alpha \in [0, 1]$ and the arithmetic mean as an aggregator *Amean* : $[0, 1]^m \longrightarrow [0, 1]$. Then for any natural m holds

$$Amean \gg Sp.$$

Proof. For all $A_1 = (a_{1,1}, a_{1,2})$, $A_2 = (a_{2,1}, a_{2,2})$, …, $A_m = (a_{m,1}, a_{m,2})$ we have

$$Sp\left(Amean\left(A_1, A_2, \ldots A_m\right)\right) =$$

$$Sp\left(\frac{a_{1,1} + a_{2,1} + \ldots + a_{m,1}}{m}, \frac{a_{1,2} + a_{2,2} + \ldots + a_{m,2}}{m}\right) =$$

$$\max\left\{\frac{a_{1,1} + a_{2,1} + \ldots + a_{m,1}}{m}, \frac{a_{1,2} + a_{2,2} + \ldots + a_{m,1}}{m}\right\} -$$

$$\alpha \min\left\{\frac{a_{1,1} + a_{2,1} + \ldots + a_{m,1}}{m}, \frac{a_{1,2} + a_{2,2} + \ldots + a_{m,2}}{m}\right\} \leq$$

$$\frac{\max\{a_{1,1} + a_{1,2}\} + \max\{a_{2,1} + a_{2,2}\} + \max\{a_{m,1} + a_{m,2}\}}{m} -$$

$$\alpha \frac{\min\{a_{1,1} + a_{1,2}\} + \min\{a_{2,1} + a_{2,2}\} + \min\{a_{m,1} + a_{m,2}\}}{m}$$

which proves the claim.

The following proposition shows that there exists an aggregator which fulfills IBP for arbitrary cardinality of finite universe with respect to all linear specificities. Remark that the maximum operator is often used as an aggregator in fuzzy modeling inference processes.

Proposition 4 Consider a linear specificity Sp: $[0, 1]^n \longrightarrow [0, 1]$ and the aggregator Agg:$[0, 1]^m \longrightarrow [0, 1]$ given by

$$Agg\ (b_1,\ b_2, \ldots,\ b_m) = \max\{b_1,\ b_2, \ldots,\ b_m\}.$$

Then for all naturals n, m

$$Agg \gg Sp.$$

Proof: The proof is trivial if $m = 1$ or $n = 1$ Consider $n, m \in \{2, 3, \ldots\}$,

$$B_1 = (b_{1,1}, b_{1,2}, \ldots, b_{1,n})$$

$$B_2 = (b_{2,1}, b_{2,2}, \ldots, b_{2,n})$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$B_m = (b_{m,1}, b_{m,2}, \ldots, b_{m,n})$$

and

$$B = \max\{B_1, B_2, \ldots,\ B_m\} = (b_1, b_1, \ldots, b_n) =$$

$(\max(b_{1,1}, b_{2,1}, \ldots, b_{m,1}), \max(b_{1,2}, b_{2,2}, \ldots, b_{m,2}),$ $\max(b_{1,n}, b_{2,n}, \ldots, b_{m,n}))$.
Then

$$Sp(Agg\ (B_1, B_2, \ldots,\ B_m)) =$$

$$Sp(\max\ (B_1, B_2, \ldots,\ B_m)) = b_i -$$

$$\text{OWA}\ v_n, \ldots, v_3, v_2\ (b_1,\ b_2, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n),$$

where $b_i = \max\{b_1, b_2, \ldots, b_n\}$ is the maximal element of the matrix $(b_{i,j})$ (maximal membership value in all considered fuzzy sets $B_1, B_2, \ldots, B_m$). Then $b_i = b_{k,i}$ for some $k \in \{1,2, \ldots, m\}$ and

$$Sp(B_k) = b_i -$$

$$\text{OWA}\ v_n, \ldots, v_3, v_2\ (b_{k,1},\ b_{k,2}, \ldots, b_{k,i-1}, b_{k,i-1}, \ldots, b_{k,n})$$

and

$$Agg(Sp(B_1), Sp(B_2), \ldots, Sp(B_m)) =$$

$$\max(Sp(B_1), Sp(B_2), \ldots, Sp(B_m)) \geq Sp(B_k)$$

Replacing $b_{k,1},\ b_{k,2}, \ldots, b_{k,i-1}, b_{k,i+1}, \ldots, b_{k,n}$ by non smaller values $b_1,\ b_2, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n$ we have

$$Agg(Sp(B_1), Sp(B_2), \ldots, Sp(B_m)) \geq Sp(B_k) \geq b_i -$$

$$\text{OWA}\ v_n, \ldots, v_3, v_2\ (b_1,\ b_2, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n) =$$

$$Sp(Agg\ (B_1, B_2, \ldots,\ B_m))$$

for any $B_1, B_2, \ldots,\ B_m \in [0, 1]^n$. Thus

$$Agg = \max \gg Sp.$$

Contrary to Proposition 4, the following one shows that there exists an aggregator which does not fulfill IBP with respect to any specificity.

Proposition 5 Consider $n, m \in \{2, 3, \ldots\}$ and the aggregator $Agg:[0, 1]^m \longrightarrow [0, 1]$ given by

$$Agg\ (b_1,\ b_2, \ldots,\ b_m) = \min\{b_1,\ b_2, \ldots,\ b_m\}.$$

Then *Agg* does not fulfill IBP with respect to any specificity $Sp: [0, 1]^n \longrightarrow [0, 1]$.

Proof: Let

$$B_1 = (1, 0, \ldots, 0)$$

$$B_2 = B_3 = \ldots = B_m = (1, ½, ½, \ldots, ½).$$

Because $Sp(B_1) = 1$, $Sp(B_2) < 1$ and

$$Agg\ (B_1, B_2, \ldots,\ B_m) = B_1$$

we have

$$Agg(Sp(B_1), Sp(B_2),..., Sp(B_m)) =$$

$$\min(Sp(B_1), Sp(B_2),..., Sp(B_m)) < 1$$

and

$$Sp(Agg\ (B_1, B_2, \ldots,\ B_m)) = Sp(B_1) = 1.$$

Agg does not dominate any specificity.

The problem of dominance $Agg \gg Sp$ evokes many questions. We formulate some of them:

Global open problem:

- to characterize aggregators $Agg: [0, 1]^m \longrightarrow [0, 1]$ and specificity measures $Sp: [0, 1]^n \longrightarrow [0, 1]$ fulfilling IBP for some (or all) $m, n \in \{2, 3, \ldots\}$.

Particular open problems

- to characterize aggregators $Agg:[0, 1]^m \longrightarrow [0, 1]$ which fulfills IBP with respect to linear specificity $Sp: [0, 1]^n \longrightarrow [0, 1]$

- to characterize the specificity $Sp : [0, 1]^n \longrightarrow [0, 1]$ for which the arithmetic mean $Amean :[0, 1]^m \longrightarrow [0, 1]$ fulfills IBP with respect to *Sp*.

- to prove or reject:

  - $Agg \gg Sp_1 \geq Sp_2 \Rightarrow Agg \gg Sp_2$

  - $Agg_1 \geq Agg_2 \gg Sp \Rightarrow Agg_1 \gg Sp$

Acknowledgements

This work was supported by grant VEGA 1/2005/05.

References

[1] P. Sarkoci and M. Šabo: Information boundedness principle in fuzzy inference process, *Kybernetika* 38 (2002), No. 3, 327-338.

[2] P. Sarkoci: Domination in the Families of Frank and Hamacher t-norms, *Kybernetika* 41 (2005), 349-360.

[3] R.M.Tardiff: On a Generalised Minkovski Inequality and Its Relation to Dominates for t-norms, Aequationes Math. 27(1984), 308-316.

[4] R. R. Yager: Global requirements for implication operators in fuzzy modus ponens, *Fuzzy Sets and Systems* 106 (1999), 3-10.

[5] R. R. Yager: Uninorms in fuzzy modeling, *Fuzzy Sets and Systems* 122 (2001), 167-175.

[6] R. R. Yager: Ordinal measure of specificity, *International Journal of General Systems* 17 (1990), 57-72.

Triple Rotation: Gymnastics for T-norms

K.C. Maes
Department of Applied Mathematics, Biometrics and Process Control,
Ghent University,
Coupure links 653,
B-9000 Gent, Belgium
Koen.Maes@Ugent.be

B. De Baets
Department of Applied Mathematics, Biometrics and Process Control,
Ghent University,
Coupure links 653,
B-9000 Gent, Belgium
Bernard.DeBaets@Ugent.be

Abstract

Given an involutive negator N and a left-continuous t-norm T whose contour line C_0 is continuous on $]0,1]$, we build a rotation-invariant t-norm from a rescaled version of T and its left, right and front rotation. Depending on the involutive negator N and the set of zero divisors of T, some reshaping of the rescaled version of T may occur during the rotation process. The rescaled version of T itself can be understood as the β-zoom of the newly constructed rotation-invariant t-norm, with β the unique fixpoint of N.

Keywords: Triple Rotation Method, Rotation-invariant T-norm, Contour Line, Companion, Zoom.

1 Introduction

Studying the structure of an increasing $[0,1]^2 \to [0,1]$ function T, the use of contour lines and zooms has proven to be very fruitful (see e.g. [8, 13, 14, 15, 16]). Also, the companion is indispensable to describe for example rotation-invariant t-norms (see e.g. [14, 16]). Note that we use the standard notations for the prototypical t-norms and t-conorms [9, 10].

Contour lines

Contour lines of an increasing $[0,1]^2 \to [0,1]$ function T are defined as the lower, upper, left or right limits of its horizontal cuts, *i.e.* the intersections of its graph by planes parallel to the domain $[0,1]^2$. Although there exist four different types of contour lines, those determined by the upper limits of the horizontal cuts are of particular interest for the study of rotation-invariant t-norms [14].

Definition 1 [13] *Let $a \in [0,1]$. The* contour line *C_a of an increasing $[0,1]^2 \to [0,1]$ function T is the $[0,1] \to [0,1]$ function defined by*

$$C_a(x) = \sup\{t \in [0,1] \mid T(x,t) \leq a\}\,. \tag{1}$$

For a left-continuous t-norm T, the contour line C_a equals the partial function $I_T(\bullet, a)$ of the *residual implicator* I_T (see e.g. [4]). In particular, the zero contour line C_0 then coincides with the *residual negator* $N_T = I_T(\bullet, 0)$. Contour lines of a continuous t-norm T are also called *level functions* [11].

The companion

A second useful tool to study an increasing $[0,1]^2 \to [0,1]$ function T is its companion Q.

Definition 2 [14] *The* companion *Q of an increasing $[0,1]^2 \to [0,1]$ function T is the $[0,1]^2 \to [0,1]$ function defined by*

$$Q(x,y) = \sup\{t \in [0,1] \mid C_t(x) \leq y\}\,.$$

We have shown in [14] that $Q(x,y) = \inf\{T(x,u) \mid u \in]y,1]\}$ (with $\inf \emptyset = 1$). This property allows to construct the graph of Q from the graph of T. Clearly, $Q(x,y) = T(x,y)$ whenever $T(x,\bullet)$ is right continuous in $y \in [0,1[$. Every left-continuous, increasing $[0,1]^2 \to [0,1]$ function T that has absorbing element 0 is determined by its companion Q.

Zooms

Finally, every increasing $[0,1]^2 \to [0,1]$ function T is trivially described by its associated set of zooms.

Definition 3 [16] *Let T be an increasing $[0,1]^2 \to [0,1]$ function and take $(a,b) \in [0,1]^2$ such that $a < b$ and $T(b,b) \leq b$. Consider an $[a,b] \to [0,1]$ isomorphism σ. The (a,b)-*zoom *$T^{(a,b)}$ of T is the $[0,1]^2 \to [0,1]$ function defined by*

$$T^{(a,b)}(x,y) = \sigma\left[\max\left(a, T(\sigma^{-1}[x], \sigma^{-1}[y])\right)\right]\,.$$

*If $b = 1$ we simply talk about the a-*zoom *T^a of T.*

The graph of $T^{(a,b)}$ is determined by the rescaling of the set $\{(x, y, T(x, y)) \mid (x, y) \in [a, b]^2 \wedge a < T(x, y)\}$ (zoom in) into the unit cube (zoom out). Whenever $T(b, b) \leq a$, the function $T^{(a,b)}$ is trivially constant: $T^{(a,b)}(x, y) = a$, for every $(x, y) \in [0, 1]^2$. For $b = 1$ the boundary condition $T(1, 1) \leq 1$ is always fulfilled such that the a-zoom of T is defined for every $a < 1$.

Zooms are extremely suited to study an increasing function T that satisfies $T \leq T_{\mathbf{M}}$. The restriction $T(b, b) \leq b$ (Definition 3) then trivially holds. Recall that a t-subnorm T is a $[0, 1]^2 \to [0, 1]$ function satisfying all t-norm properties but the neutral element. Instead $T \leq T_{\mathbf{M}}$ must hold [7].

Theorem 1 [16] *Consider $(a, b) \in [0, 1]^2$ such that $a < b$. Then the (a, b)-zoom of a t-subnorm is a t-subnorm and the a-zoom of a t-norm is a t-norm.*

2 The triple rotation method

It is well known that left-continuous t-norms ensure the definability of the t-norm-based residual implicator. Therefore they are of great interest to people working on monoidal t-norm based logic (MTL logic) [2] and involutive monoidal t-norm based logic (IMTL logic) [1, 12]. The latter requires the involutivity of the residual negator $N_T = C_0$. We have shown [13, 14] that the involutivity of C_0 is equivalent with its continuity and with the rotation invariance of the left-continuous t-norm T considered.

Definition 4 [5] *Let N be an involutive negator. An increasing $[0, 1]^2 \to [0, 1]$ function T is called* rotation invariant w.r.t. an involutive negator N *(*i.e. *an involutive decreasing $[0, 1] \to [0, 1]$ function) if for every $(x, y, z) \in [0, 1]^3$ it holds that*

$$T(x, y) \leq z \;\Leftrightarrow\; T(y, z^N) \leq x^N \,. \tag{2}$$

Jenei has proven that every t-norm T that is rotation invariant w.r.t. an involutive negator N is necessarily left continuous and $N_T = N$ [5]. Therefore, we briefly call a t-norm *rotation invariant* if it is left continuous and has a continuous contour line C_0. Note that the continuity of C_0 does not necessarily imply the left continuity of T [14].

Based on contour lines, the companion and zooms, we have presented in [16] a natural method for decomposing a rotation-invariant t-norm T. In case the contour line C_β of T, with β the unique fixpoint of C_0, is continuous on $]\beta, 1]$, there exists a unique decomposition. In this contribution we transform our decomposition method into a straightforward construction tool for rotation-invariant t-norms. The presented results extend our work from [15] and comprise to a large extent the rotation and rotation-annihilation construction of Jenei [7]. We assume the following setting:

- T: an arbitrary left-continuous t-norm (with contour lines C_a and companion Q) such that C_0 is continuous on $]0, 1]$ and Q is commutative on $[0, \alpha[^2$, with $\alpha = \inf\{t \in [0, 1] \mid C_0(t) = 0\}$.
- N: an arbitrary involutive negator with fixpoint β.
- σ: an arbitrary $[\beta, 1] \to [0, 1]$ isomorphism.
- M: the decreasing $[0, 1] \to [0, \beta]$ function defined by $x^M = 1$ whenever $x \in [0, \beta[$ and $x^M = \sigma^{-1}[C_0(\sigma[x])]$ whenever $x \in [\beta, 1]$.
- $\mathcal{D}$: the area $\{(x, y) \in [0, 1]^2 \mid x^N < y\} = \mathcal{D}_{\mathrm{I}} \cup \mathcal{D}_{\mathrm{II}} \cup \mathcal{D}_{\mathrm{III}} \cup \mathcal{D}_{\mathrm{IV}}$, with

$$\begin{aligned}
\mathcal{D}_{\mathrm{I}} &= \{(x, y) \in]\beta, 1]^2 \mid x^M < y\}\,, \\
\mathcal{D}_{\mathrm{II}} &= \{(x, y) \in]0, \beta] \times]\beta, 1] \mid x^N < y\}\,, \\
\mathcal{D}_{\mathrm{III}} &= \{(x, y) \in]\beta, 1] \times]0, \beta] \mid x^N < y\}\,, \\
\mathcal{D}_{\mathrm{IV}} &= \{(x, y) \in]\beta, 1[^2 \mid y \leq x^M\}\,.
\end{aligned}$$

Note that the choice of T, N and σ fixes M and $\mathcal{D}$.

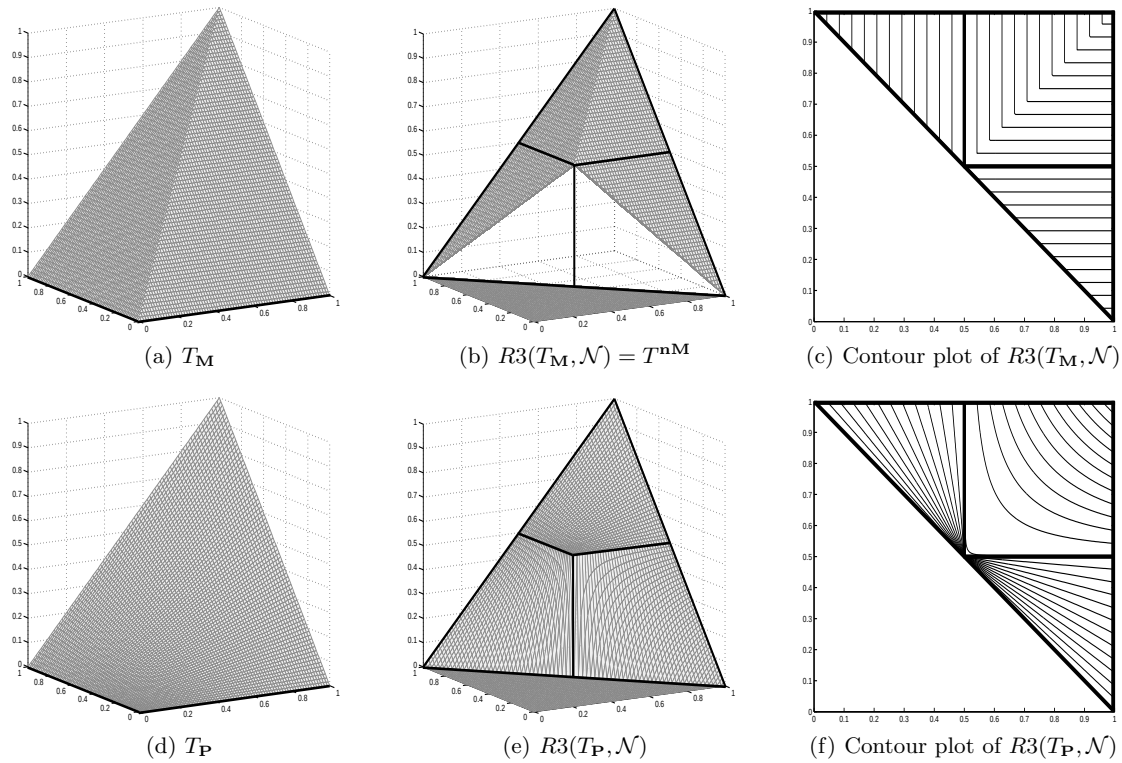
Theorem 2 *The $[0, 1]^2 \to [0, 1]$ function $R3(T, N)$ defined by $R3(T, N)(x, y) =$*

$$\begin{cases}
\sigma^{-1}\left[T(\sigma[x], \sigma[y])\right], & \text{if } (x, y) \in \mathcal{D}_{\mathrm{I}}\,, \\
\left(\sigma^{-1}\left[C_{\sigma[x^N]}(\sigma[y])\right]\right)^N, & \text{if } (x, y) \in \mathcal{D}_{\mathrm{II}}\,, \\
\left(\sigma^{-1}\left[C_{\sigma[y^N]}(\sigma[x])\right]\right)^N, & \text{if } (x, y) \in \mathcal{D}_{\mathrm{III}}\,, \\
\left(\sigma^{-1}\left[Q\big(C_0(\sigma[x]), C_0(\sigma[y])\big)\right]\right)^N, & \text{if } (x, y) \in \mathcal{D}_{\mathrm{IV}}\,, \\
0, & \text{if } (x, y) \notin \mathcal{D}\,,
\end{cases} \tag{3}$$

is a rotation-invariant t-norm. $R3(T, N)$ is the only left-continuous t-norm that has N as a contour line ($a = 0$) and has β-zoom $R3(T, N)^\beta = T$.

In [15] we showed that $R3(T, N)|_{\mathcal{D}_{\mathrm{II}}}$ and $R3(T, N)|_{\mathcal{D}_{\mathrm{III}}}$ are determined by the (transformed) left and right rotation of $R3(T, N)|_{\mathcal{D}_{\mathrm{I}}}$ around the axis through the points $(0, 0, 1)$ and $(1, 1, 0)$. As will become clear from the examples, $R3(T, N)|_{\mathcal{D}_{\mathrm{IV}}}$ is determined by the (transformed) front rotation of $R3(T, N)|_{\mathcal{D}_{\mathrm{I}} \cap]\beta, \sigma^{-1}(\alpha)]^2}$ around the axis through the points $(\beta, \sigma^{-1}[\alpha], \beta)$ and $(\sigma^{-1}[\alpha], \beta, \beta)$. Note also that $R3(T, N)|_{\mathcal{D}_{\mathrm{I}}}$ is a rescaled version of 'the non-zero part' of T. Inspired by these geometrical observations, we briefly call $R3(T, N)$ the *triple rotation of T based on N*. The construction method itself is referred to as the *triple rotation method*.

For the following examples we use the linear rescaling function $\varsigma : x \mapsto (x - \beta)/(1 - \beta)$. Any other rescaling function entails a transformation of the procured t-norm.



(a) $T_{\mathbf{M}}$ (b) $R3(T_{\mathbf{M}},\mathcal{N}) = T^{\mathbf{nM}}$ (c) Contour plot of $R3(T_{\mathbf{M}},\mathcal{N})$

(d) $T_{\mathbf{P}}$ (e) $R3(T_{\mathbf{P}},\mathcal{N})$ (f) Contour plot of $R3(T_{\mathbf{P}},\mathcal{N})$

Figure 1: The triple rotation of $T_{\mathbf{M}}$ and $T_{\mathbf{P}}$ based on $\mathcal{N}$.

A first class of examples is obtained by considering those t-norms T that have no zero-divisors (*i.e.* $0 < T(x,y)$, for every $(x,y) \in]0,1]^2$). In this case $\alpha = 0$, $\sigma^{-1}[\alpha] = \beta$ and $\mathcal{D}_{\mathrm{IV}} = \emptyset$. The triple rotation method then coincides with the rotation construction of Jenei [6]. In Fig. 1, for example, we apply the triple rotation method to the minimum operator $T_{\mathbf{M}}$ and the algebraic product $T_{\mathbf{P}}$. The triple rotation $R3(T_{\mathbf{M}},\mathcal{N})$ of $T_{\mathbf{M}}$ based on the standard negator $\mathcal{N}$ equals the nilpotent minimum $T^{\mathbf{nM}}$. The bold black lines in Figs. 1(a) and 1(d) indicate the corresponding zero contour lines. The bold black lines in all other subfigures visualize the partition $\mathcal{D} = \mathcal{D}_{\mathrm{I}} \cup \mathcal{D}_{\mathrm{II}} \cup \mathcal{D}_{\mathrm{III}} \cup \mathcal{D}_{\mathrm{IV}}$.

Secondly, the triple rotation method can be performed on most of the rotation-invariant t-norms. In this case $\alpha = 1 = \sigma^{-1}[\alpha]$ [13] and we can rewrite Eq. (3) in a more feasible form [14]: $R3(T,N)(x,y) =$

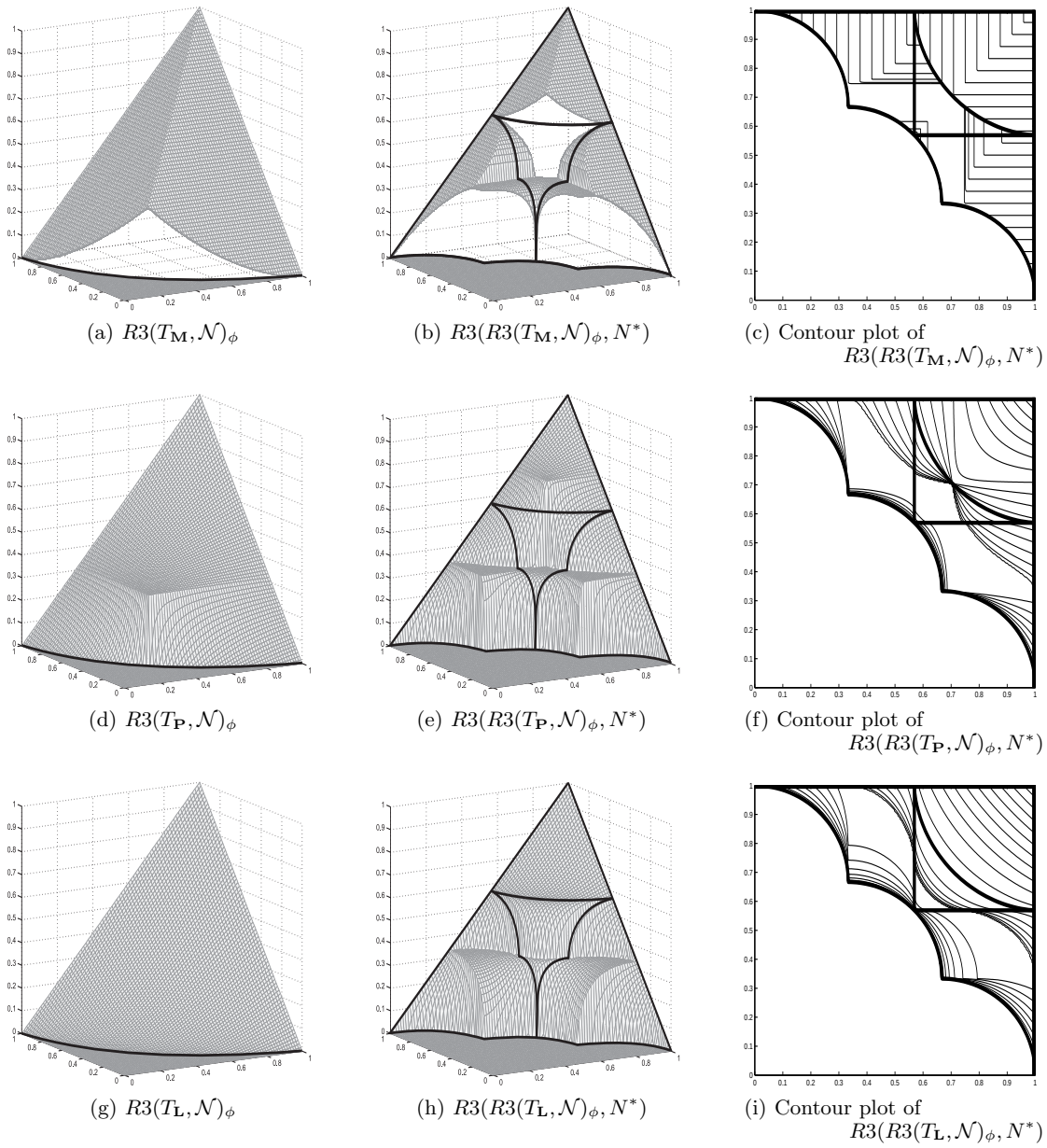
$$\begin{cases} \sigma^{-1}\left[T(\sigma[x],\sigma[y])\right], & \text{if } (x,y) \in \mathcal{D}_{\mathrm{I}}\,, \\ \left(\sigma^{-1}\left[C_0\left(T\big(C_0(\sigma[x^N]),\sigma[y]\big)\right)\right]\right)^N, & \text{if } (x,y) \in \mathcal{D}_{\mathrm{II}}\,, \\ \left(\sigma^{-1}\left[C_0\left(T\big(\sigma[x],C_0(\sigma[y^N])\big)\right)\right]\right)^N, & \text{if } (x,y) \in \mathcal{D}_{\mathrm{III}}\,, \\ \left(\sigma^{-1}\left[Q\big(C_0(\sigma[x]),C_0(\sigma[y])\big)\right]\right)^N, & \text{if } (x,y) \in \mathcal{D}_{\mathrm{IV}}\,, \\ 0, & \text{if } (x,y) \notin \mathcal{D}\,. \end{cases}$$

However, for the triple rotation method to yield a t-norm it is absolutely necessary that the companion Q of T is commutative on $[0,1[^2$ [15]. The rotation-invariant t-norms depicted in Figs. 2(a), 2(d) and 2(g) satisfy this mandatory condition. They are the *ϕ-transforms* of the triple rotations $R3(T_{\mathbf{M}},\mathcal{N})$, $R3(T_{\mathbf{P}},\mathcal{N})$ and $R3(T_{\mathbf{L}},\mathcal{N})$, with ϕ the automorphism defined by $\phi(x) := x^{3/5}$. Recall that the ϕ-transform T_ϕ of a t-norm T is the t-norm defined by $T_\phi(x,y) = \phi^{-1}[T(\phi[x],\phi[y])]$. Let N^* be the involutive negator defined by

$$x^{N^*} := \begin{cases} \frac{2}{3} + \sqrt{\frac{1}{9} - x^2}\,, & \text{if } x \in [0,\frac{1}{3}] \\ \frac{1}{3} + \sqrt{\frac{1}{9} - (x-\frac{1}{3})^2}\,, & \text{if } x \in [\frac{1}{3},\frac{2}{3}] \\ \sqrt{\frac{1}{9} - (x-\frac{2}{3})^2}\,, & \text{if } x \in [\frac{2}{3},1]\,. \end{cases} \tag{4}$$

In Figs. 2(b), 2(e) and 2(h) we apply the triple rotation method based on N^* to the ϕ-transforms $R3(T_{\mathbf{M}},\mathcal{N})_\phi$, $R3(T_{\mathbf{P}},\mathcal{N})_\phi$ and $R3(T_{\mathbf{L}},\mathcal{N})_\phi$. The obtained rotation-invariant t-norms cannot be described by the rotation construction nor by the rotation-annihilation construction of Jenei [15].

As can be seen in Figs. 2(b), 2(e) and 2(h), if $N \neq \mathcal{N}$ or $C_0 \neq \mathcal{N}$, the left, right and front rotation of $R3(T,N)|_{\mathcal{D}_{\mathrm{I}}}$ have to be reshaped to



(a) $R3(T_{\mathbf{M}}, \mathcal{N})_\phi$ (b) $R3(R3(T_{\mathbf{M}}, \mathcal{N})_\phi, N^*)$ (c) Contour plot of $R3(R3(T_{\mathbf{M}}, \mathcal{N})_\phi, N^*)$

(d) $R3(T_{\mathbf{P}}, \mathcal{N})_\phi$ (e) $R3(R3(T_{\mathbf{P}}, \mathcal{N})_\phi, N^*)$ (f) Contour plot of $R3(R3(T_{\mathbf{P}}, \mathcal{N})_\phi, N^*)$

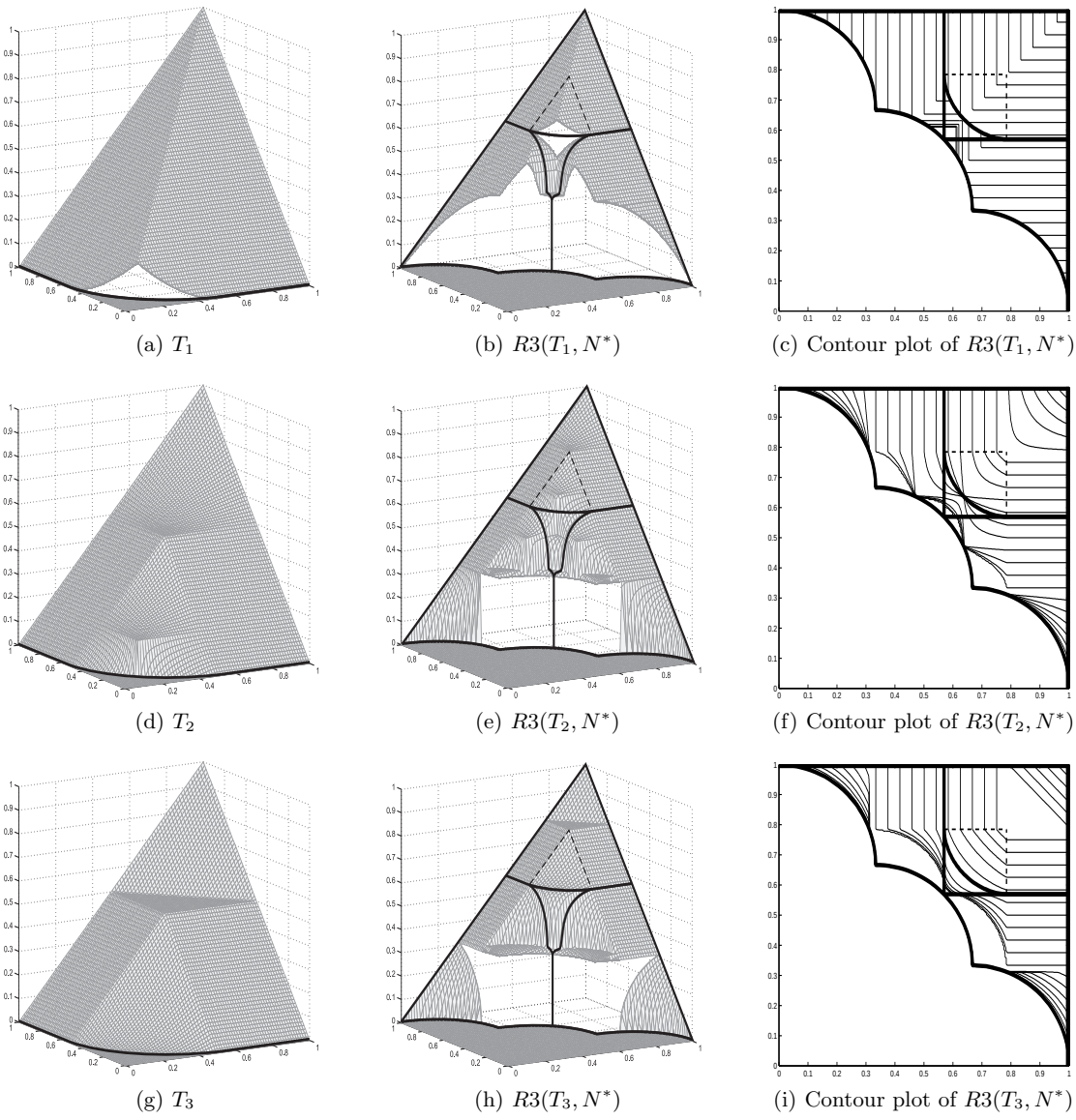
(g) $R3(T_{\mathbf{L}}, \mathcal{N})_\phi$ (h) $R3(R3(T_{\mathbf{L}}, \mathcal{N})_\phi, N^*)$ (i) Contour plot of $R3(R3(T_{\mathbf{L}}, \mathcal{N})_\phi, N^*)$

Figure 2: The triple rotation of $R3(T_{\mathbf{M}}, \mathcal{N})_\phi$, $R3(T_{\mathbf{P}}, \mathcal{N})_\phi$ and $R3(T_{\mathbf{L}}, \mathcal{N})_\phi$ based on N^*.

fit into the areas $\mathcal{D}_{\mathrm{II}}$, $\mathcal{D}_{\mathrm{III}}$ and $\mathcal{D}_{\mathrm{IV}}$, respectively. The involutive negator N and the contour line C_0 of T are responsible for this reshaping. Note also that in general $R3(T, N)(\beta, \bullet) = N \circ M = R3(T, N)(\bullet, \beta)$. Therefore, the t-norms $R3(T, N^*)$ visualized in Figs. 2(b), 2(e) and 2(h) have identical partial functions $R3(T, N^*)(\bullet, \beta) = R3(T, N^*)(\beta, \bullet)$, with $\beta = \frac{1}{3} + \frac{1}{\sqrt{18}}$ the fixpoint of N^*. Indeed, their associated functions M ($C_0 = \mathcal{N}_\phi$) and $N = N^*$ are identical.

Finally, if $\alpha \in]0, 1[$, then T is necessarily an ordinal sum of a rotation-invariant t-norm whose companion is commutative on $[0, 1[^2$ and an arbitrary left-continuous t-norm. The latter largely follows from the following characterization.

Theorem 3 [16] *Consider a left-continuous t-norm T and take $a \in [0, 1]$ such that $a < \theta := \inf\{t \in [0, 1] \mid C_a(t) = a\}$. Then the following assertions are equivalent:*



(a) T_1 (b) $R3(T_1, N^*)$ (c) Contour plot of $R3(T_1, N^*)$

(d) T_2 (e) $R3(T_2, N^*)$ (f) Contour plot of $R3(T_2, N^*)$

(g) T_3 (h) $R3(T_3, N^*)$ (i) Contour plot of $R3(T_3, N^*)$

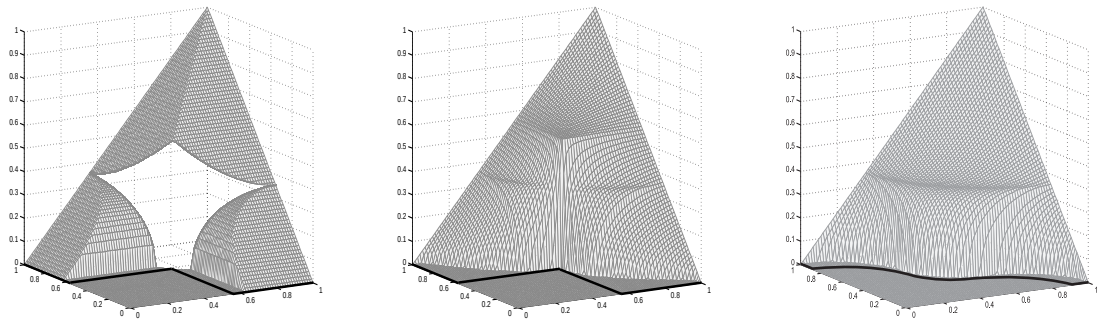
Figure 3: The triple rotation of T_1, T_2 and T_3 based on N^*.

1. *C_a is continuous on $]a, 1]$.*
2. *C_a is involutive on $]a, \theta[$.*
3. *$C_a(]a, \alpha[) =]a, \theta[$.*
4. *$T^{(a,\theta)}$ is a rotation-invariant t-norm.*

In Fig. 3 we present the triple rotation of the ordinal sums

$$T_1 := \left(\left\langle 0, \tfrac{1}{2}, R3(T_{\mathbf{M}}, \mathcal{N})_\phi \right\rangle, \left\langle \tfrac{1}{2}, 1, T_{\mathbf{M}} \right\rangle\right)$$
$$T_2 := \left(\left\langle 0, \tfrac{1}{2}, R3(T_{\mathbf{P}}, \mathcal{N})_\phi \right\rangle, \left\langle \tfrac{1}{2}, 1, T_{\mathbf{P}} \right\rangle\right)$$
$$T_3 := \left(\left\langle 0, \tfrac{1}{2}, R3(T_{\mathbf{L}}, \mathcal{N})_\phi \right\rangle, \left\langle \tfrac{1}{2}, 1, T_{\mathbf{L}} \right\rangle\right),$$

based on the involutive negator N^*. Note that here $\alpha = \frac{1}{2}$. For the t-norms $R3(T, N^*)$ visualized in Figs. 3(b), 3(e) and 3(h) it clearly holds that $R3(T, N^*)|_{\mathcal{D}_{\mathrm{IV}}}$ can be understood as a reshaped front rotation of $R3(T, N^*)|_{\mathcal{D}_{\mathrm{I}} \cap]\beta, \varsigma^{-1}(\frac{1}{2})]^2}$, with β the fixpoint of N^*. The dashed lines in the figures indicate the area $\mathcal{D}_{\mathrm{I}} \cap]\beta, \varsigma^{-1}(\frac{1}{2})]^2$. The zooms $(R3(T, N^*))^{((\varsigma^{-1}[\frac{1}{2}])^{N^*}, \varsigma^{-1}[\frac{1}{2}])}$ of these three t-norms $R3(T, N^*)$ (with $T \in \{T_1, T_2, T_3\}$) are rotation-invariant t-norms, obtained by performing the triple rotation method on the rotation-invariant t-norms $(R3(T, N^*))^{(\beta, \varsigma^{-1}[\frac{1}{2}])} = T^{(0,\frac{1}{2})}$. For this latter construction the involutive negator $\breve{\varsigma} \circ N^* \circ \breve{\varsigma}^{-1}$ is used, with $\breve{\varsigma}$ the linear rescaling function from $[(\varsigma^{-1}(\frac{1}{2}))^{N^*}, \varsigma^{-1}(\frac{1}{2}))]$ to $[0, 1]$.



(a) $R3(R3(T_{\mathbf{M}},\mathcal{N})_\phi,N^*)^a$ (b) $R3(R3(T_{\mathbf{P}},\mathcal{N})_\phi,N^*)^a$ (c) $R3(R3(T_{\mathbf{L}},\mathcal{N})_\phi,N^*)^a$

Figure 4: The a-zooms of $R3(R3(T_{\mathbf{M}},\mathcal{N})_\phi,N^*)$, $R3(R3(T_{\mathbf{P}},\mathcal{N})_\phi,N^*)$ and $R3(R3(T_{\mathbf{L}},\mathcal{N})_\phi,N^*)$, with $a = (\varsigma^{-1}[\phi^{-1}(\frac{1}{2})])^{N^*}$.

Besides the triple rotation method we can also use zooms to build new t-norms. Let a be the height of the two lowest 'jumps' in Fig. 2(e). In Fig. 4 we visualize the a-zooms of the t-norms depicted in Figs. 2(b), 2(e) and 2(h). The bold black lines indicate the corresponding zero contour lines.

References

[1] F. Esteva, J. Gispert, L. Godo, and F. Montagna, *On the standard and rational completeness of some axiomatic extensions of the monoidal t-norm logic*, Studia Logica **71** (2002), 199–226.

[2] F. Esteva and L. Godo, *Monoidal t-norm based fuzzy logic: towards a logic for left-continuous t-norms*, Fuzzy Sets and Systems **124** (2001), 271–288.

[3] J. Fodor, *A new look at fuzzy connectives*, Fuzzy Sets and Systems **57** (1993), 141–148.

[4] J. Fodor and M. Roubens, *Fuzzy Preference Modelling and Multicriteria Decision Support*, Kluwer Academic Publishers, 1994.

[5] S. Jenei, *Geometry of left-continuous t-norms with strong induced negations*, Belg. J. Oper. Res. Statist. Comput. Sci. **38** (1998), 5–16.

[6] S. Jenei, *A characterization theorem on the rotation construction for triangular norms*, Fuzzy Sets and Systems **136** (2003), 283–289.

[7] S. Jenei, *How to construct left-continuous triangular norms - state of the art*, Fuzzy Sets and Systems **143** (2004), 27–45.

[8] S. Jenei, *On the geometry of associativity*, Semigroup Forum, to appear.

[9] E.P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*, Trends in Logic, Vol. 8, Kluwer Academic Publishers, 2000.

[10] E.P. Klement, R. Mesiar, and E. Pap, *Triangular norms. Position paper I: basic analytical and algebraic properties*, Fuzzy Sets and Systems **143** (2004), 5–26.

[11] E.P. Klement, R. Mesiar, and E. Pap, *Different types of continuity of triangular norms revisited*, New Mathematics and Natural Computation **1** (2005), 195–211.

[12] L. Lianzhen and L. Kaitai, *Involutive monoidal t-norm based logic and R_0 logic*, Internat. J. Intell. Systems **19** (2004), 491–497.

[13] K.C. Maes and B. De Baets, *A contour view on uninorm properties*, Kybernetika **42** (3) (2006), 303–318.

[14] K.C. Maes and B. De Baets, *On the structure of left-continuous t-norms that have a continuous contour line*, Fuzzy Sets and Systems **158** (2007), 843–860.

[15] K.C. Maes and B. De Baets, *The triple rotation method for constructing rotation-invariant t-norms*, Fuzzy Sets and Systems, in press.

[16] K.C. Maes and B. De Baets, *Advances in the geometrical study of rotation-invariant t-norms*, Lecture Notes in Computer Science (2007), in press.

Representations of Archimedean t-norms in interval-valued fuzzy set theory

Glad Deschrijver
Fuzziness and Uncertainty Modeling Research Unit
Department of Applied Mathematics and Computer Science, Ghent University
Krijgslaan 281 (S9), B-9000 Gent, Belgium
Glad.Deschrijver@UGent.be
http://www.fuzzy.UGent.be/

Abstract

In this paper the Archimedean property and the nilpotency of t-norms on the lattice $\mathcal{L}^I$ is investigated, where $\mathcal{L}^I$ is the underlying lattice of interval-valued fuzzy set theory (Sambuc, 1975) and intuitionistic fuzzy set theory (Atanassov, 1983). We give some characterizations of continuous t-norms on $\mathcal{L}^I$ which satisfy the residuation principle, $\mathcal{T}(D,D) \subseteq D$, the Archimedean property and nilpotency. **Keywords:** interval-valued fuzzy set, intuitionistic fuzzy set, t-norm, Archimedean, nilpotent, strict, representation.

1 Introduction

Triangular norms on $[0,1]$ were introduced in [18] and play an important role in fuzzy set theory (see e.g. [9, 12, 13] for more details). One of the most important properties that can be satisfied by t-norms on the unit interval is the Archimedean property, for example continuous t-norms can be fully characterized by means of Archimedean t-norms, the Archimedean property is closely related to additive and multiplicative generators [13, 15, 16].

Interval-valued fuzzy set theory [11, 17] is an extension of fuzzy theory in which to each element of the universe a closed subinterval of the unit interval is assigned which approximates the unknown membership degree. Another extension of fuzzy set theory is intuitionistic fuzzy set theory introduced by Atanassov [1]. Intuitionistic fuzzy sets assign to each element of the universe not only a membership degree, but also a non-membership degree which is less than or equal to 1 minus the membership degree (in fuzzy set theory the non-membership degree is always equal to 1 minus the membership degree). In [7] it is shown that intuitionistic fuzzy set theory is equivalent to interval-valued fuzzy set theory and that both are equivalent to L-fuzzy set theory in the sense of Goguen [10] w.r.t. a special lattice $\mathcal{L}^I$.

In this paper we will investigate the nilpotency property and we will give some characterizations of continuous t-norms on $\mathcal{L}^I$ which satisfy the residuation principle, $\mathcal{T}(D,D) \subseteq D$ and the Archimedean property.

2 Preliminary definitions

The underlying lattice $\mathcal{L}^I$ of interval-valued fuzzy set theory is given as follows.

Definition 2.1 *We define* $\mathcal{L}^I = (L^I, \leq_{L^I})$, *where*

$$L^I = \{[x_1,x_2] \mid (x_1,x_2) \in [0,1]^2 \text{ and } x_1 \leq x_2\},$$
$$[x_1,x_2] \leq_{L^I} [y_1,y_2] \iff (x_1 \leq y_1 \text{ and } x_2 \leq y_2),$$
for all $[x_1,x_2],[y_1,y_2]$ *in* L^I.

Similarly as Lemma 2.1 in [7] it is shown that $\mathcal{L}^I$ is a complete lattice.

Let U be a universal set.

Definition 2.2 [11, 17] *An interval-valued fuzzy set on* U *is a mapping* $A : U \to L^I$.

Definition 2.3 [1] *An intuitionistic fuzzy set on* U *is a set*

$$A = \{(u, \mu_A(u), \nu_A(u)) \mid u \in U\},$$

where $\mu_A(u) \in [0,1]$ denotes the membership degree and $\nu_A(u) \in [0,1]$ the non-membership degree of u in A and where for all $u \in U$, $\mu_A(u) + \nu_A(u) \leq 1$.

An intuitionistic fuzzy set A on U can be represented by the $\mathcal{L}^I$-fuzzy set A given by

$$\begin{array}{rcl} A : U & \to & L^I : \\ u & \mapsto & [\mu_A(u), 1-\nu_A(u)], \quad \forall u \in U \end{array}$$

In Figure 1 the set L^I is shown. Note that to any element $x = [x_1, x_2]$ of L^I there corresponds a point $(x_1, x_2) \in \mathbb{R}^2$.

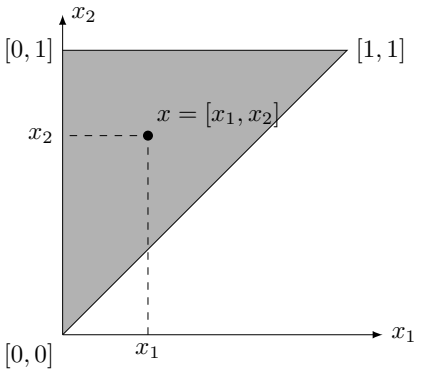


Figure 1: The grey area is L^I.

In the sequel, if $x \in L^I$, then we denote its bounds by x_1 and x_2, i.e. $x = [x_1, x_2]$. The smallest and the largest element of $\mathcal{L}^I$ are given by $0_{\mathcal{L}^I} = [0,0]$ and $1_{\mathcal{L}^I} = [1,1]$. We define for further usage the set $D = \{[x_1, x_1] \mid x_1 \in [0,1]\}$. Note that, for x, y in L^I, $x <_{L^I} y$ is equivalent to "$x \leq_{L^I} y$ and $x \neq y$", i.e. either $x_1 < y_1$ and $x_2 \leq y_2$, or $x_1 \leq y_1$ and $x_2 < y_2$. We denote by $x \ll_{L^I} y$: $x_1 < y_1$ and $x_2 < y_2$.

Definition 2.4 *A t-norm on a complete lattice $\mathcal{L} = (L, \leq_L)$ is a commutative, associative, increasing mapping $\mathcal{T} : L^2 \to L$ which satisfies $\mathcal{T}(1_{\mathcal{L}}, x) = x$, for all $x \in L$.*

Let $\mathcal{T}$ be a t-norm on a complete lattice $\mathcal{L} = (L, \leq_L)$ and $x \in L$, then we denote $x^{(n)_{\mathcal{T}}} = \mathcal{T}(x, x^{(n-1)_{\mathcal{T}}})$, for $n \in \mathbb{N} \setminus \{0,1\}$, and $x^{(1)_{\mathcal{T}}} = x$.

We say that a t-norm $\mathcal{T}$ on $\mathcal{L}$ satisfies the residuation principle if and only if, for all x, y, z in L,

$$\mathcal{T}(x,y) \leq_L z \iff y \leq_L \mathcal{I}_{\mathcal{T}}(x,z),$$

where $\mathcal{I}_{\mathcal{T}}$ is the residual implication of $\mathcal{T}$ defined by $\mathcal{I}_{\mathcal{T}}(x,z) = \sup\{\gamma \mid \gamma \in L \text{ and } \mathcal{T}(x,\gamma) \leq_L z\}$, for all x, z in L.

For t-norms on $\mathcal{L}^I$, we consider the following special classes.

Definition 2.5 *A t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is called t-representable if there exist t-norms T_1 and T_2 on $([0,1], \leq)$ such that $T_1(x,y) \leq T_2(x,y)$, for all x, y in $[0,1]$, and such that, for all x, y in L^I,*

$$\mathcal{T}(x,y) = [T_1(x_1, y_1), T_2(x_2, y_2)].$$

Then T_1 and T_2 are called the representants of $\mathcal{T}$ and we denote $\mathcal{T}$ by $\mathcal{T}_{T_1,T_2}$.

A t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is called pseudo-t-representable if there exists a t-norm T on $([0,1], \leq)$ such that, for all x, y in L^I,

$$\mathcal{T}(x,y) = [T(x_1, y_1), \max(T(x_1, y_2), T(x_2, y_1))].$$

Then T is called the representant of $\mathcal{T}$ and we denote $\mathcal{T}$ by $\mathcal{T}_T$.

Definition 2.6 *A negation on a complete lattice $\mathcal{L} = (L, \leq_L)$ is a decreasing mapping $\mathcal{N} : L \to L$ for which $\mathcal{N}(0_{\mathcal{L}}) = 1_{\mathcal{L}}$ and $\mathcal{N}(1_{\mathcal{L}}) = 0_{\mathcal{L}}$. If $\mathcal{N}(\mathcal{N}(x)) = x$, for all $x \in L$, then $\mathcal{N}$ is called involutive.*

Let $\mathcal{I}_{\mathcal{T}}$ be the residual implication of a t-norm $\mathcal{T}$ on $\mathcal{L}$. The mapping $\mathcal{N}_{\mathcal{I}_{\mathcal{T}}} : L \to L$ defined by $\mathcal{N}_{\mathcal{I}_{\mathcal{T}}}(x) = \mathcal{I}_{\mathcal{T}}(x, 0_{\mathcal{L}})$, for all $x \in L$, is a negation on $\mathcal{L}$, called the negation generated by $\mathcal{I}_{\mathcal{T}}$.

Theorem 2.1 [6] *Let $\mathcal{N}$ be a negation on $\mathcal{L}^I$. Then $\mathcal{N}$ is involutive if and only if there exists an involutive negation N on $([0,1], \leq)$ such that, for all $x \in L^I$, $\mathcal{N}(x) = [N(x_2), N(x_1)]$.*

Let $n \in \mathbb{N} \setminus \{0\}$. If for a mapping $f : [0,1]^n \to [0,1]$ and a mapping $F : (L^I)^n \to L^I$ it holds that $F(D^n) \subseteq D$, and $F([a_1, a_1], \ldots, [a_n, a_n]) = [f(a_1, \ldots, a_n), f(a_1, \ldots, a_n)]$, for all $a_1, \ldots, a_n$ in L^I, then we say that F is a natural extension of f to L^I. E.g. for any t-norm T on $([0,1], \leq)$, the t-norms $\mathcal{T}_{T,T}$ and $\mathcal{T}_T$ are natural extensions of T to L^I; if $\mathcal{N}$ is an involutive negation on $\mathcal{L}^I$, then from Theorem 2.1 it follows that there exists an involutive negation N on $([0,1], \leq)$ such that $\mathcal{N}$ is a natural extension of N.

3 The Archimedean property for t-norms on $([0,1],\leq)$

Denote $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$. Then Archimedean t-norms are defined as follows.

Definition 3.1 [13, 14] *Let T be a t-norm on $([0,1],\leq)$. We say that T is Archimedean if*

$$(\forall (x,y) \in]0,1[^2)(\exists n \in \mathbb{N}^*)(x^{(n)_T} < y).$$

An element $x \in]0,1[$ is called a nilpotent element of T if there exists an $n \in \mathbb{N}^*$ such that $x^{(n)_T} = 0$; x is called a zero-divisor of T if there exists an $y \in]0,1[$ such that $T(x,y) = 0$. A t-norm T on $([0,1],\leq)$ is called nilpotent if it is continuous and if each $x \in]0,1[$ is a nilpotent element of T; a t-norm T on $([0,1],\leq)$ is called strict if T is continuous and strictly increasing on $]0,1]^2$.

Theorem 3.1 [13] *Let T be a continuous Archimedean t-norm on $([0,1],\leq)$. Then the following are equivalent:*

(i) *T is nilpotent;*

(ii) *there exists some nilpotent element of T in $]0,1[$;*

(iii) *there exists some zero divisor of T in $]0,1[$;*

(iv) *T is not strict.*

For example, the product t-norm T_P on $([0,1],\leq)$ defined by $T_P(x,y) = xy$, for all x, y in $[0,1]$, is a strict t-norm, and the Łukasiewicz t-norm T_W defined by $T_W(x,y) = \max(0, x + y - 1)$, for all x, y in $[0,1]$, is a nilpotent t-norm.

Theorem 3.2 [13] *Let T be a t-norm on $([0,1],\leq)$.*

- *T is continuous, Archimedean and nilpotent if and only if there exists an increasing permutation φ of $([0,1],\leq)$ such that T is the φ-transform of T_W, i.e. $T = \varphi^{-1} \circ T_W \circ (\varphi \times \varphi)$, where $\times$ denotes the product operation* [8].

- *T is continuous, Archimedean and strict if and only if there exists an increasing permutation φ of $([0,1],\leq)$ such that T is the φ-transform of T_P.*

4 The Archimedean property for t-norms on $\mathcal{L}^I$

We extend the definitions from the previous section to $\mathcal{L}^I$. There are several possible extensions of the Archimedean property, which we call Archimedean, weak Archimedean and strong Archimedean property. Throughout this section we will use the sets $L_1^I = \{x \mid x \in L^I \text{ and } x_1 \in]0,1[\}$ and $L_{12}^I = \{x \mid x \in L^I \text{ and } x_1 > 0 \text{ and } x_2 < 1\}$.

Definition 4.1 [5] *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$. We say that*

- *$\mathcal{T}$ is Archimedean if*

$$(\forall (x,y) \in (L_1^I)^2)(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} <_{L^I} y);$$

- *$\mathcal{T}$ is strongly Archimedean if*

$$(\forall (x,y) \in (L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\})^2)$$
$$(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} <_{L^I} y);$$

- *$\mathcal{T}$ is weakly Archimedean if*

$$(\forall (x,y) \in (L_{12}^I)^2)(\exists n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} <_{L^I} y).$$

Obviously, if a t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is Archimedean, then it is weakly Archimedean, and if $\mathcal{T}$ is strongly Archimedean, then it is Archimedean. The converse implications do not hold (counterexamples are given in [5]).

In [2, 3] the Archimedean property is defined for t-norms on a general bounded poset. If we apply this definition to $\mathcal{L}^I$, then we obtain the following condition for a t-norm $\mathcal{T}$ on $\mathcal{L}^I$:

$$(\forall (x,y) \in (L^I)^2)((\forall n \in \mathbb{N}^*)(x^{(n)_\mathcal{T}} \geq_{L^I} y)$$
$$\implies (x = 1_{\mathcal{L}^I} \text{ or } y = 0_{\mathcal{L}^I})). \qquad (1)$$

The following theorem shows that the Archimedean property defined using (1) corresponds to the Archimedean property given in Definition 4.1.

Theorem 4.1 [5] *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$. Then $\mathcal{T}$ is Archimedean (in the sense of Definition 4.1) if and only if $\mathcal{T}$ satisfies* (1).

Now we generalize nilpotency and related concepts to $\mathcal{L}^I$.

Definition 4.2 *Let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$.*

(i) An element $a \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ is called a nilpotent element of $\mathcal{T}$ if there exists some $n \in \mathbb{N}^*$ such that $a^{(n)_{\mathcal{T}}} = 0_{\mathcal{L}^I}$.

(ii) An element $a \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ is called a zero divisor of $\mathcal{T}$ if there exists some $b \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ such that $\mathcal{T}(a, b) = 0_{\mathcal{L}^I}$.

Definition 4.3

(i) A t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is called nilpotent if it is continuous and if each $a \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$ is a nilpotent element of $\mathcal{T}$.

(ii) A t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is called weakly nilpotent if it is continuous and if each $a \in L^I_{12}$ is a nilpotent element of $\mathcal{T}$.

(iii) A t-norm $\mathcal{T}$ on $\mathcal{L}^I$ is called strict if it is continuous and strictly increasing on $(L^I \setminus \{0_{\mathcal{L}^I}\})^2$.

Theorem 4.2 [5] *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$. Then $\mathcal{T}$ satisfies the Archimedean property if and only if $\mathcal{T}(x, x) <_{L^I} x$, for all $x \in L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$.*

Theorem 4.3 *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$. Then the following are equivalent:*

(N1) $\mathcal{T}$ is nilpotent;

(N2) each $a \in L^I_1$ is a nilpotent element of $\mathcal{T}$.

5 Representations of continuous t-norms on $\mathcal{L}^I$

We first recall two representation theorems which we will need in order to represent some classes of Archimedean t-norms on $\mathcal{L}^I$.

Theorem 5.1 [4] *Consider a continuous mapping $\mathcal{T} : (L^I)^2 \to L^I$. Then $\mathcal{T}$ is a t-norm on $\mathcal{L}^I$ for which*

(T.1) $\mathcal{T}(x, \sup(y, z)) = \sup(\mathcal{T}(x, y), \mathcal{T}(x, z))$, for all x, y, z in L^I, and

(T.2) $\mathcal{T}(D, D) \subseteq D$,

if and only if there exist an element $t \in [0, 1]$, a continuous t-norm T on $([0, 1], \leq)$ and a continuous increasing mapping $\tilde{g} : [0, 1] \to [0, 1]$ such that

(T'.1) $\tilde{g}(T(y_1, z_1)) = \tilde{g}(T(\tilde{g}^{(-1)}(\tilde{g}(y_1)), z_1))$, for all y_1, z_1 in $[0, 1]$,

(T'.2) $\tilde{g}(T(y_1, z_1)) \leq T(\tilde{g}(y_1), z_1)$, for all y_1, z_1 in $[0, 1]$,

(T'.3) $\tilde{g}(T(\tilde{g}^{(-1)}(t), \tilde{g}^{(-1)}(y_1))) \leq \tilde{g}(y_1)$, for all y_1 in $[0, 1]$,

(T'.4) $\tilde{g}(1) = 1$,

(T'.5) for all x, y in $[0, 1]$,

$$\mathcal{T}(x, y) = \Big[T(x_1, y_1), \max\Big(\tilde{g}\Big(T\Big(\tilde{g}^{(-1)}(t), T(\tilde{g}^{(-1)}(x_2), \tilde{g}^{(-1)}(y_2))\Big)\Big), \tilde{g}(T(\tilde{g}^{(-1)}(x_2), y_1)), \tilde{g}(T(\tilde{g}^{(-1)}(y_2), x_1)), T(x_1, y_1)\Big)\Big],$$

where, for all z_1 in $[0, 1]$,

$$\tilde{g}^{(-1)}(z_1) = \sup\{y_1 \mid y_1 \in [0, 1] \text{ and } \tilde{g}(y_1) = z_1\}.$$

Theorem 5.2 [4] *Consider a continuous mapping $\mathcal{T} : (L^I)^2 \to L^I$. Then $\mathcal{T}$ is a t-norm on $\mathcal{L}^I$ such that*

(T".1) $\mathcal{T}$ satisfies the residuation principle, and

(T.2) $\mathcal{T}(D, D) \subseteq D$,

if and only if there exist an element $t \in [0, 1]$, a continuous t-norm T on $([0, 1], \leq)$ and a continuous increasing mapping $\tilde{g} : [0, 1] \to [0, 1]$ such that

(T'.1) $\tilde{g}(T(y_1, z_1)) = \tilde{g}(T(\tilde{g}^{(-1)}(\tilde{g}(y_1)), z_1))$, for all y_1, z_1 in $[0, 1]$,

(T'.2) $\tilde{g}(T(y_1, z_1)) \leq T(\tilde{g}(y_1), z_1)$, for all y_1, z_1 in $[0, 1]$,

(T'.3) $\tilde{g}(T(\tilde{g}^{(-1)}(t), \tilde{g}^{(-1)}(y_1))) \leq \tilde{g}(y_1)$, for all y_1 in $[0, 1]$,

(T'.4) $\tilde{g}(1) = 1$,

(T'.5) for all x, y in $[0, 1]$,

$$\mathcal{T}(x, y) = \Big[T(x_1, y_1), \max\Big(\tilde{g}\Big(T\Big(\tilde{g}^{(-1)}(t), T(\tilde{g}^{(-1)}(x_2), \tilde{g}^{(-1)}(y_2))\Big)\Big), \tilde{g}(T(\tilde{g}^{(-1)}(x_2), y_1)), \tilde{g}(T(\tilde{g}^{(-1)}(y_2), x_1)), T(x_1, y_1)\Big)\Big],$$

where, for all z_1 in $[0, 1]$,

$$\tilde{g}^{(-1)}(z_1) = \sup\{y_1 \mid y_1 \in [0, 1] \text{ and } \tilde{g}(y_1) = z_1\}.$$

In this case the residuum $\mathcal{I}_{\mathcal{T}}$ of $\mathcal{T}$ is given by, for all x, z in L^I,

$$\begin{aligned}&\mathcal{I}_{\mathcal{T}}(x,z)\\&=\Big[\inf\Big(I_T(x_1,z_1), I_T(\tilde{g}^{(-1)}(x_2),\tilde{g}^{(-1)}(z_2)),\\&\quad \tilde{g}\Big(I_T\Big(T(\tilde{g}^{(-1)}(t),\tilde{g}^{(-1)}(x_2)),\tilde{g}^{(-1)}(z_2)\Big)\Big),\\&\quad \tilde{g}(I_T(x_1,\tilde{g}^{(-1)}(z_2)))\Big),\\&\quad \inf\Big(\tilde{g}\Big(I_T\Big(T(\tilde{g}^{(-1)}(t),\tilde{g}^{(-1)}(x_2)),\\&\quad \tilde{g}^{(-1)}(z_2)\Big)\Big), \tilde{g}(I_T(x_1,\tilde{g}^{(-1)}(z_2)))\Big)\Big].\end{aligned}$$

Example 5.1 Let for all $a, b \in [0,1]$ and x, y in L^I, $T(a,b) = \max(0, a+b-1)$,

$$\tilde{g}(a) = \begin{cases} 0, & \text{if } a \le \frac{1}{2},\\ 2a-1, & \text{else,}\end{cases}$$

and $\mathcal{T}(x,y) = [\max(0, x_1+y_1-1), \max(0, 2x_1+y_2-2, 2y_1+x_2-2, x_1+y_1-1)]$. Then T, $\tilde{g}$ and $\mathcal{T}$ satisfy the conditions of Theorem 5.1.

Using Theorems 5.1 and 5.2 we obtain the following result.

Theorem 5.3 *If $\mathcal{T}$ is a continuous t-norm on $\mathcal{L}^I$ which satisfies (T.1) and (T.2), then $\mathcal{T}$ satisfies the residuation principle.*

Note that from the fact that $\tilde{g}$ is increasing and continuous, it follows that, for all $z_1 \in [0,1]$,

$$\begin{aligned}&\tilde{g}^{(-1)}(z_1)\\&= \sup\{y_1 \mid y_1 \in [0,1] \text{ and } \tilde{g}(y_1) \le z_1\},\\&\tilde{g}(\tilde{g}^{(-1)}(z_1))\\&= \tilde{g}(\sup\{y_1 \mid y_1 \in [0,1] \text{ and } \tilde{g}(y_1) = z_1\}) = z_1.\end{aligned}$$

Hence, for all x_1, z_1 in $[0,1]$,

$$\begin{aligned}&\tilde{g}(x_1) \le z_1\\&\quad\Longleftrightarrow\ x_1 \in \{y_1 \mid y_1 \in [0,1] \text{ and } \tilde{g}(y_1) \le z_1\}\\&\quad\Longleftrightarrow\ x_1 \le \tilde{g}^{(-1)}(z_1). \qquad (2)\end{aligned}$$

Theorem 5.4 *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$ which satisfies (T.1) and (T.2). The negation $\mathcal{N}_{\mathcal{I}_{\mathcal{T}}}$ generated by $\mathcal{I}_{\mathcal{T}}$ is involutive if and only if $\mathcal{T}$ is pseudo-t-representable and the negation N_{I_T} generated by the residual implication of the representant T of $\mathcal{T}$ is involutive.*

Theorem 5.4 shows that the class of pseudo-t-representable t-norms play an important role if we need t-norms for which the negation generated by the residual implication is involutive. Indeed, in the class of continuous t-norms which satisfy the residuation principle and which are a natural extension of a t-norm on the unit interval, the only t-norms for which the negation generated by their residual implication is involutive, are pseudo-t-representable.

6 Properties of Archimedean and nilpotent t-norms on $\mathcal{L}^I$

Theorem 6.1 *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$ which satisfies (T.1) and (T.2). Then $\mathcal{T}$ is weakly nilpotent if and only if t-norm T involved in the representation of $\mathcal{T}$ according to Theorem 5.1 is nilpotent.*

Theorem 6.2 *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$ which satisfies (T.1) and (T.2). If the t-norm T involved in the representation of $\mathcal{T}$ according to Theorem 5.1 is nilpotent and if $(\mathcal{T}([0,1],[0,1]))_2 < 1$, then $\mathcal{T}$ is nilpotent.*

Theorem 6.3 *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$ which satisfies (T.1) and (T.2). Then $\mathcal{T}$ is weakly Archimedean if and only if t-norm T involved in the representation of $\mathcal{T}$ according to Theorem 5.1 is Archimedean.*

Theorem 6.4 *Let $\mathcal{T}$ be a continuous t-norm on $\mathcal{L}^I$ which satisfies (T.1) and (T.2). If the t-norm T involved in the representation of $\mathcal{T}$ according to Theorem 5.1 is Archimedean and if $(\mathcal{T}([0,1],[0,1]))_2 < 1$, then $\mathcal{T}$ is Archimedean.*

Note that if $\mathcal{T}$ is t-representable, then $\mathcal{T}([0,1],[0,1]) = [0,1]$. So $[0,1]$ is an idempotent element and thus not a nilpotent element of $\mathcal{T}$. Hence, taking into consideration Theorem 6.2, we find that the only subclass of the class of t-norms represented by Theorem 5.1 that does *not* have nilpotent members is the class of t-representable t-norms. Similarly, since for t-representable t-norms $\mathcal{T}$ it holds that $\mathcal{T}([0,1],[0,1]) = [0,1]$, these t-norms are not Archimedean. Theorem 6.4 shows that the t-representable members of the class of t-norms represented by Theorem 5.1 are the only ones that can never be Archimedean.

Theorem 6.5 *Let $\mathcal{T}$ be a continuous weakly Archimedean t-norm on $\mathcal{L}^I$. Then the following are equivalent:*

(i) $\mathcal{T}$ *is weakly nilpotent;*

(ii) there exists some nilpotent element of $\mathcal{T}$ *in* L_{12}^I*;*

(iii) there exists some zero divisor of $\mathcal{T}$ *in* L_{12}^I*.*

Theorem 6.6 *Let* $\mathcal{T}$ *be a continuous strongly Archimedean t-norm on* $\mathcal{L}^I$*. Then the following are equivalent:*

(i) $\mathcal{T}$ *is nilpotent;*

(ii) there exists some nilpotent element of $\mathcal{T}$ *in* $L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$*;*

(iii) there exists some zero divisor of $\mathcal{T}$ *in* $L^I \setminus \{0_{\mathcal{L}^I}, 1_{\mathcal{L}^I}\}$*.*

7 Representations of continuous Archimedean t-norms on $\mathcal{L}^I$

In this section we will extend Theorem 3.2 to $\mathcal{L}^I$. Since from Theorem 3.1 it follows that the class of continuous Archimedean t-norms on the unit interval can be splitted into two subclasses (the nilpotent t-norms and the strict t-norms), we also split our discussion on the Archimedean t-norms on $\mathcal{L}^I$ in two parts. First we characterize the weakly Archimedean weakly nilpotent t-norms belonging to the class of continuous t-norms $\mathcal{T}$ on $\mathcal{L}^I$ which satisfy the residuation principle and for which $\mathcal{T}(D, D) \subseteq D$. From Theorem 6.1 we know that these are the t-norms on $\mathcal{L}^I$ for which the t-norm T involved in their representation (according to Theorem 5.1) is nilpotent. After that, we characterize the weakly Archimedean t-norms belonging to the previously mentioned class but which are *not* weakly nilpotent. From Theorem 6.1 we know that the t-norm T involved in their representation (see Theorem 5.1) is not nilpotent, and thus strict.

7.1 Archimedean t-norms which are weakly nilpotent

Lemma 7.1 *Let* $\tilde{g} : [0,1] \to [0,1]$ *be a continuous increasing mapping and* φ *be an automorphism of* $([0,1], \leq)$*. Let* T *be the t-norm on* $([0,1], \leq)$ *defined by* $T(x_1, y_1) = \varphi^{-1}(\max(0, \varphi(x_1) + \varphi(y_1) - 1)$*, for all* x_1, y_1 *in* $[0,1]$*. Then (T'.1) holds if and only if* $\tilde{g}_{|[\tilde{g}^{(-1)}(0),1]}$ *is a bijection from* $[\tilde{g}^{(-1)}(0), 1]$ *to* $[0,1]$*. Furthermore* $\tilde{g}^{(-1)}(z_1) = \tilde{g}^{-1}(z_1)$*, for all* $z_1 \in]0,1]$*.*

Lemma 7.2 *Let* $\tilde{g} : [0,1] \to [0,1]$ *be a continuous increasing mapping and* φ *be an automorphism of* $([0,1], \leq)$*. Let* T *be the t-norm on* $([0,1], \leq)$ *defined by* $T(x_1, y_1) = \varphi^{-1}(\max(0, \varphi(x_1) + \varphi(y_1) - 1)$*, for all* x_1, y_1 *in* $[0,1]$*. Then (T'.2) holds if and only if for all* a, b *in* $[0,1]$ *such that* $a > b > \varphi(\tilde{g}^{(-1)}(0))$*,*

$$\frac{\varphi \circ \tilde{g} \circ \varphi^{-1}(a) - \varphi \circ \tilde{g} \circ \varphi^{-1}(b)}{a - b} \geq 1.$$

Lemma 7.3 *Let* $\tilde{g} : [0,1] \to [0,1]$ *be a continuous increasing mapping and* φ *be an automorphism of* $([0,1], \leq)$*. Let* T *be the t-norm on* $([0,1], \leq)$ *defined by* $T(x_1, y_1) = \varphi^{-1}(\max(0, \varphi(x_1) + \varphi(y_1) - 1)$*, for all* x_1, y_1 *in* $[0,1]$*, and assume that (T'.1) and (T'.2) hold. Then (T'.3) holds if and only if* $\tilde{g}^{(-1)}(0) \leq \tilde{g}\big(\varphi^{(-1)}\big(I_{T_W}\big(\varphi(\tilde{g}^{(-1)}(t)), \varphi(\tilde{g}^{(-1)}(0))\big)\big)\big)$*.*

Corollary 7.4 *Let* $\tilde{g} : [0,1] \to [0,1]$ *be a continuous increasing mapping and* φ *be an automorphism of* $([0,1], \leq)$*. Let* T *be the t-norm on* $([0,1], \leq)$ *defined by* $T(x_1, y_1) = \varphi^{-1}(\max(0, \varphi(x_1) + \varphi(y_1) - 1)$*, for all* x_1, y_1 *in* $[0,1]$*, and assume that (T'.1) and (T'.2) hold. Then (T'.3) holds if and only if*

$$t \leq \tilde{g}\big(\varphi^{(-1)}\big(I_{T_W}\big(\varphi(\tilde{g}^{(-1)}(\tilde{g}^{(-1)}(0))), \varphi(\tilde{g}^{(-1)}(0))\big)\big)\big).$$

Theorem 7.5 *Consider a continuous mapping* $\mathcal{T} : (L^I)^2 \to L^I$*. Then* $\mathcal{T}$ *is a t-norm on* $\mathcal{L}^I$ *such that*

(AT.1) $\mathcal{T}$ *is weakly Archimedean,*

(AT.2) $\mathcal{T}$ *is weakly nilpotent,*

(AT.3) $\mathcal{T}$ *satisfies the residuation principle, and*

(AT.4) $\mathcal{T}(D, D) \subseteq D$*,*

if and only if there exist an element $t \in [0,1]$*, an increasing permutation* φ *of* $([0,1], \leq)$ *and a continuous increasing mapping* $\tilde{g} : [0,1] \to [0,1]$ *such that*

(AT'.1) $\tilde{g}_{|[\tilde{g}^{(-1)}(0),1]}$ *is a bijection from* $[\tilde{g}^{(-1)}(0), 1]$ *to* $[0,1]$*, and* $\tilde{g}(x_1) = 0$*, for all* $x_1 \in [0, \tilde{g}^{(-1)}(0)]$*,*

(AT'.2) for all a, b *in* $[0,1]$ *such that* $a > b > \varphi(\tilde{g}^{(-1)}(0))$*,*

$$\frac{\varphi \circ \tilde{g} \circ \varphi^{-1}(a) - \varphi \circ \tilde{g} \circ \varphi^{-1}(b)}{a - b} \geq 1,$$

(AT'.3) $\tilde{g}^{(-1)}(0) \leq \tilde{g}\big(\varphi^{(-1)}\big(I_{T_W}\big(\varphi(\tilde{g}^{(-1)}(t)), \varphi(\tilde{g}^{(-1)}(0))\big)\big)\big)$,

(AT'.4) for all x, y *in* $[0,1]$,

$$\begin{aligned}&\mathcal{T}(x,y)\\ &= \Big[\varphi^{-1}\big(\max(0,\varphi(x_1)+\varphi(y_1)-1)\big),\\ &\quad \max\Big(\tilde{g}\big(\varphi^{-1}\big(\max(0,\varphi(\tilde{g}^{(-1)}(t))+\\ &\quad \varphi(\tilde{g}^{(-1)}(x_2))+\varphi(\tilde{g}^{(-1)}(y_2))-2)\big)\big),\\ &\quad \tilde{g}\big(\varphi^{-1}\big(\max(0,\varphi(\tilde{g}^{(-1)}(x_2))+\varphi(y_1)-1)\big)\big),\\ &\quad \tilde{g}\big(\varphi^{-1}\big(\max(0,\varphi(\tilde{g}^{(-1)}(y_2))+\varphi(x_1)-1)\big)\big),\\ &\quad \varphi^{-1}\big(\max(0,\varphi(x_1)+\varphi(y_1)-1)\big)\Big)\Big],\end{aligned}$$

where, for all z_1 *in* $[0,1]$,

$$\tilde{g}^{(-1)}(z_1) = \sup\{y_1 \mid y_1 \in [0,1] \text{ and } \tilde{g}(y_1) = z_1\}.$$

From Theorems 6.1, 6.2, 6.3 and 6.4 it follows that a similar representation theorem holds when we replace (AT.1) and (AT.2) by the conditions: $\mathcal{T}$ is Archimedean and nilpotent, and $\mathcal{T}([0,1],[0,1]) <_{L^I} [0,1]$.

7.2 Archimedean t-norms which are not weakly nilpotent

Lemma 7.6 *Let* $\tilde{g} : [0,1] \to [0,1]$ *be a continuous increasing mapping and* φ *be an automorphism of* $([0,1],\leq)$. *Let* T *be the t-norm on* $([0,1],\leq)$ *defined by* $T(x_1,y_1) = \varphi^{-1}(\varphi(x_1)\varphi(y_1))$, *for all* x_1, y_1 *in* $[0,1]$. *Then (T'.1) holds if and only if* $\tilde{g}_{|[\tilde{g}^{(-1)}(0),1]}$ *is a bijection from* $[\tilde{g}^{(-1)}(0),1]$ *to* $[0,1]$. *Furthermore* $\tilde{g}^{(-1)}(z_1) = \tilde{g}^{-1}(z_1)$, *for all* $z_1 \in]0,1]$.

Lemma 7.7 *Let* $\tilde{g} : [0,1] \to [0,1]$ *be a continuous increasing mapping and* φ *be an automorphism of* $([0,1],\leq)$. *Let* T *be the t-norm on* $([0,1],\leq)$ *defined by* $T(x_1,y_1) = \varphi^{-1}(\varphi(x_1)\varphi(y_1))$, *for all* x_1, y_1 *in* $[0,1]$. *Then (T'.2) holds if and only if for all* a, b *in* $[0,1]$ *such that* $a > b > \varphi(\tilde{g}^{(-1)}(0))$,

$$\frac{a}{b} \leq \frac{\varphi\circ\tilde{g}\circ\varphi^{-1}(a)}{\varphi\circ\tilde{g}\circ\varphi^{-1}(b)}.$$

Lemma 7.8 *Let* $\tilde{g} : [0,1] \to [0,1]$ *be a continuous increasing mapping and* φ *be an automorphism of* $([0,1],\leq)$. *Let* T *be the t-norm on* $([0,1],\leq)$ *defined by* $T(x_1,y_1) = \varphi^{-1}(\varphi(x_1)\varphi(y_1))$, *for all* x_1, y_1 *in* $[0,1]$, *and assume that (T'.1) and (T'.2) hold. Then (T'.3) holds if and only if* $\tilde{g}^{(-1)}(0) \leq \tilde{g}\big(\varphi^{(-1)}\big(I_{T_P}\big(\varphi(\tilde{g}^{(-1)}(t)), \varphi(\tilde{g}^{(-1)}(0))\big)\big)\big)$.

Theorem 7.9 *Consider a continuous mapping* $\mathcal{T} : (L^I)^2 \to L^I$. *Then* $\mathcal{T}$ *is a t-norm on* $\mathcal{L}^I$ *such that*

(AT.1) $\mathcal{T}$ *is weakly Archimedean,*

(AT.2) $\mathcal{T}$ *is not weakly nilpotent,*

(AT.3) $\mathcal{T}$ *satisfies the residuation principle, and*

(AT.4) $\mathcal{T}(D,D) \subseteq D$,

if and only if there exist an element $t \in [0,1]$, *a continuous t-norm* T *on* $([0,1],\leq)$ *and a continuous increasing mapping* $\tilde{g} : [0,1] \to [0,1]$ *such that*

(AT".1) $\tilde{g}_{|[\tilde{g}^{(-1)}(0),1]}$ *is a bijection from* $[\tilde{g}^{(-1)}(0),1]$ *to* $[0,1]$, *and* $\tilde{g}(x_1) = 0$, *for all* $x_1 \in [0, \tilde{g}^{(-1)}(0)]$,

(AT".2) for all a, b *in* $[0,1]$ *such that* $a > b > \varphi(\tilde{g}^{(-1)}(0))$,

$$\frac{\varphi\circ\tilde{g}\circ\varphi^{-1}(a)}{\varphi\circ\tilde{g}\circ\varphi^{-1}(b)} \geq \frac{a}{b},$$

(AT".3) $\tilde{g}^{(-1)}(0) \leq \tilde{g}\big(\varphi^{(-1)}\big(I_{T_P}\big(\varphi(\tilde{g}^{(-1)}(t)), \varphi(\tilde{g}^{(-1)}(0))\big)\big)\big)$,

(AT".4) for all x, y *in* $[0,1]$,

$$\begin{aligned}&\mathcal{T}(x,y)\\ &= \Big[\varphi^{-1}\big(\varphi(x_1)\varphi(y_1)\big),\\ &\quad \max\Big(\tilde{g}\big(\varphi^{-1}\big(\varphi(\tilde{g}^{(-1)}(t))\\ &\quad \varphi(\tilde{g}^{(-1)}(x_2))\varphi(\tilde{g}^{(-1)}(y_2))\big)\big)\big),\\ &\quad \tilde{g}\big(\varphi^{-1}\big(\varphi(\tilde{g}^{(-1)}(x_2))\varphi(y_1)\big)\big),\\ &\quad \tilde{g}\big(\varphi^{-1}\big(\varphi(\tilde{g}^{(-1)}(y_2))\varphi(x_1)\big)\big),\\ &\quad \varphi^{-1}\big(\varphi(x_1)\varphi(y_1)\big)\Big)\Big],\end{aligned}$$

where, for all z_1 *in* $[0,1]$,

$$\tilde{g}^{(-1)}(z_1) = \sup\{y_1 \mid y_1 \in [0,1] \text{ and } \tilde{g}(y_1) = z_1\}.$$

From Theorems 6.1, 6.2, 6.3 and 6.4 it follows that a similar representation theorem holds when we replace (AT.1) and (AT.2) by the

conditions: $\mathcal{T}$ is Archimedean and not nilpotent, and $\mathcal{T}([0,1],[0,1]) <_{L^I} [0,1]$. It remains an open problem whether all continuous Archimedean t-norms on $\mathcal{L}^I$ which are not (weakly) nilpotent are strict.

8 Conclusion

In this paper we presented some properties of t-norms on $\mathcal{L}^I$ which are Archimedean and nilpotent, or which satisfy some related properties such as the weak Archimedean property and weak nilpotency. We investigated some properties of the class of t-norms $\mathcal{T}$ on $\mathcal{L}^I$ which are continuous, satisfy the residuation principle and $\mathcal{T}(D,D) \subseteq D$. We gave necessary and sufficient conditions for the members of this class so that they satisfy the Archimedean property, nilpotency, or any of the weak variants of these properties. We also gave a representation of the members of the above mentioned class of t-norms on $\mathcal{L}^I$ which are either (weakly) Archimedean and (weakly) nilpotent, or (weakly) Archimedean but *not* (weakly) nilpotent. Thus we obtained a full characterization of the class of continuous t-norms $\mathcal{T}$ on $\mathcal{L}^I$ which satisfy the residuation principle, $\mathcal{T}(D,D) \subseteq D$, and which are (weakly) Archimedean.

References

[1] K. T. Atanassov, *Intuitionistic fuzzy sets*, Physica-Verlag, Heidelberg/New York, 1999.

[2] G. Birkhoff, *Lattice Theory*, volume 25, AMS Colloquium Publications, Providence, Rhode Island, 1973.

[3] B. De Baets and R. Mesiar, Triangular norms on product lattices, *Fuzzy Sets and Systems*, **104**(1) (1999) 61–75.

[4] G. Deschrijver, A representation of t-norms in interval-valued L-fuzzy set theory, *Fuzzy Sets and Systems*, submitted.

[5] G. Deschrijver, The Archimedean property for t-norms in interval-valued fuzzy set theory, *Fuzzy Sets and Systems*, **157**(17) (2006) 2311–2327.

[6] G. Deschrijver, C. Cornelis and E. E. Kerre, On the representation of intuitionistic fuzzy t-norms and t-conorms, *IEEE Transactions on Fuzzy Systems*, **12**(1) (2004) 45–61.

[7] G. Deschrijver and E. E. Kerre, On the relationship between some extensions of fuzzy set theory, *Fuzzy Sets and Systems*, **133**(2) (2003) 227–235.

[8] M. Eisenberg, *Axiomatic theory of sets and classes*, Holt, Rinehart and Winston, Inc., New York, 1971.

[9] J. C. Fodor and M. Roubens, *Fuzzy preference modelling and multicriteria decision support*, Kluwer Academic Publishers, Dordrecht, 1994.

[10] J. Goguen, L-fuzzy sets, *Journal of Mathematical Analysis and Applications*, **18**(1) (1967) 145–174.

[11] M. B. Gorzałczany, A method of inference in approximate reasoning based on interval-valued fuzzy sets, *Fuzzy Sets and Systems*, **21**(1) (1987) 1–17.

[12] P. Hájek, *Metamathematics of fuzzy logic*, Kluwer Academic Publishers, Dordrecht, 1998.

[13] E. P. Klement, R. Mesiar and E. Pap, *Triangular norms*, Kluwer Academic Publishers, Dordrecht, 2002.

[14] E. P. Klement, R. Mesiar and E. Pap, Triangular norms. Position paper I: basic analytical and algebraic properties, *Fuzzy Sets and Systems*, **143**(1) (2004) 5–26.

[15] E. P. Klement, R. Mesiar and E. Pap, Triangular norms. Position paper II: general constructions and parametrized families, *Fuzzy Sets and Systems*, **145**(3) (2004) 411–438.

[16] E. P. Klement, R. Mesiar and E. Pap, Triangular norms. Position paper III: continuous t-norms, *Fuzzy Sets and Systems*, **145**(3) (2004) 439–454.

[17] R. Sambuc, *Fonctions Φ-floues. Application à l'aide au diagnostic en pathologie thyroidienne*, Ph. D. Thesis, Université de Marseille, France, 1975.

[18] B. Schweizer and A. Sklar, *Probabilistic metric spaces*, North-Holland, New York, 1983.

Constructions of aggregation operators that preserve ordering of the data

G. Beliakov
Deakin University
Australia
gleb@deakin.edu.au

T. Calvo
Universidad de Alcalá
Spain
tomasa.calvo@uah.es

Abstract

We address the issue of identifying various classes of aggregation operators from empirical data, which also preserves the ordering of the outputs. It is argued that the ordering of the outputs is more important than the numerical values, however the usual data fitting methods are only concerned with fitting the values. We will formulate preservation of the ordering problem as a standard mathematical programming problem, solved by standard numerical methods.

Keywords: Aggregation operators, preference ordering, decision making.

1 Introduction

Construction of aggregation operators from empirical data is very useful in practice, when a specific aggregation operator has to be chosen for a specific application. This work was pioneered in [25], where the authors introduced γ-operators, a convex combination of triangular norms and conorms. Identification of the weights of arithmetic means and OWA operators from the data was treated in [4, 10, 20, 21], identification of the coefficients of fuzzy measures for Choquet integral based aggregation was treated in [4,13,14], and identification of additive generators of t–norms, t–conorms, uninorms and nullnorms was treated in [4, 5].

In all mentioned studies, the choice of parameters was driven by how well an aggregation function predicted the observed input-output pairs, the data set $\{(\vec{x}_k, y_k)\}, k = 1, \dots K$. The goodness of fit was measured by using the least squares criterion, or the least absolute deviation criterion. In the first case the problem was set up as a standard quadratic programming problem, and in the second case as a linear programming problem.

However, in [16] it was argued that fitting the numerical outputs is not as important as preserving the ordering of the outputs. The empirical data usually comes from human subjective evaluation, and people do not reliably express their preference on a numerical scale. In contrast, people are very good at ranking the alternatives. Therefore, the authors of [16] argued that fitting methods should aim at preserving the order of empirical output values. They showed that various methods of fitting the numerical values do not preserve this ordering. However, a solution which does preserve the ordering of the outputs has never been spelled out.

The aim of this contribution is to show that preservation of outputs ordering can be achieved by a very simple technique of adding $K - 1$ linear inequalities to the least squares and least absolute deviation problems. Furthermore, in many cases, that cover all major families of aggregation operators, the structure of the resulting quadratic and linear programming problems does not change, which allows one to apply standard numerical optimization methods. We formulate the resulting mathematical programming problems explicitly in these cases. Finally, we present a new formulation of the aggregation operator identification problem, in which a weighted combination of the numerical fitness and ordering preservation criteria is optimized, as well as its solution methods. This problem is particularly useful when the data is contaminated by noise.

2 Fitting aggregation operators

Various methods of fitting parameters of an aggregation operator to the data are available [4, 10, 14, 16, 20, 21, 25]. Given a data set $\{(\vec{x}_k, y_k)\}, k = 1, \dots K$, and a class of aggregation operators parameterized by a vector $\vec{w}$, $f(\vec{x}; \vec{w})$, the mentioned methods minimize the least squares (LS) criterion

$$\sum_{k=1}^{K} (f(\vec{x}_k; \vec{w}) - y_k)^2$$

with respect to the parameters $\vec{w}$, subject to the conditions that f is an aggregation operator, i.e., it verifies at least the conditions of monotonicity and $f(\vec{0};\vec{w}) = 0, f(\vec{1};\vec{w}) = 1$. Of course, other conditions like symmetry, idempotency, existence of a neutral element, annihilator, etc., can also be added.

An alternative is to use the least absolute deviation criterion (LDA) [7], i.e., minimize

$$\sum_{k=1}^{K} |f(\vec{x}_k;\vec{w}) - y_k|$$

with respect to the weights, subject to the same conditions. The use of LDA is less sensitive to outliers, and importantly, in many cases it allows one to set up an equivalent linear programming problem, which is easily solved by the standard simplex algorithm, even if the number of weights is very large (e.g., when identifying a fuzzy measure).

If f depends on the weights $\vec{w}$ *linearly*, which is the case when f is an arithmetic mean, OWA, Choquet integral, γ-operator and some other aggregation operators, then minimization of LS criterion becomes a standard quadratic programming problem (QP), and minimization of LDA criterion becomes a linear programming problem (LP) after introducing auxiliary variables. Furthermore, linearization methods allow one to set up QP or LP problems for quasi-arithmetic means, generalized OWA and generalized Choquet integrals, see [4–6].

We note that all mentioned methods approximate, not interpolate, the empirical values y_k (except some special cases). Empirical data comes with errors, and it is pointless to fit it exactly. If the data were interpolated, then of course the order of the outputs would be preserved automatically. Thus our goal is to ensure that the order is preserved during the solution to LS or LDA problems.

3 Preservation of ordering

Without loss of generality, we assume that the outputs are ordered as $y_1 \le y_2 \le \ldots \le y_K$ (the data can always be re-ordered in this way). The condition for order preservation is

$$f(\vec{x}_i;\vec{w}) \le f(\vec{x}_j;\vec{w}), \text{ for all } i < j. \tag{1}$$

Because the data set is ordered, this condition is implied by a simpler condition

$$f(\vec{x}_i;\vec{w}) \le f(\vec{x}_{i+1};\vec{w}), \text{ for all } i = 1,\ldots,K-1. \tag{2}$$

In general, this is a system of $K-1$ nonlinear inequalities, which is very hard to solve. But in many interesting cases f depends on $\vec{w}$ linearly, and in this case we obtain a system of linear inequalities, which does not change the structure of the LS or LDA problem.

Let $f(\vec{x};\vec{w}) =< \vec{g}(\vec{x}), \vec{w} > \; = \sum_{i=1}^{n} w_i g_i(\vec{x})$, $\vec{g}$ being some basis functions. For example $g_i(\vec{x}) = x_i$ for the arithmetic means, $g_i(\vec{x}) = x_{(i)}$ for an OWA operator[1]. Then the LS problem becomes

$$\begin{aligned} \text{Minimize} \quad & \textstyle\sum_{k=1}^{K} (< \vec{g}(\vec{x}_k), \vec{w} > \; - y_k)^2, \\ \text{s.t.} \quad & < \vec{g}(\vec{x}_{k+1}), \vec{w} > \; - \; < \vec{g}(\vec{x}_k), \vec{w} > \; \ge 0, \\ & k = 1,\ldots,K-1, \\ & \text{other (linear) conditions on } \vec{w}. \end{aligned} \tag{3}$$

Problem (3) is QP, which differs from the original LS problem only by additional $K-1$ linear constraints. Consequently, standard methods of solution of QPs can be employed. In the case of LDA problem, the situation is similar, we have an additional set of linear constraints, and if LDA was converted to LP, then the additional constraints are directly transferred to the LP problem. Thus by using the auxiliary variables $r_k^+, r_k^- \ge 0$, such that $r_k^+ + r_k^- = | < \vec{g}(\vec{x}_k), \vec{w} > -y_k|$, and $r_k^+ - r_k^- =< \vec{g}(\vec{x}_k), \vec{w} > -y_k$, we have an LP

$$\begin{aligned} \text{Minimize} \quad & \textstyle\sum_{k=1}^{K} r_k^+ + r_k^-, \\ \text{s.t.} \quad & < \vec{g}(\vec{x}_k), \vec{w} > \; - r_k^+ + r_k^- = y_k, \\ & k = 1,\ldots,K, \\ & < \vec{g}(\vec{x}_{k+1}), \vec{w} > \; - \; < \vec{g}(\vec{x}_k), \vec{w} > \; \ge 0, \\ & k = 1,\ldots,K-1, \\ & r_k^+, r_k^- \ge 0, k = 1,\ldots,K, \\ & \text{other (linear) conditions on } \vec{w}. \end{aligned} \tag{4}$$

The methods of solution to problems (3) and (4) are well known, see [7, 18].

4 Special cases

In this section we will present explicit problem formulations for a number of popular families of aggregation operators, in the case of LS fitting (Problem (3)). The case of LDA (4) is treated very similarly.

4.1 Arithmetic means and OWA

Since $g_i(\vec{x}) = x_i$ and we have constraints $w_i \ge 0$, $\sum_{i=1}^{n} w_i = 1$, n is the dimension of the input vector $\vec{x}$, Problem (3) translates into

$$\begin{aligned} \text{Minimize} \quad & \textstyle\sum_{k=1}^{K} (< \vec{x}_k, \vec{w} > \; - y_k)^2, \\ \text{s.t.} \quad & < \vec{x}_{k+1} - \vec{x}_k, \vec{w}) > \; \ge 0, \\ & k = 1,\ldots,K-1, \\ & \textstyle\sum w_i = 1, w_i \ge 0. \end{aligned} \tag{5}$$

[1] As usual, $x_{(i)}$ denotes the i-th largest component of $\vec{x}$.

For OWA operators [23], let $\vec{z}_k = (x_{k,(1)}, x_{k,(2)}, \ldots, x_{k,(n)})$ be the vector obtained from $\vec{x}_k$ by arranging its components in non-increasing order. Then LS problem translates into

$$\begin{array}{rl} \text{Minimize} & \sum_{k=1}^{K}(<\vec{z}_k, \vec{w}> - y_k)^2, \\ \text{s.t.} & <\vec{z}_{k+1} - \vec{z}_k, \vec{w}) > \ \geq 0, \\ & k = 1, \ldots, K-1, \\ & \sum w_i = 1, w_i \geq 0. \end{array} \tag{6}$$

For OWA operators, a frequent additional requirement is preservation of a given measure of orness [10, 23], which translates into an additional linear constraint

$$<\vec{a}, \vec{w}> \ = \alpha,$$

where $a_i = \frac{n-i}{n-1}$, and $0 \leq \alpha \leq 1$ is specified by the user.

4.2 Quasi-arithmetic means and generalized OWA

Let $h : [0,1] \to [-\infty, \infty]$, be a given continuous strictly monotone function. A quasi-arithmetic mean is the function

$$f(\vec{x}; \vec{w}) = h^{-1}(\sum_{i=1}^{n} w_i h(x_i)).$$

This class includes geometric, harmonic, quadratic means, power means and many others.

A generalized OWA operator (also called Ordered Weighted Quasi-Arithmetic means (OWQA) in [8]) is the function

$$f(\vec{x}; \vec{w}) = h^{-1}(\sum_{i=1}^{n} w_i h(z_i)),$$

where $\vec{z} = (x_{(1)}, \ldots, x_{(n)})$.

Fitting the weights of quasi-arithmetic means and generalized OWA is done by linearizing inputs and outputs, i.e., solving

$$\begin{array}{rl} \text{Minimize} & \sum_{k=1}^{K}(<h(\vec{x}_k), \vec{w}> - h(y_k))^2, \\ \text{s.t.} & <h(\vec{x}_{k+1}) - h(\vec{x}_k), \vec{w}) > \ \geq 0, \\ & k = 1, \ldots, K-1, \\ & \sum w_i = 1, w_i \geq 0, \end{array} \tag{7}$$

where $h(\vec{x}) = (h(x_1), \ldots, h(x_n))$.

4.3 Choquet integrals

Let the set N be $N = \{1, 2, \ldots, n\}$. A fuzzy measure is a set function $v : 2^N \to [0,1]$ which is monotonic (i.e. $v(S) \leq v(T)$ whenever $S \subseteq T$) and satisfies $v(\emptyset) = 0, v(N) = 1$. The discrete Choquet integral is defined with respect to a fuzzy measure, and can be written as [12]

$$C_v(\vec{x}) = \sum_{i=1}^{n} \left[x_{(i)} - x_{(i-1)}\right] v(H_i), \tag{8}$$

where $x_{(0)} = 0$ by convention, and $H_i = \{(i), \ldots, (n)\}$ is the subset of indices of $n-i+1$ largest components of $\vec{x}$. Note that here $x_{(i)}$ denotes the i-th smallest component of $\vec{x}$. A fuzzy measure has 2^n parameters, two of which are fixed: $v(\emptyset) = 0$, $v(N) = 1$.

Let us represent Choquet integral as a dot product $<\vec{g}(\vec{x}), \vec{v}>$, where $\vec{v} \in [0,1]^{2^n}$ is the vector of coefficients of the fuzzy measure. It is convenient to use the index $j = 0, \ldots, 2^n - 1$ whose binary representation corresponds to the characteristic vector of the set $J \subseteq N$, $\vec{c} \in \{0,1\}^n$ defined by $c_{n-i+1} = 1$ if $i \in J$ and 0 otherwise. For example, let $n = 5$; for $j = 101$ (binary), $\vec{c} = (0,0,1,0,1)$ and $v_j = v(\{1,3\})$. We shall use letters K, J, etc., to denote subsets that correspond to indices k, j, etc.

Let us define the basis functions $g_j, j = 0, \ldots, 2^n - 1$ as $g_j(\vec{x}) = \max(0, \min_{i \in J}\{x_i\} - \max_{i \in N \setminus J}\{x_i\})$, where $J \subseteq N$ whose characteristic vector corresponds to the binary representation of j. Then $C_v(\vec{x}) = <\vec{g}(\vec{x}), \vec{v}>$.

Now, identification of the coefficients of the fuzzy measure v becomes a QP,

$$\begin{array}{rl} \text{Minimize} & \sum_{k=1}^{K}(<\vec{g}(\vec{x}_k), \vec{v}> - y_k)^2, \\ \text{s.t.} & <\vec{g}(\vec{x}_{k+1}) - \vec{g}(\vec{x}_k), \vec{v}> \ \geq 0, \\ & k = 1, \ldots, K-1, \\ & v_0 = 0, v_{2^n-1} = 1, \\ & v_k - v_j \geq 0 \text{ for all } k, j \text{ such that } J \subset K. \end{array} \tag{9}$$

This is a large scale (even for moderate n) QP with a sparse matrix of constraints, and there are numerical methods that exploit such a sparse structure [11]. However when using LDA criterion, it becomes an LP problem with a sparse matrix, which can be solved efficiently for a very large number of parameters.

It is well known that for additive fuzzy measures Choquet integrals become arithmetic means, and for symmetric fuzzy measures, they become OWA operators. To reduce the complexity of the problem, Grabisch introduced k-additive fuzzy measures [12], in which only combinations of at most k indices allow for interactions of variables. The condition of k-additivity is translated into a set of additional linear equality constraints on the coefficients of fuzzy measure, and these are readily included into QP or LP. Furthermore, the same applies to various other indices, such as Shapley index and its generalizations [12].

By applying a non-linear invertible transformation h

to the components of $\vec{x}$, one obtains a generalized Choquet integral [8, 22]

$$C_{vh}(\vec{x}) = h^{-1}\left(\sum_{i=1}^{n}\left[h(x_{(i)}) - h(x_{(i-1)})\right]v(H_i)\right). \tag{10}$$

The coefficients of the fuzzy measure can be fitted by linearizing, similarly to the case of quasi-arithmetic means and generalized OWA operators (by applying h to $\vec{x}_k$ and to y_k in (9)).

4.4 γ-operators

We consider a generalized version of γ-operators by Zimmermann [24, 25], which are called T-S operators in [19], defined as a linear or log-linear combination of a t–norm T and t–conorm S,

$$f(\vec{x}) = \gamma T(\vec{x}) + (1-\gamma)S(\vec{x}),$$

$\gamma \in [0,1]$, or

$$f(\vec{x}) = T(\vec{x})^{\gamma} S(\vec{x})^{1-\gamma}.$$

A more general version is obtained by using an invertible strictly monotone function h

$$f(\vec{x}) = h^{-1}\left(\gamma h(T(\vec{x})) + (1-\gamma)h(S(\vec{x}))\right).$$

The linear and log-linear combinations are the special cases corresponding to $h = Id$ and $h = \log$.

We consider the general case with an arbitrary strictly monotone function h. For a fixed pair of t–norm and t–conorm, the goal is to identify an unknown parameter γ that fits the data best. This is done by using $w_1 = \gamma, w_2 = 1-\gamma$ and writing

$$h^{-1}(f(\vec{x})) = h^{-1}(w_1 h(T(\vec{x})) + w_2 h(S(\vec{x}))) =$$

$$h^{-1}(<\vec{g}(\vec{x}), \vec{w}>),$$

$\vec{g} = (h(T), h(S))$, $w_1 + w_2 = 1, w_1, w_2 \geq 0$, and noticing that after linearization we obtain a QP problem again. In this specific case we get

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K}(w_1 h(T(\vec{x}_k)) + w_2 h(S(\vec{x}_k)) - h(y_k))^2, \\ \text{s.t.} \quad & w_1(h(T(\vec{x}_{k+1})) - h(T(\vec{x}_k))) + \\ & w_2(h(S(\vec{x}_{k+1})) - h(S(\vec{x}_k))) \geq 0, \\ & k = 1, \ldots, K-1, \\ & w_1 + w_2 = 1, w_1, w_2 \geq 0. \end{aligned} \tag{11}$$

4.5 General aggregation operators

A method of fitting general aggregation operators using tensor-product splines was proposed in [2,3]. This method is based on representing f by means of a linear combination $f(\vec{x}) =< \vec{B}(\vec{x}), \vec{c} >$, where functions $\vec{B} = \vec{B}_1(x_1)\vec{B}_2(x_2)\ldots\vec{B}_n(x_n)$ are tensor products of univariate B-splines with respect to each variable [9]. For explicit formulae we refer to [2–4]. For our discussion we only need to note that monotone tensor product splines are linear combinations of some well defined basis functions, and that the conditions of monotonicity translate into a system of linear inequalities on spline coefficients. Thus fitting tensor-product splines to the data involves a QP problem (or LP problem if we use LDA criterion).

Preservation of the ordering of the outputs, as we know, is an additional system of linear inequalities, that does not change the structure of QP or LP, thus the methods presented in [3, 4] can be applied with only a minor modification.

4.6 Fitting additive generators of t–norms/t–conorms

A method of fitting continuous Archimedean t–norms/t–conorms to empirical data was presented in [3–5]. It relies on fitting the additive generators, as pointwise convergence of a sequence of additive generators is equivalence to uniform convergence of the corresponding t–norms/t–conorms [15, 17]. In this method an additive generator is represented via a monotone spline

$$h(t) =< \vec{B}(t), \vec{c} >,$$

where $\vec{B}(t)$ is a vector of B-splines, and $\vec{c}$ is the vector of spline coefficients. The conditions of monotonicity of h are imposed through linear restrictions on spline coefficients, and the additional conditions $h(0) = 1, h(0.5) = 1$ also translate into linear equality constraints [2].

Since Archimedean t–norms satisfy

$$T(\vec{x}) = h^{(-1)}(\sum_{i=1}^{n} h(x_i)),$$

($h^{(-1)}$ denotes pseudoinverse), after linearization the least squares criterion translates into

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K}\left(\sum_{i=1}^{n} < \vec{B}(x_{kn}), \vec{c} > - < \vec{B}(y_k), \vec{c} >\right)^2 \\ \text{s.t.} \quad & \text{linear restrictions on } \vec{c}. \end{aligned} \tag{12}$$

By rearranging the terms of the sum we get

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K}\left(< \left[\sum_{i=1}^{n}\vec{B}(x_{kn}) - \vec{B}(y_k)\right], \vec{c} >\right)^2 \\ \text{s.t.} \quad & \text{linear restrictions on } \vec{c}. \end{aligned} \tag{13}$$

[2] The issue of asymptotic behaviour near $t = 0$ for strict Archimedean t–norms is solved by using "well-founded" generators [5, 15]

Next we add preservation of outputs ordering conditions, to obtain the following QP (note that the sign of inequality has changed because h is decreasing)

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K} \left(< \left[\sum_{i=1}^{n} \vec{B}(x_{kn}) - \vec{B}(y_k) \right], \vec{c} > \right)^2 \\ \text{s.t.} \quad & < \left[\sum_{i=1}^{n} \vec{B}(x_{k+1,n}) - \sum_{i=1}^{n} \vec{B}(x_{k,n}) \right], \vec{c} > \quad \leq 0, \\ & \text{linear restrictions on } \vec{c}. \end{aligned} \tag{14}$$

For t–conorms we obtain a similar problem by duality. Furthermore, a very similar procedure works for representable uninorms and nullnorms. An additional issue here is proper dealing with the neutral element/annihilator, and its identification from the data. It was resolved in [3–5], and fortunately, preservation of output ordering does not change the structure of those methods either, it only adds $K-1$ additional linear constraints.

5 Balancing ordering and fitting numerical values

In the preceding discussion we specified preservation of the output orderings as hard constraints, enforced at the expense of fitting to the data. Since empirical data has an associated noise, it may be impossible to satisfy all these constraints by using a specified class of aggregation operators. The system of constraints is said to be inconsistent. In this section we discuss modifications of the above mentioned optimization problems, that allow one to soften ordering constraints and balance them against fitting numerical data.

Consider a revised version of Problem (3).

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K} (< \vec{g}(\vec{x}_k), \vec{w} > - y_k)^2 + \\ & P \sum_{k=1}^{K-1} \max\{< \vec{g}(\vec{x}_k) - \vec{g}(\vec{x}_{k+1}), \vec{w} >, \, 0\}, \\ & \text{other linear conditions on } \vec{w}. \end{aligned} \tag{15}$$

Here P is the penalty parameter, for small values of P we emphasize fitting the numerical data, while for large values of P we emphasize preservation of ordering. Of course, the second sum may not be zero at the optimum, which indicates inconsistency of constraints.

Unfortunately, Problem (15) is no longer a quadratic programming problem, it is a nonsmooth but convex optimization problem, and there are efficient numerical methods of its solution, e.g., [1]. However, for LDA criterion, we can preserve the structure of LP, namely we convert Problem (4), by using auxiliary variables r_k^+, r_k^- and $q_k = \max\{< \vec{g}(\vec{x}_k) - \vec{g}(\vec{x}_{k+1}), \vec{w} >, \, 0\}$ into

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K} r_k^+ + r_k^- + P \sum_{k=1}^{K-1} q_k \\ \text{s.t.} \quad & < \vec{g}(\vec{x}_k), \vec{w} > - r_k^+ + r_k^- = y_k, \\ & k = 1, \ldots, K, \\ & q_k + < \vec{g}(\vec{x}_{k+1}) - \vec{g}(\vec{x}_k), \vec{w} > \quad \geq 0, \\ & k = 1, \ldots, K-1, \\ & q_k, r_k^+, r_k^- \geq 0, k = 1, \ldots, K, \\ & \text{other linear conditions on } \vec{w}. \end{aligned} \tag{16}$$

The special cases we considered in Section 4 allow such an LP formulation, and we note that the dimension of the problem increases only by K, which is not excessively large.

6 Conclusion

Fitting various families of aggregation operators to empirical data is useful for identifying the most suitable aggregation operator in practical applications. It was argued in [16] that preserving the ordering of the output values is more important than fitting actual numerical values, as human subjects — sources of such data, are more consistent with ordering the alternatives than numerical values. The authors of [16] examined a number of classes of aggregation operators and established that no class of that group preserved the ordering of outputs. However they did not set up a suitable optimization problem which would force fitted aggregation operators to preserve the outputs ordering.

In this contribution we developed a general mathematical programming problem which includes preservation of ordering as hard and soft constraints. In the first case, unless the constraints are inconsistent, our method guarantees that the ordering of the outputs is preserved. In the second case, the ordering requirement is balanced against fitting the numerical values, and a solution that minimizes discrepancy of the orderings is delivered.

We have presented specific problem formulations applicable to several broad and popular classes of aggregation operators, and in all cases kept the structure of the optimization problem, either a quadratic or linear programming problem. The advantage is that standard efficient methods of solution are applied.

Acknowledgement

This work was supported by the projects MTM2004-3175, BFM2003-05308, TIC2002-11942-E and TIC2003-09001-C02-02 from Ministerio de Educación y Ciencia, Spain, and PRIB-2004-9250, Govern de les Illes Balears.

References

[1] A. Bagirov. A method for minimization of quasidifferentiable functions. *Optimization Methods and Software*, 17:31–60, 2002.

[2] G. Beliakov. Shape preserving approximation using least squares splines. *Approximation theory and applications*, 16:80–98, 2000.

[3] G. Beliakov. Monotone approximation of aggregation operators using least squares splines. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10:659–676, 2002.

[4] G. Beliakov. How to build aggregation operators from data? *Int. J. Intelligent Systems*, 18:903–923, 2003.

[5] G. Beliakov. Fitting triangular norms to empirical data. In P.E. Klement and R. Mesiar, editors, *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*, pages 255–265. Elsevier, New York, 2005.

[6] G. Beliakov. Learning weights in the generalized OWA operators. *Fuzzy Optimization and Decision Making*, 4:119–130, 2005.

[7] P. Bloomfield and W.L. Steiger. *Least Absolute Deviations. Theory, Applications and Algorithms.* Birkhauser, Boston, Basel, Stuttgart, 1983.

[8] T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. Aggregation operators: properties, classes and construction methods. In T. Calvo, G. Mayor, and R. Mesiar, editors, *Aggregation Operators. New Trends and Applications*, pages 3–104. Physica-Verlag, Heidelberg, New York, 2002.

[9] P. Dierckx. *Curve and Surface Fitting with Splines.* Clarendon press, Oxford, 1995.

[10] D. Filev and R. Yager. On the issue of obtaining OWA operator weights. *Fuzzy Sets and Systems*, 94:157–169, 1998.

[11] E. M. Gertz and S. J. Wright. Object-oriented software for quadratic programming. *ACM Trans. Math. Software*, 29:58–81, 2003.

[12] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.

[13] M. Grabisch. The Choquet integral as a linear interpolator. In *10th Int. Conf. on Information Processing and Management of Uncertainty*, pages 373–378, Perugia, Italy, 2004.

[14] M. Grabisch, H.T. Nguyen, and E.A. Walker. *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference.* Kluwer, Dordrecht, 1995.

[15] S. Jenei. On Archimedian triangular norms. *Fuzzy Sets and Systems*, 99:179–186, 1998.

[16] U. Kaymak and H.R. van Nauta Lemke. Selecting an aggregation operator for fuzzy decision making. In *3rd IEEE Intl. Conf. on Fuzzy Systems*, volume 2, pages 1418–1422, 1994.

[17] E.P. Klement, R. Mesiar, and E. Pap. *Triangular Norms.* Kluwer, Dordrecht, 2000.

[18] C. Lawson and R. Hanson. *Solving Least Squares Problems.* SIAM, Philadelphia, 1995.

[19] A. Pradera, E. Trillas, and T. Calvo. A general class of triangular norm-based aggregation operators: quasi-linear T-S operators. *Intl. J. of Approximate Reasoning*, 30:57–72, 2002.

[20] V. Torra. Learning weights for the quasi-weighted means. *IEEE Trans. on Fuzzy Syst.*, 10:653–666, 2002.

[21] V. Torra. OWA operators in data modeling and reidentification. *IEEE Trans. on Fuzzy Syst.*, 12:652–660, 2004.

[22] R. Yager. Generalized OWA aggregation operators. *Fuzzy Optimization and Decision Making*, 3:93–107, 2004.

[23] R. Yager and J. Kacprzyk, editors. *The Ordered Weighted Averaging Operators. Theory and Applications.* Kluwer, Boston, 1997.

[24] H.-J. Zimmermann. *Fuzzy Set Theory - and its Applications.* Kluwer, Boston, 1996.

[25] H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4:37–51, 1980.

Fitting ST-OWA operators to empirical data

Gleb Beliakov
Deakin University
Australia
gleb@deakin.edu.au

Luigi Troiano
RCOST - University of Sannio
82100 Benevento - Italy
troiano@unisannio.it

Abstract

The OWA operators gained interest among researchers as they provide a continuum of aggregation operators able to cover the whole range of compensation between the minimum and the maximum. In some circumstances, it is useful to consider a wider range of values, extending below the minimum and over the maximum. ST-OWA are able to surpass the boundaries of variation of ordinary compensatory operators. Their application requires identification of the weighting vector, the t-norm, and the t-conorm. This task can be accomplished by considering both the desired analytical properties and empirical data.

Keywords: Aggregation operators, OWA, t–norms, ST-OWA.

1 Introduction

The *Ordered Weighted Averaging* (OWA) operator [19] gained interest among researchers due to its property of providing a parameterized aggregation operation that includes the min, max and arithmetic mean. Due to this ability, OWA operators have been experimented in many problems regarding decision making, information retrieval and information fusion [6, 18, 20].

An OWA operator is defined as

$$M_{[\mathbf{w}]}(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_{(i)} \tag{1}$$

where $a_{(1)}, \ldots, a_{(n)}$ is a non-increasing permutation of arguments $a_1, \ldots, a_n$, so that $a_{(i)} \geq a_{(j)} \forall i < j$.

The weighting vector $\mathbf{w} = (w_1, \ldots, w_n)$ provides the parametrization of OWA operators. Weights are such that

$$w_i \in [0, 1] \tag{2}$$

$$\sum_{i=1}^{n} w_i = 1 \tag{3}$$

Some notable examples of the weighting vector are:

- $M_{[1,0,\ldots,0]}(a_1, \ldots, a_n) = \max_{i=1..n} a_i$
- $M_{[1/n,\ldots,1/n]}(a_1, \ldots, a_n) = \frac{1}{n} \sum_{i=1}^{n} a_i$
- $M_{[0,\ldots,0,1]}(a_1, \ldots, a_n) = \min_{i=1..n} a_i$

The main property of OWA operators is that they are *internal* (i.e., compensatory, averaging) operators, as

$$\min_{i=1..n} a_i \leq M_{[\mathbf{w}]}(a_1, \ldots, a_n) \leq \max_{i=1..n} a_i \tag{4}$$

for any weighting vector $\mathbf{w}$. This property is useful in many applications as it allows to compensate lower inputs with higher inputs. Nevertheless, in some circumstances it can be limitative because (i) there is an interaction between the arguments, or (ii) some reinforcement of higher and lower scores is required together with their compensation.

To overcome these limitations, the families of T-OWA, S-OWA and ST-OWA have been proposed and investigated. We can rewrite Eq.(1) as

$$M_{[\mathbf{w}]}(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i \min(a_{(1)}, \ldots, a_{(i)}) \tag{5}$$

This suggests that Eq.(1) can be generalized by using a generic t-norm T [11], as

$$M_{T[\mathbf{w}]}(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i T(a_{(1)}, \ldots, a_{(i)}) \tag{6}$$

This is the T-OWA operator introduced in [21]. In this case, the operator is provided with an additional

parameter which is a t-norm T. Different t-norms can be used. Usual choices include the minimum (min), the product (T_P), the Lukasiewicz t-norm (T_L), and the drastic t-norm (T_D). Other possibilities entail the use of parametric t-norms, such as Schweizer and Sklar's t-norms ($T_{SS(p)}$) and Yager's t-norms ($T_{Y(v)}$) [11].

Eq.(1) can also be rewritten as

$$M_{[\mathbf{w}]}(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i \max(a_{(i)}, \ldots, a_{(n)}), \quad (7)$$

which leads to the definition of S-OWA operators [21] as

$$M_{S[\mathbf{w}]}(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i S(a_{(i)}, \ldots, a_{(n)}). \quad (8)$$

The additional parameter is provided by a t-conorm S. The usual choices are the maximum (max), the probabilistic sum (S_P), the Lukasiewicz t-conorm (S_L), and the drastic t-conorm (S_D). Also in this case, it is possible to use parametric t-conorms such as Schweizer and Sklar's t-conorms ($S_{SS(p)}$) and the Yager's t-conorm ($S_{Y(v)}$).

The S-OWA and T-OWA operators extend the range of values provided by the ordinary OWA operators, as depicted in Fig.1.

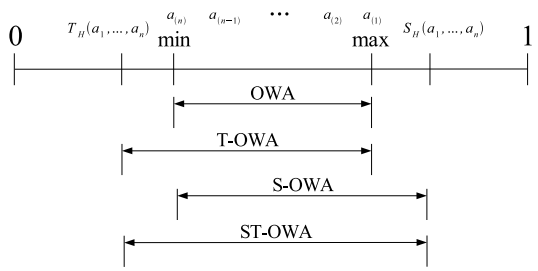


Figure 1: Range of values

In particular, the T-OWA extends the range of values below the minimum, whilst the S-OWA extends the range above the maximum. To extend the range in both directions (see Fig.1), the ST-OWA has been defined as [17]

$$M_{ST[\mathbf{w}]}(a_1, ..., a_n) = \sum_{i=1}^{n} w_i((1-\sigma)T(a_{(1)}, ..., a_{(i)}) + \sigma S(a_{(i)}, ..., a_{(n)})), \quad (9)$$

where $\sigma \in [0, 1]$ is the OWA attitudinal character, as described in the following section. In this case, the t-norm T and the t-conorm S provide the two parameters additional to the weighting vector $\mathbf{w}$.

An important issue for practical applicability of ST-OWA operators is the identification of parameters, which aims at properly choosing the t-norm, the t-conorm and the weighting vector. This task can be performed by considering both the desired analytical properties and empirical data. In particular, when parametric t–norms are considered, the identification can be translated into an optimization problem aimed at finding the parameter(s) of a t-norm (t-conorm) and the OWA weighting vector that best fit the empirical data. The remainder of this paper is devoted to the solution of this problem. Section 2 provides an overview of the main analytical properties of ST-OWA operators. Section 3 deals with the identification problem. Section 4 describes an example of an application. Section 5 briefly outlines conclusions and future work.

2 Properties

Different weighting vectors entail different emphasis to higher and lower input values. This is described by the *attitudinal character*, also known as the *orness measure*, which is a function of weights defined as

$$AC(\mathbf{w}) = \sum_{i=1}^{n} w_i \frac{n-i}{n-1} \in [0, 1]. \quad (10)$$

In particular

$$AC(\mathbf{w}) = \begin{cases} 1 & \text{if } w_1 = 1, w_{i\neq 1} = 0, \\ 0.5 & \text{if } w_i = 1/n, \\ 0 & \text{if } w_n = 1, w_{i\neq n} = 0. \end{cases} \quad (11)$$

The attitudinal character itself can be computed by using OWA operators.

Proposition 1.

$$AC(\mathbf{w}) = M_{[\mathbf{w}]}(1, \tfrac{n-2}{n-1}, \ldots, \tfrac{1}{n-1}, 0). \quad (12)$$

As discussed in [21], Prop.1 suggests a way for computing the attitudinal character of T-OWA operators as

$$AC(\mathbf{w}, T) = M_{T[\mathbf{w}]}(1, \tfrac{n-2}{n-1}, \ldots, \tfrac{1}{n-1}, 0). \quad (13)$$

In a similar way we can compute the attitudinal character of S-OWA operators as

$$AC(\mathbf{w}, S) = M_{S[\mathbf{w}]}(1, \tfrac{n-2}{n-1}, \ldots, \tfrac{1}{n-1}, 0), \quad (14)$$

and for ST-OWA operators, as

$$AC(\mathbf{w}, S, T) = M_{ST[\mathbf{w}]}(1, \tfrac{n-2}{n-1}, \ldots, \tfrac{1}{n-1}, 0). \quad (15)$$

Moreover, for any choice of T and S, we get

$$AC(\mathbf{w}, S, T) = (1-\sigma)AC(\mathbf{w}, T) + \sigma AC(\mathbf{w}, S), \quad (16)$$

where $\sigma = AC(\mathbf{w})$.

T-norms and t-conorms can be compared with respect to the aggregated value they provide.

Definition 1. *Given two t-norms T_1 and T_2, T_1 is* stronger *than T_2 iff*

$$T_1(x,y) \geq T_2(x,y)\ \forall x,y \in [0,1]. \tag{17}$$

In this case we write $T_1 \geq T_2$.

It can be easily proven that $T_D \leq T_L \leq T_P \leq \min$. The same definition can be applied to t-conorms as well. In this case, it results in $\max \leq S_P \leq S_L \leq S_D$. There is a kind of symmetry between t-norms and t-conorms, due to the duality. With respect to parametric families of t-norms (t-conorms), we note that the values of the parameter provide a natural ordering for some families (although this is not true in general). For instance, with respect to Yager's t-norm and t-conorm, defined as

$$T_{Y(v)}(x,y) = 1 - \min(1, ((1-x)^v + (1-y)^v)^{\frac{1}{v}}), \tag{18}$$

$$S_{Y(v)}(x,y) = \min(1, (x^v + y^v)^{\frac{1}{v}}), \tag{19}$$

increasing values of the parameter v entail stronger t-norms (weaker t-conorms). It is obvious that t-conorms are stronger than t-norms.

The order relation between two t-norms (t-conorms), leads to ordering of T-OWA (S-OWA) operators. Indeed

Proposition 2. *Given two t-norms (or t-conorms) R_1 and R_2, such that $R_1 \geq R_2$, it holds*

$$M_{R_1[\mathbf{w}]}(a_1,\ldots,a_n) \geq M_{R_2[\mathbf{w}]}(a_1,\ldots,a_n) \tag{20}$$

for any weighting vector $\mathbf{w}$, and any arguments $a_1,\ldots,a_n$. Then we write

$$M_{R_1} \geq M_{R_2}. \tag{21}$$

For any weighting vector $\mathbf{w}$, it holds

$$M_{T_D} \leq M_{T_L} \leq M_{T_P} \leq M \leq M_{S_P} \leq M_{S_L} \leq M_{S_D}. \tag{22}$$

For any, weighting vector $\mathbf{w}$, if $T_1 \leq T_2$ ($S_1 \geq S_2$) then

$$\begin{aligned} &AC(\mathbf{w},T_1) \leq AC(\mathbf{w},T_2), \\ &(AC(\mathbf{w},S_1) \geq AC(\mathbf{w},S_2)). \end{aligned} \tag{23}$$

Proposition 3. *For any t-norm T and t-conorm S,*

$$AC(\mathbf{w},T) \leq AC(\mathbf{w}) \leq AC(\mathbf{w},S). \tag{24}$$

An important property of the OWA operators (as well as any aggregation operator), is their *monotonicity* with respect to arguments.

Proposition 4. *Given two vectors of n arguments, $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$, such that $a_i \geq b_i \forall i = 1..n$, it holds*

$$M_{[\mathbf{w}]}(a_1,\ldots,a_n) \geq M_{[\mathbf{w}]}(b_1,\ldots,b_n) \tag{25}$$

for any weighting vector $\mathbf{w}$.

Duality of t-norms and OWA operators can be translated to the ST-OWA setting. Recall that the dual weighting vector $\hat{\mathbf{w}}$ is obtained from $\mathbf{w}$ using $\hat{w}_i = w_{n-i+1}$. Similarly to ordinary OWA operators, we can easily prove the following proposition.

Proposition 5. *Given a pair of weighting vectors $\hat{w}$ and w, if $\hat{w}$ is dual to w, and S is dual to T, then $M_{ST[\hat{\mathbf{w}}]}$ and $M_{ST[\mathbf{w}]}$ are dual.*

Definition 2. *Given an aggregation operator $R_{\mathbf{w}}$ which depends on the weighting vector $\mathbf{w}$, and given a vector $\mathbf{a} \in [0,1]^n$, its variation range at $\mathbf{a}$ is the interval $r(\mathbf{a}) = [\inf_{\mathbf{w}} R_{\mathbf{w}}(\mathbf{a}), \sup_{\mathbf{w}} R_{\mathbf{w}}(\mathbf{a})]$. The aggregation operator $R_{\mathbf{w}}$ has a greater variation range than $S_{\mathbf{w}}$ if $s(\mathbf{a}) \subset r(\mathbf{a})$ for all $\mathbf{a} \in [0,1]^n$.*

If the dual pair is $< \min, \max >$ we get the ST-OWA with the minimal variation range: it corresponds to the case of ordinary OWA operators. If the dual pair is $< T_D, S_D >$, we get the ST-OWA with maximum variation. In the case of a dual pair $< T, S >$, the attitudinal character is

$$\begin{aligned} AC(\mathbf{w},S,T) &= (1-\sigma)AC(\mathbf{w},T) + \sigma(1 - AC(\hat{\mathbf{w}},T)) = \\ &= (1-\sigma)(1 - AC(\hat{\mathbf{w}},S)) + \sigma AC(\mathbf{w},S)) \end{aligned} \tag{26}$$

In particular,

$$AC(\mathbf{w},S,T) = \begin{cases} 0, & \sigma = 0 \\ 1/2, & \sigma = 1/2 \\ 1, & \sigma = 1 \end{cases} \tag{27}$$

Example 1. *Let us consider $\mathbf{a} = [0.9, 0.7, 0.8, 0.9, 0.8]$. The range of variation by varying $\sigma \in [0,1]$ for an ordinary OWA is*

$$< \min, \max >: [0.700, 0.900]$$

If we choose a different pair $< T, S >$, the range of variation at $\mathbf{a}$ becomes

$$\begin{aligned} < T_P, S_P >&: [0.363, 0.999] \\ < T_L, S_L >&: [0.100, 1.000] \\ < T_D, S_D >&: [0.000, 1.000] \end{aligned}$$

In particular, if we choose $\mathbf{w} = [0.25, 0.25, 0.5, 0, 0]$, then

$$\sigma = AC(\mathbf{w}) = 0.687$$

and

$$\begin{aligned} AC(\mathbf{w},S_P,T_P) &= 0.738 \\ AC(\mathbf{w},S_L,T_L) &= 0.718 \\ AC(\mathbf{w},S_D,T_D) &= 0.824 \end{aligned}$$

3 Operator identification

In this section we consider various instances of the problem of fitting parameters of ST-OWA to empirical data. We assume that there is a set of input-output pairs $\mathcal{D} = \{(\mathbf{x}_k, y_k)\}, k = 1, \dots K$, with $\mathbf{x}_k \in [0,1]^n$, $y_k \in [0,1]$ and n is fixed. Our goal is to determine parameters S, T, $\mathbf{w}$ which fit the data best.

3.1 Identification with fixed S and T

In this instance of the problem we assume that both S and T have been specified. The issue is to determine the vector $\mathbf{w}$. For S-OWA and T-OWA, fitting the data in the least squares sense involves a solution to a quadratic programming problem (QP)

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K} \left(\sum_{i=1}^{n} w_i S(x_{(i)k}, \dots, x_{(n)k}) - y_k \right)^2 \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i = 1, w_i \geq 0, \end{aligned} \tag{28}$$

and similarly for the case of T-OWA. We note that the values of S at any $\mathbf{x}_k$ are fixed (do not depend on $\mathbf{w}$). This problem is very similar to that of calculating the weights of standard OWA operators from data [2, 4, 7, 16], but involves a fixed function $S(x_{(i)k}, \dots, x_{(n)k})$ rather than just $x_{(i)k}$.

If an additional requirement is to have a specified value of $AC(\mathbf{w}, S)$ and $AC(\mathbf{w}, T)$, then it becomes just an additional linear constraint, which does not change the structure of QP problem (28).

Next, consider fitting ST-OWA. Here, for a fixed value of $AC(\mathbf{w}) = \sigma$, we have the QP problem

$$\begin{aligned} \text{Minimize} \quad & \sum_{k=1}^{K} \left(\sum_{i=1}^{n} w_i ST(\mathbf{x}_k, \sigma) - y_k \right)^2 \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i = 1, w_i \geq 0, \\ & AC(\mathbf{w}) = \sigma, \end{aligned} \tag{29}$$

where

$$ST(\mathbf{x}, \sigma) = (1-\sigma)T(x_{(1)}, \dots, x_{(i)}) + \sigma S(x_{(i)}, \dots, x_{(n)}).$$

However σ may not be specified, and hence has to be also found from the data. In this case, we present a bi-level optimization problem, in which at the outer level nonlinear (possibly global) optimization is performed with respect to parameter σ, and at the inner level the problem (29) with a fixed σ,

$$\text{Minimize}_{\ \sigma \in [0,1]} \left[F(\sigma) \right], \tag{30}$$

where $F(\sigma)$ denotes the solution to (29).

We note that the inner quadratic problem (29) has a unique global minimum, which is easily found by using standard numerical methods, see [8, 12]. If a global optimization method is applied to the outer problem with respect to σ, then the globally optimal solution to (30) with respect to both σ and $\mathbf{w}$ will be obtained. Numerical solution to the outer problem with just one variable can be performed by a number of methods, including grid search, multistart local search, or Pijavski-Shubert method [9, 13].

3.2 Identification of T-OWA and S-OWA

Consider now the problem of fitting parameters of the parametric families of participating t–norm and t–conorm, simultaneously with $\mathbf{w}$ and σ. With start with S-OWA, and assume a suitable family of t–norms T has been chosen, e.g., Yager t–conorms $S_{Y(v)}$ parameterized with v. We will rely on efficient solution to problem (28) with a fixed S (i.e., fixed v). We set up a bi-level optimization problem

$$\text{Minimize}_{\ v \in [0,\infty]} \left[F_1(v) \right],$$

where $F_1(v)$ denotes solution to (28).

The outer problem is nonlinear, possibly global optimization problem, but because it has only one variable, its solution is relatively simple. We recommend Pijavski-Shubert deterministic method [13]. Identification of T is performed analogously.

The advantage of using bi-level optimization is that the nonlinear parameter (v) is separated from the vector of weights, which is found by solving a standard QP. Hence for the nonlinear problem we have just one variable, and for multivariate problem with respect to $\mathbf{w}$ we have a special structure, accommodated by efficient QP algorithms. Since the inner QP problem is convex, it has a unique global minimum, and the whole problem with respect to all parameters is solved to global optimality.

Next consider fitting ST-OWA operators. Here we have three parameters: the two parameters of the participating t–norm and t–conorm, which we will denote by v_1, v_2, and σ as in Problem (29). Of course, T and S may be chosen as dual to each other, in which case we have to fit only one parameter $v = v_1 = v_2$. To use the special structure of the problem with respect to $\mathbf{w}$ we again set up a bi-level optimization problem analogously to (30).

$$\text{Minimize}_{\ \sigma \in [0,1], v_1, v_2 \geq 0} \left[F(\sigma, v_1, v_2) \right], \tag{31}$$

where $F(\sigma, v_1, v_2)$ is the solution to QP problem

$$\text{Minimize} \quad \sum_{k=1}^{K}\left(\sum_{i=1}^{n} w_i ST(\mathbf{x}_k,\sigma,v_1,v_2)-y_k\right)^2 \quad (32)$$
$$\text{s.t.} \quad \sum_{i=1}^{n} w_i = 1, w_i \geq 0,$$
$$AC(\mathbf{w}) = \sigma,$$

and

$$ST(\mathbf{x},\sigma,v_1,v_2) = (1-\sigma)T_{Y(v_1)}(x_{(1)},\ldots,x_{(i)}) + \sigma S_{Y(v_2)}(x_{(i)},\ldots,x_{(n)}).$$

Solution to the outer problem is complicated because of the possibility of numerous local minima. One has to rely on methods of multivariate global optimization [9,14]. One such (deterministic) method is the Cutting Angle Method (CAM) developed in [1,3,15]. It allows one to solve efficiently global optimization problems in up to 10 variables.

3.3 Least absolute deviation problem

Besides the least squares approach, fitting to the data can be performed by using the Least Absolute Deviation (LDA) criterion [5], by replacing the sum of squares in (28) and (29) with the sum of absolute values. It is argued that the LDA criterion is less sensitive to outliers.

In this case the optimization problems are converted to linear programming (LP) problems by introducing auxiliary non-negative variables r_k^+, r_k^-, such that

$$r_k^+ - r_k^- = \sum_{i=1}^{n} w_i ST(\mathbf{x}_k,\sigma) - y_k,$$

and

$$r_k^+ + r_k^- = \left|\sum_{i=1}^{n} w_i ST(\mathbf{x}_k,\sigma) - y_k\right|.$$

This conversion is well known, see [5]. The counterparts of problems (28) and (29) become LP problems, which are easily solved by the simplex method. The outer nonlinear optimization problems do not change.

3.4 Preservation of ordering of the outputs

In [10] it was argued that fitting the numerical outputs is not as important as preserving the ordering of the outputs. The empirical data usually comes from human subjective evaluation, and people do not reliably express their preference on a numerical scale. In contrast, people are very good at ranking the alternatives. Therefore, the authors of [10] argued that fitting methods should aim at preserving the order of empirical output values. They showed that various methods of fitting the numerical values do not preserve this ordering.

Without loss of generality, we assume that the outputs are ordered as $y_1 \leq y_2 \leq \ldots \leq y_K$. The condition for order preservation is

$$M_{ST[\mathbf{w}]}(\mathbf{x}_k) \leq M_{ST[\mathbf{w}]}(\mathbf{x}_{k+1}), \text{ for all } k = 1,\ldots,K-1. \quad (33)$$

Because $M_{ST[\mathbf{w}]}$ depends on $\mathbf{w}$ linearly for a fixed σ, (33) is a system of linear inequalities, which does not change the structure of the QP or LP problems.

For example, problem (28) will have additional $K-1$ linear constraints for $k = 1,\ldots,K-1$:

$$\sum_{i=1}^{n} w_i \left(S(x_{(i)k+1},\ldots,x_{(n)k+1}) - S(x_{(i)k},\ldots,x_{(n)k})\right) \geq 0.$$

Problem (29) will have the constraints

$$\sum_{i=1}^{n} w_i \left(ST(\mathbf{x}_{k+1},\sigma) - ST(\mathbf{x}_k,\sigma)\right) \geq 0.$$

4 Numerical experiments and examples

As a first step of testing the correctness and suitability of the mentioned algorithms for determination of ST-OWA parameters, we have generated random data $\mathbf{x}_k, k = 1,\ldots,20$ and the values of y_k, computed by a model S-, T- and ST-OWA aggregation operators in 3-5 variables (i.e., with fixed weights and fixed participating t–norm and t–conorm from Yager, Hamacher and Frank families). Then we used these data to calculate:

- The weighting vector $\mathbf{w}$ of S-, T- and ST-OWA operators with known parameter(s) v (v_1, v_2).
- The weighting vector and the unknown parameter v of S- and T-OWA operators.
- The weighting vector and the unknown parameter v of S- and T-OWA operators with a given attitudinal character.
- The weighting vector and the unknown parameters v_1, v_2 of ST-OWA operators with a given attitudinal character σ.
- The weighting vector and the unknown parameters v_1, v_2 and σ of ST-OWA operators.

In the test cases we also included the limiting cases of ST-OWA being a pure t–norm, t–conorm or OWA. All our experiments were successful, the correct values of the parameters used in the models used to simulate the data have been found. The computing time for S-, T-

and ST-OWA operators with given v (v_1, v_2) was below 1 sec (Pentium IV 2 GHz workstation), and when parameters v, v_1, v_2, σ also had to be fitted to the data, the computing time was < 10 sec. These experiments show the robustness and efficiency of the proposed algorithms.

The algorithm robustness is provided by the non-sensitivity of precision to structural attributes, such as the number of criteria n or by the number of input values K. Indeed, the non-sensitivity of precision to the number of criteria n, guarantees the algorithm to be able to converge also with a growing dimension of the input space. In order to verify this hypothesis we considered, for each $n = 3..10$, 20 random samples, each made of 20 n-ples of input value; the input values were aggregated assuming random value of the weighting vector and t–norm/t–conorm parameters. For each sample we collected the RMSE and coefficients between the observed and calculated values. ANOVA procedure showed the null hypothesis (i.e. H_0=*there at least two samples whose RMSE, or correlation coefficient, are statistically different*) can be rejected ($p < 0.01$). This resulted for T-, S- and ST-OWA with respect to Frank, Hamacher and Yager norms. Similarly, we assumed $n = 3$ and for $K = 10, 20, \ldots, 100$ we considered 20 random samples. Again, ANOVA procedure showed that RMSE and correlation coefficients are not statistically different ($p < 0.01$).

An interesting property that emerged from experimentation was the ability of the algorithm to compensate noise, as depicted by Table 1.

Run	Noise RMSE	Approx. RMSE
1	0.05312	0.00955
2	0.06228	0.01205
3	0.05512	0.00996
4	0.06213	0.01160
5	0.05315	0.00979
6	0.04991	0.01285
7	0.06202	0.01091
8	0.05437	0.01123
9	0.05621	0.01176
10	0.05947	0.01227

Table 1: Noise compensation

For this test we assumed $n = 3$ and $K = 20$. For each input 3-ple we computed the aggregated value using the Hamacher t-norm, given the weighting vector and the t-norm parameter. Then, we added to the aggregated value uniformly distributed errors in the range $[-0.1, 0.1]$. The resulted value was given to the algorithm in order to identify the ST-OWA operator best fitting the data. Table 1 shows the result on this experimentation on a sample of 10 runs. In column 2 we report the RMSE affecting the expected values, while in column 3 the RMSE between the expected and observed values. We can note that the two series are statistically different (the Wilcoxon matched-pairs signed-ranks test on 50 runs confirmed this hypothesis). Therefore, the algorithm behaves in such a way as to identify parameters that are different from the original ones, but that provides an ST-OWA operator able to better approximate the expected values. Similar conclusions can be obtained for S- and T-OWA operators, and for Yager and Frank t-norms.

Next we took the actual experimental data (two data sets of 20 data) from [22, 23] (also replicated in [10]). For both data sets, the root mean square error of approximation were $RMSE_1 = 0.0148$ and $RMSE_2 = 0.0105$, which compares favorably with the fitted standard OWA ($RMSE_1 = 0.015$, $RMSE_2 = 0.011$) and Zimmermann's γ operators ($RMSE_1 = 0.0151$, $RMSE_2 = 0.0105$). The correlation coefficients (between the observed and calculated values) has also been higher ($Corr_1 = 0.985$, $Corr_2 = 0.974$). While the gain in numerical accuracy is not very impressive, we note that for the two data sets, we used different participating t–norm and t–conorm in γ-operators (max/min and T_P/S_P, by trial and error process) and reported the best result, while for ST-OWA operators the parameters were found automatically. We also note that both data sets have only two inputs.

Finally, we used three empirical data sets with 4 inputs, collected by the second author, described as follows. A group of students ($K = 41$) was asked to provide their numerical evaluation of the quality of three objects: public broadcast TV programs, University and town they live in. First they provided an overall score (y) and then scores with respect to four criteria $(x_1, \ldots, x_4)$, namely, quality of programs, advertisement pressure, sports, news (for TV); curriculum, potential for personal growth, quality of labs, other services (University); public events, criminality, cleanliness and services to young people (town). There were no missing data. There were clear outliers (e.g., respondent #14 has provided scores like $\mathbf{x} = (23, 25, 23, 10), y = 85$ despite the overall averaging tendency for the group), however outliers were not removed for the analysis.

For the three data sets we obtained $RMSE_1 = 0.022, RMSE_2 = 0.018$ and $RMSE_3 = 0.021$, with correlation between computed and observed values $0.76, 0.82, 0.79$ respectively. These values were the best we obtained among OWA, γ-operators (using min/max and T_P/S_D), various families of t–norms, uninorms, generalized means and Choquet integrals, although weighted generalized means gave close results. This shows the potential of ST-OWA operators in fitting experimental data.

Calculations were performed using the AOTool software available from `http://www.deakin.edu.au/~gleb/aotool.html`.

5 Conclusions and future work

We have studied the problem of identification of ST-OWA aggregation operators from empirical data. ST-OWA provide a broader range of behavior, from conjunctive via averaging to disjunctive, within one class of aggregation operators, parameterized with T, S and weighting vector $\mathbf{w}$. They are useful in the situations where it is unknown a priori which type of behavior is consistent with the data.

We have presented several mathematical programming formulations of the problem of identifying ST-OWA parameters from empirical data. An interesting feature of these problems is that they are set as nonlinear optimization problems with respect to just one or three parameters, with the vector of weights identified by standard quadratic or linear programming methods. This, of course, greatly reduces computational costs.

Our future line of research is to replace OWA with more general Choquet integral type construction, as well as to consider generalized OWA and nonparametric families of continuous t–norms.

References

[1] G. Beliakov. Geometry and combinatorics of the cutting angle method. *Optimization*, 52:379–394, 2003.

[2] G. Beliakov. How to build aggregation operators from data? *Int. J. Intelligent Systems*, 18:903–923, 2003.

[3] G. Beliakov. The cutting angle method - a tool for constrained global optimization. *Optimization Methods and Software*, 19:137–151, 2004.

[4] G. Beliakov. Learning weights in the generalized OWA operators. *Fuzzy Optimization and Decision Making*, 4:119–130, 2005.

[5] P. Bloomfield and W.L. Steiger. *Least Absolute Deviations. Theory, Applications and Algorithms.* Birkhauser, Boston, Basel, Stuttgart, 1983.

[6] B. Bouchon-Meunier. *Aggregation and Fusion of Imperfect Information.* Physica–Verlag, Heidelberg, 1998.

[7] D. Filev and R. Yager. On the issue of obtaining OWA operator weights. *Fuzzy Sets and Systems*, 94:157–169, 1998.

[8] E. M. Gertz and S. J. Wright. Object-oriented software for quadratic programming. *ACM Trans. Math. Software*, 29:58–81, 2003.

[9] R. Horst, P. Pardalos, and N. Thoai. *Introduction to Global Optimization.* Kluwer Academic Publishers, Dordrecht, 2nd edition, 2000.

[10] U. Kaymak and H.R. van Nauta Lemke. Selecting an aggregation operator for fuzzy decision making. In *3rd IEEE Intl. Conf. on Fuzzy Systems*, volume 2, pages 1418–1422, 1994.

[11] E.P. Klement, R. Mesiar, and E. Pap. *Triangular Norms.* Kluwer, Dordrecht, 2000.

[12] C. Lawson and R. Hanson. *Solving Least Squares Problems.* SIAM, Philadelphia, 1995.

[13] S.A. Pijavski. An algorithm for finding the absolute extremum of a function. *USSR Comput. Math. and Math. Phys.*, 2:57–67, 1972.

[14] J. Pinter. *Global Optimization in Action: Continuous and Lipschitz Optimization–Algorithms, Implementations, and Applications.* Kluwer Academic Publishers, Dordrecht; Boston, 1996.

[15] A.M. Rubinov. *Abstract Convexity and Global Optimization.* Kluwer Academic Publishers, Dordrecht; Boston, 2000.

[16] V. Torra. OWA operators in data modeling and reidentification. *IEEE Trans. on Fuzzy Syst.*, 12:652–660, 2004.

[17] L. Troiano and R.R. Yager. On some properties of mixing OWA operators with t-norms and t-conorms. In *Proc. EUSFLAT 2005*, volume Vol.II, pages 1206–1212, Barcellona, Spain, 2005.

[18] R. Yager and J. Kacprzyk, editors. *The Ordered Weighted Averaging Operators. Theory and Applications.* Kluwer, Boston, 1997.

[19] R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183–190, 1988.

[20] R.R. Yager. Families of OWA operators. *Fuzzy Sets and Systems*, 59:125–148, 1993.

[21] R.R. Yager. Extending multicriteria decision making by mixing t-norms and OWA operators. *Int. J. Intelligent Systems*, 20:453–474, 2005.

[22] H.-J. Zimmermann. *Fuzzy Set Theory - and its Applications.* Kluwer, Boston, 1996.

[23] H.-J. Zimmermann and P. Zysno. Latent connectives in human decision making. *Fuzzy Sets and Systems*, 4:37–51, 1980.

Ordinal Means

Anna Kolesárová
Institute IAM, FCHFT
Slovak University of Technology
Radlinského 9, 812 37 Bratislava
Slovakia
anna.kolesarova@stuba.sk

Gaspar Mayor
Department of Mathematics
and Computer Science
University of the Balearic Islands
Palma de Mallorca
Spain
gmayor@uib.es

Radko Mesiar
Department of Mathematics
and Descriptive Geometry, FCE
Slovak University of Technology
Radlinského 11, 813 68 Bratislava
Slovakia
and
ÚTIA AV ČR, 182 08 Prague,
Czech Republik
mesiar@math.sk

Abstract

The aim of the contribution is the discussion of some types and classes of means on ordinal scales, especially kernel and shift invariant ordinal means, weighted ordinal means based on weighted divisible t–conorms (t–norms) and dissimilarity based ordinal means. Moreover, several types of ordinal arithmetic means are introduced.

Keywords: Ordinal aggregation operator, Ordinal mean, Divisible t–conorm, Dissimilarity function.

1 Introduction

Typical means on the cardinal scale $[0, 1]$ are the arithmetic mean M and several of its generalizations, such as quasi–arithmetic means, weighted arithmetic means, weighted quasi–arithmetic means, OWA operators, weighted ordered quasi–arithmetic means. Recall that OWA operators can be viewed as weighted arithmetic means applied not directly to given input values, but to the ordered ones. Weighted ordered quasi–arithmetic means are defined in a similar way. Moreover, the class of weighted quasi–arithmetic means contains M, and also quasi–arithmetic and weighted arithmetic means.

All till now mentioned means are characterized by weights of single inputs (possibly reordered) and by a transformation function (generator). Several other types of means on $[0, 1]$ are characterized by certain special properties. For example, kernel aggregation operators are means with Chebyshev norm equal to 1 [4, 14, 15]. Shift invariant aggregation operators are means commuting with (acceptable) shifts [24, 29]. Comonotone additivity is the property characterizing means related to the Choquet integral, while max- and min–homogenity characterize the Sugeno integral–based means [6, 28].

In [3] Calvo and Mesiar introduced weighted (continuous) t–conorms, and their class of aggregation functions also covers all weighted quasi–arithmetic means for which 0 is not an annihilator. Observe that by duality weighted (continuous) t–norms can be introduced, covering all weighted quasi–arithmetic means for which the element 1 is not an annihilator, and thus the union of both these classes contains all quasi–arithmetic means.

Ordinal scales become more and more important, especially because of the "computerization" of several branches of human thinking [8, 22, 10, 19, 23]. Thus, practical applications of fuzzy logic are limited to a finite number of truth values. Firstly, technical implementations allow us to work only with a finite (though very large) number of values. Secondly, when representing vagueness, it is usually meaningless to distinguish a high number of truth values; only a small number suffices. Also note that the reasoning with linguistic truth degrees reduces, from a mathematical point of view, to processing on a fixed discrete ordinal scale related to the number of different truth degrees involved.

The aim of this contribution is the discussion of some types and classes of means on ordinal scales (ordinal means, for short). The paper is organized as follows. In the next section kernel and shift invariant ordinal means are defined and some of their properties are studied. In Section 3, the notions of weighted divisible t–conorms and t–norms on the discrete scale are introduced and next exploited for defining the lower and upper ordinal arithmetic (weighted arithmetic) means. Section 4 is devoted to the dissimilarity based ordinal means. Finally, in Conclusion a generalization of the results obtained in Sections 3 and 4 for arithmetic ordinal means is proposed.

2 Kernel and shift invariant ordinal means

Each finite scale with $m+1$ elements can be represented by the scale $L_m = \{0, 1, \ldots, m\}$.

Definition 1 *Let $n \in \mathbb{N}$. A non–decreasing mapping $A : L_m^n \to L_m$ is called an n–ary ordinal mean whenever*

$$\min(x_1, \ldots, x_n) \leq A(x_1, \ldots, x_n) \leq \max(x_1, \ldots, x_n) \tag{1}$$

for all $(x_1, \ldots, x_n) \in L_m^n$. An ordinal mean is a mapping $A : \bigcup_{n\in\mathbb{N}} L_m^n \to L_m$ such that $A_{|L_m^n}$ is an n–ary ordinal mean for each $n \in \mathbb{N}$.

Note that due to the monotonicity of ordinal means, the boundary condition (1) can be replaced by idempotency:

$$A(x, \ldots, x) = x \quad \text{for all } x \in L_m. \tag{2}$$

Definition 2 *An n–ary ordinal mean $A : L_m^n \to L_m$ is an n–ary kernel ordinal mean if for all $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n) \in L_m^n$*

$$\begin{aligned} |A(x_1, \ldots, x_n) &- A(y_1, \ldots, y_n)| \\ &\leq \max(|x_1 - y_1|, \ldots, |x_n - y_n|). \end{aligned} \tag{3}$$

A mapping $A : \bigcup_{n\in\mathbb{N}} L_m^n \to L_m$ is a kernel ordinal mean if $A_{|L_m^n}$ is an n–ary kernel ordinal mean for each $n \in \mathbb{N}$.

Definition 3 *An n–ary ordinal mean $A : L_m^n \to L_m$ is a shift invariant n–ary ordinal mean if*

$$A(x_1 + a, \ldots, x_n + a) = a + A(x_1, \ldots, x_n) \tag{4}$$

for each $a \in \{1, \ldots, m\}$ and all $x_1, \ldots, x_n \in \{0, 1, \ldots, m-a\}$. A mapping $A : \bigcup_{n\in\mathbb{N}} L_m^n \to L_m$ is a shift invariant ordinal mean if $A_{|L_m^n}$ is a shift invariant n–ary ordinal mean for each $n \in \mathbb{N}$.

For n–ary means shift invariantness is equivalent to the requirement $A(x_1+1, \ldots, x_n+1) = 1 + A(x_1, \ldots, x_n)$ for all $(x_1, \ldots, x_n) \in (L_m \setminus \{m\})^n$.

Following the ideas of the corresponding result for kernel aggregation operators in [4], it can be shown that an ordinal mean A is a kernel ordinal mean if and only if it is sub–shift invariant, i.e.,

$$A(x_1 + 1, \ldots, x_n + 1) \leq 1 + A(x_1, \ldots, x_n) \tag{5}$$

for all $(x_1, \ldots, x_n) \in (L_m \setminus \{m\})^n$.

In decision making an important and desirable property of aggregation operators is their *joint strict monotonicity* which in the case of ordinal means can be characterized by the inequality

$$A(x_1, \ldots, x_n) < A(x_1 + 1, \ldots, x_n + 1) \tag{6}$$

for all $(x_1, \ldots, x_n) \in (L_m \setminus \{m\})^n$ and for each $n \in \mathbb{N}$.

Note that for each shift invariant ordinal mean $A : L_m^n \to L_m$ and all $(x_1, \ldots, x_n) \in L_m^n$ it holds

$$A(x_1, \ldots, x_n) = a + A(x_1 - a, \ldots, x_n - a),$$

where $a = \min(x_1, \ldots, x_n)$. Thus to know A it is enough to know its values at points $(x_1, \ldots, x_n) \in L_m^n$ such that $0 \in \{x_1, \ldots, x_n\}$, i.e., $(x_1, \ldots, x_n) \in L_m^n \setminus (L_m \setminus \{0\})^n = L_{m,0}^n$. Vice–versa, each non–decreasing mapping $B : L_{m,0}^n \to L_m$ bounded from above by the max–operator can be extended to a shift invariant mapping $A_B : L_m^n \to L_m$ defined by

$$A_B(x_1, \ldots, x_n) = a + B(x_1 - a, \ldots, x_n - a), \tag{7}$$

where $a = \min\{x_1, \ldots, x_n\}$. Unfortunately, A_B need not be monotone, thus not a mean. Based on the results from [24, 29] and their proofs, we can derive the next representation.

Proposition 1 *A mapping $A_B : L_m^n \to L_m$ given by (7) is a shift invariant ordinal mean if and only if B possesses the zero–kernel property*

$$\begin{aligned} |B(x_1, \ldots, x_n) &- B(y_1, \ldots, y_n)| \\ &\leq \max(|x_1 - y_1|, \ldots, |x_n - y_n|) \end{aligned}$$

for all $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n) \in L_{m,0}^n$ such that $x_i = y_i = 0$ for some $i \in \{1, \ldots, n\}$.

It is evident that each jointly strictly monotone ordinal mapping $A : L_m^n \to L_m$ which is non–decreasing, is an n–ary ordinal mean. Mordelová and Muel proved [27] that each binary jointly strictly monotone kernel ordinal mean is exactly a shift invariant ordinal mean. This result can easily be extended for arbitrary $n \in \mathbb{N}$.

Proposition 2 *Let $n \in \mathbb{N}$. For a mapping $A : L_m^n \to L_m$ the following claims are equivalent*

(i) *A is an n–ary shift invariant ordinal mean.*

(ii) *A is an n–ary jointly strictly monotone kernel ordinal mean.*

Proof. In the light of Proposition 1, the proof that (i) $\Rightarrow$ (ii) is trivial. Conversely, suppose that A is an n–ary jointly strictly monotone kernel ordinal mean. Due to the kernel property of A, for each $(x_1, \ldots, x_n) \in (L_m \setminus \{m\})^n$ it holds $A(x_1 +$

$1, \ldots, x_n + 1) - A(x_1, \ldots, x_n) \in \{0, 1\}$. The joint strictly monotonicity of A ensures $A(x_1 + 1, \ldots, x_n + 1) - A(x_1, \ldots, x_n) > 0$, and thus $A(x_1 + 1, \ldots, x_n + 1) - A(x_1, \ldots, x_n) = 1$, which implies the shift invariantness of A.

□

3 Weighted ordinal means

In this section we will present ordinal means corresponding to weighted (quasi-)arithmetic means, OWA–operators and weighted ordered quasi–arithmetic means. They are based on the original ideas of Godo and Torra [7] exploiting ordinal divisible t–conorms (t–norms) on L_m which were modified in [17] in such a way that the procedure always results in an ordinal mean. We introduce the formula for weighted ordinal means in an equivalent form based on the special representation of divisible ordinal t–conorms.

Recall that continuous t–conorms on the scale $[0, 1]$ were characterized as follows (for more details we refer to [1, 30, 18, 12]).

Proposition 3 *A function $S : [0, 1]^2 \to [0, 1]$ is a continuous t–conorm if and only if, there is a finite or countably infinite set K, a family $(\,]a_k, b_k[\,)_{k\in K}$ of non–empty, parwise disjoint open subintervals of $[0, 1]$, and a family $(g_k)_{k\in K}$, $g_k : [a_k, b_k] \to [0, \infty]$, of continuous, strictly increasing functions with $g_k(a_k) = 0$, for each $k \in K$, such that*

$$S(x, y) = \begin{cases} g_k^{(-1)}\left(g_k(x) + g_k(y)\right) \\ \qquad\qquad if\ (x, y) \in]a_k, b_k]^2, \\ \\ \max\{x, y\} \qquad otherwise, \end{cases} \tag{8}$$

where $g_k^{(-1)} : [0, \infty] \to [a_k, b_k]$ is the pseudo–inverse of g_k, see [13], given by

$$g_k^{(-1)}(x) = \sup\{z \in [a_k, b_k] \,|\, g_k(z) \le x\}.$$

The notion of weighted continuous t–conorms on $[0, 1]$ was introduced in [3].

Definition 4 *Let $S : [0, 1]^2 \to [0, 1]$ be a continuous t–conorm given by (8), and let $\mathbf{w} = (w_1, \ldots, w_n) \in [0, 1]^n$ be a normal weighting vector, i.e., with weights satisfying the property $\sum_{i=1}^{n} w_i = 1$. The weighted t–conorm $S_{\mathbf{w}} : [0, 1]^n \to [0, 1]$ is given by*

$$S_{\mathbf{w}}(x_1, \ldots, x_n) = \begin{cases} g_k^{(-1)}\left(\sum_{i=1}^{n} w_i\, g_k\left(\max\{a_k, x_i\}\right)\right) \\ \quad if\ \max\{x_i \,|\, w_i > 0\} \in]a_k, b_k], \\ \\ \max\{x_i \,|\, w_i > 0\} \qquad otherwise. \end{cases} \tag{9}$$

Example 1

(i) Let $S_L : [0, 1]^2 \to [0, 1]$ be the Łukasiewicz t–conorm, $S_L(x, y) = \min\{x + y, 1\}$, i.e., $K = \{1\}$, $a_1 = 0$, $b_1 = 1$, $g_1 : [0, 1] \to [0, \infty]$ is given by $g_1(x) = x$. Then for the uniform weighting vector $\mathbf{w}_u = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$ we have $(S_L)_{\mathbf{w}_u}(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$, i.e., $(S_L)_{\mathbf{w}_u} = M$ is the arithmetic mean.

For a general normal weighting vector $\mathbf{w} = (w_1, \ldots, w_n)$, i.e., $\mathbf{w} \in [0, 1]^n$, $\sum_{i=1}^{n} w_i = 1$, $(S_L)_{\mathbf{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i\, x_i$, i.e., $(S_L)_{\mathbf{w}}$ is a weighted arithmetic mean. Note that each weighted arithmetic mean W can be represented in this way.

(ii) Let $S : [0, 1]^2 \to [0, 1]$ be a continuous Archimedean t–conorm, i.e.,

$$S(x, y) = g^{-1}(\min\{g(1), g(x) + g(y)\})$$

for some strictly increasing continuous function $g : [0, 1] \to [0, \infty]$ with $g(0) = 0$. Then for any normal weighting vector $\mathbf{w}$,

$$S_{\mathbf{w}}(x_1, \ldots, x_n) = g^{-1}\left(\sum_{i=1}^{n} w_i\, g(x_i)\right),$$

which means that $S_{\mathbf{w}}$ is a weighted quasi–arithmetic mean. It has annihilator 1 whenever $g(1) = \infty$ (i.e., when S is a strict t–conorm) and each weight w_i is positive. For $\mathbf{w} = \mathbf{w}_u$, $S_{\mathbf{w}_u}$ is a quasi–arithmetic mean.

Weighted continuous t–norms $T_{\mathbf{w}}$ are defined in a similar way as weighted t–conorms.

Observe that each t–conorm on L_m is continuous. However, the property equivalent to the continuity of t–conorms on $[0, 1]$ is the divisibility, see, e.g., [11], and thus in the framework of discrete t–conorms we will deal with their divisibility [8, 23]. Divisible t–conorms on L_m were characterized in [21]:

Proposition 4 *A function $S : L_m^2 \to L_m$ is a divisible t–conorm on L_m if and only if, there is a set $\{b_0, \ldots, b_j\} \subset L_m$, $b_0 = 0 < b_1 < \ldots < b_j = m$ such*

that

$$S(x,y) = \begin{cases} \min\{b_k, x+y-b_{k-1}\} \\ \qquad\qquad if\ (x,y) \in]b_{k-1}, b_k]^2, \\ \max\{x,y\} \qquad otherwise. \end{cases} \tag{10}$$

Divisible t–norms on L_m are in a one–to–one correspondence with divisible t–conorms on L_m throughout Frank's functional equation

$$T(x,y) + S(x,y) = x + y. \tag{11}$$

Each divisible t–conorm S on L_m given by (10) can also be represented in form (9), putting $a_k = b_{k-1}$ and $g_k : \{a_k, a_k+1, \ldots, b_k\} \to [0,\infty]$ given by $g_k(x) = x - a_k$, and the pseudo–inverse $g_k^{(-1)} : [0,\infty] \to \{a_k, \ldots, b_k\}$ given by

$$g_k^{(-1)}(x) = \sup\left\{z \in \{a_k, \ldots, b_k\} \,|\, g_k(z) \le x\right\}.$$

Now, to define a divisible weighted t–conorm $S_{\mathbf{w}}$ on L_m, we can formally repeat Definition 4.

Definition 5 *Let $\mathbf{w} \in [0,1]^n$ be a normal weighting vector and let $S : L_m^2 \to L_m$ be a divisible t–conorm. The weighted divisible t–conorm $S_{\mathbf{w}} : L_m^n \to L_m$ is given by*

$$S_{\mathbf{w}}(x_1, \ldots, x_n) = \begin{cases} g_k^{(-1)}\left(w_i \max\{x_i, b_{k-1}\}\right) \\ \qquad\qquad if\ \max\{x_i \,|\, w_i > 0\} \in]b_{k-1}, b_k], \\ 0 \qquad\qquad otherwise. \end{cases} \tag{12}$$

From equation (11) we can introduce weighted divisible t–norms on L_m, by

$$T_{\mathbf{w}}(x_1, \ldots, x_n) = \begin{cases} \inf\{z \in \{b_{k-1}, \ldots, b_k\} \,| \\ \qquad\qquad z \ge \sum_{i=1}^{n} w_i \min\{x_i, b_k\}\} \\ \qquad\qquad \text{if } \min\{x_i \,|\, w_i > 0\} \in [b_{k-1}, b_k[, \\ 1 \qquad\qquad \text{otherwise.} \end{cases} \tag{13}$$

The only divisible Archimedean t–conorm on L_m is the Łukasiewicz t-conorm, given by $S_L(x,y) = \min\{x+y, m\}$, and the corresponding weighted t–conorm for the uniform weighting vector $\mathbf{w}_u$, $(S_L)_{\mathbf{w}_u} : L_m^n \to L_m$, given by

$$\begin{aligned} &(S_L)_{\mathbf{w}_u}(x_1, \ldots, x_n) \\ &\quad = \sup\left\{z \in L_m \,|\, z \le \frac{1}{n}\sum_{i=1}^{n} x_i\right\} \\ &\quad = \lfloor M(x_1, \ldots, x_n) \rfloor, \end{aligned}$$

can be understood as the lower arithmetic mean on L_m (and denoted by M_L). Here, $\lfloor x \rfloor : \mathbb{R} \to \mathbb{Z}$ is the floor of a real x and M is the standard arithmetic mean on $\mathbb{R}$.

Similarly, $(T_L)_{\mathbf{w}_u} : L_m^n \to L_m$, defines the upper arithmetic mean on L_m,

$$\begin{aligned} &(T_L)_{\mathbf{w}_u}(x_1, \ldots, x_n) \\ &\quad = \inf\left\{z \in L_m \,|\, z \ge \frac{1}{n}\sum_{i=1}^{n} x_i\right\} \\ &\quad = \lceil M(x_1, \ldots, x_n) \rceil \end{aligned}$$

(it will be denoted by M_U), where $\lceil x \rceil : \mathbb{R} \to \mathbb{Z}$ is the ceiling of a real x.

The lower weighted arithmetic means on L_m can be introduced as $(S_L)_{\mathbf{w}}$, while the upper ones as $(T_L)_{\mathbf{w}}$,

$$(S_L)_{\mathbf{w}}(x_1, \ldots, x_n) = \lfloor W(x_1, \ldots, x_n) \rfloor,$$

$$(T_L)_{\mathbf{w}}(x_1, \ldots, x_n) = \lceil W(x_1, \ldots, x_n) \rceil.$$

Obviously, lower and upper ordinal OWA operators on L_m can be introduced by

$$(OWA_m)_* = \lfloor OWA \rfloor \text{ and } (OWA_m)^* = \lceil OWA \rceil.$$

Observe that formula (12) can also be written in the form

$$\lfloor W(\max\{x_1, b_{k-1}\}, \ldots, \max\{x_n, b_{k-1}\}) \rfloor,$$

whenever $\max\{x_i \,|\, w_i > 0\} \in]b_{k-1}, b_k]$.

Formulae (12) and (13) present ordinal forms of weighted divisible t–conorms and t–norms. The fact that there is a unique divisible Archimedean t–conorm (t–norm) on L_m, namely S_L (T_L), excludes the possibility of introducing proper quasi–arithmetic (weighted quasi–arithmetic) means on L_m using the above approach.

If we consider a general weighting vector $\mathbf{v} = (v_1, \ldots, v_n) \in [0,\infty[^n$, $\sum_{i=1}^{n} v_i > 0$, we first normalize it, $\mathbf{w} = \frac{\mathbf{v}}{\sum_{i=1}^{n} v_i}$, and then we put $S_{\mathbf{v}} = S_{\mathbf{w}}$ ($T_{\mathbf{v}} = T_{\mathbf{w}}$).

4 Dissimilarity based ordinal means

Dissimilarity based means on real intervals were introduced and studied in [25], compare also [5]. Here we adopt one of the approaches discussed in the cited papers to the ordinal scales L_m.

Definition 6 *Let $K : \mathbb{R} \to \mathbb{R}$ be a convex function with unique minimum $K(0) = 0$ and let $f : L_m \to \mathbb{R}$ be a strictly monotone function. The function $D_{K,f} : L_m^2 \to \mathbb{R}$ given by*

$$D_{K,f}(i,j) = K(f(i) - f(j))$$

is called a dissimilarity function.

To define a symmetric ordinal mean exploiting a dissimilarity function $D_{K,f}$, one first needs to define a "middle point" of any interval $\{i, i+1, \ldots, j\}$, $0 \leq i \leq j \leq m$. Though there are several consistent possibilities, we will deal with two approaches only, namely, with the "lower middle point" $LMP(i,j) = \lfloor \frac{i+j}{2} \rfloor$ and the "upper middle point" $UMP(i,j) = \lceil \frac{i+j}{2} \rceil$.

Definition 7 *Let $D_{K,f}$ be a given dissimilarity function.*

(i) *The mapping $M_{K,f,L} : L_m^n \to L_m$ given by*

$$M_{K,f,L}(x_1, \ldots, x_n) = LMP(l,u) \tag{14}$$

where

$$l = \min\{k \in L_m \mid \sum_{i=1}^{n} K(f(x_i) - f(k)) = \min\}$$

and

$$u = \max\{k \in L_m \mid \sum_{i=1}^{n} K(f(x_i) - f(k)) = \min\}$$

is called a lower (K,f)–ordinal mean.

(ii) *Similarly, the mapping $M_{K,f,U} : L_m^n \to L_m$ given by*

$$M_{K,f,U}(x_1, \ldots, x_n) = UMP(l,u) \tag{15}$$

is called an upper (K,f)–ordinal mean

Proposition 5 *Both (K,f)–ordinal means defined in Definition 7 are symmetric ordinal means.*

If K is an even function, then

$$M_{K,id,U}(x_1, \ldots, x_n) = m - M_{K,id,L}(m - x_1, \ldots, m - x_n).$$

Remark 1 Observe that for any dissimilarity function $D_{K,f}$ and weights $\mathbf{w} = (w_1, \ldots, w_n) > \mathbf{0}$ we can introduce a lower (an upper) (K,f)–ordinal mean minimizing the expression

$$\sum_{i=1}^{n} w_i \, K(f(x_i) - f(k))$$

in Definition 7.

Recall that for the function $K : \mathbb{R} \to \mathbb{R}$ given by $K(x) = |x|$ ($K = Abs$) and $f : L_m \to \mathbb{R}$ given by $f(i) = i$ ($f = id$), the corresponding (Abs, id)–ordinal means are ordinal medians which were discussed first in [20]. Note that for each odd m both, lower and upper ordinal medians coincide with the classical median on L_m. Observe that on any real interval I the (Abs, id)–mean is exactly the median operator, see [25, 5]. Similarly, putting $K = Q$, where $Q(x) = x^2$, $x \in \mathbb{R}$, (Q, id)–mean on any real interval I yields the standard arithmetic mean. Thus we can introduce the lower arithmetic mean on L_m as the lower (Q, id)–ordinal mean, and the upper arithmetic mean on L_m as the upper (Q, id)–ordinal mean. Due to Remark 1, weighted ordinal arithmetic means can be defined. Obviously, then also ordinal OWA's on L_m can be introduced. Coming back to a real interval I, (Q, f)–mean for a continuous and strictly monotone function $f : I \to \mathbb{R}$ yields a quasi–arithmetic mean on I generated by f. Thus (Q, f)–ordinal means on L_m can be understood as ordinal quasi–arithmetic means. Moreover, we can introduce weighted ordinal quasi–arithmetic means.

Example 2 Let $m = 3$, i.e., let us work on the scale $L_3 = \{0, 1, 2, 3\}$. Put $n = 4$ and consider the input $\mathbf{x} = (2, 0, 3, 0)$. Then

(i) If $K = Abs$, $f = id$, then $l = 0$, $u = 2$ and $M_{Abs,id,L}(\mathbf{x}) = M_{Abs,id,U}(\mathbf{x}) = 1$.

(ii) If $K = Q$, $f = id$, then $l = u = 1$ and $M_{Q,id,L}(\mathbf{x}) = M_{Q,id,U}(\mathbf{x}) = 1$.

(iii) If $K = Q$, $f = Q$, then $l = u = 2$ and $M_{Q,Q,L}(\mathbf{x}) = M_{Q,Q,U}(\mathbf{x}) = 2$.

(iv) If $K(x) = \begin{cases} x & \text{if } x \geq 0, \\ \frac{2}{7}\, x^2 & \text{if } x \leq 0 \end{cases}$, $f = id$, then $l = 1$, $u = 2$ and $M_{K,id,L}(\mathbf{x}) = 1$, $M_{K,id,U}(\mathbf{x}) = 2$.

Note that all dissimilarity based ordinal means, lower and upper ones, are shift invariant, i.e., jointly strictly monotone kernel ordinal means.

5 Conclusion

The lower ordinal arithmetic mean M_L derived in Section 3 is defined for all $\mathbf{x} = (x_1, \ldots, x_n) \in L_m^n$ by $M_L(\mathbf{x}) = \left\lfloor \sum_{i=1}^{n} \frac{x_i}{n} \right\rfloor$. Similarly, the upper ordinal arithmetic mean M_U is given by $M_U(\mathbf{x}) = \left\lceil \sum_{i=1}^{n} \frac{x_i}{n} \right\rceil$.

In Section 4, for dissimilarity based ordinal arithmetic

means it holds

$$M_{Q,id,L}(\mathbf{x}) = \begin{cases} \left\lfloor \sum_{i=1}^{n} \frac{x_i}{n} \right\rfloor \\ \quad \text{if} \quad \sum_{i=1}^{n} \frac{x_i}{n} - \left\lfloor \sum_{i=1}^{n} \frac{x_i}{n} \right\rfloor \leq \frac{1}{2}, \\ \left\lceil \sum_{i=1}^{n} \frac{x_i}{n} \right\rceil \quad \text{otherwise,} \end{cases}$$

and

$$M_{Q,id,U}(\mathbf{x}) = \begin{cases} \left\lfloor \sum_{i=1}^{n} \frac{x_i}{n} \right\rfloor \\ \quad \text{if} \quad \sum_{i=1}^{n} \frac{x_i}{n} - \left\lfloor \sum_{i=1}^{n} \frac{x_i}{n} \right\rfloor < \frac{1}{2}, \\ \left\lceil \sum_{i=1}^{n} \frac{x_i}{n} \right\rceil \quad \text{otherwise.} \end{cases}$$

These results lead to introducing two classes of ordinal arithmetic means, namely, $(M_{c,L})_{c\in[0,1[}$ and $(M_{c,U})_{c\in\,]0,1]}$, where

$$M_{c,L}(\mathbf{x}) = \begin{cases} i & \text{if} \quad i \leq \sum_{i=1}^{n} \frac{x_i}{n} \leq i + c, \\ i+1 & \text{if} \quad i + c < \sum_{i=1}^{n} \frac{x_i}{n} < i + 1, \end{cases}$$

$$M_{c,U}(\mathbf{x}) = \begin{cases} i & \text{if} \quad i \leq \sum_{i=1}^{n} \frac{x_i}{n} < i + c, \\ i+1 & \text{if} \quad i + c \leq \sum_{i=1}^{n} \frac{x_i}{n} < i + 1. \end{cases}$$

It holds:

$$M_L = M_{1,U} \quad \text{and} \quad M_U = M_{0,L},$$
$$M_{Q,id,L} = M_{\frac{1}{2},L} \quad \text{and} \quad M_{Q,id,U} = M_{\frac{1}{2},U}$$

and

$$M_{c,L}(x_1, \ldots, x_n) = m - M_{1-c,U}(m - x_1, \ldots, m - x_n)$$

for all $(x_1, \ldots, x_n) \in L_m^n$.

Evidently, all these ordinal means are symmetric, shift invariant, thus jointly strictly monotone kernel ordinal means on L_m

Acknowledgement

The first author kindly acknowledges the support of the grants VEGA 1/3012/06 and 1/3014/06 and the third author the support of the grant VEGA 1/4209/07. Both these authors also acknowledge the support of Science and Technology Assistance Agency under the contract No. APVV–0375–06. The second author acknowledges the support of the Govern Balear grant PCTIB2005GC1–07 and the Spanish DGI grants TIN2004–21700–E and MTM2006-08322.

References

[1] J. Aczél, Sur les opérations definies pour des nombres réels. Bull. Soc. Math. France 76 (1949) 69–64.

[2] T. Calvo, A. Kolesárová, M. Komorníková, R. Mesiar, A Review of Aggregation Operators. University of Alcalá Press, Alcalá de Henares, Madrid, 2001.

[3] T. Calvo, R. Mesiar, Weighted means based on triangular conorms. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9 (2001) 183–196.

[4] T. Calvo, R. Mesiar, Stability of aggregation operators. Proc of Int. Conference in Fuzzy Logic and Technology, Leicester, 2001, pp. 475–478.

[5] Calvo T., Mesiar R., Yager R.R., Qualitative weights and aggregation. IEEE Transactions on Fuzzy Systems 12 (2004) 62-69.

[6] D. Denneberg , Non-additive Measure and Integral. Kluwer, Dordrecht, 1994.

[7] L. Godo, V. Torra, On aggregation operators for qualitative information. IEEE Transactions on Fuzzy Systems 8, Vol. 2 (2000) 143–154.

[8] L. Godo, C. Sierra, A new approach to connective generation in the framework of expert systems using fuzzy logic. Proc. IEEE–ISMVL'88, Computer Society Press, 2000, pp. 157–162.

[9] L. Godo, V. Torra, Extending Choquet integral for aggregation of ordinal values. Proc. IPMU'2000, Madrid, 2000, pp. 410–416.

[10] J.C. Fodor, Preference Structures and Aggregation Procedures, D. Sc. Thesis, Hungarian Academy of Sciences, Budapest, 2002.

[11] P. Hájek, Metamathematics of Fuzzy Logic. Kluwer Academic Publishers, Dordrecht, 1998.

[12] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.

[13] E.P. Klement, R. Mesiar, E. Pap, Quasi– and pseudo–inverses of monotone functions, and the construction of t-norms. Fuzzy Sets and Systems 104 (1999) 3-13.

[14] A. Kolesárová A., J. Mordelová, 1–Lipschitz and kernel aggregation operators. Proc. of Summer School on Aggregation Operators AGOP'2001, Oviedo, Spain, 2001, pp. 71–75.

[15] A. Kolesárová A., J. Mordelová, E. Muel, A review of kernel aggregation operators. Proc. of Summer School on Agregation Operators AGOP' 2003, Alcal de Henares, Spain, 2003, pp. 97-102.

[16] A. Kolesárová, Aggregation of qualitative information: Weighted means on ordinal scales. Proc. 12th Zittau East–West Colloquium, Zittau, Germany, 2005, pp. 46–52.

[17] A. Kolesárová, G. Mayor, R. Mesiar, Weighted ordinal means, Information Sciences (2006), accepted.

[18] C.M. Ling, Representation of associative functions. Publ. Math. Debrecen 12 (1965) 189–212.

[19] J.-L. Marichal, Aggregation operators for multicriteria decision aid, Ph. D. Thesis, University of Liége, Belgium, 1998.

[20] J. Martín, G. Mayor, J. Suñer, Symmetric aggregation operators on a finite chain. Proc. IPMU'2002, Annecy, France, 2002, pp. 1365–1372.

[21] G. Mayor, J. Torrens, On a class of operators for expert systems. International Journal of Intelligent Systems 8 (1993) 771–778.

[22] M. Mas, G. Mayor, J. Torrens, t–operators and uninorms on a finite totally ordered set. International Journal of Intelligent Systems 14 (1999) 909–922.

[23] G. Mayor, J. Torrens, Triangular norms on discrete settings. In: Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms, E.P. Klement and R. Mesiar, eds., Elsevier B.V., 2005, pp. 189–230.

[24] R. Mesiar, T. Rüchschlossová, Characterization of invariant aggregation operators. Fuzzy Sets and Systems 142 (2004) 63–74.

[25] R. Mesiar, Fuzzy set approach to the utility preference relations, and aggregation operators. European J. of Opeartional Research 176 (2007) 414–422.

[26] J. Mordelová, E. Muel, Binary strictly monotone kernel aggregation operators on finite scales. Proc. of Int. Conference Uncertainty Modelling' 2001, Bratislava, 2001, pp. 127–131.

[27] J. Mordelová, E. Muel, Binary jointly strictly kernel aggregation operators on finite scales. Proc. Uncertainty Modelling' 2001, Bratislava, 2001, pp. 127–131.

[28] E. Pap, Null-Additive Set Functions. Kluwer, Dordrecht, 1995.

[29] T. Rüchschlossová, Aggregation Operators and Invariantness, PhD. thesis, Slovak Technical University, Bratislava, 2003.

[30] B. Schweizer, A. Sklar, Probabilistic Metric Spaces, North–Holland, New York, 1983.

[31] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Trans. Syst. Man Cybern. 18 (1988) 183–190.

A new look on discrete quasi-copulas

I. Aguiló
Dpt. de Matemàtiques i Inf.
Universitat de les Illes Balears.
07122 Palma de Mallorca. Spain
isabel.aguilo@uib.es

J. Suñer
Dpt. de Matemàtiques i Inf.
Universitat de les Illes Balears.
07122 Palma de Mallorca. Spain
jaume.sunyer@uib.es

J. Torrens
Dpt. de Matemàtiques i Inf.
Universitat de les Illes Balears.
07122 Palma de Mallorca. Spain
dmijts0@uib.es

Abstract

In this paper discrete quasi-copulas (defined on a square grid I_n^2 of [0,1]) are studied and it is proved that they can be represented by means of a special class of matrices with entries in [-1,1]. Special considerations are made for the case of irreducible discrete quasi-copulas (those with range I_n) defined on the finite chain I_n, showing that they can be represented through Alternating-Sign Matrices and that they generate all discrete quasi-copulas through convex sums.

Keywords: Copula, quasi-copula, discrete scale, irreducible discrete quasi-copula, ASM matrix.

1 Introduction

It is well known the importance that copulas have in Statistics as a consequence of Sklar's theorem. This theorem shows that the joint distribution function H of two random variables X and Y having marginal distribution functions F and G, respectively, can be obtained as

$$H(x,y) = C(F(x), G(y)) \qquad \text{for all} \quad x, y \in \overline{I\!R} \quad (1)$$

where $C : [0,1]^2 \to [0,1]$ is a copula uniquely determined on $\{(F(x), G(y)) \mid x, y \in \overline{I\!R}\}$, and vice versa, given two (one dimensional) distribution functions F and G and a copula C, the function H defined by (1) is a two dimensional distribution with marginals F and G (see [15] and [12]). In fact, from this theorem, the concept of subcopula becomes essential and the problem of extending subcopulas to a copula naturally arises. A special case of subcopulas are the so-called *discrete copulas* defined on a finite subset of $[0,1]^2$, usually $I_{n,m} = I_n \times I_m$ where

$$I_n = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}.$$

Discrete copulas are interesting because they are related through Sklar's theorem with bivariate discrete random variables (see [10]) and they apply specially in cases where discrete ordinal structures of statistical data are essential.

Discrete copulas with domain I_n^2 are studied in [7] where it is proved that they can be represented by means of bistochastic matrices. The particular case of discrete copulas on I_n that take values again in I_n was introduced in [9] where it is proved that they can be represented through permutation matrices. Moreover, this particular type of copulas are called in [7] *irreducible discrete copulas* because, as it is proved there, all discrete copulas on I_n^2 are convex sums of irreducible discrete copulas.

On the other hand, quasi-copulas were introduced in [1] as a generalization of copulas and they were characterized in operational terms in [4]. Recently, discrete quasi-copulas defined on $I_{n,m}$ are introduced in [13] where they are represented by means of $(n \times m)$ matrices that are simply their operation table. Moreover, some methods to extend a discrete quasi-copula to a quasi-copula are presented. The special case of internal discrete quasi-copulas $C : I_n^2 \to I_n$, called here *irreducible discrete quasi-copulas* similarly to the case of copulas, has been recently introduced in [8].

Copulas and quasi-copulas can also be viewed as conjunctive aggregation operators and many papers have appeared in last years dealing with them, from this point of view (see [2], [3], [5], [7], [8], [9]).

In this paper we deal with discrete quasi-copulas defined on I_n^2 presenting a new look on them following a similar study to the one given for discrete copulas in [7] and [9]. We begin with section 2 where some preliminaries are given that will be used along the paper. In section 3 we prove that discrete quasi-copulas can be represented by means of a class of matrices with entries in the interval $[-1,1]$ that we will call *Generalized Bistochastic Matrices* or *GBM* in short. We

deal first with the special case of internal discrete copulas on I_n. Discrete quasi-copulas in this special case are called *irreducible discrete quasi-copulas*, similarly to the case of copulas, and it is proved that they can be represented by the so-called *Alternating-Sign Matrices* or *ASM* matrices in short (see [14]). We devote section 4 to prove that all discrete quasi-copulas are convex sums of irreducible ones.

2 Preliminaries

In this section we give only basic definitions and properties of copulas and subcopulas. A more extensive study can be found in [12] and [4] and, for the discrete case, in [7] and [9].

Definition 1 *([12]) A (two dimensional) copula C is a binary operation on $[0,1]$, i.e., $C : [0,1] \times [0,1] \longrightarrow [0,1]$ such that*

(C1) $C(x,0) = C(0,x) = 0 \ \forall x \in [0,1]$

(C2) $C(x,n) = C(n,x) = x \ \forall x \in [0,1]$

(C3) $C(x,y) + C(x',y') \geq C(x,y') + C(x',y)$ *for all $x, x', y, y' \in [0,1]$ with $x \leq x', y \leq y'$ (2-increasing condition)*

Although the notion of quasi-copula was introduced in a different way, it is proved in [4] the following equivalent definition.

Definition 2 *A (two dimensional) quasi-copula Q is a binary operation on $[0,1]$, i.e., $Q : [0,1] \times [0,1] \longrightarrow [0,1]$ such that*

(Q1) $Q(x,0) = Q(0,x) = 0 \ \forall x \in [0,1]$ *and* $Q(x,n) = Q(n,x) = x \ \forall x \in [0,1]$

(Q2) *Q is non-decreasing in each component*

(Q3) *Q satisfies the Lipschitz condition with constant 1:*

$$| Q(x',y') - Q(x,y) | \leq |x' - x| + |y' - y|$$

for all $x, x', y, y' \in [0,1]$

The following characterization of quasi-copulas was given again in [4]:

Proposition 1 *A binary operation $Q : [0,1] \times [0,1] \longrightarrow [0,1]$ is a quasi-copula if, and only if, it satisfies $(Q1)$ and the following property:*

(Q4) $Q(x,y) + Q(x',y') \geq Q(x,y') + Q(x',y)$ *for all $x, x', y, y' \in [0,1]$ such that $x \leq x', y \leq y'$ where at least one of these four elements is equal either to 0 or to 1 (2-increasing condition on the border of $[0,1]^2$)*

Clearly each copula is a quasi-copula but not vice versa. Quasi-copulas which are not copulas are called *proper quasi-copulas.*

Definition 3 *A function $C : D \times D' \to [0,1]$, where D, D' are subsets of [0,1] containing $\{0,1\}$, is called a subcopula if it satisfies the properties of a copula for all $(x,y) \in D \times D'$.*

A special case of subcopulas are the so-called *discrete copulas* defined on a finite subset of $[0,1]^2$, usually $I_{n,m} = I_n \times I_m$ where

$$I_n = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}.$$

In this paper we will consider discrete copulas defined on I_n^2. Thus, in our context, we have the following definition (see, for instance, [7]).

Definition 4 *A discrete copula C on I_n is a binary operation $C : I_n \times I_n \longrightarrow [0,1]$ satisfying properties $(C1 - C3)$ for all $x, y \in I_n$.*

The range of a discrete copula C on I_n always contains I_n. When the range of C is exactly I_n we deal in fact with *internal discrete copulas* $C : I_n \times I_n \to I_n$, called *irreducible discrete copulas* in [7]. In fact, this class of discrete copulas were already introduced and studied in [9], and it is proved in [7] that all discrete copulas are convex sums of irreducible ones.

3 Discrete quasi-copulas

Following the idea of discrete copulas, we want to deal in this paper with the notion of discrete quasi-copula already introduced in [13][1].

Definition 5 *A discrete quasi-copula Q on I_n is a binary operation $Q : I_n \times I_n \longrightarrow [0,1]$ such that*

(DQ1) $Q(\frac{i}{n}, 0) = Q(0, \frac{i}{n}) = 0 \ \forall \frac{i}{n} \in I_n$ *and* $Q(\frac{i}{n}, 1) = Q(1, \frac{i}{n}) = \frac{i}{n} \ \forall \frac{i}{n} \in I_n$

(DQ2) *Q is non-decreasing in each component*

(DQ3) *Q satisfies the Lipschitz condition with constant 1.*

[1] In fact the definition could be clearly given for discrete quasi-copulas on $I_n \times I_m$, but we are interested here in the most usual case $n = m$.

It is clear that any discrete copula is a discrete quasi-copula but not vice versa and, in this sense, we will call *proper discrete quasi-copulas* to discrete quasi-copulas that are not discrete copulas.

The following characterizations can be easily derived (see [13]).

Proposition 2 *Let $Q : I_n \times I_n \to [0,1]$ be any binary operator. The following statements are equivalent:*

(i) *Q is a discrete quasi-copula on I_n.*

(ii) *Q satisfies condition (DQ1) and also*

$$0 \leq Q\left(\frac{i}{n}, \frac{j}{n}\right) - Q\left(\frac{i-1}{n}, \frac{j}{n}\right) \leq \frac{1}{n}$$

and

$$0 \leq Q\left(\frac{i}{n}, \frac{j}{n}\right) - Q\left(\frac{i}{n}, \frac{j-1}{n}\right) \leq \frac{1}{n}$$

for all $1 \leq i, j \leq n$.

(iii) *Q satisfies condition (DQ1) and*

$$Q\left(\frac{i}{n}, \frac{j}{n}\right) + Q\left(\frac{i'}{n}, \frac{j'}{n}\right) \geq Q\left(\frac{i}{n}, \frac{j'}{n}\right) + Q\left(\frac{i'}{n}, \frac{j}{n}\right)$$

for all $0 \leq i, j, i', j' \leq n$ such that $i \leq i', j \leq j'$ where at least one of these four elements is equal either to 0 or to n.

3.1 Irreducible discrete quasi-copulas

It is clear from condition $(DQ1)$ that the range of any discrete quasi-copula on I_n contains I_n. Let us consider, like in the case of copulas, the special case of discrete quasi-copulas with range equal to I_n, that is, discrete quasi-copulas with minimal range.

Definition 6 *An irreducible discrete quasi-copula on I_n is a discrete quasi-copula Q on I_n with range I_n.*

Thus, an irreducible discrete quasi-copula on I_n is an internal binary operation $Q : I_n \times I_n \to I_n$ with properties $(DQ1) - (DQ3)$. Note that, in order to study operations defined on a finite chain like I_n, the only important thing is the number of elements of the chain, $n+1$ (see [11]). Then, this study is usually given for the most simple of these chains, that is,

$$L_n = \{0, 1, \ldots, n\}.$$

In this sense, we will study irreducible discrete quasi-copulas on L_n, that is, $Q : L_n \times L_n \to L_n$.

Remark 1 *It is clear that, given any irreducible discrete quasi-copula Q on L_n, the operator $Q' : I_n \times I_n \to I_n$ given by*

$$Q'\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{1}{n} \cdot Q(i,j)$$

is an irreducible discrete quasi-copula on I_n. And vice versa, for any irreducible quasi-copula on I_n, Q', the operator $Q : L_n \times L_n \to L_n$ defined by

$$Q(i,j) = n \cdot Q'\left(\frac{i}{n}, \frac{j}{n}\right)$$

is an irreducible discrete quasi-copula on L_n.

Thus, from now on, we will deal in this subsection with discrete quasi-copulas on L_n. As one of the main results, we will give an useful characterization of irreducible discrete quasi-copulas in terms of a special class of matrices, generalizing the representation of copulas given in [9]. In this case, we use the so-called Alternating-Sign Matrices. For more details on this class of matrices, see [14] and the web page: http://www.research.att.com//njas/sequences/

Definition 7 *An $n \times n$ Alternating-Sign Matrix (ASM matrix) is an $n \times n$ matrix $A = (a_{ij})$ such that*

1. *$a_{ij} \in \{-1, 0, 1\} \quad \forall i, j \in \{1, 2, \ldots, n\}$*
2. *The first and the last elements $a_{ij} \neq 0$ of each row and each column are 1.*
3. *All the elements $a_{ij} \neq 0$ of each row and each column have alternating signs.*

Remark 2 *In particular, the sum of each row and each column equals 1. Observe also that a permutation matrix is an ASM matrix.*

Next we give the characterization of an irreducible quasi-copula in terms of ASM matrices.

Proposition 3 *A binary operator $Q : L_n \times L_n \longrightarrow L_n$ is a discrete quasi-copula if, and only if, there exists an $n \times n$ ASM matrix $A = (a_{ij})$ such that, for all $r, s \in L_n$,*

$$Q(r,s) = \begin{cases} 0 & \text{if } r = 0 \text{ or } s = 0 \\ \displaystyle\sum_{\substack{i \leq r \\ j \leq s}} a_{ij} & \text{otherwise} \end{cases}$$

Given the quasi-copula Q, the matrix A is obtained as

$$a_{ij} = Q(i,j) + Q(i-1, j-1) - Q(i, j-1) - Q(i-1, j) \tag{2}$$

Remark 3 *Note that, following Remark 1, the corresponding irreducible discrete quasi-copula Q' on I_n will be given by*

$$Q\left(\frac{r}{n},\frac{s}{n}\right)=\begin{cases}0 & \text{if } r=0 \text{ or } s=0\\ \dfrac{1}{n}\displaystyle\sum_{\substack{i\le r\\ j\le s}} a_{ij} & \text{otherwise}\end{cases}$$

Corollary 1 *There is a one-to-one correspondence between the set of all irreducible discrete quasi-copulas on L_n and the set of all $n\times n$ ASM matrices. This correspondence assigns to an irreducible discrete quasi-copula Q, the ASM matrix $A=(A_{ij})$ given by (2), that will be called from now on, the associated ASM matrix of Q.*

It can be proved that if the 2-increasing condition fails for an irreducible discrete quasi-copula Q, it must fail in a square given by two consecutive rows and two consecutive columns, that is, a square determined by vertices (i,j) and $(i+1,j+1)$, for some i,j such that $0<i,j<n-1$. Moreover, the correspondence given in Corollary 1 is such that each non-fulfilment of the 2-increasing condition in a square given by vertices (i,j) and $(i+1,j+1)$ corresponds to a negative entry (-1) in position $(i+1,j+1)$ of the associated ASM matrix and vice versa.

Thus, this matrix representation is useful in finding proper quasi-copulas simply by constructing an ASM matrix with at least one negative entry.

Example 1 *Let us consider the irreducible quasi-copula Q defined on $L_4=\{0,1,2,3,4\}$ by the following table:*

Q	*0*	*1*	*2*	*3*	*4*
0	*0*	*0*	*0*	*0*	*0*
1	*0*	*0*	**0**	**1**	*1*
2	*0*	*0*	**1**	**1**	*2*
3	*0*	*1*	*2*	*2*	*3*
4	*0*	*1*	*2*	*3*	*4*

Then the associated ASM of Q is:

$$A=\begin{pmatrix}0&0&1&0\\0&1&-1&1\\1&0&0&0\\0&0&1&0\end{pmatrix}$$

Note that the negative entry $a_{2,3}=-1$ corresponds to the non-fulfilment of the 2-increasing condition quoted in bold face in the table of Q.

On the other hand, there are more useful applications of the given representation. For instance, the open question of finding the number of irreducible quasi-copulas on L_n was posed in [8]. From the bijection stated in Corollary 1 this question can be easily answered. The number of irreducible discrete quasi-copulas on L_n is equal to the number of $n\times n$ ASM matrices, which is known as the Robbins number (see the web page: http://www.research.att.com//njas/sequences/). That is:

Proposition 4 *The number of irreducible discrete quasi-copulas on L_n is*

$$\prod_{k=0}^{n-1}\frac{(3k+1)!}{(n+k)!}=\prod_{k=1}^{n}\frac{(3k-2)!}{(2n-k)!}$$

As we know (see [9]), $n!$ of them are copulas, and the other are proper quasi-copulas. For example, for $n=3$, we have 7 quasi-copulas, one of them being a proper quasi-copula, and for $n=4$, we have 42 quasi-copulas, and 18 of them are proper quasi-copulas. All of them can be easily listed by constructing their associated ASM matrices.

Finally, another application of our representation is due to the easy manipulation of matrices. As an example of this easy manipulation we can derive the following two properties of irreducible discrete quasi-copulas.

Proposition 5 *There is one and only one $n\times n$ ASM matrix with the diagonal formed by -1 in all positions except the first and the last ones. This ASM matrix is given by:*

$$\begin{pmatrix}0&1&0&\cdots&0&0&0\\1&-1&1&\cdots&0&0&0\\0&1&-1&\cdots&0&0&0\\\vdots&\vdots&\vdots&\ddots&\vdots&\vdots&\vdots\\0&0&0&\cdots&-1&1&0\\0&0&0&\cdots&1&-1&1\\0&0&0&\cdots&0&1&0\end{pmatrix}$$

and it corresponds to the maximum proper Archimedean discrete quasi-copula on L_n, given by:

$$Q(x,y)=\begin{cases}x-1 & \text{if } 0<x=y<n\\ \min\{x,y\} & \text{otherwise}\end{cases}$$

Proposition 6 *There is one and only one $n\times n$ ASM matrix with the inverse diagonal formed by -1 in all positions except the first and the last ones. This ASM*

matrix is given by:

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 & 1 \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 1 & -1 & \cdots & 0 & 0 & 0 \\ 1 & -1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}$$

and it corresponds to the proper commutative discrete quasi-copula Q on L_n (close to the Łukasiewicz copula) given by:

$$Q(x,y) = \begin{cases} 1 & \text{if } x+y=n, x \neq 0, n \\ \max\{x+y-n, 0\} & \text{otherwise.} \end{cases}$$

3.2 Discrete quasi-copulas

In this subsection, we give a characterization of discrete quasi-copulas $Q : I_n \times I_n \longrightarrow [0,1]$ in terms of a new kind of matrices, that we will call Generalized Bistochastic Matrices. Let us see the definition of this class of matrices:

Definition 8 *A Generalized Bistochastic Matrix (GBM matrix) is an $n \times n$ matrix $A = (a_{ij})$ such that*

1. $\forall i,j \in \{1,2,\ldots,n\},\ a_{ij} \in [-1,1]$
2. $\forall j = 1,\ldots,n,\ \sum_{i=1}^{n} a_{ij} = 1,$
   $\forall i = 1,\ldots,n,\ \sum_{j=1}^{n} a_{ij} = 1.$
3. $\forall j = 1,\ldots,n,\ \forall k = 1,\ldots,n,\ 0 \leq \sum_{i=1}^{k} a_{ij} \leq 1$
   $\forall i = 1,\ldots,n,\ \forall k = 1,\ldots,n,\ 0 \leq \sum_{j=1}^{k} a_{ij} \leq 1$

Remark 4 *In particular, the first and the last elements $a_{ij} \neq 0$ of each row and each column must be strictly positive. But, contrary to the case of ASM matrices, there is not alternation of signs, that is, there can be two consecutive negative (or positive) elements.*

Now, we have the representation of quasi-copulas:

Proposition 7 *A binary operator $Q : I_n \times I_n \longrightarrow [0,1]$ is a discrete quasi-copula if, and only if, there exists a GBM matrix $A = (a_{ij})$ such that, for all $r,s \in L_n$,*

$$Q\left(\frac{r}{n}, \frac{s}{n}\right) = \begin{cases} 0 & \text{if } r = 0 \text{ or } s = 0 \\ \frac{1}{n} \cdot \sum_{\substack{i \leq r \\ j \leq s}} a_{ij} & \text{otherwise} \end{cases}$$

Given the quasi-copula Q, the matrix A is obtained as

$$a_{ij} = Q\left(\frac{i}{n}, \frac{j}{n}\right) + Q\left(\frac{i-1}{n}, \frac{j-1}{n}\right) - \left(\frac{i}{n}, \frac{j-1}{n}\right) - Q\left(\frac{i-1}{n}, \frac{j}{n}\right)$$

Example 2 *Let us consider the quasi-copula Q defined on $I_4 = \{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$ by the following table:*

Q	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	1
0	0	0	0	0	0
$\frac{1}{4}$	0	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{1}{4}$
$\frac{2}{4}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{2}{4}$
$\frac{3}{4}$	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{2}{4}$	$\frac{3}{4}$
1	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	1

Then the associated GBM matrix is:

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{-1}{4} & \frac{-1}{4} & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Remark 5 *Again, any non-fulfilment of Q of the 2-increasing condition, for vertices of the form $(i,j),(i+1,j+1)$, corresponds to a negative entry of A and vice versa.*

Next result follows immediately:

Corollary 2 *A discrete quasi-copula Q is commutative if and only if its associated GBM matrix is symmetric.*

Similarly to the case of copulas, idempotent elements of discrete quasi-copulas play an important role in their application. Recall that a discrete quasi-copula Q on I_n is *Archimedean* if the only idempotent elements of Q are 0 and 1.

Given an $n \times n$ matrix A, we will denote by $A^{(r)}$ the submatrix formed by the first r rows and r columns of A. Then we have the following results:

Proposition 8 *An element r is an idempotent element of a discrete quasi-copula Q if and only if $A^{(r)}$ is an $r \times r$ GBM matrix.*

Proposition 9 *Let Q be a discrete quasi-copula on I_n with associated GBM matrix A. Then Q is Archimedian if and only if $A^{(r)}$ is not a GBM matrix for any $r \in \{1, \ldots, n-1\}$.*

It is proved in [8] that any irreducible discrete quasi-copula can be represented as an ordinal sum of Archimedean irreducible discrete quasi-copulas. The same proof given there applies also for general discrete quasi-copulas obtaining the following.

Proposition 10 *Let Q be a discrete quasi-copula on I_n with the following idempotent elements: $0 = x_0 < x_1 < \ldots < x_{k-1} < x_k = 1$. Then Q is an ordinal sum*

$$Q = (\langle x_{i-1}, x_i, Q_i \rangle \mid i \in \{1, \ldots, k\}),$$

i.e., $Q(x, y) =$

$$\begin{cases} x_{i-1} + Q_i(x - x_{i-1}, y - x_{i-1}) & \text{if } (x, y) \in [x_{i-1}, x_i]^2 \\ & \text{for some } 1 \leq i \leq k \\ \min\{x, y\} & \text{otherwise.} \end{cases}$$

where Q_i is an Archimedean quasi-copula on $L_{(x_i - x_{i-1})}$ for each $i \in \{1, \ldots, k\}$.

Remark 6 *Note that in the above proposition, consecutive idempotents $j, j+1, \ldots, j+s$ correspond to trivial ordinal summands Q_i on L_1 where the only possible quasi-copula is the minimum. Thus, in this case Q is also given by the minimum in $[j, j+s]^2$.*

Let us now introduce the concept of ordinal sum of GBM matrices closely related to the ordinal sum of discrete quasi-copulas.

Definition 9 *Let n_i be a positive integer and A_i an $n_i \times n_i$ GBM matrix for $i = 1, \ldots, k$. Let $n = n_1 + \ldots + n_k$ and define an $n \times n$ GBM matrix, A, as*

$$A = \begin{pmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ 0 & 0 & A_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & A_k \end{pmatrix}$$

We will call A the ordinal sum of the matrices $A_1, \ldots, A_k$ and we will denote it by $A = A_1 \oplus \ldots \oplus A_k$.

Then, we have the following correspondence between both concepts of ordinal sums.

Proposition 11 *Let Q be an ordinal sum of discrete Archimedean quasi-copulas of the form*

$$Q = (\langle x_{i-1}, x_i, Q_i \rangle \mid i \in \{1, \ldots, k\}).$$

Then, its associated GBM matrix is the ordinal sum of the GBM matrices A_i, where each A_i is the associated $(x_i - x_{i-1}) \times (x_i - x_{i-1})$ GBM matrix of Q_i for $i = 1, \ldots, k$. Moreover, the same is true for irreducible discrete quasi-copulas and their associated ASM matrices.

Note that, in the above correspondence between ordinal sums of discrete quasi-copulas and matrices, if we have a trivial ordinal summand Q_i of a quasi-copula Q, then we will also have the trivial (1×1) GBM matrix associated to this summand, in the ordinal sum of GBM matrices corresponding to Q.

4 Convex set of discrete quasi-copulas

It is proved in [7] that the set of all discrete copulas is the convex closure of all irreducible discrete copulas. The same statement is true for discrete quasi-copulas:

Proposition 12 *The class of all discrete quasi-copulas is the smallest convex set containing the class of all irreducible discrete quasi-copulas.*

A consequence of this proposition is that any discrete quasi-copula is a convex linear combination of irreducible discrete quasi-copulas. Let us give next the algorithm that, given the GBM matrix A of a discrete quasi-copula Q, produces the ASM matrices corresponding to the irreducible quasi-copulas of the convex linear combination.

Algorithm

Let $A = (a_{ij})$ be a GBM matrix.

1) Let $c_1 = \min\{|a_{ij}| : a_{ij} \neq 0\}$
2) If $c_1 = 1$, then A is an ASM matrix and the process has finished.

   If not, let us construct an ASM matrix A_1 with only the following restrictions: We put a 1 (if $a_{ij} > 0$) or a -1 (if $a_{ij} < 0$) in position (i, j) and 0's in all positions where A has 0's. If the minimum c_1 is reached in more than one position, then we can put a 1 or a -1 (depending on wether $a_{ij} > 0$ or < 0) in any one of these positions.
3) Let $A_1^* = \dfrac{1}{1 - c_1}(A - c_1 \cdot A_1)$. Note that equivalently, we will have

$$A = c_1 \cdot A_1 + (1 - c_1) \cdot A_1^*.$$

A_1^* is a GBM matrix. If it is an ASM matrix, the algorithm is finished. If not, we apply steps 1, 2 and 3 to the matrix A_1^*.

This process ends up with the list of ASM matrices and the coefficients of the convex linear combination are easily obtained from the algorithm.

Example 3 *Let us consider the quasi-copula Q given in example 2. Its associated GBM matrix is:*

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{-1}{4} & \frac{-1}{4} & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

The first step of the algorithm gives $c_1 = \frac{1}{4}$, reached for instance in position a_{22}. Thus, we can choose the following ASM matrix:

$$A_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Then, the matrix $A_1^* = \dfrac{1}{1-c_1}(A - c_1 \cdot A_1)$ is

$$\begin{pmatrix} \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{-1}{3} & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

A_1^* is a GBM matrix, but it is not an ASM matrix. Thus we apply steps 1, 2 and 3 to this matrix. Now we have $c_2 = \frac{1}{3}$, reached in a_{23}, and we choose an appropriate A_2:

$$A_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Next, we calculate the corresponding A_2^*:

$$A_2^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

which results to be an ASM matrix. Thus the algorithm has finished. Finally, we obtain the convex combination:

$$A = \frac{1}{4}A_1 + \frac{3}{4}A_1^* = \frac{1}{4}A_1 + \frac{3}{4}\left(\frac{1}{3}A_2 + \frac{2}{3}A_2^*\right) =$$

$$\frac{1}{4}A_1 + \frac{1}{4}A_2 + \frac{1}{2}A_2^*.$$

Consequently, the discrete quasi-copula Q is given by the convex linear combination:

$$Q = \frac{1}{4}Q_1 + \frac{1}{4}Q_2 + \frac{1}{2}Q_2^*$$

where Q_1, Q_2 and Q_2^* are the irreducible discrete quasi-copulas with associated ASM matrices A_1, A_2 and A_2^*, respectively.

Acknowledgement

This paper has been partially supported by the Spanish grants MTM2006-05540 and MTM2006-08322, and the Government of the Balearic Islands grant PCTIB-2005GC1-07.

References

[1] C. Alsina, R.B. Nelsen and B. Schweiser, On the characterization of a class of binary operations on distribution functions, Statist. Probab. Lett. 17 (1993) 85–89.

[2] F. Durante, Solution of an open problem for associative copulas, Fuzzy Sets and Systems 152 n. 2 (2005) 411–415.

[3] F. Durante, R. Mesiar and C. Sempi, On a family of copulas constructed from the diagonal section, Soft Computing 10 n. 6 (2006) 490–494.

[4] C. Genest, J.J. Quesada Molina, J.A. Rodríguez Lallena and C. Sempi, A characterization of quasi-copulas, Journal of Multivariate Analysis 69 (1999) 193–205.

[5] E.P. Klement and A. Kolesárová, Extension to copulas and quasi-copulas as special 1-Lipschitz aggregation operators, Kybernetika 41 n. 3 (2005) 329–348.

[6] E.P. Klement, R. Mesiar and E. Pap, Triangular norms, Kluwer Academic Publishers, Dordrecht, 2000.

[7] A. Kolesárová, R. Mesiar, J. Mordelová and C. Sempi, Discrete copulas, IEEE Transaction on Fuzzy System 14 n. 5 (2006) 698–705.

[8] A. Kolesárová and J. Mordelová, Quasi-copulas and copulas on a discrete scale, Soft Computing 10 (2006) 495–501.

[9] G. Mayor, J. Suñer and J. Torrens, Copula-like operations on finite settings, IEEE Transaction on Fuzzy Systems 13 n. 4 (2005) 468–477.

[10] G. Mayor, J. Suñer and J. Torrens, Sklar's Theorem in Finite Settings, IEEE Transaction on Fuzzy Systems, to be published.

[11] G. Mayor and J. Torrens, Triangular norms in discrete settings, in: E.P. Klement and R. Mesiar (Eds.), *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, Amsterdam, 2005, pp. 193–236.

[12] R.B. Nelsen, An introduction to copulas, Springer-Verlag, Lecture Notes in Statistics, vol. 139, New York, 1999.

[13] J.J. Quesada Molina and C. Sempi, Discrete quasi-copulas, Insur. Math. Econ. 37 n. 1 (2005) 27–41

[14] D.P. Robbins and H. Rumsey, Determinants and alternating-sign matrices, Advances in Math. 62 (1986) 169–184.

[15] A. Sklar, Fonctions de répartition à n dimensions et leur marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

Introduction to Asymptotically Idempotent Aggregation Operators

Roberto Ghiselli Ricci
D.E.I.R.
Via Torre Tonda, 34
Università di Sassari
07100 Sassari, Italy
ghiselli@unimo.it

Abstract

This paper deals with aggregation operators. A new class of aggregation operators, called asymptotically idempotent, is introduced. A generalization of the basic notion of aggregation operator is provided, with an in-depth discussion of the notion of idempotency. Some general contruction methods of commutative, asymptotically idempotent aggregation operators admitting a neutral element are illustrated.

Keywords: Aggregation Operator, Idempotency, Commutativity, Neutral Element.

1 Introduction

The mathematical process of fusion of several input real values into a single output real value is crucial in many fields. According to the various applications, different properties are requested to the aggregation. In particular, dealing with problems of ranking, the basic characteristics requested of the aggregation are *anonymity*, which occurs when the knowledge of the order of the input values is irrelevant, and *unanimity*, which reads as follows: when all the partial scores are equal to a certain value, this must be also the global score. Anonymity and unanimity are mathematically translated into commutativity and idempotency of the aggregation operator. However, especially when the number of the inputs is very large, less common aspects or demands assume a certain importance: many data could be devoid of significance (problematical aspect) or the possibility that a group of many positive scores may have a greater weight than one of a few positive scores (a refined request of the aggregation). In order to take into account the problematical aspect, it might be useful to introduce a neutral element, i.e. an element which has the same effect as its omitting. The presence of a neutral element contravenes the idempotency, but, if we relax idempotency in a suitable way, quite surprisingly, we can meet also the above mentioned refined request. In this work, we introduce a new class of aggregation operators, called asympotically idempotent, which, in presence of a neutral element, permit the output value to be sensitive to the number of input values, so undoubtedly improving the quality of ranking. We discuss the properties of this class and provide some general methods of construction of these operators.

2 Basic Concepts

In this work, we are interested in aggregation of input values, as well as outputs, belonging to some closed interval $[a, b] \subset \mathbb{R}$.

Definition 1 *A mapping*

$$g : [a, b]^n \to [a, b],\ n \in \mathbb{N}$$

is called an n-ary aggregation function (AF) acting on $[a, b]$ *if it is non-decreasing monotone in its components, that is*

$$g(x_1, ..., x_n) \leq g(x'_1, ..., x'_n), \tag{1}$$

whenever $a \leq x_i \leq x'_i \leq b$ *for all* $i \in \{1, ..., n\}$. *Moreover, g is strict if (1) holds with the strict inequality provided that* $(x_1, ..., x_n) \neq (x'_1, ..., x'_n)$. *Finally, g is commutative if*

$$g(x_1, ..., x_n) = g(x^*_1, ..., x^*_n) \tag{2}$$

for any permutation $(x^*_1, ..., x^*_n)$ *of an arbitrary tuple* $(x_1, ..., x_n) \in [a, b]^n$.

Remark 1 Regarding the property of continuity, according to (1), any AF is continuous if and only if it is continuous in its components.

Definition 2 *Let g be an n-ary AF acting on $[a,b]$. Fixed any $x \in [a,b]$, an element $(x,...,x) \in [a,b]^n$ is called an idempotent element for g if*

$$g(x,...,x) = x. \tag{3}$$

The n-ary AF g is idempotent, if g fulfills (3) for any $x \in [a,b]$.

Definition 3 *A sequence $\boldsymbol{G}=\{G_n\}_n$ of n-ary AFs acting on $[a,b]$ is called an aggregation operator on $[a,b]$ (briefly, AO on $[a,b]$).*

Definition 4 *An AO $\boldsymbol{G}=\{G_n\}_n$ on $[a,b]$ is called asymptotically idempotent (AI) if*

$$\lim_{n\to\infty} G_n(x,...,x) = x \quad \text{for all } x \in [a,b]. \tag{4}$$

Remark 2 It is our opinion that a sequence of n-ary AFs satisfying (4) is qualified for deserving the "title" of AO. In fact, from the theoretical point of view, the idempotent, "standard" AOs are a particular case of the AI ones; from the practical point of view, condition (4) assures the sensitivity of the output to the number of inputs, a refined property which is recommended in many applications, as told in the introduction. However, on the one hand, the AI AOs could not form a subclass of AOs, as to be expected, if we maintained the classical, commonly used in literature, definition of AO on a real closed interval $[a,b]$, which, in addition, requires the following two conditions:

$$G_n(a,...,a) = a,\ G_n(b,...,b) = b \tag{5}$$

and

$$G_1(x) = x \quad \text{for all } x \in [a,b]. \tag{6}$$

Indeed, there exist AI AOs which do not meet (5) and (6), as shown in the following example, where $[a,b] = [0,1]$ and the n-ary AF is

$$G_n(x_1,...,x_n) = \max_{i=1,...,n}\{x_i\} \cdot \frac{\sqrt{\sum_{i=1}^{n} {x_i}^2}}{1+\sqrt{\sum_{i=1}^{n} {x_i}^2}}. \tag{7}$$

On the other hand, there exist AOs, under the classical definition, which do not satify (4), as shown in the following example, where $[a,b] = [0,1]$ and the n-ary AF is

$$G_n(x_1,...,x_n) = \begin{cases} x_1, & \text{if } n = 1; \\ 0, & \text{if } x_1 = \cdots = x_n = 0 \\ & \text{for all } n; \\ 1, & \text{if } x_1 = \cdots = x_n = 1 \\ & \text{for all } n; \\ 0, & \text{otherwise} \\ & \text{and } n \text{ is even}; \\ 1, & \text{otherwise} \\ & \text{and } n > 2 \text{ is odd}. \end{cases}$$

A way which seems to be reasonable for bypassing this "cul-de-sac" is to weaken the definition of AO, omitting conditions (5) and (6).

Definition 5 *Let $\boldsymbol{G}=\{G_n\}_n$ be an AO on $[a,b]$. We say that $\boldsymbol{G}$ is commutative, idempotent, strict or continuous if, for each $n \in \mathbb{N}$ ($n \geq 2$ in case of commutativity), any G_n is commutative, idempotent, strict or continuous respectively.*

Definition 6 *An AO $\boldsymbol{G}=\{G_n\}_n$ on $[a,b]$ is called associative if for all $m,n \in \mathbb{N}$ and for all tuples $(x_1,...,x_m) \in [a,b]^m$ and $(y_1,...,y_n) \in [a,b]^n$*

$$G_{n+m}(x_1,...,x_m,y_1,...,y_n) =$$
$$= G_2(G_m(x_1,...,x_m), G_n(y_1,...,y_n)).$$

From the structural point of view, an associative AO $\boldsymbol{G}$ is uniquely determined by the corresponding binary AF G_2, hence, with abuse of notation, we will use the same symbol for $\boldsymbol{G}$ and G_2.

Definition 7 *Let $\boldsymbol{G}=\{G_n\}_n$ be an AO on $[a,b]$. Then an element $e \in [a,b]$ is called a neutral element (NE) for $\boldsymbol{G}$ if, for each $n \geq 2$, for each $k \in \{1,2,...,n\}$ and for all $x_1,...,x_{k-1},x_{k+1},...,x_n \in [a,b]$, we have*

$$G_n(x_1,...,x_{k-1},e,x_{k+1},...,x_n) =$$
$$= G_{n-1}(x_1,...,x_{k-1},x_{k+1},...,x_n). \tag{8}$$

Remark 3 Observe that if $\boldsymbol{G}=\{G_n\}_n$ is an AO on $[a,b]$ which admits $e \in [a,b]$ as NE, the binary AF G_2, according to (8), satisfies

$$G_2(e,x) = G_2(x,e) = G_1(x),$$

which reduces to the standard form when (6) holds. What is interesting is that unicity of neutral elements is preserved also in case of AI AOs, as stated by the following lemma

Proposition 1 *Let $\boldsymbol{G}=\{G_n\}_n$ be an AI AO on $[a,b]$ admitting $e \in [a,b]$ as NE. Then, e is the unique NE for $\boldsymbol{G}$.*

The next is a well-known result regarding the transformations of AOs by means of a strictly monotone bijection.

Proposition 2 *Let $\boldsymbol{G}=\{G_n\}_n$ be an AO on $[a,b]$ and $\varphi:[c,d]\to[a,b]$ a strictly monotone bijection, where $c,d\in\mathbb{R}$, with $c<d$. Then $\boldsymbol{G}^{\varphi}=\{G_n^{\varphi}\}_n$, where the n-ary AF G_n^{φ} acting on $[c,d]$ is defined by*

$$G_n^{\varphi}(u_1,...,u_n)=\varphi^{-1}\Big(G_n(\varphi(u_1),...,\varphi(u_n))\Big) \qquad (9)$$

for all $u_1,...,u_n\in[c,d]$, is an AO on $[c,d]$. Moreover, if $e\in[a,b]$ is a NE for $\boldsymbol{G}$, then $\varphi^{-1}(e)\in[c,d]$ is a NE for $\boldsymbol{G}^{\varphi}$. Finally, if $\boldsymbol{G}$ is commutative or continuous, then $\boldsymbol{G}^{\varphi}$ is commutative or continuous respectively.

This result suggests us to introduce a notion of *isomorphism* between AOs acting on the same interval.

Definition 8 *Let $\boldsymbol{G}=\{G_n\}_n$ and $\boldsymbol{G}^*=\{G_n^*\}_n$ be a pair of AOs on $[a,b]$. Then, we will say that $\boldsymbol{G}$ and $\boldsymbol{G}^*$ are isomorphic if there exists a strictly monotone bijection $\varphi:[a,b]\to[a,b]$ such that*

$$\boldsymbol{G}^*\equiv\boldsymbol{G}^{\varphi}.$$

Finally, we conclude this section with a generalized notion of idempotency for an AO.

Definition 9 *Let $\boldsymbol{G}=\{G_n\}_n$ be an AO on $[a,b]$. We will say that $\boldsymbol{G}$ is quasi-idempotent if for each $x_0\in[a,b]$ there exists an $n_0=n_0(x_0)\in\mathbb{N}$ such that $G_n(x_0,...,x_0)=x_0$ for all $n\geq n_0$.*

3 Commutative AOs with a NE

Let us denote by $\mathcal{A}$ and $\mathcal{B}$ the families of AOs on $[a,b]$, for any pair $(a,b)\in\mathbb{R}^2$ with $a<b$, admitting $e\in\{a,b\}$ and $e\in\,]a,b[$ as NE, respectively. Proposition 1 and Definition 8 allow us, without loss of generality and up to isomorphisms, to fix for both classes $e=0$ as NE and the domains $[0,1]$ and $[-1,1]$ respectively. Then, we set $\mathcal{E}:=\mathcal{A}\cup\mathcal{B}$ and we will denote by $\boldsymbol{A}=\{A_n\}_n$, $\boldsymbol{B}=\{B_n\}_n$ or $\boldsymbol{E}=\{E_n\}_n$ an arbitrary element of $\mathcal{A}$, $\mathcal{B}$ or $\mathcal{E}$ respectively: in the last case, we will denote by D the domain, where D may be indifferently $[0,1]$ or $[-1,1]$. Finally, given any $\boldsymbol{E}=\{E_n\}_n\in\mathcal{E}$, we set $d_n(x):=E_n(x,...,x)$, where $d_n:D\to D$ for all $n\in\mathbb{N}$.

Proposition 3 *Given any $\boldsymbol{E}=\{E_n\}_n\in\mathcal{E}$ and fixed any $x\in D$, the sequence $\{d_n(x)\}_n$ is monotone (strictly monotone if $\boldsymbol{E}$ is strict and $x\neq 0$).*

Proof If $x=0$, we have that $d_n(0)=E_n(0,...,0)=$ (according to (8)) $E_{n+1}(0,...,0)=d_{n+1}(0)$, so that $d_n(0)=d_1(0)$ for all $n\in\mathbb{N}$. If $x>0$, $d_n(x)=E_n(x,...,x)=$(according to (8)) $E_{n+1}(x,...,x,0)\leq$ (according to (1), with the strict inequality, if the operator is strict) $E_{n+1}(x,...,x,x)=d_{n+1}(x)$ for all $n\in\mathbb{N}$. The case $x<0$ may be shown in complete analogy. □

An immediate consequence is that the sequence $\{d_n(x)\}_n$ converges on D to a function we will denote by $d(x)$, where $d:D\to D$. Hence, it is clear that any $\boldsymbol{E}\in\mathcal{E}$ is AI if and only if $d(x)=x$ for all $x\in D$.

Proposition 4 *Given any $\boldsymbol{E}=\{E_n\}_n\in\mathcal{E}$, we have that $E_n(x_1,...,x_n)\geq 0$ (>0 if $\boldsymbol{E}$ is strict) for all $x_1,...,x_n\geq 0$ (and $x_j>0$ for some $j\in\{1,...,n\}$), while $E_n(x_1,...,x_n)\leq 0$ (<0 if $\boldsymbol{E}$ is strict) for all $x_1,...,x_n\leq 0$ (and $x_j<0$ for some $j\in\{1,...,n\}$).*

Remark 4 Note that, starting from an arbitrary $\boldsymbol{B}\in\mathcal{B}$, we can always generate an AO $\boldsymbol{A}^{\boldsymbol{B}}\in\mathcal{A}$, simply restricting the domain of $\boldsymbol{B}$ to the real unit interval, i.e. $\boldsymbol{A}^{\boldsymbol{B}}:=\boldsymbol{B}|_{[0,1]}$.

Definition 10 *We will say that $\boldsymbol{B}=\{B_n\}_n\in\mathcal{B}$ is symmetrical with respect to $e=0$ (e-symm, for short) if, for each $n\in\mathbb{N}$, we have*

$$B_n(x_1,...,x_n)=-B_n(|x_1|,...,|x_n|)$$

$$\text{for all } x_1,...,x_n\in[-1,0[.$$

Now, we set $\mathcal{CA}:=\{\boldsymbol{A}\in\mathcal{A}:\boldsymbol{A}\text{ is commutative}\}$ and $\mathcal{CB}:=\{\boldsymbol{B}\in\mathcal{B}:\boldsymbol{B}\text{ is commutative}\}$.

Remark 5 Observe that, starting from an arbitrary $\boldsymbol{A}\in\mathcal{CA}$, we can always generate an AO $\boldsymbol{B}^{\boldsymbol{A}}=\{B_n^A\}_n\in\mathcal{CB}$, where the n-ary AF is defined as follows:

$$B_n^A(x_1,...,x_n)=B_n^A(x_1^*,...,x_k^*,x_{k+1}^*,...,x_n^*)=$$

$$=A_k(x_1^*,...,x_k^*)-A_{n-k}(|x_{k+1}^*|,...,|x_n^*|),$$

where $(x_1^*,...,x_k^*,x_{k+1}^*,...,x_n^*)$ is any permutation of an arbitrary tuple $(x_1,...,x_n)\in[-1,1]^n$ such that $x_1^*,...,x_k^*\geq 0$, while $x_{k+1}^*,...,x_n^*<0$, for some $k\in\{0,...,n\}$, with the convention $A_0=0$. The only point deserving to be shown is the monotonicity of the arbitrary n-ary AF: given any $(x_1,...,x_n)\in[-1,1]^n$, without loss of generality, we can suppose that $x_1,...,x_k\geq 0$ and $x_{k+1},...,x_n<0$, where $k\in\{0,...,n-1\}$. Now, fixed any $i\in\{1,...,n\}$, we have to prove that

$$B_n^A(x_1,...,x_i,...,x_n)\leq B_n^A(x_1,...,x_i',...,x_n) \qquad (10)$$

for all $x_i'\in[x_i,1]$. The case $i\in\{1,...,k\}$ for $k\neq 0$ is trivial, so assume that $i\in\{k+1,...,n\}$ for any $k\in\{0,...,n-1\}$. If $x_i'\geq 0$, (10) becomes

$$A_k(x_1,...,x_k)-$$

$$-A_{n-k}(|x_{k+1}|,...,|x_{i-1}|,|x_i|,|x_{i+1}|,...,|x_n|) \leq$$
$$\leq A_{k+1}(x_1,...,x_k,x_i') -$$
$$-A_{n-k-1}(|x_{k+1}|,...,|x_{i-1}|,|x_{i+1}|,...,|x_n|). \quad (11)$$

Note that $A_k(x_1,...,x_k) = A_{k+1}(x_1,...,x_k,0) \leq A_{k+1}(x_1,...,x_k,x_i')$ by (8) and (1) respectively, and, by the same reasons, $A_{n-k-1}(|x_{k+1}|,...,|x_{i-1}|,|x_{i+1}|,...,|x_n|) = A_{n-k}(|x_{k+1}|,...,|x_{i-1}|,0,|x_{i+1}|,...,|x_n|) \leq A_{n-k}(|x_{k+1}|,...,|x_{i-1}|,|x_i|,|x_{i+1}|,...,|x_n|)$, so that (11) is assured. Finally, if $x_i' < 0$, (10) becomes

$$A_{n-k}(|x_{k+1}|,...,|x_i|,...,|x_n|) \geq$$
$$\geq A_{n-k}(|x_{k+1}|,...,|x_i'|,...,|x_n|),$$

which follows from (1) and the fact that $|x_i| \geq |x_i'|$.

Note that any $\boldsymbol{B^A}$ is $e-symm$ and further it fulfills the following stronger property:

$$B^A_{2n}(x_1,...,x_n,-x_1,...,-x_n) = 0 \quad (12)$$

for all $x_1,...,x_n \in [-1,0[\cup]0,1]$ and for each $n \in \mathbb{N}$.

Remark 6 Observe that, starting from an arbitrary $\boldsymbol{A} \in \mathcal{CA}$, if we generate, as shown in the previous remark, $\boldsymbol{B^A} \in \mathcal{CB}$ and subsequently $\boldsymbol{A^{B^A}} \in \mathcal{CA}$, as illustrated in Remark 4, we easily get $\boldsymbol{A} \equiv \boldsymbol{A^{B^A}}$. This is equivalent to saying that $\mathcal{CB}|_{[0,1]} = \mathcal{CA}$, where $\mathcal{CB}|_{[0,1]} := \{ \boldsymbol{A^B}: \boldsymbol{B} \in \mathcal{CB}\}$. The same does not occur if, starting from an arbitrary $\boldsymbol{B} \in \mathcal{CB}$, we consider first $\boldsymbol{A^B}$ and then $\boldsymbol{B^{A^B}}$, because we cannot generally conclude that $\boldsymbol{B} = \boldsymbol{B^{A^B}}$. Indeed, if this were true, $\boldsymbol{B}$ would necessarily satisfy (12), but, as we will see in the next section, there exist AOs belonging to $\mathcal{CB}$ which do not verify (12).

4 Construction methods of commutative, AI AOs with a NE

In this section, we are interested in providing some procedures for building different models of AI AOs belonging to $\mathcal{CE} := \mathcal{CA} \cup \mathcal{CB}$. First of all, we emphasize the fact that there are not many examples of idempotent AOs $\in \mathcal{CE}$ and generally they have a rather poor structure. For instance, if we restrict to the associative ones, the only associative, idempotent AO $\in \mathcal{CA}$ is the $t-$conorm given by the maximum (and, at the same time, if we fixed $e = 1$ as NE, we would find as unique example the t-norm given by the minimum). Otherwise, the associative, idempotent AOs $\in \mathcal{CB}$ form the more substantial family of idempotent uninorms: however, observe that, if $\boldsymbol{U}$ belongs to this class, necessarily $\boldsymbol{U}|_{[0,1]^2} \equiv \max$ and $\boldsymbol{U}|_{[-1,0]^2} \equiv \min$.

Now we present the first model of AI AO $\in \mathcal{CA}$, in which there is no necessity of permutation, or rearrangement in more general sense, of the input values.

Example 1 Let $\boldsymbol{A^{H,\psi}} = \{A_n^{\mathbf{H},\psi}\}_n$ be a sequence of n-ary AFs so described:

$$A_n^{\mathbf{H},\psi}(x_1,...,x_n) =$$
$$= \max\{x_1,...,x_n\} \cdot \psi\big(h_n(x_1,...,x_n)\big),$$

where $\psi : [0,\infty] \to [0,1]$ is a non-decreasing mapping such that $\psi(0) = 0$ and $\psi(\infty) = 1$, while $\mathbf{H} = \{h_n\}_n$ is a commutative AO acting on the interval $[0,\infty[$, with $e = 0$ as NE, such that $h_1(0) = 0$ and $\sup_{n\in\mathbb{N}} h_n(x,...,x) = \infty$ for any $x > 0$. The fact that $\boldsymbol{A^{H,\psi}}$ actually belongs to $\mathcal{CA}$ is quite easy to show. The set of mappings which behave as ψ is very large; what is more interesting is to investigate some simple ways to construct explicit examples of $\mathbf{H}$. For instance, if we consider any non-decreasing function $\mu : [0,\infty[\to [0,\infty[$ such that $\mu(0) = 0$, and $\mu(t) > 0$ as $t > 0$, it is trivial to see that $\mathbf{H}^\mu = \{h_n^\mu\}_n$, where the n-ary AF is defined

$$h_n^\mu(x_1,...,x_n) = \sum_{i=1}^n \mu(x_i),$$

fulfills all the required properties on $[0,\infty[$. Note that (7) is a particular case of this model, with $\psi(t) = \frac{\sqrt{t}}{1+\sqrt{t}}$ and $\mu(t) = t^2$. Finally, observe that $\boldsymbol{A^{H,\psi}}$ is continuous if ψ and $\mathbf{H}$ are continuous.

The next model of AI AO $\in \mathcal{CA}$ is a sort of generalization of ordered weighted average operator (OWA, for short), which requires a permutation of the input values.

Example 2 Let $\boldsymbol{W^{\Delta_\infty}} = \{W_n^{\Delta_\infty}\}_n$ be a sequence of n-ary AFs so described:

$$W_n^{\Delta_\infty}(x_1,...,x_n) = \sum_{i=1}^n w_i \cdot x_i',$$

where $(x_1',...,x_n')$ is a non-decreasing permutation of any arbitrary input n-tuple $(x_1,...,x_n)$, while $\Delta_\infty = \{w_n\}_n$ is a fixed sequence of real, non negative numbers such that

$$\sum_{n=1}^\infty w_n = 1. \quad (13)$$

It is not difficult to show that $\boldsymbol{W^{\Delta_\infty}}$ is a continuous AO belonging to $\mathcal{CA}$. Observe that $\boldsymbol{W^{\Delta_\infty}}$ is not absolutely reducible to the classical OWA, because, in that case, any n-ary AF has its own weighting triangle $\Delta_n = \{w_{1,n},...,w_{n,n}\}$, where $\sum_{i=1}^n w_{i,n} = 1$. However,

no weighting triangle is generally associated with any Δ_∞. Finally, note that, if we choose Δ_∞ such that $w_1 = 1$ and $w_i = 0$ as $i \geq 2$, then $\boldsymbol{W}^{\Delta_\infty} \equiv \max$, but, on the contrary, $\boldsymbol{W}^{\Delta_\infty} \not\equiv \min$, whatever is the Δ_∞ chosen.

Proposition 5 *The AO $\boldsymbol{A}^{\boldsymbol{H},\psi}$ is quasi-idempotent if and only if there exists $t_0 > 0$ such that $\psi(t_0) = 1$. The AO $\boldsymbol{W}^{\Delta_\infty}$ is quasi-idempotent if and only if there exists a $k \in \mathbb{N}$ such that $w_n = 0$ for all $n > k$.*

The final, general construction method we present regards a class of $e - symm$ AI AOs belonging to $\mathcal{CB}$. The philosophy of this method is that, according to Remark 6, the subdomain $[0,1]^n$ of the n-ary AF we have to define may be covered by the respective AF of any AI AO $\in \mathcal{CA}$, hence, by the symmetry, also the subdomain $[-1,0]^n$ is covered. The most interesting part is the rest of the domain, more precisely $cl([-1,1]^n \setminus I_n)$, i.e.the topological closure of the set $[-1,1]^n \setminus I_n$, where $I_n := [0,1]^n \cup [-1,0]^n$.

Example 3 Given any AI $\boldsymbol{A} \in \mathcal{CA}$, we set

$$\boldsymbol{B^*}|_{[0,1]\cup[-1,0]} := \boldsymbol{B^A}|_{[0,1]\cup[-1,0]},$$

i.e. for each $n \in \mathbb{N}$ we have $B^*_n(x_1,...,x_n) = A_n(x_1,...,x_n)$, if $(x_1,...,x_n) \in [0,1]^n$, while $B^*_n(x_1,...,x_n) = -A_n(|x_1|,...,|x_n|)$, if $(x_1,...,x_n) \in [-1,0]^n$. Let $f : [-1,1] \to [-1,1]$ be an arbitrary strictly increasing bijection such that $f(0) = 0$. Consider then a sequence $\{g_n\}_n$ of mappings from I_n to I_n defined

$$g_n(x_1,...,x_n) = f(B^*_n(x_1,...,x_n))$$

Evidently, any g_n is non-decreasing and commutative: further, for every $n \geq 2$, we have $g_n(x_1,...,x_{n-1},0) = g_{n-1}(x_1,...,x_{n-1})$ for all $(x_1,...,x_{n-1}) \in I_{n-1}$. Now, for each $n \geq 2$, we can define the n-ary AF B^*_n on $cl([-1,1]^n \setminus I_n)$ as follows:

$$B^*_n(x_1,...,x_n) = B^*_n(x^*_1,...,x^*_k,x^*_{k+1},...,x^*_n) =$$

$$= f^{-1}\big(g_k(x^*_1,...,x^*_k) + g_{n-k}(|x^*_{k+1}|,...,|x^*_n|)\big),$$

recalling that $(x^*_1,...,x^*_k,x^*_{k+1},...,x^*_n)$ is any permutation of an arbitrary input n-tuple $(x_1,...,x_n)$ such that $x^*_1,...,x^*_k \geq 0$, while $x^*_{k+1},...,x^*_n < 0$, for some $k \in \{0,...,n\}$. It is not difficult to check that B^*_n is well defined on $cl([-1,1]^n \setminus I_n)$ and the whole $\boldsymbol{B^*}$ so obtained is actually an $e - symm$ AI AO belonging to $\mathcal{CB}$. Note that if f is not symmetrical with respect to zero, unlike B^*_1, also $g_1(x) = f(B_1(x))$ is not symmetrical, hence, for some $x \neq 0$, by definition of B_2, we get $B_2(x,-x) = f^{-1}(g_1(x) + g_1(-x)) \neq 0$, so proving that such a kind of $\boldsymbol{B^*}$ generally does not fulfill (12).

References

[1] J. Aczel, Lectures on Functional Equations and Applications, Academic Press, New York, 1966.

[2] T. Calvo, G. Mayor, R. Mesiar (eds), Aggregation Operators. New trends and applications, Physica-Verlag, Heidelberg, New York, 2002

[3] J.C. Fodor, M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[4] J.C. Fodor, R.R. Yager, A. Rybalov, Structure of uninorms, Int. Jour. of Uncertainty, Fuzziness and Knowledge-Based Systems 5 (1975) 411-427.

[5] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.

[6] A.N. Kolmogorov, Sur la notion de la moyenne, Accad. Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. 12 (1930) 388-391.

[7] M. Nagumo, Űber eine Klasse der Mittelwerte, Japanese Jour. of Mathematics 6 (1930) 71-79.

[8] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Trans. on Systems, Man and Cybernetics 18 (1988) 183-190.

[9] R.R. Yager, A. Rybalov, Uninorm aggregation operators, Fuzzy Sets and Systems 80 (1996) 111-120.

MP and MT-implications on a finite scale

M. Mas
Dpt. de Matemàtiques i Inf.
Universitat de les Illes Balears.
07122 Palma de Mallorca. Spain
dmimmg0@uib.es

M. Monserrat
Dpt. de Matemàtiques i Inf.
Universitat de les Illes Balears.
07122 Palma de Mallorca. Spain
dmimma0@uib.es

J. Torrens
Dpt. de Matemàtiques i Inf.
Universitat de les Illes Balears.
07122 Palma de Mallorca. Spain
dmijts0@uib.es

Abstract

This paper is devoted to the study of discrete implications that satisfy modus ponens (MP), modus tollens (MT) or both (MPT). The main goal is to characterize all R, S, QL and D-implications on a finite chain L satisfying these properties for a given smooth t-norm T_1. The non-smooth case is also discussed for a special family of t-norms.

Keywords: Discrete implications, modus ponens, modus tollens, smooth t-norms, finite scale.

1 Introduction

It is well known that fuzzy implication functions (see the survey [12]) are used in approximate reasoning, not only to represent fuzzy conditional statements of the form "If p then q" (with p, q fuzzy statements), but also to perform inferences in any fuzzy rule based system. In this inference process, the two main classical rules are *modus ponens* (MP) and *modus tollens* (MT) that allow to perform, respectively, forward and backward inferences. In terms of fuzzy logic, these implications are operators $I : [0,1]^2 \to [0,1]$ extending the classical material implication, that is, satisfying $I(0,0) = I(0,1) = I(1,1) = 1$ and $I(1,0) = 0$.

Since conjunctions, disjunctions and negations are usually performed by t-norms (T), t-conorms (S) and strong negations (N), in fuzzy set theory as much as in fuzzy logic and approximate reasoning, the majority of the known implication functions are directly derived from these operators[1]. The four most usual ways to define these implication functions are:

i) *R-implications* defined by

$$I(x,y) = \sup\{z \in [0,1] \mid T(x,z) \leq y\} \quad (1)$$

for all $x, y \in [0,1]$.

ii) *S-implications* defined by

$$I(x,y) = S(N(x),y), \qquad x,y \in [0,1]. \quad (2)$$

iii) *QL-implications* defined by

$$I(x,y) = S(N(x),T(x,y)), \qquad x,y \in [0,1]. \quad (3)$$

iv) *D-implications*, that are the contraposition with respect to N of QL-implications and are given by

$$I(x,y) = S(T(N(x),N(y)),y), \qquad x,y \in [0,1]. \quad (4)$$

Moreover, in this context, given any t-norm T_1 and any strong negation N_1, modus ponens and modus tollens for an implication I can be written as

$$T_1(x, I(x,y)) \leq y, \qquad x,y \in [0,1] \quad (5)$$

and

$$T_1(N_1(y), I(x,y)) \leq N_1(x), \qquad x,y \in [0,1] \quad (6)$$

respectively. These two equations have been recently solved in [15] for the first three mentioned classes of implications.

On the other hand, the study of operators defined on finite scales is an area of increasing interest (see [2], [4], [5], [8], [9], [10], and [13]). Mainly, because it allows to deal with finite families of linguistic labels avoiding numerical interpretations (necessaries in the fuzzy logic approach). In this context, the book chapter [13] brings a survey on (smooth) discrete t-norms and t-conorms on a finite chain L, and the four classes of implications, R, S, QL and D-implications derived from discrete t-norms, are studied in [9] and

[1] Although some authors have derived also implications from other aggregation functions, specially uninorms (see [1], [11],[14]).

[10]. From these discrete implications, new possibilities for approximate reasoning with finite families of linguistic labels appear with their consequent applications in computing with words.

However, in this line, the study of MP and MT rules for discrete implications, equivalent to the one given in [15] for fuzzy implications, is essential. This is precisely the main goal of this paper. After some preliminaries given in section 2, we devote section 3 to characterize those R, S, QL and D-implications on a finite chain L, derived from smooth t-norms, that satisfy equation (5), equation (6) or both. Finally, section 4 is devoted to the non-smooth case for a special family of t-norms.

2 Preliminaries

We recall here the smooth t-norms and the smooth t-conorms on a finite chain L and their characterization, that will be used along the paper. It is well known that for our purposes (see [13]) all finite chains with the same number of elements are equivalent and then, from now on, we will deal with the simplest finite chain of $n+1$ elements:

$$L = \{0, 1, 2 \ldots, n-1, n\}$$

where $n \geq 1$. Such an L can be understood as a set of linguistic terms or "labels".

The following two definitions are adapted from [5] (see also [13]).

Definition 1 *A function $f : L \to L$ is said to be smooth if it satisfies one of the following conditions:*

- *f is nondecreasing and $f(x) - f(x-1) \leq 1$ for all $x \in L$ with $x \geq 1$.*
- *f is nonincreasing and $f(x-1) - f(x) \leq 1$ for all $x \in L$ with $x \geq 1$.*

Definition 2 *A binary operator F on L is said to be smooth if it is smooth in each variable.*

The importance of the smoothness condition lies in the fact that it is generally used as a discrete counterpart of continuity on [0,1].

Although t-norms, t-conorms and strong negations are usually binary operators on [0,1], they can be defined as in [2] on any bounded partially ordered set and, in particular, on L. In this last case, they are usually known as discrete t-norms and discrete t-conorms. In this way, recall that smoothness for discrete t-norms (and also for t-conorms) is equivalent to the *divisibility* condition, that is, $x \leq y$ if and only if there exists $z \in L$ such that $T(y, z) = x$, (see [13]).

Proposition 1 *There is one and only one strong negation on L that is given by*

$$N(x) = n - x \quad \textit{for all } x \in L. \tag{7}$$

From now on, N will always denote the negation on L given by (7). Smooth t-norms have been characterized as ordinal sums of Archimedean ones as follows.

Proposition 2 (See [13]). *There is one and only one Archimedean smooth t-norm on L, denoted by $T_{\mathbf{L}}$, given by*

$$T_{\mathbf{L}}(x, y) = \max\{0, x + y - n\} \quad \textit{for all } x, y \in L \tag{8}$$

which is known as the Łukasiewicz t-norm. Moreover, given any subset J of L containing $0, n$, there is one and only one smooth t-norm on L that has J as the set of idempotent elements, that will be denoted by T_J. In fact, if J is the set

$$J = \{0 = i_0 < i_1 < \ldots < i_{m-1} < i_m = n\}$$

then T_J is given by:

$$T_J(x, y) = \begin{cases} \max\{i_k, x + y - i_{k+1}\} & \textit{if } x, y \in [i_k, i_{k+1}] \\ & \textit{for some } i_k \in J \\ \min\{x, y\} & \textit{otherwise} \end{cases} \tag{9}$$

Smooth t-conorms have a classification theorem like the above one for t-norms which can be easily deduced by N-duality. The expression of the only Archimedean smooth t-conorm on L is given by

$$S_{\mathbf{L}}(x, y) = \min\{n, x + y\} \quad \text{for all } x, y \in L \tag{10}$$

which is also known as the Łukasiewicz t-conorm. In general, we have

Proposition 3 (See [13]). *Given any subset J of L containing $0, n$, there is one and only one smooth t-conorm on L, S_J, that has J as the set of idempotent elements. In fact, if J is the set*

$$J = \{0 = i_0 < i_1 < \ldots < i_{m-1} < i_m = n\}$$

then S_J is given by:

$$S_J(x, y) = \begin{cases} \min\{i_{k+1}, x + y - i_k\} & \textit{if } x, y \in [i_k, i_{k+1}] \\ & \textit{for some } i_k \in J \\ \max\{x, y\} & \textit{otherwise} \end{cases} \tag{11}$$

Note that with these notations, the Łukasiewicz t-norm and t-conorm can be written, respectively as $T_{\mathbf{L}} = T_{\{0,n\}}$ and $S_{\mathbf{L}} = S_{\{0,n\}}$. The following result follows from the previous propositions.

Proposition 4 (See [13]). *There are exactly 2^{n-1} different smooth t-norms (t-conorms) on L.*

Definition 3 *A binary operator $I : L \times L \to L$ is said to be a (discrete) implication if it satisfies:*

(I1) I is nonincreasing in the first variable and nondecreasing in the second one.

(I2) $I(0,0) = I(n,n) = n$ and $I(n,0) = 0$.

Note that, from the definition, it follows that $I(0,x) = n$ and $I(x,n) = n$ for all $x \in L$.

The four ways to define fuzzy implications apply here to define discrete implications. However, since the only strong negation on L is the one given by (7) and we deal with a finite scale, in our case they can be rewritten as follows:

$$I(x,y) = \max\{z \in L \mid T(x,z) \leq y\}, \quad x,y \in L. \tag{12}$$

$$I(x,y) = S(n-x,y), \qquad x,y \in L. \tag{13}$$

$$I(x,y) = S(n-x,T(x,y)), \qquad x,y \in L. \tag{14}$$

$$I(x,y) = S(T(n-x,n-y),y), \qquad x,y \in L. \tag{15}$$

All these classes of discrete implications have been already studied: R and S-implications in [9], and QL and D-implications in [10]. Thus we refer to these cited papers for details on these kinds of discrete implications that we will use in the paper. Although the non-smooth case is considered in these references, in the present work we will deal only with R, S, QL and D-operators derived from smooth t-norms and smooth t-conorms.

3 Main results

In this section we deal with (implication) operators on the finite chain L that satisfy the modus ponens, the modus tollens or both, with respect to a smooth t-norm T_1. Again, since we have only one strong negation on L, in our case equation (6) can be rewritten depending only on the t-norm T_1 and thus, we can adopt the following definitions:

Definition 4 *Let T_1 be a t-norm on L. A function $I : L^2 \to L$ will be called:*

- *an MP-operator for T_1 whenever it satisfies*

$$T_1(x, I(x,y)) \leq y \quad \text{for all } x,y \in L \tag{16}$$

- *an MT-operator for T_1 whenever it satisfies*

$$T_1(n-y, I(x,y)) \leq n-x \quad \text{for all } x,y \in L \tag{17}$$

- *an MPT-operator for T_1 whenever it is both, an MP and an MT-operator.*

Moreover, we will say that I is an MP-implication (or MT, or MPT-implication) if it is an implication and also an MP-operator (or MT or MPT-operator, respectively).

Note that whereas all R and S-operators (given by equations (12) and (13), respectively) are always implications, this is not the case for QL and D-operators (given by equations (14) and (15), respectively), see for instance [10]. Thus, we will divide our study of properties MP, MT and MPT in three subsections, the first one devoted to R-implications, the second one devoted to S-implications and, finally, the third one devoted to QL and D-operators in general, including of course QL and D-implications.

Before this, let us begin with two easy but important propositions in the discussion of the mentioned properties. The first one deals with MP-operators.

Proposition 5 *Let T_1 be a t-norm on L and $I : L^2 \to L$ an MP-operator for T_1. Then,*

$$T_1(x, I(x,0)) = 0 \quad \text{for all } x \in L.$$

The second one deals with MT-operators.

Proposition 6 *Let T_1 be a t-norm on L and $I : L^2 \to L$ an MT-operator for T_1. Then,*

$$T_1(n-y, I(n,y)) = 0 \quad \text{for all } y \in L.$$

3.1 R-implications

Given any t-norm T we will denote by I_T its residual implication, that is, the operator given by equation (12):

$$I_T(x,y) = \max\{z \in L \mid T(x,z) \leq y\}.$$

We deal in this subsection with implications $I_T : L^2 \to L$, where T is a smooth t-norm. Let us recall here the structure of these implications. For an easier understanding we give their graphical structure in Figure 1 instead of their formulas, that can be found in [9]. Thus, R and S-implications can be viewed in Figures 1 and 2, respectively.

We begin with the property of MP-implications. Given any t-norm T, let us denote also by $Idemp_T$ the set of all idempotent elements of the t-norm T, that is:

$$Idemp_T = \{x \in L \mid T(x,x) = x\}.$$

Then we have the following proposition.

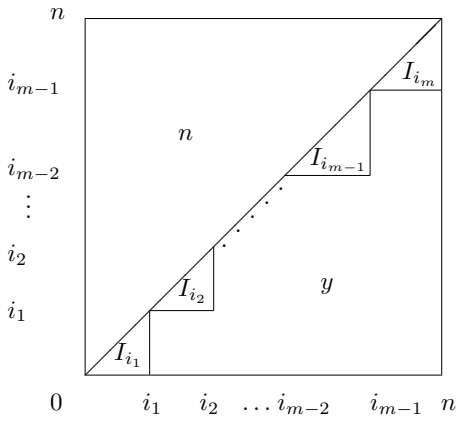


Figure 1. The structure of the R-implication derived from T_J where $J = \{0 = i_0 < i_1 < \ldots < i_{m-1} < i_m = n\}$ and $I_{i_{k+1}}(x, y) = i_{k+1} + y - x$ for $k = 0, ..., m-1$.

Proposition 7 *Let T, T_1 be smooth t-norms on L and $I_T : L^2 \to L$ the R-implication associated to T. The following statements are equivalent:*

i) *I_T is an MP-implication for T_1.*

ii) *$I_T(x, y) \leq I_{T_1}(x, y)$ for all $x, y \in L$.*

iii) *$Idemp_{T_1} \subseteq Idemp_T$.*

Thus, it is clear that we can deduce the following particular cases.

Corollary 1 *Let T, T_1 be smooth t-norms on L and $I_T : L^2 \to L$ the R-implication associated to T. Then,*

i) *If $T = \min$, $I_{\min}$ is an MP-implication for any smooth t-norm T_1.*

ii) *If $T = T_{\mathbf{L}}$, $I_{T_{\mathbf{L}}}$ is an MP-implication for T_1 if and only if $T_1 = T_{\mathbf{L}}$.*

iii) *If $T_1 = T_{\mathbf{L}}$, I_T is an MP-implication for any smooth t-norm T.*

iv) *If $T_1 = \min$, I_T is an MP-implication for* min *if and only if $T = \min$.*

With respect to MT-implications we obtain solutions only when T_1 is the Łukasiewicz t-norm and then I_T works for any smooth t-norm T. This can be proved through the following two propositions.

Proposition 8 *Let T, T_1 be t-norms on L with T smooth and let $I_T : L^2 \to L$ be the R-implication associated to T. If I_T is an MT-implication for T_1 then necessarily $T_1(x, n - x) = 0$ for all $x \in L$.*

Remark 1 *It is proved in Lemma 1 of [9] that, for smooth t-norms, the previous condition is equivalent to be T_1 the Łukasiewicz t-norm. That is, for smooth t-norms,*

$$T_1(x, n - x) = 0 \ \text{ for all } \ x \in L \iff T_1 = T_{\mathbf{L}}.$$

Proposition 9 *Let T, T_1 be smooth t-norms on L and $I_T : L^2 \to L$ the R-implication associated to T. Then I_T is an MT-implication for T_1 if and only if $T_1 = T_{\mathbf{L}}$.*

Now, jointly the obtained results for MP and for MT-implications we obtain the following corollary.

Corollary 2 *Let T, T_1 be smooth t-norms on L and $I_T : L^2 \to L$ the R-implication associated to T. Then,I_T is an MPT-implication for T_1 if and only if $T_1 = T_{\mathbf{L}}$.*

3.2 S-implications

Let us now deal with S-implications. Given any t-conorm S we will denote by I_S the corresponding S-implication given by equation (13). That is:

$$I_S(x, y) = S(n - x, y) \quad \text{for all } x, y \in L.$$

The structure of S-implications can be viewed in Figure 2 (their expression can be found in [9]).

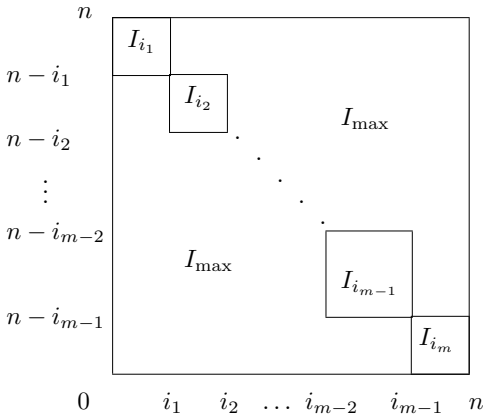


Figure 2. The structure of the S-implication derived from the dual t-conorm of T_J, where $J = \{0 = i_0 < i_1 < \ldots < i_{m-1} < i_m = n\}$, $I_{\max}(x, y) = \max\{n - x, y\}$ and $I_{i_{k+1}}(x, y) = \min\{n - i_k, i_{k+1} + y - x\}$ for $k = 0, ..., m-1$.

For S-implications, the study of the modus ponens is solved by the following two propositions.

Proposition 10 *Let T_1 be a t-norm, S a smooth t-conorm on L and $I_S : L^2 \to L$ the corresponding S-*

implication. If I_S is an MP-implication for T_1 then necessarily $T_1(x, n-x) = 0$ for all $x \in L$.

Proposition 11 *Let T_1 be a smooth t-norm and S a smooth t-conorm on L and $I_S : L^2 \to L$ the corresponding S-implication. Then I_S is an MP-implication for T_1 if and only if $T_1 = T_{\mathbf{L}}$.*

In this case, the study of modus tollens gives exactly the same solutions. Specifically,

Proposition 12 *Let T_1 be a t-norm and S a smooth t-conorm on L and $I_S : L^2 \to L$ the corresponding S-implication. If I_S is an MT-implication for T_1 then necessarily $T_1(x, n-x) = 0$ for all $x \in L$.*

Proposition 13 *Let T_1 be a smooth t-norm and S a smooth t-conorm on L and $I_S : L^2 \to L$ the corresponding S-implication. Then I_S is an MT-implication for T_1 if and only if $T_1 = T_{\mathbf{L}}$.*

And consequently we have:

Corollary 3 *Let T_1 be a smooth t-norm and S a smooth t-conorm on L and $I_S : L^2 \to L$ the corresponding S-implication. Then I_S is an MPT-implication for T_1 if and only if $T_1 = T_{\mathbf{L}}$.*

3.3 QL and D-operators

In this subsection we deal with QL and D-operators given by equations (14) and (15), respectively. As we have commented, not all of them are implications in the sense of Definition 3. In fact, it is proved in [10] that this occurs in the smooth case (for both QL and D) if and only if $S = S_{\mathbf{L}}$. However we will study the properties MP, MT and MPT in general for all QL and D-operators, regardless of they are or they are not implications.

Again we can begin with this first proposition.

Proposition 14 *Let T_1 be a t-norm, I_{QL} a QL-operator and I_D a D-operator. Then:*

i) *If I_{QL} (I_D) is an MP-operator for T_1 then necessarily $T_1(x, n-x) = 0$ for all $x \in L$.*

ii) *If I_{QL} (I_D) is an MT-operator for T_1 then necessarily $T_1(x, n-x) = 0$ for all $x \in L$.*

When we deal with QL and D-implications, since then the t-conorm S in equations (14) and (15) must be $S = S_{\mathbf{L}}$, they are derived simply from a smooth t-norm as:

$$I_{QL}(x, y) = S_{\mathbf{L}}(n - x, T(x, y)) = n - x + T(x, y) \quad (18)$$

for all $x, y \in L$, and

$$I_D(x, y) = S_{\mathbf{L}}(T(n - x, n - y), y) = y + T(n - x, n - y) \quad (19)$$

for all $x, y \in L$, respectively. Moreover, it is proved in [10] that the set of QL-implications and the set of D-implications coincide when we derive them from smooth t-norms and t-conorms, and consequently we can study both kind of implications at the same time. Let us recall the structure of QL and D-implications in Figure 3 (again their formulas can be found in [10]).

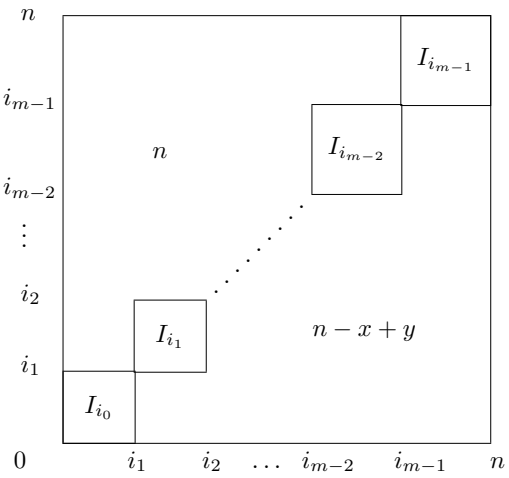


Figure 3. The structure of the QL-implication derived from T_J where $J = \{0 = i_0 < i_1 < \ldots < i_{m-1} < i_m = n\}$. For $j = 0, ..., m-1$, each I_{i_j} is given by $I_{i_j}(x, y) = \max\{n - x + i_j, n + y - i_{j+1}\}$ for all $x, y \in [i_j, i_{j+1}]$. The same structure corresponds to the D-implication derived from $T_{N(J)}$.

In the study of QL and D-implications that are MP, MT and MPT for T_1 we obtain always the same result. Any of these conditions is satisfied if and only if $T_1 = T_{\mathbf{L}}$. For QL-operators and D-operators in general, the result is the same, but now we need to study both kinds of operators separately. In any case, we have:

Proposition 15 *Let T_1 be a smooth t-norm, I_{QL} a QL-operator and I_D a D-operator. Then, the following statements are equivalent:*

i) *I_{QL} (I_D) is an MP-operator for T_1.*

ii) *I_{QL} (I_D) is an MT-operator for T_1.*

iii) *I_{QL} (I_D) is an MPT-operator for T_1.*

iv) *$T_1 = T_{\mathbf{L}}$.*

A table summarizing all results in this section can be viewed in Table 1.

	MP for T_1	MT for T_1	MPT for T_1
I_T	$\substack{Idemp_{T_1} \\ \subseteq Idemp_T}$	$T_1 = T_{\mathbf{L}}$	$T_1 = T_{\mathbf{L}}$
I_S	$T_1 = T_{\mathbf{L}}$	$T_1 = T_{\mathbf{L}}$	$T_1 = T_{\mathbf{L}}$
I_{QL}	$T_1 = T_{\mathbf{L}}$	$T_1 = T_{\mathbf{L}}$	$T_1 = T_{\mathbf{L}}$
I_D	$T_1 = T_{\mathbf{L}}$	$T_1 = T_{\mathbf{L}}$	$T_1 = T_{\mathbf{L}}$

Table 1. Characterization of R, S, QL and D-operators that are MP, MT and MPT-operators for a smooth t-norm T_1.

4 The non-smooth case

In our previous study we have seen that, given any t-norm T_1 and any binary operator $I : L^2 \to L$, to be I an MP, MT, or MPT-operator, the condition

$$T_1(x, n-x) = 0 \quad \text{for all} \quad x \in L \tag{20}$$

is necessary in almost all cases. From Remark 1 we know that in the smooth case this is equivalent to be T_1 the Łukasiewicz t-norm. However, in the non-smooth case we have many others t-norms on L satisfying such condition. Namely, for any $k \in L$ such that $n - k \leq k$ we have the following indexed family of t-norms T^k given by

$$T^k(x,y) = \begin{cases} 0 & \text{if } x+y \leq n \\ x+y-k & \text{if } x+y > n \text{ and} \\ & \quad n-k \leq x,y \leq k \\ \min\{x,y\} & \text{otherwise.} \end{cases} \tag{21}$$

Each t-norm of this family satisfies equation (20), see for instance [9]. In fact, these t-norms T^k are the discrete counterpart of the family of t-norms introduced by J. C. Fodor in [3] (see also [6]) when he studies contrapositive symmetry (genuine property of S-implications) for R-implications. For each t-norm of this family the corresponding R and S-implications coincide (see also [7], Jenei family of t-norms in pages 97–98). This fact is also true for our family T^k in the discrete case (see [9]).

Note that T^k is non-smooth except for the case $k = n$ and in this extreme case T^n agrees with the Łukasiewicz t-norm $T_{\mathbf{L}}$. Note also that the nilpotent minimum (see again [9]) is obtained in the other extreme case given by $n - k = \lfloor n/2 \rfloor$ where $\lfloor n/2 \rfloor$ means the floor of $n/2$, that is, the greatest integer which is smaller than or equal to $n/2$.

Remark 2 *Note that, in the particular case when $n - k = k$, n must be an even number and $k = n/2$. Moreover, in this case both T^k and T^{k+1} coincide with the nilpotent minimum. Thus, from now on, we will consider only the cases when $n - k < k$ without any loose of generality.*

We will consider also $n \geq 3$ since for the cases $n = 1, 2$ it is always $T^k = T_{\mathbf{L}}$.

The indexed family of t-norms T^k can be viewed in Figure 4.

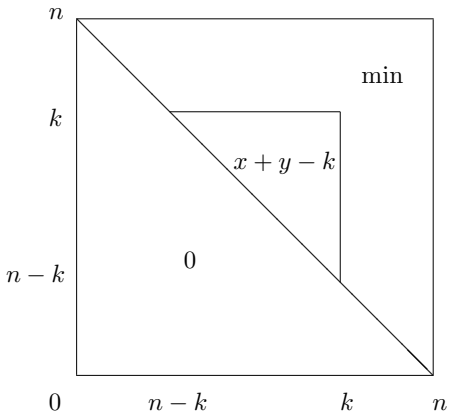


Figure 4. The structure of the t-norm T^k.

Thus, since each T^k satisfies the necessary condition (20), we can study which R, S, QL and D-implications derived from smooth t-norms are MP, MT or MPT-implications for each T^k.

In this line we have the following results with respect to the modus ponens. We begin with R-implications.

Proposition 16 *Let T be a smooth t-norm and I_T its corresponding R-implication. Then, I_T is an MP-implication for T^k if and only if $Idemp_T$ contains the set $[k, n] = \{x \in L \mid k \leq x\}$.*

The case of S-implications is solved in the following proposition.

Proposition 17 *Let S be a smooth t-conorm and I_S its corresponding S-implication. Then, I_S is an MP-implication for T^k if and only if $Idemp_S$ contains the set $[0, n-k]$.*

The case of QL and D-implications is also easy. Since they are derived from $S_{\mathbf{L}}$ and a smooth t-norm T, they are given by equations (18) and (19) respectively.

We will only deal with T^k when $n - k < k < n$ since the case $k = n$ corresponds to the Łukasiewicz t-norm that has been studied in the previous section.

Proposition 18 *Suppose $n - k < k < n$. Let T be a smooth t-norm and I_{QL} and I_D the corresponding QL*

and D-implications derived from T. Then, the following statements are equivalent:

i) *I_{QL} is an MP-implication for T^k*

ii) *I_D is an MP-implication for T^k*

iii) *T is the Łukasiewicz t-norm $T_{\mathbf{L}}$.*

iv) *$I_{QL} = I_D = I$ is the Kleene-Dienes implication given by*

$$I(x, y) = \max(n - x, y).$$

In table 2 we can view summarized all results in this section concerning MP-implications for the t-norm T^k.

	MP-implication for T^k
R-implication from T	$[k, n] \subseteq Idemp_T$
S-implications from S	$[0, n-k] \subseteq Idemp_S$
QL-implications from T	$T = T_{\mathbf{L}}$
D-implications from T	$T = T_{\mathbf{L}}$

Table 2. Characterization of MP-implications for the t-norm T^k with $n - k < k < n$.

On the other hand, the general case of QL and D-operators is not so easy. That is, when these operators are derived from a smooth t-conorm S different from $S_{\mathbf{L}}$. In this case, we can give only some partial results, as follows:

- When $S = \max$, for any smooth t-norm T, the QL-operator given by

  $$I_{QL}(x, y) = \max(n - x, T(x, y))$$

  and the D-operator given by

  $$I_D(x, y) = \max(T(n - x, n - y), y)$$

  are MP-operators for T^k $(n - k \leq k \leq n)$.
- When $S(x, x) = x$ for all $x \leq n-k$, both the QL and the D-operator derived from S and any smooth t-norm T are MP-operators for T^k.

Finally, we want to deal with the MT-property for T^k.

From the duality between MP and MT, we will be able to derive identical results for the case of modus tollens, to the ones obtained for the modus ponens, just by contraposition.

The only exception of this is for R-implications. In this case the results can not be derived from contraposition and we need to study MT independently of MP. However, we also obtain an identical result to the one obtained for modus ponens.

Proposition 19 *Let T be a smooth t-norm and I_T its corresponding R-implication. Then, I_T is an MT-implication for T^k if and only if $Idemp_T$ contains the set $[k, n]$.*

In all remaining cases all results can be derived from contraposition. First of all, note that S-implications always satisfy contraposition with respect to the unique negation $N(x) = n - x$. On the other hand, D-operators are the contraposition (with respect to $N(x) = n - x$) of QL-operators and vice versa. Using these facts and the duality between MP and MT we can easily prove the results concerning MT.

Proposition 20 *Let S be a smooth t-conorm and I_S its corresponding S-implication. Then, the following statements are equivalent:*

i) *I_S is an MP-implication for T^k*

ii) *I_S is an MT-implication for T^k*

iii) *I_S is an MPT-implication for T^k*

iv) *$Idemp_S$ contains the set $[0, n - k]$.*

Proposition 21 *Suppose $n - k < k < n$. Let T be a smooth t-norm and I_{QL} and I_D the corresponding QL and D-implications derived from T. Then, the following statements are equivalent:*

i) *I_{QL} (and I_D) is an MP-implication for T^k*

ii) *I_{QL} (and I_D) is an MT-implication for T^k*

iii) *I_{QL} (and I_D) is an MPT-implication for T^k*

iv) *T is the Łukasiewicz t-norm $T_{\mathbf{L}}$.*

v) *$I_{QL} = I_D = I$ is the Kleene-Dienes implication.*

The results for MT are summarized now in table 3.

	MT-implication for T^k
R-implication from T	$[k, n] \subseteq Idemp_T$
S-implications from S	$[0, n-k] \subseteq Idemp_S$
QL-implications from T	$T = T_{\mathbf{L}}$
D-implications from T	$T = T_{\mathbf{L}}$

Table 3. Characterization of MT-implications for the t-norm T^k with $n - k < k < n$.

To finish, note that the general case of modus ponens for QL and D-operators can be translated also for modus tollens via duality, obtaining again exactly the same results.

In view of the characterizations of MP and MT conditions, since both coincide in all four cases, it is clear that in each case the corresponding characterization also works in fact for the MPT-condition.

5 Conclusion

The two main inference rules, modus ponens (MP) and modus tollens (MT), are studied for the four most usual classes of discrete implications: R, S, QL and D-implications. A characterization of MP and a characterization of MT is given for all these kinds of implications, obtaining in the majority of cases the condition $T_1(x, n-x) = 0$, which directly derives (in the smooth case) into the Łukasiewicz t-norm. For this reason, the non-smooth case is also studied for a general class of discrete t-norms T_1 that satisfy the condition above. In this study, a lot of new solutions among R, S, QL and D-implications, derived from smooth t-norms, is obtained for both properties MP and MT.

Acknowledgement

This paper has been partially supported by the Spanish grant MTM2006-05540 and the Government of the Balearic Islands grant PCTIB-2005GC1-07.

References

[1] B. De Baets and J. C. Fodor, Residual operators of uninorms, Soft Computing 3 (1999) 89–100.

[2] B. De Baets and R. Mesiar, Triangular norms on product lattices, Fuzzy Sets and Systems 104 (1999) 61–75.

[3] J.C. Fodor, Contrapositive symmetry on fuzzy implications, Fuzzy Sets and Systems 69 (1995) 141–156.

[4] J. C. Fodor, Smooth associative operations on finite ordinal scales, IEEE Trans. on Fuzzy Systems 8 (2000) 791–795.

[5] Ll. Godo and C. Sierra, A new approach to connective generation in the framework of expert systems using fuzzy logic, in: *Proceedings of the XVIIth ISMVL*, Palma de Mallorca, 1988, pp. 157–162.

[6] S. Jenei (2000), New family of triangular norms via contrapositive symmetrization of residuated implications, *Fuzzy Sets and Systems*, Vol. 110, 157-174.

[7] E.P. Klement, R. Mesiar and E. Pap, Triangular norms, Kluwer Academic Publishers, Dordrecht (2000).

[8] M. Mas, G. Mayor and J. Torrens, $t-$Operators and uninorms on a finite totally ordered set, Int. J. of Intelligent Systems 14 (1999) 909–922.

[9] M. Mas, M. Monserrat and J. Torrens, S-implications and R-implications on a finite chain, Kybernetika 40 (2004) 3–20.

[10] M. Mas, M. Monserrat and J. Torrens, On two types of discrete implications, Int. J. of Approximate Reasoning 40 (2005) 262–279.

[11] M. Mas, M. Monserrat and J. Torrens, QL and D-Implications from uninorms, in : *Proceedings of IPMU-2006*, Paris, France, 2006, pp. 622–629.

[12] M. Mas, M. Monserrat, J. Torrens and E. Trillas, A survey on fuzzy implication functions, IEEE Transactions on Fuzzy Systems, accepted.

[13] G. Mayor and J. Torrens, Triangular norms in discrete settings, in: E.P. Klement and R. Mesiar (Eds.), *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, Amsterdam, 2005, pp. 193–236.

[14] D. Ruiz and J. Torrens, Residual implications and co-implications from idempotent uninorms, Kybernetika 40 (2004) 21–38.

[15] E. Trillas, C. Alsina and A. Pradera, On MPT-implication functions for fuzzy logic, Rev. R. Acad. Cien. Serie A. Mat. 98 (2004) 259–271.

(U, N)-Implications and their Characterizations

Balasubramaniam Jayaram
Dept. of Mathematics and Computer Sciences,
Sri Sathya Sai University,
Prasanthi Nilayam, A.P-515134, INDIA.
jbala@ieee.org

Michał Baczyński
Institute of Mathematics,
University of Silesia,
40-007 Katowice, ul. Bankowa 14, POLAND.
michal.baczynski@us.edu.pl

Abstract

In this work we characterize (U, N)-implications obtained from disjunctive uninorms and continuous negations.

Keywords: Fuzzy implication, Uninorm, Fuzzy negation, (U, N)-implication.

1 Introduction

(U, N)-implications are some generalizations of (S, N)-implications, where a t-conorm S is replaced by a uninorm U. A similar generalization of residual implications from the setting of t-norms to the setting of uninorms has been done by De Baets and Fodor in [3]. Ruiz and Torrens have investigated quite extensively on fuzzy implications from uninorms [11] and their distributivity [10], [12].

Despite this interest, fuzzy implications obtained from uninorms are yet to be characterized. Recently, some characterizations of (S, N)-implications were given by the authors in [2]. In this work, along similar lines, we investigate and characterize (U, N)-implications obtained from continuous negations N.

After introducing the necessary preliminaries on the basic fuzzy logic operations, we list out some of the most desirable - but relevant to this work - properties of fuzzy implications and investigate their interdependencies. Following this we discuss the class of (U, N)-operations and the properties they satisfy. Finally, based on the above analysis, we derive a characterization for (U, N)-implications generated from continuous negations.

2 Basic Fuzzy Logic Operations

To make this work self-contained, we briefly mention some of the concepts and results employed in the rest of the work.

Definition 1 (see [4, 7])**.** A decreasing function $N\colon [0,1] \to [0,1]$ is called a fuzzy negation if $N(0) = 1$ and $N(1) = 0$. A fuzzy negation N is called

- strict if it is both strictly decreasing and continuous;
- strong if it is an involution, i.e., $N(N(x)) = x$ for all $x \in [0,1]$.

It is well-known that if $[a, b]$ and $[c, d]$ are two closed subintervals of $[-\infty, +\infty]$ and $f\colon [a, b] \to [c, d]$ is a monotone function, then the set of discontinuous points of f is a countable subset of $[a, b]$ (see [9]). In this case we will use the pseudo-inverse $f^{(-1)}\colon [c, d] \to [a, b]$ of a decreasing and non-constant function f defined by (see [7, Sect. 3.1])

$$f^{(-1)}(y) = \sup\{x \in [a, b] \mid f(x) > y\}, \qquad y \in [c, d].$$

Lemma 1 ([2], Proposition 28)**.** *If N is a continuous fuzzy negation, then the function $\mathfrak{N}\colon [0,1] \to [0,1]$ defined by*

$$\mathfrak{N}(x) = \begin{cases} N^{(-1)}(x), & \text{if } x \in (0,1], \\ 1, & \text{if } x = 0, \end{cases} \tag{1}$$

is a strictly decreasing fuzzy negation. Moreover

$$\mathfrak{N}^{(-1)} = N, \tag{2}$$

$$N \circ \mathfrak{N} = \mathrm{id}_{[0,1]}, \tag{3}$$

$$\mathfrak{N} \circ N\,|_{\mathrm{Ran}(\mathfrak{N})} = \mathrm{id}_{\mathrm{Ran}(\mathfrak{N})}. \tag{4}$$

Definition 2 (see [5])**.** An associative, commutative, increasing operation $U\colon [0,1]^2 \to [0,1]$ is called a uninorm, if there exists an $e \in [0,1]$ (called the neutral element) such that

$$U(e, x) = x, \qquad x \in [0,1].$$

Remark 1. (i) If $e = 0$, then U is a t-conorm and if $e = 1$, then U is a t-norm.

(ii) It can be easily showed, that the neutral element e corresponding to a uninorm U is unique.

(iii) For any uninorm U we have $U(0,1) \in \{0,1\}$.

(iv) A uninorm U such that $U(0,1) = 0$ is called a conjunctive uninorm and if $U(0,1) = 1$ it is called a disjunctive uninorm.

Examples of fuzzy negations, uninorms as well as the different classes of uninorms (the classes U_{min}, U_{max}, representable uninorms, idempotent uninorms) can be found in recent literature (see [4, Chap. 1], [7, Sect. 10.2], [3, 5]).

3 Fuzzy Implications

3.1 Definition and Properties

In this work the following equivalent definition proposed by Fodor and Roubens [4] is used.

Definition 3. A function $I\colon [0,1]^2 \to [0,1]$ is called a fuzzy implication operation, or a fuzzy implication, if it satisfies the following conditions:

$$I \text{ is decreasing in the first variable,} \tag{I1}$$
$$I \text{ is increasing in the second variable,} \tag{I2}$$
$$I(0,0) = 1, \tag{I3}$$
$$I(1,1) = 1, \tag{I4}$$
$$I(1,0) = 0. \tag{I5}$$

The set of all fuzzy implications will be denoted by $\mathcal{FI}$.

Directly from the above definition we see that each fuzzy implication I satisfies the following left and right boundary condition, respectively:

$$I(0,y) = 1, \qquad y \in [0,1], \tag{LB}$$
$$I(x,1) = 1, \qquad x \in [0,1]. \tag{RB}$$

Therefore, I satisfies also the normality condition

$$I(0,1) = 1. \tag{NC}$$

Consequently, every fuzzy implication restricted to the set $\{0,1\}^2$ coincides with the classical implication.

Definition 4. Let $I\colon [0,1]^2 \to [0,1]$ be any function and $\alpha \in [0,1)$. The function N_I^α given by

$$N_I^\alpha(x) = I(x,\alpha), \qquad x \in [0,1]$$

is called the natural negation of I with respect to α.

Lemma 2. *Let $I\colon [0,1]^2 \to [0,1]$ be any function and $\alpha \in [0,1)$ be arbitrary but fixed. Then the following statements are equivalent:*

(i) N_I^α is a fuzzy negation.

(ii) $I(0,\alpha) = 1$ and $I(1,\alpha) = 0$.

Proof. (i) $\Longrightarrow$ (ii) Since N_I^α is a fuzzy negation, $I(0,\alpha) = N_I^\alpha(0) = 1$ and $I(1,\alpha) = N_I^\alpha(1) = 0$.

(ii) $\Longrightarrow$ (i) This implication is obvious from the definition of a fuzzy negation.

□

It should be noted that for any $I \in \mathcal{FI}$ we have (I5), so for $\alpha = 0$ we have the natural negation $N_I = N_I^0$ of I. Also α should be less than 1, since $I(1,1) = 1$.

In the following we list out some of the desirable properties of fuzzy implications:

Definition 5. Let I be a fuzzy implication and N a fuzzy negation.

(i) I is said to have the exchange principle, if

$$I(x, I(y,z)) = I(y, I(x,z)), \tag{EP}$$

for all $x, y, z \in [0,1]$,

(ii) I is said to satisfy the law of left contraposition with respect to N if, for any $x, y \in [0,1]$,

$$I(N(x), y) = I(N(y), x). \tag{L-CP}$$

(iii) I is said to satisfy the law of right contraposition with respect to N, if, for any $x, y \in [0,1]$,

$$I(x, N(y)) = I(y, N(x)). \tag{R-CP}$$

(iv) I is said to satisfy the law of contraposition with respect to N, if, for any $x, y \in [0,1]$,

$$I(x,y) = I(N(y), N(x)). \tag{CP}$$

Lemma 3 ([2], Lemma 17)**.** *Let $I\colon [0,1]^2 \to [0,1]$ be any function and N a continuous fuzzy negation.*

(i) If I satisfies (I1) *and R-CP(N), then I satisfies* (I2).

(ii) If I satisfies (I2) *and R-CP(N), then I satisfies* (I1).

Lemma 4. *Let $I\colon [0,1]^2 \to [0,1]$ and N_I^α be a fuzzy negation for an arbitrary but fixed $\alpha \in [0,1)$.*

(i) If I satisfies (I2)*, then I satisfies* (I5).

(ii) Let I have (I2) *and* (EP). *Then I satisfies* (I3) *if and only if I satisfies* (I4).

(iii) If I satisfies (EP)*, then I satisfies R-CP(N_I^α),*

Proof. (i) Since N_I^α is a fuzzy negation and I satisfies (I2) we get $I(1,0) \le I(1,\alpha) = N_I^\alpha(1) = 0$.

(ii) Let I have (I2) and (EP). If I satisfies (I4), then since $N_I^\alpha(0) = 1$ we have $1 = I(1,1) = I(1, N_I^\alpha(0)) = I(1, I(0,\alpha)) = I(0, I(1,\alpha)) = I(0, N_I^\alpha(1)) = I(0,0) = 1$, i.e., I satisfies (I3). The reverse implication can be shown similarly.

(iii) Since I satisfies (EP), we have $I(x, N_I^\alpha(y)) = I(x, I(y,\alpha)) = I(y, I(x,\alpha)) = I(y, N_I^\alpha(x))$, i.e., I has R-CP(N_I^α).

□

Lemma 5. *Let I be any fuzzy implication and N_I^α be a continuous fuzzy negation for an arbitrary but fixed $\alpha \in [0,1)$. If N is a strictly decreasing fuzzy negation such that $N_I^\alpha \circ N = \mathrm{id}_{[0,1]}$ and I satisfies* (EP)*, then I satisfies L-CP(N).*

Proof. By our assumptions we get

$$\begin{aligned} I(N(x),y) &= I(N(x), N_I^\alpha \circ N(y)) \\ &= I(N(x), I(N(y),\alpha)) \\ &= I(N(y), I(N(x),\alpha)) \\ &= I(N(y), N_I^\alpha \circ N(x)) \\ &= I(N(y),x), \end{aligned}$$

for any $x, y \in [0,1]$ □

Remark 2. Under the assumptions of Lemma 5, we have:

(i) If N_I^α is a strict negation, then I satisfies L-CP($(N_I^\alpha)^{-1}$).

(ii) If N_I^α is a strong negation, then I satisfies CP(N_I^α).

3.2 (S,N)-Implications and their Characterization

In this section, we give a brief introduction to one of the families of fuzzy implications that is very well studied in the fuzzy literature.

Definition 6 (cf. [1, 4, 13])**.** A function $I\colon [0,1]^2 \to [0,1]$ is called an (S,N)-implication, if there exist a t-conorm S and a fuzzy negation N such that

$$I(x,y) = S(N(x),y), \qquad x,y \in [0,1]. \tag{5}$$

If N is a strong negation, then I is called a strong implication (S-implication, for short).

The following characterization of some subclasses of (S,N)-implications is from [2], which is an extension of a result in [13].

Theorem 1 ([2])**.** *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is an (S,N)-implication generated from some t-conorm S and some continuous (strict, strong) fuzzy negation N.

(ii) I satisfies (I1), (EP) *and the function N_I is a continuous (strict, strong) fuzzy negation.*

Moreover, the representation of (S,N)-implication is unique in this case.

In Theorem 1, the property (I1) can be substituted by (I2). Moreover, axioms in the above theorem are independent from each other.

4 (U,N)-Operations and (U,N)-Implications

A natural generalization of (S,N)-implications in the uninorm framework is to consider a uninorm in the place of a t-conorm.

4.1 Definition and Properties

Definition 7. A function $I\colon [0,1]^2 \to [0,1]$ is called a (U,N)-operation, if there exist a uninorm U and a fuzzy negation N such that

$$I_{U,N}(x,y) = U(N(x),y), \qquad x,y \in [0,1]. \tag{6}$$

If a (U,N)-operation is generated from U and N, then we will often denote this by $I_{U,N}$.

Proposition 1. *If $I_{U,N}$ is a (U,N)-operation, then*

(i) $I_{U,N}$ satisfies (I1), (I2), (I5), (NC) *and* (EP),

(ii) $N_{I_{U,N}}^e = N$ and $I_{U,N}$ satisfies R-CP(N),

(iii) if N is strict, then $I_{U,N}$ satisfies L-CP(N^{-1}),

(iv) if N is strong, then $I_{U,N}$ satisfies CP(N).

Proof. (i) By the monotonicity of U and N we get that $I_{U,N}$ satisfies (I1) and (I2). Moreover, it can be easily verified that $I_{U,N}$ satisfies (I5) and (NC). Finally, from the associativity and the commutativity of U we have also (EP).

(ii) For any $x \in [0,1]$ we have

$$N_{I_{U,N}}^e(x) = I_{U,N}(x,e) = U(N(x),e) = N(x).$$

Next, since $I_{U,N}$ satisfies (EP), from Lemma 4(iii) with $\alpha = e$ we have that $I_{U,N}$ satisfies R-CP(N).

(iii) If N is a strict negation, then because of Remark 2(i) we can deduce, that $I_{U,N}$ satisfies L-CP(N^{-1}).

(iv) If N is a strong negation, then because of Remark 2(ii) we can deduce, that $I_{U,N}$ satisfies CP(N).

□

If $e = 0$, then U is a t-conorm and $I_{U,N}$, as an (S, N)-implications, is always a fuzzy implication. If $e = 1$, then U is a t-norm and $I_{U,N}$ is not a fuzzy implication, since (I3) is violated. If $e \in (0, 1)$, then not for every uninorm U the (U, N)-operation is a fuzzy implication. Next results characterize these (U, N)-operation, which satisfy (I3) and (I4).

Theorem 2 (cf. [3])**.** *Let U be a uninorm with the neutral element $e \in (0, 1)$. Then the following statements are equivalent:*

(i) The function $I_{U,N}$ as defined in (6) *is a fuzzy implication.*

(ii) U is a disjunctive uninorm, i.e., $U(0, 1) = 1$.

Proof. Let U be a uninorm with the neutral element $e \in (0, 1)$.
(i) $\Longrightarrow$ *(ii)* If $I_{U,N}$ as defined in (6) is a fuzzy implication, then from (I3) we have $U(0, 1) = U(1, 0) = I_{U,N}(0, 0) = 1$.
(ii) $\Longrightarrow$ *(i)* Assume that $U(0, 1) = 1$. From Proposition 1 it is enough to show only (I3) and (I4):

$$I_{U,N}(0, 0) = U(N(0), 0) = U(1, 0) = U(0, 1) = 1,$$
$$I_{U,N}(1, 1) = U(N(1), 1) = U(0, 1) = 1.$$

□

Following the terminology used by Mas *et al.* [8] for QL-implications, only if the (U, N)-operation $I_{U,N}$ is a fuzzy implication we use the term (U, N)-implication.

Lemma 6. *Let $I_{U,N}$ be a (U, N)-implication obtained from a uninorm U with $e \in (0, 1)$ as its neutral element and continuous negation N. Let $\alpha \in (0, 1)$ be an arbitrary but fixed number. Then the following statements are equivalent:*

(i) $N^{\alpha}_{I_{U,N}} = N$.

(ii) $\alpha = e$.

Proof. Let $e \in (0, 1)$ be the neutral element of U and $\alpha \in (0, 1)$ be an arbitrary but fixed number.

(i) $\Longrightarrow$ *(ii)* If $N^{\alpha}_{I_{U,N}} = N$, then since N is continuous there exists an e' such that $e = N(e')$ and $N^{\alpha}_{I_{U,N}}(e') = I_{U,N}(e', \alpha) = U(N(e'), \alpha) = N(e') = e$. But $U(N(e'), \alpha) = U(e, \alpha) = \alpha$, because e is the neutral element of U. Hence $\alpha = e$.

(ii) $\Longrightarrow$ *(i)* On the other hand, if $\alpha = e$, then

$$N^{\alpha}_{I_{U,N}}(x) = I_{U,N}(x, \alpha) = I_{U,N}(x, e) = U(N(x), e) = N(x)$$

for all $x \in [0, 1]$, i.e., $N^{\alpha}_{I_{U,N}} = N$.

□

4.2 Characterizations of (U, N)-Implications

We start our presentation with following result.

Proposition 2. *Let I be a fuzzy implication and N any fuzzy negation. If we define a binary operation $U_{I,N}$ on $[0, 1]$ as follows*

$$U_{I,N}(x, y) = I(N(x), y), \qquad x, y \in [0, 1], \tag{7}$$

then

(i) $U_{I,N}(x, 1) = U_{I,N}(1, x) = 1$ for all $x \in [0, 1]$, in particular $U_{I,N}(0, 1) = 1$,

(ii) $U_{I,N}$ is increasing in both the variables,

(iii) $U_{I,N}$ is commutative if and only if I has L-CP(N).

In addition, if I has L-CP(N), then

(iv) $U_{I,N}$ is associative if and only if I satisfies the exchange property (EP)*.*

(v) an arbitrary $\alpha \in (0, 1)$ is the neutral element of $U_{I,N}$ if and only if $N^{\alpha}_{I} \circ N = \mathrm{id}_{[0,1]}$.

Proof. (i) Let $x \in [0, 1]$. By the boundary condition (RB) on I we have $U_{I,N}(x, 1) = I(N(x), 1) = 1$. Also, $U_{I,N}(1, x) = I(N(1), x) = I(0, x) = 1$ again by (LB) of I.

(ii) That $U_{I,N}$ is increasing in both variables is a direct consequence of the monotonicity of I and N.

(iii) If $U_{I,N}$ is commutative, then for all $x, y \in [0, 1]$ we get $I(N(x), y) = U_{I,N}(x, y) = U_{I,N}(y, x) = I(N(y), x)$, i.e., I satisfies L-CP(N). The reverse implication can be obtained by retracing the above steps.

(iv) Let $x,y,z\in[0,1]$. If I satisfies (EP), then

$$\begin{aligned}U_{I,N}(x,U_{I,N}(y,z)) &= I(N(x),I(N(y),z))\\ &= I(N(x),I(N(z),y))\\ &= I(N(z),I(N(x),y))\\ &= I(N(I(N(x),y)),z)\\ &= I(N(U_{I,N}(x,y)),z)\\ &= U_I(U_{I,N}(x,y),z).\end{aligned}$$

On the other hand, if $U_{I,N}$ is associative, then

$$\begin{aligned}I(x,I(y,z)) &= U_{I,N}(N(x),U_{I,N}(N(y),z))\\ &= U_{I,N}(U_{I,N}(N(x),N(y)),z))\\ &= U_{I,N}(U_{I,N}(N(y),N(x)),z))\\ &= U_{I,N}(N(y),U_{I,N}(N(x),z))\\ &= I(y,I(x,z)).\end{aligned}$$

(v) Let $\alpha\in(0,1)$ be arbitrary fixed. If α is the neutral element of $U_{I,N}$, then, for any $x\in[0,1]$, we have $x=U_{I,N}(x,\alpha)=I(N(x),\alpha)=N_I^\alpha(N(x))$. Conversely, if $N_I^\alpha\circ N=\mathrm{id}_{[0,1]}$, then, for any $x\in[0,1]$ we get $U_{I,N}(\alpha,x)=U_{I,N}(x,\alpha)=I(N(x),\alpha)=N_I^\alpha(N(x))=x$ and α is the neutral element of $U_{I,N}$.

□

If N_I^α is a continuous fuzzy negation for an arbitrary but fixed $\alpha\in(0,1)$, then by Lemma 1 and previous results we can consider the modified pseudo-inverse $\mathfrak{N}_I^\alpha$ given by

$$\mathfrak{N}_I^\alpha(x)=\begin{cases}(N_I^\alpha)^{(-1)}(x), & \text{if } x\in(0,1],\\ 1, & \text{if } x=0,\end{cases}\tag{8}$$

as the potential candidate for the fuzzy negation N in (7). Hence from Lemma 5 with $N=\mathfrak{N}_I^\alpha$ we obtain the following result.

Corollary 1 (cf. [2], Corollary 29)**.** *If a fuzzy implication I satisfies* (EP) *and N_I^α, the natural negation of I with respect to an arbitrary but fixed $\alpha\in(0,1)$, is a continuous fuzzy negation, then I satisfies* (L-CP) *with $\mathfrak{N}_I^\alpha$ from* (8).

Hence, if a fuzzy implication I satisfies (EP) and N_I^α is a continuous fuzzy negation for some $\alpha\in(0,1)$, then we conclude, that the formula (7) can be considered for the modified pseudo-inverse of the natural negation of I.

Corollary 2. *If $I\in\mathcal{FI}$ satisfies* (EP) *and N_I^α is a continuous fuzzy negation with respect to an arbitrary but fixed $\alpha\in(0,1)$, then the function U_I defined by*

$$U_I(x,y)=I(\mathfrak{N}_I^\alpha(x),y),\qquad x,y\in[0,1]\tag{9}$$

is a disjunctive uninorm with neutral element α, where $\mathfrak{N}_I$ is as defined in (8).

Theorem 3. *For a function $I\colon[0,1]^2\to[0,1]$ the following statements are equivalent:*

(i) I is an (U,N)-operation generated from some disjunctive uninorm U with neutral element $e\in(0,1)$ and some continuous fuzzy negation N.

(ii) I is an (U,N)-implication generated from some uninorm U with neutral element $e\in(0,1)$ and some continuous fuzzy negation N.

(iii) I satisfies (I1), (I3), (EP) *and the function N_I^e is a continuous negation for some $e\in(0,1)$.*

Moreover, the representation (6) *of (U,N)-implication is unique in this case.*

Proof. That (i) is equivalent to (ii) follows immediately from Theorem 2.

(ii) $\Longrightarrow$ *(iii)* Assume, that I is an (U,N)-implication based on a uninorm U with neutral element $e\in(0,1)$ and a continuous negation N. Since every (U,N)-implication is a fuzzy implication, I satisfies (I1) and (I3). Moreover, by Proposition 1 it satisfies (EP) and $N_I^e=N$. In particular N_I^e is continuous.

(iii) $\Longrightarrow$ *(ii)* Firstly see, that from Lemma 4(iii) it follows that I satisfies (R-CP) with respect to the continuous N_I^e. Next, Lemma 3(i) implies that I satisfies (I2). Once again from Lemma 4(i) and (ii) we have that I satisfies (I3), (I4) and (I5), and hence $I\in\mathcal{FI}$. Further, by virtue of Lemmas 1 and 5 the implication I satisfies L-CP($\mathfrak{N}_I^e$). Because of Corollary 2 the function U_I defined by (9) is a disjunctive uninorm with the neutral element e.
We will show that $I_{U_I,N_I^e}=I$. Fix arbitrarily $x,y\in[0,1]$. If $x\in\mathrm{Ran}(\mathfrak{N}_I^e)$, then by (4) we have

$$\begin{aligned}I_{U_I,N_I^e}(x,y) &= U_I(N_I^e(x),y)\\ &= I(\mathfrak{N}_I^e\circ N_I^e(x),y)=I(x,y).\end{aligned}$$

If $x\notin\mathrm{Ran}(\mathfrak{N}_I^e)$, then from the continuity of N_I^e there exists $x_0\in\mathrm{Ran}(\mathfrak{N}_I^e)$ such that $N_I^e(x)=N_I^e(x_0)$. Firstly see, that $I(x,y)=I(x_0,y)$ for all $y\in[0,1]$. Indeed, let us fix arbitrarily $y\in[0,1]$. From the continuity of N_I^e there exists $y'\in[0,1]$ such that $N_I^e(y')=y$, so

$$\begin{aligned}I(x,y) &= I(x,N_I^e(y'))=I(y',N_I^e(x))\\ &= I(y',N_I^e(x_0))=I(x_0,N_I^e(y'))=I(x_0,y).\end{aligned}$$

From the above fact we get

$$\begin{aligned}I_{U_I,N_I^e}(x,y) &= U_I(N_I^e(x),y)\\ &= U_I(N_I^e(x_0),y)=I(x_0,y)=I(x,y),\end{aligned}$$

so I is an (U,N)-implication.
Finally, assume that there exist two continuous fuzzy

negations N_1, N_2 and two uninorms U_1, U_2 with neutral elements $e, e' \in (0,1)$, respectively, such that $I(x,y) = U_1(N_1(x), y) = U_2(N_2(x), y)$ for all $x, y \in [0,1]$. Fix arbitrarily $x_0, y_0 \in [0,1]$. Firstly observe that from Proposition 1 we get $N_1 = N_2 = N_I^e = N_I^{e'}$. By virtue of Lemma 6 we get, that $e' = e$. Now, since N_I^e is a continuous negation there exists $x_1 \in [0,1]$ such that $N_I^e(x_1) = x_0$. Thus $U_1(x_0, y_0) = U_1(N_I^e(x_1), y_0) = U_2(N_I^e(x_1), y_0) = U_2(x_0, y_0)$, i.e., $U_1 = U_2$. Hence N and U are uniquely determined. In fact $U = U_I$ defined by (9). □

In above theorem the property (I1) can be substituted by (I2) and the property (I3) can be substituted by (I4). Moreover, the above axioms are independent from each other.

Now, the following result easily follows:

Theorem 4. *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is an (U,N)-implication generated from some disjunctive uninorm U with neutral element $e \in (0,1)$ and some strict (strong) fuzzy negation N.

(ii) I satisfies (I1), (I3), (EP) *and the function N_I^e is a strict (strong) negation.*

Once again, the representations of the (U,N)-implications described above are unique and the presented axioms are independent from each other. It is interesting, that using similar methods as in this section we are able to obtain the following characterization of (U,N)-operations.

Theorem 5. *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is an (U,N)-operation generated from some uninorm U with neutral element $e \in (0,1)$ and some continuous fuzzy negation N.

(ii) I satisfies (I1), (EP) *and the function N_I^e is a continuous negation for some $e \in (0,1)$.*

Once again, in the above theorems, the property (I1) can be substituted by (I2).

5 Concluding Remarks

In this work, we characterize (U,N)-implications obtained from disjunctive uninorms U and continuous negations N. Toward this end, we have investigated some desirable algebraic properties of fuzzy implications and obtained some characterization results. It should be noted, that (U,N)-implications are closely related with e-implications investigated in [6], whose representation is still unknown.

References

[1] C. Alsina, E. Trillas, When (S,N)-implications are (T,T_1)-conditional functions? *Fuzzy Sets and Systems* **134** (2003) 305–310.

[2] M. Baczyński, B. Jayaram, On the characterizations of (S,N)-implications, *Fuzzy Sets and Systems*, (2007), doi:10.1016/j.fss.2007.02.010.

[3] B. De Baets, J. Fodor, Residual operators of uninorms, *Soft Computing* **3** (1999) 89–100.

[4] J. Fodor, M. Roubens, *Fuzzy preference modeling and multicriteria decision support*, Kluwer, Dordrecht, 1994.

[5] J. Fodor, R. Yager, A. Rybalov, Structure of uninorms, *Internat. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **5** (1997) 411–427.

[6] Gh. Khaledi, M. Mashinchi, S.A. Ziaie, The monoid structure of e-implications and pseudo-e-implications, *Inform. Sci.* **174** (2005) 103–122.

[7] E.P. Klement, R. Mesiar, E. Pap, *Triangular norms,* Kluwer, Dordrecht, 2000.

[8] M. Mas, M. Monserrat, J. Torrens, QL-implications versus D-implications, *Kybernetika* **42** (2006) 351–366.

[9] W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, New York, 1976.

[10] D. Ruiz, J. Torrens, Distributive idempotent uninorms, *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems* **11** (2003) 413–428.

[11] D. Ruiz, J. Torrens, Residual implications and co-implications from idempotent uninorms, *Kybernetika* **40** (2004) 21–38.

[12] D. Ruiz, J. Torrens, Distributivity of strong implications over conjunctive and disjunctive uninorms, *Kybernetika* **42** (2006) 319–336.

[13] E. Trillas, L. Valverde, On some functionally expressable implications for fuzzy set theory, in: E.P. Klement (Ed.) *Proc. of the 3rd Inter. Seminar on Fuzzy Set Theory, Linz*, Austria, 1981, pp. 173–190.

Intersections between Basic Families of Fuzzy Implications: (S,N)-, R- and QL-Implications

Balasubramaniam Jayaram
Dept. of Mathematics and Computer Sciences,
Sri Sathya Sai University,
Prasanthi Nilayam, A.P-515134, INDIA.
jbala@ieee.org

Michal Baczyński
Institute of Mathematics,
University of Silesia,
40-007 Katowice, ul. Bankowa 14, POLAND.
michal.baczynski@us.edu.pl

Abstract

In this work, our primary focus is to determine the intersections that exist between the family of QL-implications and the families of (S,N)- and R-implications. Toward this end, firstly, we investigate the conditions under which a QL-operation becomes a fuzzy implication. Since the exchange principle and the ordering property of fuzzy implications play an important role in this study, we propose some necessary and/or sufficient conditions on the underlying operations under which QL-operations satisfy them. As part of this attempt some interesting results pertaining to natural negations from t-conorms and the exact intersection between (S,N)- and R-implications have been obtained. We also mention some open problems relating to QL-implications.

Keywords: Fuzzy implication, QL-implication, R-implication, S-implication, (S,N)-implication, Law of excluded middle.

1 Introduction

The natural generalization of the implication in quantum logic to fuzzy logic – QL-operations – has not received as much attention as (S,N)- and R-implications. Perhaps, one of the reasons can be attributed to the fact that not all members of this family satisfy one of the main properties expected of a fuzzy implication, viz., left antitonicity. Also, in the earlier works, some conditions imposed on the fuzzy logic operations employed in the definition of QL-operations restricted both the class of operations from which they could be obtained and the properties these implications satisfied (see Remark 5 in Section 6.2 for details).

In this work, we study the family of QL-operations in fuzzy logic, without any restrictions on the underlying operations. We propose some necessary and/or sufficient conditions on the underlying operations under which QL-operations satisfy some of the most desirable algebraic properties. Finally, we return to the prime focus of this work, viz., the intersections that exist between the family of QL-implications and the families of (S,N)- and R-implications. Toward this end, we have also precisely determined the intersection of the families of (S,N)- and R-implications.

2 Preliminaries

Firstly we briefly mention some of the concepts and results employed in the rest of the work.

Definition 1 (see [5, 7, 8])**.** A decreasing function $N\colon [0,1] \to [0,1]$ is called a fuzzy negation if $N(0) = 1$ and $N(1) = 0$. A fuzzy negation N is said to be

- strict if it is strictly decreasing and continuous;
- strong if it is an involution, i.e., $N(N(x)) = x$ for all $x \in [0,1]$;
- non-vanishing if $N(x) = 0 \iff x = 1$;
- non-filling if $N(x) = 1 \iff x = 0$.

The classical negation $N_{\mathbf{C}}(x) = 1 - x$ is a strong negation, while $N_{\mathbf{K}}(x) = 1 - x^2$ is only strict, whereas $N_{\mathbf{D1}}$ and $N_{\mathbf{D2}}$ are non-filling and non-vanishing negations, respectively, where:

$$N_{\mathbf{D1}}(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{if } x > 0, \end{cases} \qquad N_{\mathbf{D2}}(x) = \begin{cases} 1, & \text{if } x < 1, \\ 0, & \text{if } x = 1. \end{cases}$$

Definition 2 (see [8])**.** An associative, commutative, increasing operation $T\colon [0,1]^2 \to [0,1]$ is called t-norm if it has neutral element equal to 1. An associative, commutative, increasing operation $S\colon [0,1]^2 \to [0,1]$ is called t-conorm if it has neutral element equal to 0. A t-norm T is said to be positive if $T(x,y) = 0 \iff x = 0$ or $y = 0$. A t-conorm S is said to be negative if $S(x,y) = 1 \iff x = 1$ or $y = 1$.

Table 1: Examples of t-norms

Name	**Formula**
$T_{\mathbf{M}}$: minimum	$\min(x, y)$
$T_{\mathbf{P}}$: product	$x \cdot y$
$T_{\mathbf{L}}$: Łukasiewicz	$\max(x + y - 1, 0)$,
$T_{\mathbf{D}}$: drastic product	$\begin{cases} 0, & \text{if } x, y \in [0, 1) \\ \min(x, y), & \text{otherwise} \end{cases}$
$T_{\mathbf{nM}}$: nilpotent minimum	$\begin{cases} 0, & \text{if } x + y \leq 1 \\ \min(x, y), & \text{otherwise} \end{cases}$

Table 2: Examples of t-conorms

NAME	**FORMULA**
$S_{\mathbf{M}}$: maximum	$\max(x, y)$
$S_{\mathbf{P}}$: algebraic sum	$x + y - x \cdot y$
$S_{\mathbf{L}}$: Łukasiewicz	$\min(x + y, 1)$
$S_{\mathbf{D}}$: drastic sum	$\begin{cases} 1, & \text{if } x, y \in (0, 1] \\ \max(x, y), & \text{otherwise} \end{cases}$
$S_{\mathbf{nM}}$: nilpotent maximum	$\begin{cases} 1, & \text{if } x + y \geq 1 \\ \max(x, y), & \text{otherwise} \end{cases}$

Firstly, we show that one can obtain a fuzzy negation from any t-conorm S and discuss its relevance vis-á-vis the law of excluded middle, which in the classical case has the following form: $p \vee \neg p$.

Definition 3 (cf. [8], Definition 5.5.2)**.** Let S be a t-conorm. A function $N_S \colon [0, 1] \to [0, 1]$ defined as

$$N_S(x) = \inf\{y \in [0, 1] \mid S(x, y) = 1\}, \quad x \in [0, 1], \quad (1)$$

is called the natural negation of S.

Remark 1. (i) It is easy to prove that N_S is a fuzzy negation for every t-conorm S.

(ii) If S is a negative t-conorm, then $N_S = N_{\mathbf{D2}}$.

(iii) Let S be a t-conorm and $x \in [0, 1]$ be fixed. Let us denote $A_x = \{y \in [0, 1] \mid S(x, y) = 1\}$. Then $1 \in A_x$ and hence $A_x \neq \emptyset$. If $x_0 = \inf A_x$, then, by the monotonicity of S, we have that either $A_x = [x_0, 1]$ or $A_x = (x_0, 1]$.

(iv) If $S(x, y) = 1$ for some $x, y \in [0, 1]$, then $y \geq N_S(x)$. On the other hand, since A_x is always an interval for every fixed $x \in [0, 1]$, if $y > N_S(x)$, then $S(x, y) = 1$.

Definition 4 (cf. [5])**.** Let S be a t-conorm and N a fuzzy negation. We say that the pair (S, N) satisfies the law of excluded middle if

$$S(N(x), x) = 1, \qquad x \in [0, 1]. \qquad \text{(LEM)}$$

A graphical interpretation of (LEM) is as follows: the graph of the negation N demarcates the region on the unit square $[0, 1]^2$ above which $S = 1$. It is possible that there are a few more points below the graph of N on whom S assumes the value 1. For example, consider the Łukasiewicz t-conorm $S_{\mathbf{L}}$ and the strict negation $N_{\mathbf{K}}(x) = 1 - x^2$. Then $S_{\mathbf{L}}(0.5, N(0.5)) = S_{\mathbf{L}}(0.5, 0.75) = 1$. Also notice that $S_{\mathbf{L}}(0.5, 0.5) = 1$. On the other hand, it should be emphasized that $S = S_{\mathbf{D}}$ does not satisfy (LEM) with its natural negation $N_{S_{\mathbf{D}}} = N_{\mathbf{D1}}$.

Lemma 1. *If a t-conorm S and a fuzzy negation N satisfy* (LEM)*, then*

(i) $N(x) \geq N_S(x)$, for all $x \in [0, 1]$;

(ii) $N_S \circ N(x) \leq x$, for all $x \in [0, 1]$.

Proof. (i) On the contrary, if for some $x_0 \in [0, 1]$ we have $N(x_0) < N_S(x_0)$, then $S(N(x_0), x_0) < 1$ by Remark 1 (iv).
(ii) From Definition 3 we have, for any $x \in [0, 1]$,

$$N_S(N(x)) = \inf\{y \in [0, 1] \mid S(N(x), y) = 1\}.$$

Now, since $S(N(x), x) = 1$, we get $x \geq N_S(N(x))$. □

It follows from Remark 1 (ii) that if S is a negative t-conorm, then S satisfies (LEM) only with the greatest fuzzy negation $N_{\mathbf{D2}}$.

Proposition 1. *If S is a right-continuous t-conorm with the natural negation N_S, then the following statements are equivalent:*

(i) N_S is continuous.

(ii) N_S is strong.

Proof. By the right-continuity of S we can show that the infimum in (1) reduces to minimum and thus the pair (S, N_S) satisfies (LEM). Hence $x \geq N_S \circ N_S(x)$, so $N_S(x) \leq N_S \circ N_S \circ N_S(x)$. On the other hand, since (LEM) holds for every $x \in [0, 1]$ we have

$$S(N_S \circ N_S(x), N_S(x)) = S(N_S(x), N_S \circ N_S(x)) = 1,$$

which implies that $N_S(x) \geq N_S \circ N_S \circ N_S(x)$. Since N_S is continuous, for every $y \in [0, 1]$ there exists an $x \in [0, 1]$ such that $y = N_S(x)$. Therefore, from the above inequalities, we get that $y = N_S \circ N_S(y)$ for every $y \in [0, 1]$, i.e., N_S is a strong negation.
The reverse implication is obvious. □

Remark 2. Just as one can obtain the natural negation N_S from a t-conorm, the natural negation of a t-norm T can be obtained as follows:

$$N_T(x) = \sup\{y \in [0, 1] \mid T(x, y) = 0\}, \quad x \in [0, 1].$$

The counterpart of the law of excluded middle is the law of contradiction

$$T(N(x),x)=0, \qquad x\in[0,1], \tag{LC}$$

where T is a t-norm and N is a fuzzy negation. For more on the above laws of excluded middle and contradiction, see for example [5].

It should be noted, that the following relation exists between N_T and N_S.

Theorem 1 ([4])**.** *Let T be a left-continuous t-norm with N_T being strong. If (T,N_T,S) form a De Morgan triple, i.e., S is the N_T-dual of T, then S is right-continuous and $N_S=N_T$.*

3 Fuzzy Implications

In this work the following equivalent definition proposed by Fodor and Roubens [5] is used.

Definition 5. A function $I\colon[0,1]^2\to[0,1]$ is called a fuzzy implication if it satisfies the following conditions:

$$I \text{ is decreasing in the first variable,} \tag{I1}$$
$$I \text{ is increasing in the second variable,} \tag{I2}$$
$$I(0,0)=1, \qquad I(1,1)=1, \qquad I(1,0)=0. \tag{I3}$$

Table 3: Examples of some fuzzy implications

Name	**Formula**
$I_{\mathbf{KD}}$: Kleene-Dienes	$\max(1-x,y)$
$I_{\mathbf{LK}}$: Łukasiewicz	$\min(1,1-x+y)$
$I_{\mathbf{FD}}$: Fodor	$\begin{cases}1, & \text{if } x\le y\\ \max(1-x,y), & \text{if } x>y\end{cases}$
$I_{\mathbf{SD}}$	$\begin{cases}y, & \text{if } x=1\\ N(x), & \text{if } y=0\\ 1, & \text{otherwise}\end{cases}$
$I_{\mathbf{TD}}$	$\begin{cases}1, & \text{if } x<1\\ y, & \text{if } x=1\end{cases}$
$I_{\mathbf{TM}}$	$\begin{cases}1, & \text{if } x\le y\\ S(N(x),y), & \text{if } x>y\end{cases}$
$I_{\mathbf{KP}}$	$\min(1,1-x^2+xy)$

Directly from Definition 5 we see that each fuzzy implication I satisfies the following left and right boundary condition, respectively:

$$I(0,y)=1, \qquad y\in[0,1], \tag{LB}$$
$$I(x,1)=1, \qquad x\in[0,1]. \tag{RB}$$

Therefore, I satisfies also the normality condition:

$$I(0,1)=1. \tag{NC}$$

Definition 6. A fuzzy implication I is said to satisfy

- the left neutrality property, if
$$I(1,y)=y, \qquad y\in[0,1]; \tag{NP}$$
- the exchange principle, if for all $x,y,z\in[0,1]$,
$$I(x,I(y,z))=I(y,I(x,z)); \tag{EP}$$
- the identity principle, if
$$I(x,x)=1, \qquad x\in[0,1]; \tag{IP}$$
- the ordering property, if
$$I(x,y)=1 \iff x\le y, \qquad x,y\in[0,1]. \tag{OP}$$

Definition 7. Let I be a fuzzy implication. The function N_I defined as $N_I(x):=I(x,0)$ for all $x\in[0,1]$, is called the natural negation of I.

It can be easily shown that N_I is a fuzzy negation.

Proposition 2 (cf. [5], Corollary 1.1; [2], Lemma 14)**.** *If a function $I\colon[0,1]^2\to[0,1]$ satisfies* (EP) *and* (OP)*, then N_I is either a strong negation or a discontinuous negation.*

4 (S,N)-Implications and R-Implications

In this section, we give a brief introduction to two of the families of fuzzy implications that are very well studied in the literature and present some characterizations and results that will be useful in the sequel.

Definition 8 (see [1, 3, 5, 10])**.** A function $I\colon[0,1]^2\to[0,1]$ is called an (S,N)-implication if there exist a t-conorm S and a fuzzy negation N such that

$$I(x,y)=S(N(x),y), \qquad x,y\in[0,1].$$

If N is a strong negation, then I is called an S-implication. Moreover, if I is generated from S and N, then we will often write $I_{S,N}$.

The following characterization of (S,N)-implications are from [3], which is an extension of a result in [10].

Theorem 2. *For a function $I\colon[0,1]^2\to[0,1]$ the following statements are equivalent:*

(i) I is an (S,N)-implication generated from some t-conorm S and some continuous (strict, strong) fuzzy negation N.

(ii) I satisfies (I2)*,* (EP) *and the function N_I is a continuous (strict, strong) fuzzy negation.*

Definition 9 (see [5, 7])**.** A function $I\colon [0,1]^2 \to [0,1]$ is called an R-implication if there exists a t-norm T such that for all $x, y \in [0,1]$,

$$I(x,y) = \sup\{t \in [0,1] : T(x,t) \leq y\}. \qquad (2)$$

If I is generated from T, then we will often write I_T.

In this work we only consider I_T from left-continuous t-norms, in which case the supremum in (2) reduces to maximum (see [7]).

Theorem 3 ([5], Theorem 1.14)**.** *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I is an R-implication based on some left-continuous t-norm T.

(ii) I satisfies (I2)*,* (EP)*,* (OP) *and $I(x,\cdot)$ is right-continuous for any fixed $x \in [0,1]$.*

It can be immediately noted that $N_T(\cdot) = I_T(\cdot, 0)$, where I_T is obtained from a t-norm T. From the above theorem we see that for a left-continuous t-norm T, the fuzzy negation N_T is either strong or discontinuous. Therefore Proposition 1 can also be seen as the dual of above result.

We also have the following connections between a left-continuous t-norm T and the R-implication I_T.

Lemma 2 (cf. [5, 7])**.** *(i) If T is a left-continuous t-norm, then $T = T_{I_T}$, where, for all $x, y \in [0,1]$,*

$$T_{I_T}(x,y) = \min\{t \in [0,1] \mid I_T(x,t) \geq y\}.$$

(ii) T is a left-continuous t-norm if and only T and I_T form an adjoint (residual) pair, i.e.,

$$T(x,y) \leq z \iff I_T(x,z) \geq y. \qquad (\mathrm{RP})$$

Theorem 4 ([2], Theorem 15)**.** *If a function $I\colon [0,1]^2 \to [0,1]$ satisfies* (EP)*,* (OP) *and N_I is a strong negation, then $T\colon [0,1]^2 \to [0,1]$ defined as*

$$T(x,y) = N_I(I(x, N_I(y))), \qquad x, y \in [0,1],$$

is a t-norm. Additionally, T and I satisfy (RP)*.*

5 QL-Operations and QL-Implications

While (S,N)- and R-implications are the generalizations of a material and intuitionistic-logic implications, in this section we deal with yet another popular way of obtaining fuzzy implications - as the generalization of the following implication defined in quantum logic:

$$p \to q \equiv \neg p \vee (p \wedge q).$$

Needless to state, when the truth values are restricted to $\{0,1\}$ its truth table coincides with the classical implication. In this section we deal with the generalization of the above implication.

Definition 10 (cf. [5, 9])**.** A function $I\colon [0,1]^2 \to [0,1]$ is called a QL-operation if there exist a t-norm T, a t-conorm S and a fuzzy negation N such that

$$I(x,y) = S(N(x), T(x,y)), \qquad x, y \in [0,1].$$

If I is generated from the triple (T,S,N), then we will often write $I_{T,S,N}$ instead of I.

Firstly, we investigate some properties of QL-operations. We will see that not all QL-operation are fuzzy implications in the sense of Definition 5. The proof of the following proposition can be obtained in a straightforward manner.

Proposition 3. *If $I_{T,S,N}$ is a QL-operation, then $I_{T,S,N}$ satisfies* (I2)*,* (I3)*,* (NC)*,* (LB)*,* (NP) *and $N_{I_{T,S,N}} = N$.*

Remark 3. $I_{T,S,N}$ does not always satisfy (I1). For example, consider the function $I_{\mathbf{Z}}(x,y) = \max(1 - x, \min(x,y))$, called the Zadeh implication in the literature (see [5]). Let $x_1 = 0.7 < 0.8 = x_2$ and $y = 0.9$. Then $I_{\mathbf{Z}}(x_1, y) = 0.7 < 0.8 = I_{\mathbf{Z}}(x_2, y)$ and hence $I_{\mathbf{Z}}$ does not satisfy (I1), but it is a QL-operation obtained from the triple $(T_{\mathbf{M}}, S_{\mathbf{M}}, N_{\mathbf{C}})$. On the other hand, the Kleene-Dienes fuzzy implication $I_{\mathbf{KD}}$ is a QL-operation obtained from the triple $(T_{\mathbf{L}}, S_{\mathbf{L}}, N_{\mathbf{C}})$ (see Table 4).

Therefore the first main problem is connected with the characterization of those QL-operations which satisfy (I1). Unfortunately, only partial results are known in the literature. A characterization of QL-operations satisfying (I1) is given in [9] for some continuous cases.

Lemma 3. *If a QL-operation $I_{T,S,N}$ obtained from a triple (T,S,N) is a fuzzy implication, then the pair (S,N) satisfies the law of excluded middle* (LEM)*.*

That the condition in the above Lemma is only necessary but not sufficient can be seen from the $I_{T,S,N}$ obtained from the triple $(T_{\mathbf{P}}, S_{\mathbf{nM}}, N_{\mathbf{C}})$ which is not a fuzzy implication.

The following results are easy to obtain from Lemma 1.

Proposition 4. *If a fuzzy negation N in a triple (T,S,N) is less than N_S, then the pair (S,N) does not satisfy* (LEM) *and hence the QL-operation $I_{T,S,N}$ is not a fuzzy implication.*

Proposition 5. *A QL-operation $I_{T,S,N}$ obtained from a triple (T,S,N), where S is a negative t-conorm, is a fuzzy implication if and only if $N = N_{\mathbf{D2}}$. Moreover, $I_{T,S,N} = I_{\mathbf{TD}}$, in this case (see Tables 3 and 4).*

Following the terminology used by Mas *et al.* [9], only if the QL-operation $I_{T,S,N}$ is a fuzzy implication we use the term QL-implication.

Table 4: Examples of some QL-implications. (NV = Non-Vanishing; P = Positive)

$\mathbf{S}$	$\mathbf{N}$	$\mathbf{T}$	$\mathbf{I_{T,S,N}}$
$S_{\mathbf{D}}$	NV	P	$I_{\mathbf{SD}}$
any	$N_{\mathbf{D2}}$	any	$I_{\mathbf{TD}}$
any	any	$T_{\mathbf{M}}$	$I_{\mathbf{TM}}$
$S_{\mathbf{L}}$	$N_{\mathbf{K}}$	$T_{\mathbf{P}}$	$I_{\mathbf{KP}}$
$S_{\mathbf{L}}$	$N_{\mathbf{C}}$	$T_{\mathbf{L}}$	$I_{\mathbf{KD}}$
$S_{\mathbf{L}}$	$N_{\mathbf{C}}$	$T_{\mathbf{M}}$	$I_{\mathbf{LK}}$
$S_{\mathbf{nM}}$	$N_{\mathbf{C}}$	$T_{\mathbf{M}}$	$I_{\mathbf{FD}}$

6 QL-Operations and other Properties

Keeping with our main aim of this note, in this section, we investigate the conditions under which QL-operations satisfy the properties introduced in Section 3.

6.1 QL-Operations and the Exchange Property

Next result follows from Theorem 2.

Theorem 5 (cf. [9] Proposition 8)**.** *For a QL-operation $I_{T,S,N}$ obtained from a triple (T,S,N), where N a continuous negation, the following statements are equivalent:*

(i) $I_{T,S,N}$ satisfies (EP)*.*

(ii) $I_{T,S,N}$ is an (S,N)-implication.

Remark 4. (i) Theorem 5 also gives a sufficient condition for a QL-operation $I_{T,S,N}$ obtained from a triple (T,S,N) with N a continuous negation to be a fuzzy implication. On the other hand, $I_{\mathbf{SD}}$ and $I_{\mathbf{TD}}$ show that the continuity of N is not necessary for an $I_{T,S,N}$ to satisfy (EP).

(ii) It is interesting to note that the QL-implications $I_{\mathbf{SD}}$ and $I_{\mathbf{TD}}$ are also (S,N)-implications - $I_{\mathbf{TD}}$ is an (S,N)-implication obtained from any t-conorm S and $N = N_{\mathbf{D2}}$, i.e., $I_{\mathbf{TD}} = I_{\mathbf{S},\mathbf{N_{D2}}}$, while $I_{\mathbf{SD}}$ is an (S,N)-implication where $S = S_{\mathbf{D}}$ and N is any non-vanishing negation, i.e., $I_{\mathbf{SD}} = I_{\mathbf{S_D},\mathbf{N}}$.

6.2 QL-Operations and the Identity Principle

We start our investigations with the following result.

Theorem 6 (cf. [10], Theorem 3.2)**.** *For a QL-operation $I_{T,S,N}$ obtained from a triple (T,S,N), where S is a right-continuous t-conorm, the following statements are equivalent:*

(i) $I_{T,S,N}$ satisfies (IP)*.*

(ii) $T(x,x) \geq N_S \circ N(x)$, for all $x \in [0,1]$.

Proof. (i) $\Longrightarrow$ (ii) If $I_{T,S,N}$ satisfies (IP), then for any $x \in [0,1]$ we have $I_{T,S,N}(x,x) = S(N(x),T(x,x)) = 1$. From Remark 1 (iv) we have that $T(x,x) \geq N_S \circ N(x)$ for all $x \in [0,1]$.
(ii) $\Longrightarrow$ (i) By Definition 3 and by right-continuity of S, for any $x \in [0,1]$, $N_S \circ N(x) = \min A_{N(x)}$. From Remark 1 (iv), $T(x,x) \geq N_S \circ N(x)$ for all $x \in [0,1]$ implies that $T(x,x) \in A_{N(x)}$ and hence $S(N(x),T(x,x)) = 1$, i.e., for any $x \in [0,1]$, $I_{T,S,N}(x,x) = S(N(x),T(x,x)) = 1$, so $I_{T,S,N}$ satisfies (IP). □

Example 1. Let us consider the QL-implication $I_{\mathbf{KP}}$ obtained from the triple $(T_{\mathbf{P}}, S_{\mathbf{L}}, N_{\mathbf{K}})$. Since $N_{\mathbf{S_L}}(x) = 1 - x$, note also that, $N_S \circ N_{\mathbf{K}}(x) = 1 - N_{\mathbf{K}}(x) = 1 - (1 - x^2) = x^2$ and hence $T_{\mathbf{P}}(x,x) = N_S \circ N_{\mathbf{K}}(x)$ for all $x \in [0,1]$. It is easy to note that $I_{\mathbf{KP}}$ has (IP).

Observe, that if S is negative, then from Proposition 5 we note that the QL-implications $I_{T,S,N}$ obtained, viz., $I_{\mathbf{TD}}$, satisfy (IP). Also if

- T is any t-norm, S a t-conorm and $N = N_{\mathbf{D2}}$, or
- $T = T_{\mathbf{M}}$, S is a t-conorm and N a fuzzy negation such that they satisfy (LEM), or
- T is a positive t-norm, $S = S_{\mathbf{D}}$ and N a non-vanishing negation,

then a QL-implication $I_{T,S,N}$ obtained from the triple (T,S,N) satisfies (IP).

Remark 5. Trillas and Valverde in [10] require the negation N in the definition of a QL-implication to be strong. Also the t-norm T and t-conorm S are continuous, and are expected to form a De Morgan triple with the negation N. In fact, in Theorem 3.2 of the same work, under these restrictions, condition (ii) of Theorem 6 has been obtained. From their proof, it is clear that the considered T and S are both continuous and Archimedean and hence either they are strict or nilpotent, in which case they show that condition (ii) is not satisfied and hence the claim that "QL-implications never satisfy (IP)". Whereas from the QL-implications $I_{\mathbf{TD}}$ and $I_{\mathbf{KP}}$ we see that $I_{T,S,N}$ does have (IP).

6.3 QL-Operations and the Ordering Property

Proposition 6. *Let S be a t-conorm and the QL-operation $I_{T,S,N}$ be obtained from a triple (T,S,N). If $I_{T,S,N}$ satisfies* (OP)*, then N is strictly decreasing.*

Proof. To see this, if possible, let there exist $x, y \in [0,1]$ such that $x < y$, but $N(x) = N(y)$. By (OP) we have

$$\begin{aligned} I_{T,S,N}(x,y) = 1 &\Longrightarrow S(N(x), T(x,y)) = 1 \\ &\Longrightarrow S(N(y), T(y,x)) = 1 \\ &\Longrightarrow I_{T,S,N}(y,x) = 1 \\ &\Longrightarrow y \leq x, \end{aligned}$$

a contradiction. □

Therefore, from the previous sections, it is clear that if S is a negative t-conorm, then the QL-implication $I_{T,S,N}$ obtained from a triple (T,S,N), i.e., $I_{\mathbf{TD}}$, does not have (OP). Further, if we assume that $I_{T,S,N}$ is a QL-implication, then from Proposition 4 we see that $N \geq N_{\mathbf{S}}$, which implies that a t-conorm S should be such that its natural negation N_S should be non-filling. From Definition 3 this can happen only if every $x \in (0,1)$ has a $y \in (0,1)$ such that $S(x,y) = 1$. Noting that a fuzzy implication that satisfies (OP) also satisfies (IP), using also Theorem 6, we summarize the above discussion in the following result.

Theorem 7. *If a QL-implication $I_{T,S,N}$ obtained from a triple (T,S,N) satisfies* (OP)*, then*

(i) $T(x,x) \geq N_S \circ N(x)$ *for all* $x \in [0,1]$*;*

(ii) N *is a strictly decreasing negation;*

(iii) S *is a non-negative t-conorm such that for every* $x \in (0,1)$ *there exists a* $y \in (0,1)$ *such that* $S(x,y) = 1$.

That the above conditions are not sufficient can be seen from $I_{\mathbf{SD}}$ (see Tables 3 and 4).

Example 2. The QL-implication $I_{\mathbf{KP}}$ obtained from the triple $(T_{\mathbf{P}}, S_{\mathbf{L}}, N)$ satisfies (OP).

In the case when T is the minimum t-norm $T_{\mathbf{M}}$, we have the following stronger result.

Theorem 8. *For a QL-implication $I_{T,S,N}$ obtained from a triple (T,S,N) that satisfies* (OP) *the following statements are equivalent:*

(i) T *is the minimum t-norm* $T_{\mathbf{M}}$.

(ii) $N_S \circ N = \mathrm{id}_{[0,1]}$.

7 QL-Implications with (S,N)- and R-Implications

In this section we discuss the intersection of the family of QL-implications with R- and (S,N)-implications. We give some sufficient conditions under which a QL-implication becomes an (S,N)-implication (in the case when the considered N is strong we show that some stronger results can be obtained). On the other hand, we determine precisely the conditions on the underlying operations T, S, N for a QL-implication to be an R-implication.

Theorem 9 (cf. [4])**.** *For a function $I\colon [0,1]^2 \to [0,1]$ the following statements are equivalent:*

(i) I *is an (S,N)-implication, which satisfies the ordering property* (OP).

(ii) I *is an S-implication obtained from the strong negation N_S.*

Theorem 10 (cf. [6], Section 2.1)**.** *For a fuzzy implication I the following statements are equivalent:*

(i) I *is both an R-implication obtained from a left-continuous t-norm T and also an (S,N)-implication generated from a t-conorm S and a fuzzy negation N.*

(ii) *(a)* $N = N_S = N_T$ *is strong;*
(b) T *is the N_S dual of S;*
(c) $S(x,y) = I(N(x),y)$ *is a right-continuous t-conorm.*

Proof. (i) $\Longrightarrow$ (ii) Let T be a left-continuous t-norm, S a t-conorm and N a fuzzy negation. Without any loss of generality assume that $I = I_T = I_{S,N}$.

(a) Since I is an R-implication, from Theorem 3 it satisfies (OP). Since I is also an (S,N)-implication, by Theorem 9 above, we have that $N = N_S$ is strong. Further, $N(x) = N_{I_{S,N}}(x) = I_{S,N}(x,0) = I_T(x,0) = N_T(x)$, for all $x \in [0,1]$.

(b) Since I satisfies (EP), (OP) and N_I is strong, from Theorem 4 and above facts we have $T(x,y) = N_I(I(x, N_I(y))) = N_S(S(N_S(x), N_S(y)))$.

(c) It is easy to see that $S(x,y) = I(N(x),y)$ for all $x, y \in [0,1]$. Since N_S is strong and S is the N_S-dual of a left-continuous T, S is right-continuous.

(ii) $\Longrightarrow$ (i) Let $S(x,y) = I(N(x),y)$ be a right-continuous t-conorm such that $N = N_S$ is strong. Since T is the N_S dual of the right-continuous S it is left-continuous. Let I_T be the R-implication obtained from T and I_{S,N_S} be the (S,N)-implication obtained from S and N_S. For any $x, y \in [0,1]$ we have $I_{S,N_S}(x,y) = S(N_S(x),y) = I(N_S(N_S(x)),y) = I(x,y)$. Since I_T satisfies (I1), (EP) and its natural negation $N_{I_T} = N_T$ is a strong negation, by Theorem 2 we see that I_T is an (S,N)-implication, i.e.,

$I_T = I_{S',N'}$ for an appropriate t-conorm S' and a strong N'. But $N' = N = N_S$ and hence $I_T = I_{S',N_S}$. Finally, from the proof of (i) $\Longrightarrow$ (ii) above we know that T is the N_S dual of S' and by our assumption T is the N_S dual of S. Hence $S = S'$, i.e., $I = I_T = I_{S,N_S}$. □

7.1 QL-Implications and (S,N)-Implications

We divide our investigation into two parts, based on whether the considered t-conorm S is negative or not. The following result is obvious from Proposition 5.

Theorem 11. *If S is a negative t-conorm, then the obtained QL-implication $I_{T,S,N} = I_{\mathbf{TD}}$ is an (S,N)-implication.*

Proposition 7. *Let $I_{T,S,N}$ be a QL-implication obtained from a triple (T,S,N) where S is a non-negative t-conorm. If $T = T_{\mathbf{M}}$, then $I_{T,S,N}$ is an (S,N)-implication obtained from the same t-conorm S and the same negation N, i.e., $I_{T,S,N} = I_{S,N}$.*

Proof. If $T = T_{\mathbf{M}}$, then $I_{T,S,N}$ is $I_{\mathbf{TM}}$ (see Table 4). Also, since $I_{T,S,N}$ is a fuzzy implication, we have that the pair (S,N) satisfies (LEM) and hence, by Lemma 1(i), $N \geq N_S$. Now, if $x \leq y$, then

$$I_{T,S,N}(x,y) = S(N(x), T_{\mathbf{M}}(x,y)) = S(N(x),x) = 1,$$

and $I_{S,N}(x,y) = S(N(x),y) \geq S(N_S(x),x) = 1$. On the other hand, if $x > y$, then $I_{T,S,N}(x,y) = S(N(x),y) = I_{S,N}(x,y)$. □

Theorem 12. *Let $I_{T,S,N}$ be the QL-implication obtained from a triple (T,S,N) where S is a non-negative t-conorm such that its natural negation N_S is strong. Consider the following statements:*

(i) $I_{T,S,N}$ is an (S,N)-implication obtained from the same S and N, i.e., $I_{T,S,N} = I_{S,N}$.

(ii) $T = T_{\mathbf{M}}$.

(iii) $N = N_S$.

Then the following relationships exist among the above statements:
(i) and (ii) $\Longrightarrow$ (iii).
(ii) and (iii) $\Longrightarrow$ (i).
(iii) and (i) $\Longrightarrow$ (ii).

Proof. (i) and (ii) $\Longrightarrow$ (iii) Since $T = T_{\mathbf{M}}$, $I_{T,S,N}$ is equal to $I_{\mathbf{TM}}$. If $x \in [0,1]$, then $1 = I_{T,S,N}(x,x) = I_{S,N}(x,x) = S(N(x),x)$, i.e., $S(N(x),x) = 1$. Hence from Lemma 1 (ii) we have $x \geq N_S \circ N(x)$ and by the strongness of N_S we have $N_S(x) \leq N(x)$. On the other hand, by Lemma 1 (ii) again, we have $S(N(x),x) = 1 \Longrightarrow N(x) \geq N_S(x)$. From the above inequalities we find that $N(x) = N_S(x)$, for all $x \in [0,1]$.
(ii) and (iii) $\Longrightarrow$ (i) This follows from Proposition 7.
(iii) and (i) $\Longrightarrow$ (ii) Let $I_{T,S,N_S} = I_{S,N_S}$. We know $T(x,x) \leq x$ for all $x \in [0,1]$. Now for any $x \in (0,1)$, we have $N_S(x) \neq 1$ and

$$S(N_S(x), T(x,x)) = S(N_S(x),x) = 1 \Longrightarrow T(x,x) \geq N_S \circ N_S(x) = x.$$

from whence we obtain $T(x,x) = x$ for all $x \in [0,1]$, i.e., $T = T_{\mathbf{M}}$. □

Table 5: Some QL-implications that are also (S,N)-implications. See Remark 6 for more details.

$\mathbf{S}$	$\mathbf{T}$	$\mathbf{N}$	$\mathbf{N_S}$	$\mathbf{I_{T,S,N}}$
Negative	any	$N_{\mathbf{D2}}$	$N_{\mathbf{D2}}$	$I_{\mathbf{TD}}$
$S_{\mathbf{B}}$	$T_{\mathbf{M}}$	$N_{\mathbf{D2}}$	$N_{S_{\mathbf{B}}}$	$I_{\mathbf{TD}}$
$S_{\mathbf{D}}$	$T_{\mathbf{M}}$	$N_{\mathbf{C}}$	$N_{\mathbf{D1}}$	$I_{\mathbf{S_D,N_C}}$
$S_{\mathbf{L}}$	$T_{\mathbf{L}}$	$N_{\mathbf{C}}$	$N_{\mathbf{C}}$	$I_{\mathbf{KD}}$
$S_{\mathbf{L}}$	$T_{\mathbf{P}}$	$N_{\mathbf{C}}$	$N_{\mathbf{C}}$	$I_{\mathbf{RC}}$
$S_{\mathbf{L}}$	$T_{\mathbf{M}}$	$N_{\mathbf{C}}$	$N_{\mathbf{C}}$	$I_{\mathbf{LK}}$
$S_{\mathbf{nM}}$	$T_{\mathbf{M}}$	$N_{\mathbf{C}}$	$N_{\mathbf{C}}$	$I_{\mathbf{FD}}$

Remark 6. Let us consider a t-conorm S whose natural negation N_S is discontinuous. From Proposition 7 we always have that (ii) $\Longrightarrow$ (i). Let us define a lenient version of (i) as follows:

(i') $I_{T,S,N}$ is an (S,N)-implication obtained from (possibly different) t-conorm S' and negation N', i.e., $I_{T,S,N} = I_{S',N'}$.

Obviously, when $S = S'$, $N = N'$ we have (i') = (i). Then from Table 5 the following observations can be made:

- From the first entry we notice that $N = N_S$ is not strong and $I_{T,S,N} = I_{S,N}$, but $T \neq T_{\mathbf{M}}$, i.e., (iii) & (i) $\not\Longrightarrow$ (ii), if N is not strong.
- From the second and third entries it is clear that even if $I_{T,S,N} = I_{S,N}$ and $T = T_{\mathbf{M}}$ we can have $N_S \neq N$, i.e., (i) & (ii) $\not\Longrightarrow$ (iii), when N is not strong.
- From the fourth and fifth entries it is clear that even if $I_{T,S,N} = I_{S',N'}$ and $N = N' = N_C$ we can have $T \neq T_{\mathbf{M}}$, i.e., (i') & (iii) $\not\Longrightarrow$ (ii).

7.2 QL-Implications and R-Implications

Firstly, if S is a negative t-conorm or if $N = N_{\mathbf{D2}}$, then the QL-implication $I_{T,S,N}$ obtained from the triple (T,S,N) is the R-implication $I_{\mathbf{TD}}$ obtained from the non-left-continuous t-norm $T_{\mathbf{D}}$.

From Theorem 3 we see that an R-implication I_T from a left-continuous t-norm T has both the exchange principle (EP) and the ordering property (OP). Now, if an $I_{T,S,N}$ is also an R-implication obtained from a left-continuous t-norm, then from Proposition 2 we know that $N_I = N$ is either strong or discontinuous. But from Theorem 7 (ii) we have N is strictly decreasing and hence $N_I = N$ is either strong or discontinuous, but strictly decreasing.

In the case when N is strong, since $I_{T,S,N}$ has (EP) we know from Theorem 5 that the $I_{T,S,N}$ is also an S-implication. Hence we have the following result:

Proposition 8. *Let N be a strong negation. If a QL-implication $I_{T,S,N}$ obtained from a triple (T,S,N) is an R-implication obtained from a left-continuous t-norm T, then $I_{T,S,N}$ is also an S-implication.*

The reverse implication of Proposition 8 is not valid. To see this, consider the QL-implication $I_{T,S,N}$ obtained from the triple $(T_{\mathbf{M}}, S_{\mathbf{D}}, N)$, where N is strong and hence is non-vanishing. Then $I_{T,S,N}$ is an S-implication (see Remark 4 (ii)) but not an R-implication since it does not have the ordering property (OP). In fact, this is true even if N is a strict negation.

Theorem 13. *If N is a strong negation, then the following statements are equivalent:*

(i) $I_{T,S,N} = I_{T^*}$, for some left-continuous t-norm T^*.

(ii) $T = T_{\mathbf{M}}$, $N = N_{T^*}$ and S is the right-continuous t-conorm that is the N-dual of T^* with $N = N_S$.

Proof. Let N be a strong negation.
(i) $\Longrightarrow$ (ii) If $I_{T,S,N} = I_{T^*}$, then $I_{T,S,N} = I_{S,N}$ from Proposition 8. Further, since $I_{S,N} = I_{T^*}$, Theorem 10 implies that $N = N_{I_{T^*}} = N_{T^*} = N_S$. Also, S is the right-continuous t-conorm that is the N-dual of T^*. Now, $N = N_S$ is strong and by Theorem 12 we see that $T = T_{\mathbf{M}}$.
(ii) $\Longrightarrow$ (i) On the other hand, let $T = T_{\mathbf{M}}$ and S be a right-continuous t-conorm with $N = N_S$. By virtue of Proposition 7 we get $I_{T,S,N} = I_{S,N}$. Let T^* be the N-dual t-norm of the right-continuous t-conorm S with $N = N_S$, in which case T^* is left-continuous. Now, from Theorem 10 we have that $I_{S,N} = I_{T^*}$ and hence $I_{T,S,N} = I_{T^*}$. □

The QL-implication $I_{\mathbf{KP}}$ obtained from the triple $(T_{\mathbf{P}}, S_{\mathbf{L}}, N_{\mathbf{K}})$ as given in Example 1 is a fuzzy implication that is neither an (S,N)-implication nor an R-implication obtained from a left-continuous t-norm, since it does not have (EP).

8 Conclusion

In this note the intersections that exist between the families of (S,N)-, R- and QL-implications are determined using existing characterization results. As part of this attempt some interesting results pertaining to natural negations from t-conorms, properties of QL-implications and the exact intersection between (S,N)- and R-implications have been obtained (Theorem 10). From this work the following problems arise.

Problem 1. (i) Characterize a triple (T,S,N) such that $I_{T,S,N}$ satisfies (I1). It should be noted that a characterization is already known for some continuous cases (see [9]).

(ii) Prove or give a counter example: Any QL-operation that satisfies (EP) is an (S,N)-implication.

(iii) Give an equivalent condition for an $I_{T,S,N}$ to satisfy the ordering property (OP).

References

[1] C. Alsina, E. Trillas, When (S,N)-implications are (T,T_1)-conditional functions? *Fuzzy Sets and Systems* **134** (2003) 305–310.

[2] M. Baczyński, Residual implications revisited. Notes on the Smets-Magrez Theorem, *Fuzzy Sets and Systems* **145** (2004) 267–277.

[3] M. Baczyński, B. Jayaram, On the characterizations of (S,N)-implications, *Fuzzy Sets and Systems*, (2007), doi:10.1016/j.fss.2007.02.010.

[4] M. Baczyński, B. Jayaram, (S,N)- and R-implications. Position paper: properties, characterizations, representations and intersections, *Fuzzy Sets and Systems*, *submitted*.

[5] J.C. Fodor, M. Roubens, *Fuzzy preference modeling and multicriteria decision support*, Kluwer, Dordrecht, 1994.

[6] J.C. Fodor, Contrapositive symmetry of fuzzy implications, *Fuzzy Sets and Systems* **69** (1995) 141-156.

[7] S. Gottwald, *A treatise on many-valued logics*, Research Studies Press, Baldock 2001.

[8] E.P. Klement, R. Mesiar, E. Pap, *Triangular norms*, Kluwer, Dordrecht, 2000.

[9] M. Mas, M. Monserrat, J. Torrens, QL-implications versus D-implications, *Kybernetika* **42** (2006) 351–366.

[10] E. Trillas, L. Valverde, On some functionally expressable implications for fuzzy set theory. in: E.P. Klement (Ed.) *Proc. of the 3rd Inter. Seminar on Fuzzy Set Theory, Linz*, Austria, 1981, pp. 173–190.

Logical Aggregation based on interpolative realization of Boolean algebra

Dragan G. Radojević
Mihajlo Pupin Institute, Volgina 15, 11000 Belgrade, Serbia
e-mail: dragan.radojevic@imp-automatika.bg.ac.yu

Abstract

In this paper, aggregation is treated as a logical and/or pseudo-logical operation what is important from many points of view such as adequacy and interpretations.

Keywords: Aggregation, Logical aggregation, Interpolative realization of Boolean algebra, Generalized Boolean polynomial, Pseudo Boolean polynomial, Generalized Choquet integral, OWA

1 Introduction

A very important problem in many fields of applications is the aggregation (fusion) of many partial aspects (attributes) into one global representative aspect. In the existing practice the weighted sum of partial aspects is used most often as an aggregation tool. This approach is additive and for all effects of interest which are not additive in their nature it is inadequate. For example: using a weighted sum as an aggregation tool even in the case of only two attributes (a, b), one can't to realize a simple and natural demand such as *a and b* is important. In multi-attribute decision making community this problem was recognized [2, 9] and as a solution they use theory of capacity [1] known in fuzzy community as fuzzy measure and fuzzy integrals [9]. In this approach additivity is relaxed by monotonicity, for which additivity is only a special case. As a consequence, the possible domain of application of this approach is much wider.

But from a logical point of view monotonicity is a superfluously strong constraint since many of logical functions are non monotone in their nature. A generalized discrete Choquet integral [8] is defined for a general measure – non monotone in a general case. This approach includes all logical and/or pseudo-logical functions but for only one arithmetic operator for interpolation intention, *min* function. *Interpolative realization of Boolean* algebra (IBA) [5] includes all logical functions and all interpolative operators – generalized product operators.

Logical aggregation as an adequate tool for aggregation in a general case is based on IBA. IBA is technically based on *generalized Boolean polynomials* (GBP) [5, 6].

GBP is described in Section 2. A representative example of logical aggregation is given in Section 4.

2 Generalized Boolean Polynomial

Primary attributes (properties) define a finite set $\mathbf{\Omega}=\{a_1,...,a_n\}$. No one of primary attributes can be calculated as a Boolean function of the remaining primary attributes from $\mathbf{\Omega}$. Set $BA(\mathbf{\Omega})$ of all the possible attributes generated by the set of primary attributes $\mathbf{\Omega}$ by application of Boolean operators is a partially ordered set – Boolean algebra of attributes:

$$BA(\mathbf{\Omega})=\mathbf{P}\left(\mathbf{P}(\mathbf{\Omega})\right).$$

A partial order is based on the relation of inclusion and it is value irrelevant. The following structure with two binary and one unary operators is Boolean algebra:

$$\left\langle BA(\mathbf{\Omega}),\cup,\cap,C\right\rangle.$$

Any element of Boolean algebra $\varphi \in BA(\boldsymbol{\Omega})$ is a corresponding attribute and it can be represented by the disjunctive normal form:

$$\varphi = \bigcup_{S \in \mathbf{P}(\boldsymbol{\Omega})} \left(\sigma_{\varphi}(S) \cap \alpha(S) \right), \qquad (1)$$

Atomic attributes $\alpha(S)$, $\left(S \in \mathbf{P}(\boldsymbol{\Omega})\right)$ are the simplest elements of $BA(\boldsymbol{\Omega})$ in the sense that they do not include in themselves anything except for a trivial Boolean constant **0**. The atomic attributes are described by the following expressions:

$$\alpha(S) = \bigcap_{a_i \in S} a_i \bigcap_{a_j \in \boldsymbol{\Omega} \setminus S} Ca_j, \quad S \in \mathbf{P}(\boldsymbol{\Omega}). \qquad (1.1)$$

Structural function $\sigma_{\varphi} : \mathbf{P}(\boldsymbol{\Omega}) \to \{0,1\}$ of analyzed attribute (element of Boolean algebra) $\varphi \in BA(\boldsymbol{\Omega})$:

$$\sigma_{\varphi}(S) = \begin{cases} \overline{1}, & \alpha(S) \subseteq \varphi; \\ \overline{0}, & \alpha(S) \not\subset \varphi; \end{cases} ; \qquad (1.2)$$
$$\left(S \in \mathrm{P}(\Omega); \quad \overline{0}, \overline{1}, \varphi \in BA(\Omega)\right)$$

determines which atomic elements (attributes) are included in it $(\sigma_{\varphi}(S) = \overline{1} \Leftrightarrow \alpha(S) \subseteq \varphi)$ and/or which are not included $(\sigma_{\varphi}(S) = \overline{0} \Leftrightarrow \alpha(S) \not\subset \varphi)$, where:

$$\alpha(S) \subseteq \varphi \Leftrightarrow \left(\alpha(S) \cap \varphi = \alpha(S)\right)$$
$$\alpha(S) \not\subset \varphi \Leftrightarrow \left(\alpha(S) \cap \varphi = \overline{0}\right)$$
$$\left(S \in \mathrm{P}(\Omega); \quad \overline{0}, \overline{1}, \varphi \in BA(\Omega)\right)$$

Structural function of primary attribute $a_i \in \boldsymbol{\Omega}$ is given by the following expression

$$\sigma_{a_i}(S) = \begin{cases} \overline{1}, & a_i \in S \\ \overline{0}, & a_i \notin S \end{cases}; \quad \left(S \in \mathbf{P}(\boldsymbol{\Omega})\right)$$

Determination of structure of any attribute is based on the expression above and on the following rules:

$$\sigma_{\varphi \cap \psi}(S) = \sigma_{\varphi}(S) \cap \sigma_{\psi}(S),$$
$$\sigma_{\varphi \cup \psi}(S) = \sigma_{\varphi}(S) \cup \sigma_{\psi}(S),$$
$$\sigma_{C\varphi}(S) = C\sigma_{\varphi}(S).$$

where: $S \in \boldsymbol{\Omega}$, $\varphi, \psi \in BA(\boldsymbol{\Omega})$.

Equation (1) can be described in the following form:

$$\varphi = \bigcup_{S \in \mathbf{P}(\boldsymbol{\Omega})} \left(\sigma_{\varphi}(S) \cap \left(\bigcap_{a_i \in S} a_i \bigcap_{a_j \in \boldsymbol{\Omega} \setminus S} Ca_j \right) \right).$$

Any attribute has its value realization on a valued level. A valued level is defined as a set of analyzed elements, objects, actions etc.

Any element of Boolean algebra of attributes can be represented by a *generalized Boolean polynomial*:

$$\varphi^{\otimes}\left(a_1^v, ..., a_n^v\right) = \sum_{S \in \mathbf{P}(\boldsymbol{\Omega})} \sigma_{\varphi}^v(S) \otimes \alpha^{\otimes}(S)\left(a_1^v, ..., a_n^v\right)$$
$$\varphi^{\otimes}\left(a_1^v, ..., a_n^v\right) = \sum_{S \in \mathbf{P}(\boldsymbol{\Omega})} \sigma_{\varphi}^v(S) \alpha^{\otimes}(S)\left(a_1^v, ..., a_n^v\right)$$
$$\left(a_i^v \in [0,1], \quad a_i \in \boldsymbol{\Omega}\right) \qquad (2)$$

Where: $\sigma_{\varphi}^v(S)$, $S \in \mathbf{P}(\boldsymbol{\Omega})$ is value realization of structural function $\sigma_{\varphi}(S)$, $S \in \mathbf{P}(\boldsymbol{\Omega})$, which is given by the following expression:

$$\sigma_{\varphi}^v(S) = \begin{cases} 1, & \sigma_{\varphi}(S) = \overline{1} \\ 0, & \sigma_{\varphi}(S) = \overline{0} \end{cases};$$
$$\left(S \in \mathbf{P}(\boldsymbol{\Omega}); \ \overline{0}, \overline{1}, \varphi \in BA(\boldsymbol{\Omega}); \ 0,1 \in N_0\right)$$

A generalized Boolean polynomial $\varphi^{\otimes}\left(a_1^v, ..., a_n^v\right)$ enables calculation of the value of a corresponding attribute $\varphi \in BA(\boldsymbol{\Omega})$ (element of Boolean algebra) for an analyzed object.

$\alpha^{\otimes}(S)\left(a_1^v, ..., a_n^v\right)$, $S \in \mathbf{P}(\boldsymbol{\Omega})$, are Boolean polynomial of atomic elements defined by the following expression:

$$\alpha^{\otimes}(S)\left(a_1^v, ..., a_n^v\right) = \sum_{K \in \mathbf{P}(\boldsymbol{\Omega} \setminus S)} (-1)^{|K|} \bigotimes_{a_i \in K \cup S} a_i^v \qquad (3)$$
$$\left(S \in \mathbf{P}(\boldsymbol{\Omega}), \quad a_i \in \boldsymbol{\Omega}, \quad a_i^v \in [0,1], i = 1,..,n\right).$$

Example: *Atomic Boolean polynomials for the case when set of primary attributes is* $\Omega = \{a, b\}$, *are given in the following table:*

***Table 1**: Example of Atomic Boolean polynomials*

S	$\alpha(S)$	$\alpha^{\otimes}(S)(a^{v},b^{v})$
$\varnothing$	$Ca\cap Cb$	$1-a^{v}-b^{v}+a^{v}\otimes b^{v}$
$\{a\}$	$a\cap Cb$	$a^{v}-a^{v}\otimes b^{v}$
$\{b\}$	$Ca\cap b$	$b^{v}-a^{v}\otimes b^{v}$
$\{a,b\}$	$a\cap b$	$a^{v}\otimes b^{v}$

In atomic Boolean polynomials the following operators +, - and $\otimes$ figure.

Operator $\otimes$ is *a generalized product*, defined in the same way as *T*-norms [4] with one additional axiom – *no-negativity* [5].

$$\otimes:[0,1]\times[0,1]\to[0,1]$$

1. $a_i^{v}\otimes a_j^{v}=a_j^{v}\otimes a_i^{v}$
2. $a_i^{v}\otimes\left(a_j^{v}\otimes a_k^{v}\right)=\left(a_i^{v}\otimes a_j^{v}\right)\otimes a_k^{v}$
3. $a_i^{v}\le a_j^{v}\ \Rightarrow\ a_i^{v}\otimes a_k^{v}\le a_j^{v}\otimes a_k^{v}$
4. $a_i^{v}\otimes 1=a_i^{v}$
5. $\boxed{\sum_{K\in\mathbf{P}(\mathbf{\Omega}\setminus S)}(-1)^{|K|}\bigotimes_{a_i\in S\cup K}a_i^{v}\ge 0,}\quad \forall S\in\mathbf{P}(\mathbf{\Omega})$

$\left(\mathbf{\Omega}=\{a_1,...,a_n\},\quad a_i^{v}\in[0,1],i=1,..,n\right)$

Additional axiom "non-negativity" ensures that the values of atomic Boolean polynomials are non-negative: $\alpha^{\otimes}(S)\ge 0,\quad (S\in\mathbf{P}(\mathbf{\Omega}))$. As a consequence all elements of Boolean algebra are non-negative: $\varphi^{\otimes}\left(a_1^{v},\cdots a_1^{v}\right)\ge 0,\quad \varphi\in BA(\mathbf{\Omega})$.

***Example*:** *In the case* $\Omega=\{a,b\}$ *generalized product, according to axioms of non-negativity can be in the following interval:*

$$\max(a+b-1,0)\le a\otimes b\le\min(a,b)\ .$$

In spite of formal similarity between a *T*-norm and a generalized product, their roles are crucially different: while a *T*-norm in conventional fuzzy approaches has the role of a logical operator (which is impossible in a general case) a generalized product $\otimes$ is only an arithmetic operator on a value level.

A generalized Boolean polynomial given by the expression (2) can be represented in the following way:

$$\begin{aligned}\varphi^{\otimes}\left(a_1^{v},...,a_n^{v}\right)&=\\&=\sum_{S\in\mathbf{P}(\mathbf{\Omega})}\sigma_{\varphi}^{v}(S)\otimes\sum_{K\in\mathbf{P}(\mathbf{\Omega}\setminus S)}(-1)^{|K|}\bigotimes_{a_i\in K\cup S}a_i^{v}\\&=\sum_{S\in\mathbf{P}(\mathbf{\Omega})}\sigma_{\varphi}^{v}(S)\sum_{K\in\mathbf{P}(\mathbf{\Omega}\setminus S)}(-1)^{|K|}\bigotimes_{a_i\in K\cup S}a_i^{v},\end{aligned}$$

$\left(\varphi\in BA(\mathbf{\Omega}),\quad a_i^{v}\in[0,1],\quad a_i\in\mathbf{\Omega}\right).$

A generalized Boolean polynomial can be represented as a scalar product of the following two vectors: (a) *structural vector* of analyzed Boolean algebra element – attribute

$$\vec{\sigma}_{\varphi}=\left[\sigma_{\varphi}^{v}(S)\middle|S\in\mathbf{P}(\mathbf{\Omega})\right] \tag{4}$$

where: $\mathbf{\Omega}=\{a_1,...,a_n\},\quad \varphi\in BA(\mathbf{\Omega})$,

and (b) *vector of atomic Boolean polynomials*

$$\begin{aligned}&\vec{\alpha}^{\otimes}\left(a_1^{v},...,a_n^{v}\right)=\left[\alpha^{\otimes}(S)\left(a_1^{v},...,a_n^{v}\right)\middle|S\in\mathbf{P}(\mathbf{\Omega})\right]^{T}\\&\left(a_i\in\mathbf{\Omega},\quad a_i^{v}\in[0,1],i=1,..,n\right).\end{aligned} \tag{5}$$

So, a generalized Boolean polynomial is a scalar product of the above defined two vectors:

$$\varphi^{\otimes}\left(a_1^{v},...,a_n^{v}\right)=\vec{\sigma}_{\varphi}\vec{\alpha}^{\otimes}\left(a_1^{v},...,a_n^{v}\right) \tag{6}$$

where: $\varphi\in BA(\mathbf{\Omega}),\quad a_i^{v}\in[0,1],\quad a_i\in\mathbf{\Omega}$.

For structural vectors all Boolean axioms are valid: Associativity, Commutativity, Absorption, Distributivity, Excluded middle and Contradiction

$$\vec{\sigma}_{\varphi\cup(\psi\cup\phi)}=\vec{\sigma}_{(\varphi\cup\psi)\cup\phi}\qquad \vec{\sigma}_{\varphi\cap(\psi\cap\phi)}=\vec{\sigma}_{(\varphi\cap\psi)\cap\phi}$$

$$\vec{\sigma}_{\varphi\cup\psi}=\vec{\sigma}_{\psi\cup\varphi}\qquad \vec{\sigma}_{\varphi\cap\psi}=\vec{\sigma}_{\psi\cap\varphi}$$

$$\vec{\sigma}_{\varphi\cup(\varphi\cap\psi)}=\vec{\sigma}_{\varphi}\qquad \vec{\sigma}_{\varphi\cap(\varphi\cup\psi)}=\vec{\sigma}_{\varphi}$$

$$\vec{\sigma}_{\varphi\cup(\psi\cap\phi)}=\vec{\sigma}_{(\varphi\cup\psi)\cap(\varphi\cup\phi)}\qquad \vec{\sigma}_{\varphi\cap(\psi\cup\phi)}=\vec{\sigma}_{(\varphi\cap\psi)\cup(\varphi\cap\phi)}$$

$$\vec{\sigma}_{\varphi\cup C\varphi}=\vec{1}\qquad \vec{\sigma}_{\varphi\cap C\varphi}=\vec{0}$$

respectively; and all Boolean theorems: Idempotency, Boundedness, 0 and 1 are complements, De Morgan's laws and Involution:

$$\vec{\sigma}_{\varphi\cup\varphi}=\vec{\sigma}_{\varphi} \qquad \vec{\sigma}_{\varphi\cap\varphi}=\vec{\sigma}_{\varphi}$$
$$\vec{\sigma}_{\varphi\cup 0}=\vec{\sigma}_{\varphi} \qquad \vec{\sigma}_{\varphi\cap 1}=\vec{\sigma}_{\varphi}$$
$$\vec{\sigma}_{\varphi\cup 1}=\vec{1} \qquad \vec{\sigma}_{\varphi\cap 0}=\vec{0}$$
$$\vec{\sigma}_{C0}=\vec{1} \qquad \vec{\sigma}_{C1}=\vec{0}$$
$$\vec{\sigma}_{C(\varphi\cup\psi)}=\vec{\sigma}_{C\varphi\cap C\psi} \qquad \vec{\sigma}_{C(\varphi\cap\psi)}=\vec{\sigma}_{C\varphi\cup C\psi}$$
$$\vec{\sigma}_{CC\varphi}=\vec{\sigma}_{\varphi}$$

respectively; where: $\varphi,\psi,\phi\in BA(\boldsymbol{\Omega})$.

So, the structure of a Boolean algebra element preserves Boolean properties in a generalized case described by Boolean polynomials.

As a consequence for any two elements of Boolean algebra $\varphi,\psi\in BA(\boldsymbol{\Omega})$ the following equations are valid:

$$(\varphi\cap\psi)^{\otimes}\left(a_1^v,...,a_n^v\right)=\vec{\sigma}_{\varphi\cap\psi}\vec{\alpha}^{\otimes}(x)$$
$$=\left(\vec{\sigma}_{\varphi}\wedge\vec{\sigma}_{\psi}\right)\vec{\alpha}^{\otimes}\left(a_1^v,...,a_n^v\right)$$
$$(\varphi\cup\psi)^{\otimes}\left(a_1^v,...,a_n^v\right)=\vec{\sigma}_{\varphi\cup\psi}\vec{\alpha}^{\otimes}\left(a_1^v,...,a_n^v\right)$$
$$=\left(\vec{\sigma}_{\varphi}\vee\vec{\sigma}_{\psi}\right)\vec{\alpha}^{\otimes}\left(a_1^v,...,a_n^v\right)$$
$$(C\varphi)^{\otimes}\left(a_1^v,...,a_n^v\right)=\vec{\sigma}_{C\varphi}\vec{\alpha}^{\otimes}\left(a_1^v,...,a_n^v\right)$$
$$=\left(\vec{1}-\vec{\sigma}_{\varphi}\right)\vec{\alpha}^{\otimes}\left(a_1^v,...,a_n^v\right)$$
$$=1-(\varphi)^{\otimes}\left(a_1^v,...,a_n^v\right)$$

Actually, Boolean polynomial maps a corresponding element of Boolean algebra into its value from the real unit interval [0, 1] on the value level so that a partial order on the value level is preserved. Since a partial order is based on Boolean laws, they are preserved on the value level in a general case too, contrary to other approaches.

3. Generalized Pseudo-Boolean Polynomial

To every element of IBA corresponds a generalized Boolean polynomial with the ability to process all values of primary variables from a real unit interval [0, 1]. A pseudo-Interpolative Boolean polynomial is a linear convex combination of analyzed elements of IBA – generalized Boolean polynomials:

$$\pi\varphi^{\otimes}\left(a_1^v,...,a_n^v\right)=\sum_{i=1}^{m}w_i\varphi_i^{\otimes}\left(a_1^v,...,a_n^v\right),$$
$$\sum_{i=1}^{m}w_i=1,\quad \left(w_i\geq 0,\ i=1,...,m\right), \qquad (7)$$
$$\left(a_i\in\boldsymbol{\Omega},\quad a_i^v\in[0,1],i=1,..,n\right).$$

From the definition of generalized Boolean polynomials, an *interpolative pseudo-Boolean polynomial* is given by the following expression:

$$\pi\varphi_{\mu}^{\otimes}\left(a_1^v,...,a_n^v\right)=$$
$$=\sum_{i=1}^{m}w_i\sum_{S\in\mathbf{P}(\boldsymbol{\Omega})}\sigma_{\varphi_i}^v(S)\sum_{C\in\mathbf{P}(\boldsymbol{\Omega}\setminus S)}(-1)^{|C|}\bigotimes_{a_i\in S\cup C}a_i^v$$
$$=\sum_{S\in\mathbf{P}(\boldsymbol{\Omega})}\mu(S)\sum_{C\in\mathbf{P}(\boldsymbol{\Omega}\setminus S)}(-1)^{|C|}\bigotimes_{a_i\in S\cup C}a_i^v. \quad (7.1)$$

Structure function μ of interpolative pseudo-Boolean polynomial $\pi\varphi_{\mu}^{\otimes}$ is a set function

$$\mu:\mathbf{P}(\boldsymbol{\Omega})\to[0,1],\quad \boldsymbol{\Omega}=\{a_1,...,a_n\}$$

defined by the following expression, [7]:

$$\mu(S)=\sum_{i=1}^{m}w_i\sigma_{\varphi_i}^v(S),$$
$$\left(S\in\mathbf{P}(\boldsymbol{\Omega}),\quad \varphi_i\in BA(\boldsymbol{\Omega})\right) \quad . \qquad (8)$$
$$\sum_{i=1}^{m}w_i=0,\quad w_i\geq 0,\quad i=1,...,m$$

Where: $\sigma_{\varphi_i}^v$, $i=1,...,m$ are structure functions of the corresponding Boolean functions $\varphi_i\in BA(\boldsymbol{\Omega})$, $i=1,...,m$.

The characteristics of pseudo-Boolean polynomial depend on the generalized product, and its structure function. Structure functions can be classified into: (a) additive, (b) monotone and (c) generalized ($(a)\subset(b)\subset(c)$).

4. Logical Aggregation

A starting point is a finite set of primary attributes $\boldsymbol{\Omega}=\{a_1,...,a_n\}$. The task of logical aggregation (LA) [7] is the fusion of primary quality attribute values into one resulting globally representative value using logical tools. In a general case LA has two steps: (1) Normalization of primary attributes' values:

$$\cdot^{v}:\Omega\to[0,\,1].$$

The result of normalization is a generalized logical and/or [0, 1] value of analyzed primary attribute, and

(2) Aggregation of normalized values of primary attributes into one resulting value by pseudo-logical function as a logical aggregation operator:

$$Aggr:[0,1]^{n}\to[0,1].$$

A Boolean logical function φ is transformed into a corresponding generalized Boolean polynomial (GBP), [5], $\varphi^{\otimes}:[0,1]^{n}\to[0,1]$. Actually, to any element of Boolean algebra of attributes $\varphi_i\in BA(\mathbf{\Omega})$ there corresponds uniquely GBP $\varphi_i^{\otimes}\left(a_1^{v},...,a_n^{v}\right)$. GBP is defined by expression (2) and/or (2.1).

Pseudo-logical function is a linear convex combination of generalized Boolean polynomials [5], defined by expression (7) and/or (7.1).

Operator of logical aggregation in a general case is a pseudo-logical function:

$$Agg_{\mu}^{\otimes}\left(a_1^{v},...,a_n^{v}\right)=\pi\varphi_{\mu}^{\otimes}\left(a_1^{v},...,a_n^{v}\right) \tag{9}$$

or

$$Agg_{\mu}^{\otimes}\left(a_1^{v},...,a_n^{v}\right)=\sum_{S\in\mathbf{P}(\mathbf{\Omega})}\mu(S)\sum_{C\in\mathbf{P}(\mathbf{\Omega}\setminus S)}(-1)^{|C|}\bigotimes_{a_i\in S\cup C}a_i^{v}$$

Aggregation measure is a structural function of pseudo-logical function – logical aggregation operator (9). So, *Aggregation measure* is a set function $\mu:\mathbf{P}(\mathbf{\Omega})\to[0,1]$, which in a general case is not a monotone function (generalized capacity), defined by the following expression:

$$\begin{aligned}&\mu(S)=\sum_{i=1}^{m}w_i\sigma_{\varphi_i}^{v}(S),\\&\left(S\in\mathbf{P}(\mathbf{\Omega}),\quad\varphi_i\in BA(\mathbf{\Omega})\right)\\&\sum_{i=1}^{m}w_i=0,\quad w_i\geq 0,\quad i=1,...,m\end{aligned} \tag{10}$$

As a consequence, a logical aggregation operator depends on the chosen: (a) measure of aggregation and (b) operator of generalized product. By a corresponding choice of the measure of aggregation μ and generalized product $\otimes$ the known aggregation operators can be obtained as special cases:

Weighted sum

For the aggregation measure and generalized product:

$$\mu_{add}(S)=\sum_{i=1}^{n}w_i\sigma_{a_i}^{v}(S),\quad S\in\mathbf{P}(\mathbf{\Omega});\quad\otimes:=min.$$

Logical aggregation operator is a weighted sum:

$$Agg_{\mu_{add}}^{min}\left(a_1^{v},...,a_n^{v}\right)=\sum_{a_i\in\mathbf{\Omega}}w_ia_i^{v}$$

Arithmetic mean

For the aggregation measure and generalized product:

$$w_i=\frac{1}{n},\ \mu_{mean}(S)=\frac{|S|}{|\mathbf{\Omega}|};\quad\otimes:=min$$

Logical aggregation operator is the arithmetic mean:

$$Agg_{\mu_{mean}}^{min}\left(a_1^{v},...,a_n^{v}\right)=\frac{1}{n}\sum_{a_i\in\mathbf{\Omega}}a_i^{v}$$

K-th attribute only

For the aggregation measure and generalized product:

$$w_i=\begin{cases}1 & i=k\\0 & i\neq k\end{cases};\quad\mu_k(S)=\begin{cases}1 & a_k\in S\\0 & a_k\notin S\end{cases};\quad\otimes:=min$$

Logical aggregation operator is the k-th attribute only:

$$Agg_{\mu_k}^{\otimes}\left(a_1^{v},...,a_n^{v}\right)=a_k^{v}$$

Discrete Choquet integral

For any monotone aggregation measure μ_{mon} and generalized product:

$$\mu_{mon},\qquad\otimes:=min$$

Logical aggregation operator is a discrete Choquet integral:

$$Agg_{\mu_{mon}}^{\otimes}\left(a_1^{v},...,a_n^{v}\right)=C_{\mu_{mon}}\left(a_1^{v},...,a_n^{v}\right).$$

A discrete Choquet integral is defined by the following expression:

$$C_{\mu_{mon}}\left(a_1^v,...,a_n^v\right)=\sum_{k=1}^{n}\left(a_{(k)}^v-a_{(k-1)}^v\right)\mu_{mon}\left(A_{(k)}\right),$$

where:

$$a_{(1)}^v\le...\le a_{(n)}^v;\quad A_{(k)}=\left\{a_{(k)}^v,...,a_{(n)}^v\right\}.$$

Minimal value of attributes

For the aggregation measure and generalized product:

$$\mu_{AND}(S)=\begin{cases}1, & S=\mathbf{\Omega}\\ 0, & S\neq\mathbf{\Omega}\end{cases};\quad \otimes:=min.$$

Logical aggregation operator is the *min* function

$$Agg_{\mu_{AND}}^{min}\left(a_1^v,...,a_n^v\right)=min\left\{a_1^v,...,a_n^v\right\}.$$

Maximal value of attributes

For the aggregation measure and generalized product:

$$\mu_{OR}(S)=\begin{cases}1, & S\neq\varnothing\\ 0, & S=\varnothing\end{cases};\quad \otimes;=min$$

Logical aggregation operator is the *max* function

$$Agg_{\mu_{OR}}^{min}\left(a_1^v,...,a_n^v\right)=max\left\{a_1^v,...,a_n^v\right\}.$$

OWA-ordered weight aggregation

For the aggregation measure and generalized product:

$$\mu_{OWA}(S)=\begin{cases}0, & S=\varnothing\\ \sum_{i=1}^{m}w_i, & |S|=m\end{cases};\quad \otimes:=min$$

Logical aggregation operator is an OWA aggregation operator

$$Agg_{\mu_{OWA}}^{min}\left(a_1^v,...,a_n^v\right)=OWA\left(a_1^v,...,a_n^v\right).$$

OWA is defined by the following expression:

$$OWA\left(a_1^v,...,a_n^v\right)=\sum_{i=1}^{n}w_i a_{(i)}^v$$

$$a_{(1)}^v\le a_{(2)}^v\le...\le a_{(n)}^v,\quad \sum_{i=1}^{n}w_i=1,\quad w_i\ge 0$$

.

k-th order statistics

For the aggregation measure and generalized product:

$$\mu_{k^{th}}(S)=\begin{cases}0, & |S|<k\\ 1, & |S|\ge k\end{cases};\quad \otimes:=min$$

Logical aggregation operator is the k-th order statistics

$$Agg_{\mu_{k^{th}}}^{min}\left(a_1^v,...,a_n^v\right)=a_{(k)}^v,$$

where:

$$a_{(1)}^v\le a_{(2)}^v\le...\le a_{(n)}^v.$$

5. Example of Logical Aggregation Application

A modified example from [3] is analyzed here.

Example: Objects A, B, C and D are described by quality attributes, whose values are from real unit interval [0, 1], given in the following table:

Table 2: Values of quality attributes

Object	*a*	*b*	*c*
A	.75	.9	.3
B	.75	.8	.4
C	.3	.65	.1
D	.3	.55	.2

An object should be compared on the base of a global quality. A global quality is actually aggregation of attributes so the following aspects should be incorporated: (a) the average value of quality attributes and (b) if the analyzed object is good by attribute *a* then attribute *c* is more important than *b* and if analyzed object is not good by attribute *a* then attribute *b* is more important than *c*.

A partial demand (a) is given by the following trivial expression:

$$\frac{a+b+c}{3}$$

A partial demand (b) is given by the following logical expression:

$$\varphi(a,b,c)=(a\cap c)\cup(Ca\cap b) \qquad (11)$$

A generalized Boolean polynomial of logical expression (11) is:

$$\begin{aligned}\varphi^{\otimes}(a,b,c)&=\left((a\cap c)\cup(Ca\cap b)\right)^{\otimes}\\&=b+a\otimes c-a\otimes b\end{aligned}$$

A possible logical aggregation operator is:

$$\begin{aligned}Aggr^{\otimes}(a,b,c)&=\frac{1}{2}\frac{a+b+c}{3}+\frac{1}{2}\varphi^{\otimes}(a,b,c)\\&=\frac{1}{2}\frac{a+b+c}{3}+\frac{1}{2}(b+a\otimes c-a\otimes b)\end{aligned}$$

A corresponding measure of aggregation is:

$$\mu=\frac{1}{6}(\sigma_a+\sigma_b+\sigma_c)+\frac{1}{2}(\sigma_a\wedge\sigma_c)\vee(C\sigma_a\wedge\sigma_b).$$

or given as a table:

Table 3: Measure of aggregation

S	$\mu(S)$
$\varnothing$	0
$\{a\}$	1/6
$\{b\}$	2/3
$\{c\}$	1/6
$\{a,b\}$	5/6
$\{a,c\}$	5/6
$\{b,c\}$	1/3
$\{a,b,c\}$	1

It is clear that the measure is non-monotone since $\mu(\{b\})\geq\mu(\{b,c\})$, and as a consequence it is not possible to use a standard Choquet integral.

In the case $\otimes := min$ function $\varphi^{min}(a,b,c)$ is actually a generalized discrete Choquet integral and its values are given in the following table:

Table 4: Values for $\otimes := min$

Object	$\varphi^{min}(a,b,c)$
A	.45
B	.45
C	.45
D	.45

So, these results without discrimination are not adequate.

In the case when a generalized product is an ordinary product, $\otimes := *$, quitting conventional approaches, the corresponding values of function $\varphi^{*}(a,b,c)$ are given in the following table:

Table 5: Values for $\otimes := *$

Object	$\varphi^{*}(a,b,c)$
A	.450
B	.500
C	.485
D	.445

The values of aggregation function, for given aggregation measure, table 1, and for $\otimes := *$, are presented in the following table:

Table 6: Values of resulting aggregation function

Object	$Aggr^{*}(a,b,c)$
A	.5500
B	.5750
C	.4175
D	.3725

These results completely reflect all specified demands.

Comment: *All demands, defined in this example for aggregations of analyzed quality attributes, cannot be realized using the approaches which are conventional in the field of quality control.*

6. Conclusion

The aggregation of partial goals - attributes, into one representative global goal is a very important task. Conventional aggregation tools are very often inadequate. Partial demands for aggregation can be and usually are logical demand which can be adequately described only by logical expressions. In this paper logical aggregation as a tool for aggregation is analyzed. Logical aggregation has multiple advantages among others from the stand point of its possibility and interpretability. The new approach treats logical functions – partial aggregation demand, as a generalized Boolean polynomial which can process values from the whole real unit interval [0, 1]. Logical aggregation in a general case is a weighted sum of partial demands. Therefore, aggregation in a general case is a generalized pseudo-logic function. It is interesting that conventional aggregation operators are only a special case of logical aggregation operators and, as a consequence of using LA, one can do much more in an adequate direction than before.

References

[1] G. Choquet, Theory of capacities, *Annals de l'Institute Fourier,* 5:131-295, 1953.

[2] M. Grabisch , The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89(3):445-456, 1996.

[3] M. Grabisch, Labreuche, Ch., Bi-capacities for decision making on bipolar scales. *EUROFUSE Workshop on Information Systems*, Varenna, Italy, 185-190, September 2002.

[4] E.P Klement., Mesiar R., Pap E., *Triangular norms*, Trends in Logic, Studia Logica Library, Vol 8, Kluwer Academic Publishers, 2000.

[5] D. Radojević, Interpolative realization of Boolean algebra as a consistent frame for gradation and/or fuzziness, *Forging New Frontiers: Fuzzy Pioneers II*, Ed. M. Nikravesh, J. Kacprzyk L. A. Zadeh, Studies in Fuzziness and Soft Computing, pp. 326-351, Springer, 2007.

[6] D. Radojević, Interpolative realization of Boolean algebra, NEUREL 2006, Eight Seminar on Neural Network Applications in Electrical Engineering, 201-206, 2006.

[7] D. Radojević, Logical aggregation based on Interpolative realization of Boolean algebra, submitted for publication

[8] D. Radojević, New [0,1]-valued logic: A natural generalization of Boolean logic, *Yugoslav Journal of Operational Research – YUJOR*, Belgrade, Vol 10, No 2, 185-216, 2000.

[9] M. Sugeno, *Theory of fuzzy integrals and its applications*. Doctoral Thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974

[10] R. R Yager, and Kacprzyk, J., *The Ordered Weighted Averaging Operation: Theory, Methodology and Applications*, Kluwer: Norwell, MA, 1997.

E_T-Lipschitzian Aggregation Operators

J. Jacas
Sec Matemàtiques i Informàtica
ETS Arquitecura de Barcelona
Universitat Politècnica de Catalunya
Diagonal 649
08028 Barcelona
Spain
joan.jacas@upc.edu

J. Recasens
Secció Matemàtiques i Informàtica
ETS Arquitectura del Vallès
Universitat Politècnica de Catalunya
Pere Serra 1-15
08190 Sant Cugat del Vallès
Spain
j.recasens@upc.edu

Abstract

Lipschitzian and kernel aggregation operators with respect to the natural T-indistinguishability operator E_T and their powers are studied. A t-norm T is proved to be E_T-lipschitzian, and is interpreted as a fuzzy point and a fuzzy map as well. Given an archimedean t-norm T with additive generator t, the quasi-arithmetic mean generated by t is proved to be the more stable aggregation operator with respect to T.

Keywords: Aggregation Opoerator, T-indistinguishability Operator, Lipschitzian, Kernel.

1 Introduction

Lipschitzian aggregation operators have been studied in [4] [5] [13] [14] by considering the usual metric on the unit interval. In this paper we study the lipschitzian condition of aggregation operators with respect to the natural indistinguishability operator E_T and their powers E_T^p (see definitions below) so that an aggregation operator h is E_T^P-lipschitzian when for all $x_1, x_2, ..., x_n, y_1, y_2, ..., y_n \in [0,1]$ $T(E_T^p(x_1,y_1), ..., E_T^p(x_n,y_n)) \leq E_T(h(x_1,x_2,...,x_n), h(y_1,y_2,...,y_n))$. This means that from similar inputs we obtain similar aggregations. The use of E_T and E_T^p assumes the election of a specific t-norm T and therefore the selection of a particular family of logics where the semantics of the conjunction and the biimplication are given by T and E_T.

As it will be seen in this paper, when T is the Lukasiewicz t-norm, the E_T-lipschitzian condition coincides with the 1-lipschitzian condition with the usual metric on [0,1] and the definition of [13] is recovered. This is not a surprising result, due to the relation between E_T and the usual metric on [0,1] in this case.

It is worth noticing the relation between the lipschitzian condition of an aggregation operator E_T and its extensionality with respect to the integral powers $T(\overbrace{E_T, ..., E_T}^{n \text{ times}})$ (Proposition 3.9).

Among other results, it will be proved in this paper that if T is a continuous archimedean t-norm with an additive generator t and m_t the quasi-arithmetic mean generated by t ($m_t(x_1,x_2,...,x_n) = t^{-1}\left(\frac{t(x_1)+t(x_2)+...+t(x_n)}{n}\right)$), then m_t is the more stable aggregation operator with respect to T (Proposition 3.21).

Also the t-norm T is not only lipschitzian with respect to E_T, but it can be seen as a fuzzy point and a fuzzy map as well (Proposition 3.23, Proposition 3.25) and an aggregation operator h is greater than or equal to T if and only if h is 1-E_T-lipschitzian.

In the definition of E_T-lipschitzianity we replace the t-norm T by the minimum, i.e. $(Min(E_T^p(x_1,y_1), ..., E_T^p(x_n,y_n)) \leq E_T(h(x_1,x_2,...,x_n), h(y_1,y_2,...,y_n)))$, then we obtain a generalization of the kernel aggregation operators studied in [17] [13]. Again, if T is the Lukasiewicz t-norm this definition is equivalent to the one given in the above mentioned references.

2 Preliminaries

This section contains some results on t-norms and indistinguishability operators that will be needed later on in the paper. Besides well known definitions and theorems, the power T^n of a t-norm is generalized to irrational exponents in Definition 2.3 and given explicitly for continuous archimedean t-norms in Proposition 2.4.

For the sake of simplicity we will assume continuity for the t-norms throughout the paper.

Since a t-norm T is associative, we can extend it to an

n-ary operation in the standard way:

$$T(x) = x$$
$$T(x_1, x_2, ...x_n) = T(x_1, T(x_2, ..., x_n)).$$

In particular, following the notation in [18], $T(\overbrace{x, x, ..., x}^{n\ times})$ will be denoted by $x_T^{(n)}$.

The n-th root $x_T^{(\frac{1}{n})}$ of x with respect to T is defined by

$$x_T^{(\frac{1}{n})} = sup\{z \in [0,1] \mid z_T^{(n)} \leq x\}$$

and for $m, n \in N$, $x_T^{(\frac{m}{n})} = \left(x_T^{(\frac{1}{n})}\right)_T^{(m)}$.

Lemma 2.1. [18] *If $k, m, n \in N$, $k, n \neq 0$ then $x_T^{(\frac{km}{kn})} = x_T^{(\frac{m}{n})}$.*

Lemma 2.2. *Let $x_1, ..., x_n \in (0,1]$ and $n \in N$. $T(x_{1_T}^{(\frac{1}{n})}, ..., x_{n_T}^{(\frac{1}{n})}) \neq 0$.*

The powers $x_T^{(\frac{m}{n})}$ can be extended to irrational exponents in a straightforward way.

Definition 2.3. *If $r \in R^+$ is a positive real number, let $\{a_n\}_{n\in N}$ be a sequence of rational numbers with $lim_{n\to\infty} a_n = r$. For any $x \in [0,1]$, the power $x_T^{(r)}$ is*

$$x_T^{(r)} = lim_{n\to\infty} x_T^{(a_n)}.$$

Continuity assures the existence of limit and independence of limit from the selection of the sequence $\{a_n\}_{n\in N}$.

Proposition 2.4. *Let T be an archimedean t-norm with additive generator t, $x \in [0,1]$ and $r \in R^+$. Then*

$$x_T^{(r)} = t^{[-1]}(rt(x)).$$

Proof. Due to continuity of t we need to prove it only for rational r.

If r is a natural number m, then trivially $x_T^{(m)} = t^{[-1]}(mt(x))$.

If $r = \frac{1}{n}$ with $n \in N$, then $x_T^{(\frac{1}{n})} = z$ with $z_T^{(n)} = x$ or $t^{[-1]}(nt(z)) = x$ and $x_T^{(\frac{1}{n})} = t^{[-1]}\left(\frac{t(x)}{n}\right)$.

For a rational number $\frac{m}{n}$,

$$x_T^{(\frac{m}{n})} = \left(x_T^{(\frac{1}{n})}\right)_T^{(m)} = t^{[-1]}\left(mt\left(x_T^{(\frac{1}{n})}\right)\right) =$$
$$t^{[-1]}\left(mt\left(t^{[-1]}\left(\frac{t(x)}{n}\right)\right)\right) = t^{[-1]}\left(\frac{m}{n}t(x)\right).$$

□

Theorem 2.5. *Ling* [15] *A continuous t-norm T is archimedean if and only if there exists a continuous decreasing map $t : [0,1] \to [0,\infty]$ with $t(1) = 0$ such that*

$$T(x,y) = t^{[-1]}(t(x) + t(y))$$

where $t^{[-1]}$ stands for the pseudo-inverse of t defined by

$$t^{[-1]}(x) = \begin{cases} 1 & \text{if } x < 0 \\ t^{-1}(x) & \text{if } x \in [0, t(0)] \\ 0 & \text{otherwise.} \end{cases}$$

T is strict if $t(0) = \infty$ and non-strict otherwise.

t is called an additive generator of T and two additive generators of the same t-norm differ only by a multiplicative constant.

Definition 2.6. *The residuation $\overrightarrow{T}$ of a t-norm T is defined by*

$$\overrightarrow{T}(x|y) = sup\{\alpha \in [0,1] \mid T(x,\alpha) \leq y).$$

Definition 2.7. *The natural T-indistinguishability operator E_T associated to a given t-norm T is the fuzzy relation on* [0,1] *defined by*

$$E_T(x,y) = T(\overrightarrow{T}(x|y), \overrightarrow{T}(y|x)) = Min(\overrightarrow{T}(x|y), \overrightarrow{T}(y|x)).$$

Example 2.8.

1. *If T is an archimedean t-norm with additive generator t, then $E_T(x,y) = t^{-1}(|t(x) - t(y)|)$ for all $x, y \in [0,1]$.*
2. *If T is the Lukasiewicz t-norm, then $E_T(x,y) = 1 - |x - y|$ for all $x, y \in [0,1]$.*
3. *If T is the Product t-norm, then $E_T(x,y) = \begin{cases} \frac{Min(x,y)}{Max(x,y)} & \text{if } x \neq y \\ 1 & \text{otherwise.} \end{cases}$*
4. *If T is the Minimum t-norm, then $E_T(x,y) = \begin{cases} Min(x,y) & \text{if } x \neq y \\ 1 & \text{otherwise.} \end{cases}$*

E_T is indeed a special kind of (one-dimensional) T-indistinguishability operator (Definition 2.9) [3] and in a logical context where T plays the role of the conjunction, E_T is interpreted as the bi-implication associated to T [7].

The general definition of T-indistinguishability operator is

Definition 2.9. *Given a t-norm T, a T-indistinguishability operator E on a set X is a fuzzy relation $E : X \times X \to [0,1]$ satisfying for all $x, y, z \in X$*

1. $E(x,x) = 1$ *(Reflexivity)*
2. $E(x,y) = E(y,x)$ *(Symmetry)*
3. $T(E(x,y), E(y,z)) \leq E(x,z)$ *(T-transitivity).*

Proposition 2.10. [21] *Let μ be a fuzzy subset of X and T a continuous t-norm. The fuzzy relation E_μ on X defined for all $x, y \in X$ by*

$$E_\mu(x,y) = E_T(\mu(x), \mu(y))$$

is a T-indistinguishability operator on X.

Definition 2.11. *Let E be a T-indistinguishability operator on a set X. A fuzzy subset μ of X is extensional with respect to E if and only if for all $x, y \in X$*

$$T(E(x,y), \mu(y)) \leq \mu(x).$$

Proposition 2.12. *Let E be a T-indistinguishability operator on a set X. A fuzzy subset μ of X is extensional with respect to E if and only if for all $x, y \in X$*

$$E(x,y) \leq E_T(\mu(x), \mu(y)).$$

Finally, let us recall the definition of aggregation operator.

Definition 2.13. [4] *An aggregation operator is a map $h : \bigcup_{n\in N}[0,1]^n \to [0,1]$ satisfying*

1. $h(0,...,0) = 0$ and $h(1,...,1) = 1$
2. $h(x) = x \ \forall x \in [0,1]$
3. $h(x_1,...,x_n) \leq h(y_1,...,y_n)$ if $x_1 \leq y_1, ..., x_n \leq y_n$ *(monotonicity).*

The restriction of h to $[0,1]^n$ will be denoted by $h_{(n)}$ so that a global aggregation operator h can be split into the family of n-ary operators $(h_{(n)})_{n\in N}$.

3 E_T-Lipschitzian and E_T-kernel aggregation operators

Lipschitzian and kernel aggregation operators with respect to the natural T-indistinguishability operator E_T and their powers are a special kind of aggregation operators that generalize the definitions of [13], [17]. Their interest is in the fact that they are stable operators in the sense that the similarity between the aggregation of two n-tuples is bounded by the similarity between them.

It is interesting to point out that the lipschitzian and kernel conditions are equivalent to extensionality (Proposition 3.9, Proposition 3.27).

Among other results, it will be proved that a t-norm T is E_T-lipschitzian and moreover the maps $T_{(n)}$ can be interpreted as fuzzy points of $[0,1]^n$ and a fuzzy maps from $[0,1]^k$ to $[0,1]^{n-k}$.

Also quasi-arithmetic means are proved to be the more stable aggregation operators.

Proposition 3.1. *Let E be a T indistinguishability operator on a set X. The fuzzy relation E^n defined by*

$$E^n(x,y) = T(\overbrace{E(x,y),...E(x,y)}^{n \text{ times}}) \ \forall x, y \in X$$

is a T-indistinguishability operator.

The powers E_T^n of the natural T-indistinguishability operators have been studied in relation with antonymy and fuzzy partitions in [20].

Proposition 3.2. [11] *Let E be a T-indistinguishability operator on a set X. $E^{\frac{1}{n}}$ is a T-indistinguishability operator on X.*

Corollary 3.3. *Let E be a T-indistinguishability operator on a set X. $E^{\frac{m}{n}}$ is a T-indistinguishability operator on X.*

Proof. Propositions 3.1. and 3.2. □

Corollary 3.4. *Let E_T be the natural T-indistinguishability operator on [0,1] associated to T. $E_T^{\frac{m}{n}}$ is a T-indistinguishability operator.*

Continuity of the t-norm T allows us to extend the powers of a T-indistinguishability operator to positive irrational numbers in the same way as in Definition 2.3.

Example 3.5.

1. *If T is continuous archimedean with additive generator t, then $E_T^p(x,y) = t^{[-1]}(p|t(x) - t(y)|)$ for all $x, y \in [0,1]$.*
2. *If T is the Lukasiewicz t-norm, then $E_T^p(x,y) = Max(0, 1 - p|x - y|)$ for all $x, y \in [0,1]$.*
3. *If T is the Product t-norm, then* $E_T^p(x,y) = \begin{cases} \frac{Min(x^p,y^p)}{Max(x^p,y^p)} & \text{if } x \neq y \\ 1 & \text{otherwise.} \end{cases}$
4. *If T is the Minimum t-norm, then $E_T^p(x,y) = E_T(x,y)$ for all $x, y \in [0,1]$.*

Proposition 3.6. *Let T be a t-norm and $p, q > 0$. $E_T^p \leq E_T^q$ if and only if $p \geq q$.*

Definition 3.7. *Let E be a T-indistinguishability operator on $[0,1]$. h is E-lipschitzian if and only if $\forall n \in N$, $\forall x_1, ..., x_n, y_1, ..., y_n \in [0,1]$*

$$T(E(x_1, y_1), ..., E(x_n, y_n)) \leq$$

$$E_T(h(x_1, ..., x_n), h(y_1, ..., y_n)).$$

If $E_1, ..., E_n$ are T-indistinguishability operators defined on the universes $X_1, ..., X_n$ respectively, there are at least two natural ways to define a T-indistinguishability operator on $X_1 \times ... \times X_n$.

Proposition 3.8. *Let $E_1, ..., E_n$ be T-indistinguishability operators on $X_1, ..., X_n$ respectively. Then the two fuzzy relations $T(E_1, ..., E_n)$ and $Min(E_1, ..., E_n)$ on $X_1 \times ... \times X_n$ defined for all $(x_1, ..., x_n), (y_1, , , , y_n) \in X_1 \times ... \times X_n$ by*

$$T(E_1, ..., E_n)((x_1, ..., x_n), (y_1, , , , y_n)) =$$

$$T(E_1(x_1, y_1), ..., E_n(x_n, y_n))$$

and

$$Min(E_1, ..., E_n)((x_1, ..., x_n), (y_1, , , , y_n)) =$$

$$Min(E_1(x_1, y_1), ..., E_n(x_n, y_n))$$

are T-indistinguishability operators on $X_1 \times ... \times X_n$.

Proposition 3.9. *Let E be a T-indistinguishability on $[0,1]$ and h an aggregation operator. h is E-lipschitzian if and only if $h_{(n)}$ (as a fuzzy subset of $[0,1]^n$) is extensional with respect to $T(\overbrace{E, ..., E}^{n \text{ times}})$ for all $n \in N$.*

Proof. Proposition 2.12 □

Lemma 3.10. *Let T be a continuous t-norm. The for all $x, y \in [0,1]$ $x \geq y$*

$$T(x, \overrightarrow{T}(x|y) = y.$$

Next Proposition shows that a t-norm T is an E_T-lipschitzian aggregation operator.

Proposition 3.11. *Let T be a continuous t-norm. Then T is an E_T-lipschitzian aggregation operator.*

Note that if $x_i \leq y_i$ for all $i = 1, ...n$, then $T(E_T(x_1, y_1), ..., E_T(x_n, y_n)) = E_T(T(x_1, ..., x_n), T(y_1, ..., y_n))$. Since for every t-norm different from the Minimum $E_T^p < E_T^q$ if $p > q$, we have that $T \neq Min$ is not E_T^p-lipschitzian for $p < 1$.

If T is a continuous archimedean t-norm, the E_T^p-lipschitzian property translates to a classical lipschitzian condition.

Proposition 3.12. *Let T be a continuous archimedean t-norm with additive generator t, $p \in [0,1]$ and h an aggregation operator. h is E_T^p if and only if $\forall n \in N$, $\forall x_1, ..., x_n, y_1, ..., y_n \in [0,1]$*

$$p|t(x_1) - t(y_1)| + ... + p|t(x_n) - t(y_n)| \geq$$

$$|t(h(x_1, ..., x_n)) - t(h(y_1, ..., y_n))| \quad (1).$$

Last Proposition says that for all $n \in N$ the map $H : [0, t(0)]^n \to [[0, t(0)]$ defined by

$$H(x_1, ..., x_n) = t(h(t^{-1}(x_1), ..., t^{-1}(x_n)))$$

is a p-lipschitzian map.

Also note that if T is the Lukasiewicz t-norm, then (1) is the definition of the Lipschitz property in [13], so that Definition 3.7 contains the one in [13] as a particular case.

If an aggregation operator h is E_T^p-lipschitzian, it may happen that for different values of n the corresponding n-ary operators $h_{(n)}$ may satisfy the lipschitzian conditions for different values of p ([4] p. 12).

Definition 3.13. *An aggregation operator is sub idempotent if and only if for all $x \in [0,1]$ and $n \in N$, $h(\overbrace{x, ..., x}^{n \text{ times}}) \leq x$*

Proposition 3.14. *Let $T \neq Min$ be a t-norm, h a sub idempotent aggregation operator and $n \in N$. If $h_{(n)}$ is E_T^p-lipschitzian, then $p \geq \frac{1}{n}$.*

Proof. If $h_{(n)}$ is E_T^p-lipschitzian, then in particular, for $x \in X$

$$T(\overbrace{E_T^p(1, x), ..., E_T^p(1, x)}^{n \text{ times}}) \leq E_T(h(\overbrace{1, ..., 1}^{n \text{ times}}), h(\overbrace{x, ..., x}^{n \text{ times}}))$$

and so

$$x_T^{(pn)} \leq h(\overbrace{x, ..., x}^{n \text{ times}}) \leq x$$

which holds if and only if $pn \geq 1$ or equivalently, if and only of $p \geq \frac{1}{n}$ □

If T is a non-strict continuous archimedean t-norm the sub idempotent property can be dropped.

Proposition 3.15. *Let T be a non-strict continuous archimedean t-norm with additive generator t, h an aggregation operator and $n \in N$. If $h_{(n)}$ is E_T^p-lipschitzian, then $p \geq \frac{1}{n}$.*

Proof. Putting $x_i = 1$ and $y_i = 0$ for all $i = 1, ..., n$ in Proposition 3.12, we get

$$p|t(1) - t(0)| + ... + p|t(1) - t(0)| \geq |t(1) - t(0)|.$$

$$npt(0) \geq t(0)$$

or

$$p \geq \frac{1}{n}.$$

□

In [4] it has been proved that the arithmetic mean is the only aggregation operator h whose n-ary maps $h_{(n)}$ are $\frac{1}{n}$-lipschitzian. Proposition 3.21 generalizes this result to arbitrary quasi-arithmetic means.

Next Proposition is well known.

Proposition 3.16. [1], [18] *m is a quasi-arithmetic mean in [0,1] if and only if there exists a continuous monotonic map $t : [0,1] \to [-\infty, \infty]$ such that for all $n \in N$ and $x_1, ..., x_n \in [0,1]$*

$$m(x_1, ...x_n) = t^{-1}\left(\frac{t(x_1) + ... + t(x_n)}{n}\right).$$

m is continuous if and only if Ran $t \neq [-\infty, \infty]$.

t will be called a generator of m and if m is generated by t we will denote it by m_t.

Lemma 3.17. [11] *Let $t, t' : [0,1] \to [-\infty, \infty]$ be two continuous strict monotonic maps with Ran t, Ran $t' \neq [-\infty, \infty]$ differing only by a non-zero multiplicative constant α ($t' = \alpha\ t$) and $m_t, m_{t'}$ the quasi-arithmetic means generated by them respectively. Then $m_t = m_{t'}$.*

Lemma 3.18. [11] *Let $t, t' : [0,1] \to [-\infty, \infty]$ be two continuous strict monotonic maps with Ran t, Ran $t' \neq [-\infty, \infty]$ differing only by an additive constant and $m_t, m_{t'}$ the quasi-arithmetic means generated by them respectively. Then $m_t = m_{t'}$.*

Lemma 3.19. [11] *Let $t : [0,1] \to [-\infty, \infty]$ be a continuous strict monotonic map. Then $m_t = m_{-t}$.*

Proposition 3.20. [11] *The map assigning to every continuous Archimedean t-norm T with generator t the mean m_t generated by t is a bijection between the set of continuous Archimedean t-norms and the set of continuous quasi-arithmetic means.*

Proposition 3.21. *Let T be a continuous archimedean t-norm with additive generator t and m_t the quasi-arithmetic mean generated by t.*

- *(a) For every $n \in N$ $m_{t(n)}$ is E_T^p-lipschitzian if and only if $p \geq \frac{1}{n}$.*
- *(b) m_t is the only aggregation operator fulfilling (a)*

In Proposition 3.11 we have proved that a t-norm T is E_T-lipschitzian. In fact, $T_{(n)}$ can also be seen as a fuzzy point of $[0,1]^n$ and a fuzzy map from $[0,1]^{n-1}$ into $[0,1]$.

Definition 3.22. *Let E be a T-indistinguishability operator on a set X and μ a fuzzy subset of X. μ is a fuzzy point of X with respect to E if and only if for all $x, y \in X$*

$$T(\mu(x), \mu(y)) \leq E(x, y).$$

Proposition 3.23. *Let T be a continuous t-norm. $T_{(n)}$ is a fuzzy point of $[0,1]^n$ with respect to $T(\overbrace{E_T, ..., E_T}^{n \text{ times}})$.*

Proof. We have to prove that

$$T(T(x_1, ..., x_n), T(y_1, ..., y_n))$$
$$\leq T(E_T(x_1, y_1), ..., E_T(x_n, y_n))$$

which is an immediate consequence of

$$T(x_i, y_i) \leq E_T(x_i, y_i) \text{ for all } i = 1, ..., n.$$

□

Definition 3.24. *Let E, F be two T-indistinguishability operators on X and Y respectively and R a fuzzy set of $X \times Y$ (i.e.: $R : X \times Y \to [0,1]$). R is a fuzzy map from X to Y if and only if for all $x, x' \in X$, $y, y' \in Y$*

- *(a) $T(E(x, x'), F(y, y'), R(x, y)) \leq R(x', y')$*
- *(b) $T(R(x, y), R(x, y')) \leq F(y, y')$.*

Proposition 3.25. *Let T be a continuous t-norm. $T_{(n)}$ is a fuzzy map from $[0,1]^{n-1}$ to $[0,1]$ endowed with the T indistinguishability operators $T(\overbrace{E_T, ..., E_T}^{n-1 \text{ times}})$ and E_T respectively.*

In fact, it can be proved in the same way that $T_{(n)}$ is a fuzzy map from $[0,1]^k$ to $[0,1]^{n-k}$ ($2 \leq k \leq n-1$) endowed with the T indistinguishability operators $T(\overbrace{E_T, ..., E_T}^{k \text{ times}})$ and $T(\overbrace{E_T, ..., E_T}^{n-k \text{ times}})$ respectively.

Kernel aggregation operators are a family of aggregation operators tightly related to lipschitzian ones. They were introduced in [17] (see also [13], [4]). As the lipschitzian condition, the condition for being a kernel operator was related to the usual metric on the unit interval. It can be extended using natural indistinguishability operators in the same way as it has been done in this paper with the lipschitzian condition. Again, if the T norm is the Lukasiewicz one, the original definition of [17] is recovered.

Definition 3.26. *Let E be a T-indistinguishability operator on [0,1] and h an aggregation operator. h is an E-kernel aggregation operator if and only if $\forall n \in N$, $\forall x_1, ..., x_n, y_1, ..., y_n \in [0,1]$*

$$Min(E(x_1, y_1), ..., E(x_n, y_n)) \leq$$

$$E_T(h(x_1, ..., x_n), h(y_1, ..., y_n)).$$

Proposition 3.27. *Let E be a T-indistinguishability operator on [0,1] and h an aggregation operator. h is an E-kernel aggregation operator if and only if $h_{(n)}$ (as a fuzzy subset of $[0,1]^n$) is extensional with respect to $Min(\overbrace{E, ..., E}^{n \text{ times}})$ for all $n \in N$.*

Proof. Proposition 2.12 □

For archimedean t-norms, the kernel property can be written as in the follows.

Proposition 3.28. *Let T be a continuous archimedean t-norm with additive generator t, $p \in [0,1]$ and h an aggregation operator. h is E_T^p-kernel aggregation operator if and only if $\forall n \in N$, $\forall x_1, ..., x_n, y_1, ..., y_n \in [0,1]$*

$$Max(p|t(x_1) - t(y_1)|, ..., p|t(x_n) - t(y_n)|) \geq$$

$$|t\left(h(x_1, ..., x_n)\right) - t\left(h(y_1, ..., y_n)\right)| \quad (2).$$

Proof.

$$Min(t^{-1}(p|t(x_1) - t(y_1)|), ..., t^{-1}(p|t(x_n) - t(y_n)|)) \leq$$

$$t^{-1}(|t(h(x_1, ..., x_n)) - t(h(x_1, ..., x_n))|)$$

$$t^{-1}\left(Max(p|t(x_1) - t(y_1)|, ..., p|t(x_n) - t(y_n)|)\right) \leq$$

$$t^{-1}(|t(h(x_1, ..., x_n)) - t(h(x_1, ..., x_n))|)$$

$$Max(p|t(x_1) - t(y_1)|, ..., p|t(x_n) - t(y_n)|) \geq$$

$$|t(h(x_1, ..., x_n)) - t(h(x_1, ..., x_n))|.$$

□

If T is the Lukasiewicz t-norm and $p = 1$, then (2) is the definition of the kernel aggregation operator in [17].

4 Concluding Remarks

In this paper Lipschitzian and kernel aggregation operators with respect to the natural T-indistinguishability operator E_T and their powers have been studied.

It has been proved that a t-norm T is E_T-lipschitzian, and a fuzzy point and a fuzzy map as well.

Quasi-arithmetic means m_t play an important role since they are the more stable aggregation operator with respect to T, meaning that the corresponding n-ary operators $m_{t(n)}$ are $E_T^{\frac{1}{n}}$-lipschitzian maps.

Lipschitzian and kernel properties are not only interesting for aggregation operators, but in almost any part of fuzzy reasoning and they deserve a deep study.

Acknowledgement

Research partially supported by DGICYT project number TIN2006-14311.

References

[1] J. Aczél, *Lectures on functional equations and their applications*. Academic Press, New York/London 1966.

[2] D. Boixader, Some Properties Concerning the Quasi-inverse of a t-norm, Mathware and Soft Computing 5 (1998) 5-12.

[3] D. Boixader, J. Jacas, J. Recasens, Fuzzy Equivalence Relations: Advanced Material. In Dubois, Prade Eds. *Fundamentals of Fuzzy Sets*, Kluwer, (2000), 261-290.

[4] T. Calvo, A. Kolesárova, M. Komorníková, R. Mesiar, Aggregation Operators: Properties, Classes and Construction Methods. In Mesiar, Calvo, Mayor Eds. *Aggregation Operators: New Trends and Applications*. Studies in Fuzziness and Soft Computing. Springer, (2002), 3-104.

[5] T. Calvo, R. Mesiar, Stability of aggregation operators. Proc EUSFLAT 2001, Leicester, (2001) 457-458.

[6] M. Demirci, Fundamentals of M-vague algebra and M-vague arithmetic operations, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10 (2002) 25-75.

[7] P. Hájek, *Metamathematics of Fuzzy Logic*. Kluwer, Dordrecht, 1998.

[8] J. Jacas, Fuzzy topologies induced by S-metrics, The Journal of Fuzzy Mathematics 1 (1993) 173-191.

[9] J. Jacas, J. Recasens, One dimensional indistinguishability operators. Fuzzy Sets and Systems 109 (2000) 447-451.

[10] J. Jacas, J. Recasens, Aggregation of T-Transitive Relations, Int. J. of Intelligent Systems 18 (2003) 1193-1214.

[11] J. Jacas, J. Recasens, Aggregation Operators Based on Indistinguishability Operators, Int. J. of Intelligent Systems 21 (2006) 857-873.

[12] J. Jacas, J. Recasens, Linguistic Modifiers, Fuzzy Maps and the lipschitzian Condition. Proc IPMU 2006, Paris (2006).

[13] A. Kolesárová, J. Mordelová, 1-Lipschitz and kernel aggregation operators. Proc AGOP 2001, Oviedo (2001) 71-76.

[14] A. Kolesárová, J. Mordelová, E.Muel, Kernel aggregation. operators and their marginals, Fuzzy Sets and Systems (2004) 35-50.

[15] C. M. Ling, Representation of associative functions, Publ. Math. Debrecen 12 (1965) 189-212.

[16] V, Novák, R. Mesiar, Operations fitting triangular-norm-based biresiduation, Fuzzy Sets and Systems 104 (1999) 77-84.

[17] J. Mordelová, E. Muel, Kernel aggregation operators. Proc AGOP 2001, Oviedo (2001) 95-98.

[18] E. P. Klement, R. Mesiar, E. Pap, *Triangular norms*. Kluwer, Dordrecht, 2000.

[19] B. Schweizer, A. Sklar *Probabilistic Metric Spaces*. North-Holland, Amsterdam, 1983.

[20] A.R. De Soto, J. Recasens, Modeling a Linguistic Variable as a Hierarchical Family of Partitions induced by an Indistinguishability Operator, Fuzzy Sets and Systems 121 (2000) 427-437.

[21] L.Valverde, On the structure of F-indistinguishability operators, Fuzzy Sets and Systems 17 (1985) 313–328.

[22] L.A.Zadeh, Similarity relations and fuzzy orderings, Information Science 3 (1971) 177–200.

Session 2

Axiomatic Fuzzy Mathematics – P. Cintula and L. Běhounek

Fuzzy Class Theory: Some Advanced Topics

Petr Cintula
Institute of Computer Science,
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2,
187 02 Prague 8, Czech Republic
cintula@cs.cas.cz

Rostislav Horčík
Institute of Computer Science,
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2,
187 02 Prague 8, Czech Republic
horcik@cs.cas.cz

Abstract

The goal of this paper is to push forward the development of the apparatus of the Fuzzy Class theory. We concentrate on three areas: strengthening the universal quantifier, formalizing the idea that 'similar' fuzzy sets fulfill their properties to 'similar' degrees, and embedding of classical crisp theories into Fuzzy Class theory.

Keywords: Fuzzy Class Theory, fuzzy relations, graded properties, quantifiers, logic MTL

1 Introduction

Fuzzy Class Theory has the aim to axiomatize the notion of fuzzy set. A brief overview of FCT can be found in Appendix A, where also all necessary defined predicates of the theory, freely used in the following sections, are introduced. For a detailed account of the theory we refer the reader to the original paper [1] or a freely available primer [2].

The goal of this paper is to advance the apparatus of FCT in three ways. First, we introduce a class of quantifiers strengthening the universal quantifier, the so-called super-universal quantifiers (note that the classical theory of generalized quantifiers studies just 'weaker' variants of universal quantifier, like 'many' or 'almost all', as in the classical logic the universal quantifier is the only super-universal quantifier). The usual motivations for studies of generalized quantifiers are linguistical, but our motivation is purely 'logical' and 'mathematical'. Recall that for any multiset M of terms we can show:

$$(\forall x)\varphi \to \bigwedge_{t\in M} \varphi(t)$$

The most notable super-universal quantifier is the so-called multiplicative quantifier. It is, roughly speaking, a quantifier $\mathcal{Q}$ such that for each multiset M of terms holds:

$$(\mathcal{Q}x)\varphi \to \bigotimes_{t\in M} \varphi(t)$$

Observe that $\forall$ clearly does not satisfy the above formula, but the quantifier $\triangle\forall$ does, and so does some other ones (e.g., the one assigning to each formula a largest idempotent bellow all instances of φ).

Then, in Section 3, we study the so-called weak congruence property, generalizing the work of Bělohlávek ([4, Lemma 4.8]). There the author shows that if two sets are 'equal' in the high degree, they share the same properties, formally written as:

$$X \approx^n Y \to (\varphi(X) \leftrightarrow \varphi(Y))$$

The degree n depends on the complexity of formula (property) φ. The generalization we offer is based on the crucial observation that in FCT we can define the following two different notions of set equality (both used in fuzzy set theory papers):

$$A \approx B \equiv_{\mathrm{df}} (\forall x)(x \in A \leftrightarrow x \in B)$$

$$A \cong B \equiv_{\mathrm{df}} (A \subseteq B)\ \&\ (B \subseteq A)$$

By a finer syntactic analysis of the form of formula φ we manage to prove theorem of the form:

$$X \cong^{m,n} Y \to (\varphi(X) \leftrightarrow \varphi(Y))$$

note that [4, Lemma 4.8] would give us just:

$$X \approx^{m+n} Y \to (\varphi(X) \leftrightarrow \varphi(Y))$$

We can show the following implication and non-validity of the converse one:

$$A \approx^{m+n} B \to A \cong^{m,n} B$$

Thus our result is indeed a non-trivial generalization.

Finally, in Section 4, we study how to embed classical crisp theories into FCT. The results in this section are

not surprising, they are just thorough formalizations of the fact that we can translate classical theories formalized in the classical type theory to FCT since FCT 'contains' classical simple type theory. The first step of this formalization, the case of first-order theories, was done already in the original paper [1].

2 Super-universal quantifiers

A unary quantifier is a fuzzy class of the second order and a binary quantifier is a binary fuzzy relation of the second order. We shall use the usual conventions to write $\mathcal{Q}X$ instead of $X \in \mathcal{Q}$. Here we restrict ourselves to unary quantifiers, for simplicity, the more general study of binary ones will be a subject of some future papers. Let us now mention how the quantifiers, as we defined them, relate to the 'usual' ones. We set:

$$(\mathcal{Q}x)\varphi \equiv_{\mathrm{df}} \mathcal{Q}\{x \mid \varphi\} \tag{1}$$

Other way round, if $\mathcal{Q}$ is a 'usual' quantifier of (an extension of) FCT, we can write $\mathcal{Q}X \equiv_{\mathrm{df}} (\mathcal{Q}x)(x \in X)$. In particular we have fuzzy class of fuzzy classes $\forall$ defined as $\forall X \equiv_{\mathrm{df}} (\forall x)(x \in X)$. Now we single out natural properties for "super-universal" quantifiers. Let us, for now, work on the meta-level and understand the following conditions as new axioms or rules added to FCT.

1. from $(\varphi \to \psi)$ infer $(\mathcal{Q}x)\varphi \to (\mathcal{Q}x)\psi$
2. from φ infer $(\mathcal{Q}x)\varphi$
3. $(\mathcal{Q}x)\varphi \to \varphi(t)$

Let us rewrite these conditions to the "class" form.

1. $(X \sqsubseteq Y) \to (\mathcal{Q}X \to \mathcal{Q}Y)$
2. $\mathcal{Q}\mathrm{V}$
3. $\mathcal{Q} \subseteq \forall$

Recall the meaning of the super-universal quantifier $(\mathcal{Q}x)\varphi$ can be roughly described as: "closeness of the class $\{x \mid \varphi\}$ to the universal class V". The first condition expresses monotonicity of this concept ("bigger" classes are closer to V than the "smaller" ones), the second one expresses the fact that V is fully close to itself. The final one is the condition of super-universality: in classical logic it would lead to $\mathcal{Q} = \forall$ and so this condition is always omitted when interpreting universal-like quantifiers like 'many', 'almost all', etc. Finally observe that if we would replace condition 1. with:

$$(X \subseteq Y) \to (\mathcal{Q}X \to \mathcal{Q}Y)$$

then by setting $X = \mathrm{V}$ we would get (by condition 2.) that $\forall Y \to \mathcal{Q}Y$; thus together with condition 3. we would get $\forall = \mathcal{Q}$. Other feasible option would be to replace condition 1. by:

$$(\mathcal{Q}x)(\varphi \to \psi) \to ((\mathcal{Q}x)\varphi \to (\mathcal{Q}x)\psi)$$

This condition (together with condition 2.) implies the original condition 1. and so it would constitute an interesting subclass of super-universal quantifiers, we left the study of this class to some upcoming paper.

Let us step down from the meta-level of FCT and start building a theory of super-universal quantifiers *inside* FCT. We define a (third order) class $\mathbb{SUQ}$:

$$\begin{aligned}\mathbb{SUQ}^{m,l,u}(\mathcal{Q}) \equiv_{\mathrm{df}}\\ ((\forall X, Y)((X \sqsubseteq Y) \to (\mathcal{Q}X \to \mathcal{Q}Y)))^m\\ \& \, (\mathcal{Q}\mathrm{V})^l \,\&\, \mathcal{Q} \subseteq^u \forall\end{aligned}$$

Let us recall the usual conventions:

$$\mathbb{SUQ}^m(\mathcal{Q}) \equiv_{\mathrm{df}} \mathbb{SUQ}^{m,m,m}(\mathcal{Q}) \quad \mathbb{SUQ}(\mathcal{Q}) \equiv_{\mathrm{df}} \mathbb{SUQ}^1(\mathcal{Q})$$

Lemma 2.1 FCT *proves:*

$$\mathrm{Crisp}(\mathbb{I}) \,\&\, \mathbb{I} \neq \emptyset \,\&\, (\forall \mathcal{Q} \in \mathbb{I})(\mathbb{SUQ}(\mathcal{Q})) \to \mathbb{SUQ}(\bigcap \mathbb{I})$$
$$\mathrm{Crisp}(\mathbb{I}) \,\&\, \mathbb{I} \neq \emptyset \,\&\, (\forall \mathcal{Q} \in \mathbb{I})(\mathbb{SUQ}(\mathcal{Q})) \to \mathbb{SUQ}(\bigcup \mathbb{I})$$

Obviously, in FCT we can prove, $\triangle\mathbb{SUQ}(\forall)$, $\triangle\mathbb{SUQ}(\mathrm{Ker}(\forall))$, and $\mathrm{Ker}(\forall) = \{V\}$. In fact we can prove more:

Lemma 2.2 FCT *proves:*

$$\forall = \bigcup\{\mathcal{Q} \mid \mathbb{SUQ}(\mathcal{Q})\} = \bigcup\{\mathcal{Q} \mid \triangle\mathbb{SUQ}(\mathcal{Q})\}$$
$$\mathrm{Ker}(\forall) = \bigcap\{\mathcal{Q} \mid \mathbb{SUQ}(\mathcal{Q})\} = \bigcap\{\mathcal{Q} \mid \triangle\mathbb{SUQ}(\mathcal{Q})\}$$

I.e. the universal quantifier is the largest (fully) super-universal quantifier and its kernel is the least (fully) super-universal quantifier.

Let us define other notable quantifiers, first the largest idempotent super-universal quantifier:

$$\mathcal{S} = \bigcup\{\mathcal{Q} \mid \triangle\mathbb{SUQ}(\mathcal{Q}) \,\&\, \mathcal{Q} = \mathcal{Q} \cap \mathcal{Q}\} \tag{2}$$

It can be easily shown that this quantifier corresponds to the storage quantifier introduced in [8]. Clearly, as $\mathrm{Ker}(\forall) = \mathrm{Ker}(\forall) \cap \mathrm{Ker}(\forall)$ we can use Lemma 2.1 to get:

Lemma 2.3 FCT *proves* $\mathbb{SUQ}(\mathcal{S})$

Semantically speaking we get:

Lemma 2.4 *Let* $\mathbf{A}$ *be an* $\mathrm{MTL}_\triangle$*-algebra and* $\mathbf{M}$ *an* $\mathbf{A}$*-model of* FCT*. Then* $\|\mathcal{S}X\|$ *is the largest idempotent smaller than* $\|a \in X\|$ *for each* $a \in M$*.*

Lemma 2.5 *Let* φ *be a formula,* x *a variable,* M *a multiset of terms. Then in* FCT *we can prove:*

$$(\mathcal{S}x)\varphi \to \mathop{\&}_{t\in M} \varphi(t) \tag{3}$$

$$(\mathcal{S}x)\varphi \to (\mathcal{S}x)\varphi \,\&\, \varphi(t) \tag{4}$$

Now we define another 'multiplicative' quantifier.

$$\mathcal{M} = \bigcup\{\mathcal{Q} \mid \triangle\mathbb{SUQ}(\mathcal{Q}) \,\&\, (\forall X, x)$$
$$\triangle\Big(\big((x \in X \to \mathcal{Q}X) \to \mathcal{Q}X\big) \vee \big(\mathcal{Q}X \leftrightarrow \mathcal{Q}X \,\&\, \mathcal{Q}X\big)\Big)\}$$

As clearly $\mathcal{S}$ fulfills the defining condition, we can use Lemma 2.1 to get:

Lemma 2.6 FCT *proves* $\mathbb{SUQ}(\mathcal{M})$*.*

Now we show the semantical interpretation of quantifier. Let $\mathbf{A}$ be an $\mathrm{MTL}_\triangle$-algebra. We define for a set $B \subseteq A$ a 'multiplication' of B as:

$$\mathop{\&} B = \inf_{n\in\mathbb{N}, a_1,\ldots,a_n\in B} a_1 \& \ldots \& a_n\,.$$

Notice that we do not assume that $a_i \neq a_j$ (for $i \neq j$) in the above definition, i.e., $\mathop{\&} B$ is the infimum of products $a_1 \& \ldots \& a_n$ for all finite sequences $a_1, \ldots, a_n$ of elements from B. Due to the space restriction we have to skip the formal proofs of the following two lemmata, but the second one is a clear semantical consequence of the first one.

Lemma 2.7 *Let* $\mathbf{A}$ *be an* $\mathrm{MTL}_\triangle$*-algebra and* $\mathbf{M}$ *an* $\mathbf{A}$*-model of* FCT*. Then*

$$\|\mathcal{M}X\| = \mathop{\&}\{\|a \in X\| \mid a \in M\}$$

Lemma 2.8 *Let* φ *be a formula,* x *a variable, and* M *a finite multiset of terms. Then in* FCT *we can prove:*

$$(\mathcal{S}x)\varphi \longrightarrow (\mathcal{M}x)\varphi \longrightarrow \mathop{\&}_{t\in M} \varphi(t) \tag{5}$$

Thus we can say that both $\mathcal{M}$ and $\mathcal{S}$ are 'multiplicative' quantifiers. As the following lemma shows the first implication cannot be reversed, i.e., $\mathcal{M}$ is 'better' multiplicative quantifier as it gives us a better bound to $\mathop{\&}_{t\in M} \varphi(t)$. In fact its semantics shows that $\mathcal{M}$ gives the best bound.

Lemma 2.9 *In* FCT *we cannot prove*

$$(\mathcal{M}x)\varphi \to (\mathcal{S}x)\varphi \tag{6}$$

$$(\mathcal{M}x)\varphi \to (\mathcal{M}x)\varphi \,\&\, \varphi(t) \tag{7}$$

Proof: We just show the second claim, as the first one is its simple consequence (taking in account the previous lemma and Lemma 2.5). We just construct an $\mathrm{MTL}_\triangle$-chain $\mathbf{A}$ which is a counterexample to the claim that

$$\mathop{\&} B \le \Big(\mathop{\&} B\Big) \,\&\, a\,,$$

where $B \subseteq A$ and $a \in B$. The construction of a particular model of FCT which refutes $(\mathcal{M}x)\varphi \to (\mathcal{M}x)\varphi \,\&\, \varphi(t)$ is then simple.

Let $A' = \mathbb{Z}^- \times \mathbb{Z}^-$, where $\mathbb{Z}^-$ is the set of non-positive integers. The set A' is ordered lexicographically, i.e., $\langle k, r\rangle \le \langle m, s\rangle$ iff $k < m$ or $k = m$ and $r \le s$. The operations are defined as follows:

$$\langle k, r\rangle \,\&\, \langle m, s\rangle = \langle k + m, r + s\rangle\,,$$

$$\langle k, r\rangle \to \langle m, s\rangle = \begin{cases} \langle 0, 0\rangle \text{ if } \langle k, r\rangle \le \langle m, s\rangle\,, \\ \langle m - k, \min\{0, s - r\}\rangle \text{ otherwise.} \end{cases}$$

Then $\mathbf{A'} = (A', \&, \to, \le, \langle 0, 0\rangle)$ is an integral commutative residuated chain and the ordinal sum $\mathbf{A} = \mathbf{2} \oplus \mathbf{A'}$ is an MTL-chain (where $\mathbf{2}$ is the two-element Boolean algebra). The MTL-chain $\mathbf{A}$ can be clearly viewed as an $\mathrm{MTL}_\triangle$-chain.

Now, let $B = \{\langle 0, -1\rangle\} \subseteq A$. Then

$$\mathop{\&} B = \inf_{n\in\mathbb{N}} \langle 0, -1\rangle^n = \langle -1, 0\rangle\,.$$

However $\langle -1, 0\rangle \,\&\, \langle 0, -1\rangle = \langle -1, -1\rangle < \langle -1, 0\rangle$. Q.E.D.

3 Weak congruence property

We define a notion of positive and negative occurrence of a subformula ψ in a formula φ. Let us denote it by $o^\varphi_\psi \in \{+1, -1\}$. This notation is rather relaxed as the subformula ψ can appear in φ several times. Let us understand ψ is this notion as a subformula together with its fixed occurrence rather than just a subformula. We define it by induction, let φ is a predicate formula in the language of FCT, without the propositional connective $\leftrightarrow$ and ψ its subformula.

- $o^\psi_\psi = +1$
- $o^{\varphi\circ\chi}_\psi = o^{\chi\circ\varphi}_\psi = o^\varphi_\psi$ for $\circ \in \{\&, \wedge, \vee\}$ and any formula χ
- $o^{\triangle\varphi}_\psi = o^{(\forall x)\varphi}_\psi = o^{(\exists x)\varphi}_\psi = o^\varphi_\psi$
- $o^{\chi\to\varphi}_\psi = o^\varphi_\psi$ for any formula χ
- $o^{\neg\varphi}_\psi = o^{\varphi\to\chi}_\psi = -o^\varphi_\psi$ for any formula χ

We say that an occurrence of subformula ψ of φ is $\triangle$-bound if it lies in the scope of $\triangle$ connective.

Lemma 3.1 *Let $\varphi, \psi, \chi, \delta$ be formulas, δ a subformula of χ, let us pick one occurrence of δ. Let us define formulas $\chi(\delta!\varphi)$ and $\chi(\delta!\psi)$ as formulas resulting from χ by replacing the chosen occurrence of δ by φ (ψ respectively). Finally, let $x_1, \dots x_n$ be free variables of φ and ψ which become bound in $\chi(\delta!\varphi)$ or $\chi(\delta!\psi)$*

In first-order $\mathrm{MTL}_\triangle$ *we can prove:*

- $(\forall x_1, \dots x_n)\triangle(\varphi \to \psi) \to (\chi(\delta!\varphi) \to \chi(\delta!\psi))$ *if that occurrence is positive*
- $(\forall x_1, \dots x_n)\triangle(\psi \to \varphi) \to (\chi(\delta!\varphi) \to \chi(\delta!\psi))$ *if that occurrence is negative*

Furthermore if that occurrence is not $\triangle$-bound we can prove:

- $(\forall x_1, \dots x_n)(\varphi \to \psi) \to (\chi(\delta!\varphi) \to \chi(\delta!\psi))$ *if the chosen occurrence is positive*
- $(\forall x_1, \dots x_n)(\psi \to \varphi) \to (\chi(\delta!\varphi) \to \chi(\delta!\psi))$ *if the chosen occurrence is negative*

The proof is almost straightforward but very tedious, so we skip it.

Let us denote by $\chi(\delta : \varphi)$ the formula resulting from χ by replacing *all* occurrences of its subformula δ by φ. Finally by $+^\varphi_\psi$ we denote the number of positive occurrences of ψ in φ, if any of those occurrences is $\triangle$-bound define $+^\varphi_\psi = \triangle$ (analogously we define $-^\varphi_\psi$ for negative occurrences). Recall the convention that $\varphi^\triangle = \triangle\varphi$.

Corollary 3.2 *Let $\varphi, \psi, \chi, \delta$ be formulas, δ a subformula of χ. Let $x_1, \dots x_n$ be free variables in φ and ψ which become bound in $\chi(\delta : \varphi)$ or $\chi(\delta : \psi)$. In first-order* $\mathrm{MTL}_\triangle$ *we can prove:*

$$((\forall x_1, \dots x_n)(\varphi \to \psi))^{+^\chi_\delta} \,\&$$
$$((\forall x_1, \dots x_n)(\psi \to \varphi))^{-^\chi_\delta} \to (\chi(\delta : \varphi) \to \chi(\delta : \psi))$$

Corollary 3.3 *Let φ, ψ be sentences and χ, δ be formulas, δ a subformula of χ. In first-order* $\mathrm{MTL}_\triangle$ *we can prove:*

$$(\varphi \to \psi)^{+^\chi_\delta} \,\&\, (\psi \to \varphi)^{-^\chi_\delta} \to (\chi(\delta : \varphi) \to \chi(\delta : \psi))$$

As a corollary we obtain the main theorem of this section. But first we define:

Definition 3.4 *Let φ be a formula of* FCT *and A a free variable of any order of φ. We say that A is purely extensional in φ if it occurs in φ just in subformulas of the form $X_i \in A$ for some object variables X_i. Then by $+^\varphi_A$ we denote the number of positive occurrences of formulas of the form $X_i \in A$ in φ (analogously we define $-^\varphi_A$), again we set it as $\triangle$ if any of those occurrences if $\triangle$-bound.*

Finally by $\varphi(A : X)$ we denote the formula resulting from φ by replacing all occurrences of class term A by term X of the same order as A.

Theorem 3.5 *Let φ be a formula of* FCT *and A a purely extensional free variable of φ. Then we can prove:*

$$X \approxeq^{+^\varphi_A, -^\varphi_A} Y \to (\varphi(A : X) \to \varphi(A : Y))$$

Observe that if no occurrence of A is $\triangle$-bound then $n = +^\varphi_A + -^\varphi_A$ is the number of occurrences of A in φ and as FCT trivially proves

$$X \approx^n Y \to X \approxeq^{+^\varphi_A, -^\varphi_A} Y,$$

we get:

$$X \approx^n Y \to (\varphi(A : X) \to \varphi(A : Y))$$

This form was proven already in [4, Lemma 4.8.] (in a slightly different framework). Notice that our result is indeed stronger as:

Lemma 3.6 FCT *does not prove:*

$$X \approxeq^{+^\varphi_A, -^\varphi_A} \to X \approx^n Y$$

Example 3.7 *In* FCT*, we can prove*

- $R \subseteq S \to (\mathrm{Refl}(R) \to \mathrm{Refl}(S))$
- $R \approxeq S \to (\mathrm{Sym}(R) \to \mathrm{Sym}(S))$
- $R \approxeq^{1,2} S \to (\mathrm{Trans}(R) \to \mathrm{Trans}(S))$

Compare with [4, Lemma 4.8], which would give us only:

- $R \approx S \to (\mathrm{Refl}(R) \to \mathrm{Refl}(S))$
- $R \approx^2 S \to (\mathrm{Sym}(R) \to \mathrm{Sym}(S))$
- $R \approx^3 S \to (\mathrm{Trans}(R) \to \mathrm{Trans}(S))$

4 Containment of crisp theories

We define a notion of hereditary crisp set by induction as:

$$\mathrm{HCrisp}^{(2)}(X) \equiv_{\mathrm{df}} \mathrm{Crisp}(X)$$
$$\mathrm{HCrisp}^{(n+1)}(X) \equiv_{\mathrm{df}} \mathrm{Crisp}(X) \,\& \, (\forall Z \in X)(\mathrm{HCrisp}^{(n)}(Z))$$

Sometimes we write just $\mathrm{HCrisp}(X)$ when the type is known. Notice that $\{X \mid \mathrm{HCrisp}^{(n+1)}(X)\}$ is indeed a fuzzy class of the $(n+1)$-st order.

Since FCT contains the classical theory of classes (for classes which are crisp), we can introduce all concepts which are definable in classical class theory (i.e., in classical simple type theory, or Boolean higher-order logic). The only thing we need to do is adding new predicate and functional symbols of the appropriate sorts and axioms saying that all predicates and functions appearing in the theory are crisp.

Definition 4.1 *Let $\mathbf{\Gamma}$ be a higher-order language and T a $\mathbf{\Gamma}$-theory. We define the language* $\mathrm{FCT}(\mathbf{\Gamma})$ *as the language of* FCT *extended by $\mathbf{\Gamma}$. We define translation $'$ of $\mathbf{\Gamma}$-formulas and terms into* $\mathrm{FCT}(\mathbf{\Gamma})$*-formulas and terms in the following way:*

- $t' = t$ *for each term* t
- $\varphi' = \varphi$ *for each atomic formula* φ
- $(\varphi \diamond \psi)' = \varphi' \diamond \psi'$ *for* $\diamond \in \{\&, \rightarrow, \wedge\}$
- $((\forall X)(\varphi))' = (\forall X)(\mathrm{HCrisp}(X) \rightarrow \varphi')$
- $((\exists X)(\varphi))' = (\exists X)(\mathrm{HCrisp}(X)\ \&\ \varphi')$

We define the theory $\mathrm{FCT}(T)$ *in the language* $\mathrm{FCT}(\mathbf{\Gamma})$ *as the theory with the following axioms:*

- *The axioms of* FCT
- *The axioms of the form* φ' *for each* $\varphi \in T$
- $\mathrm{HCrisp}(Q)$ *for each predicate symbol* $Q \in \mathbf{\Gamma}$

The proof of the following lemma is almost straightforward.

Lemma 4.2 *Let $\mathbf{\Gamma}$ be a higher-order language, T a $\mathbf{\Gamma}$-theory, $\mathbf{L}$ an $\mathrm{MTL}_\triangle$-algebra. If $\mathbf{M}$ is an $\mathbf{L}$-model of* $\mathrm{FCT}(T)$*, then the classical model $\mathbf{M}^c$ in the language $\mathbf{\Gamma}$ with the domain M and $S_{\mathbf{M}^c} = S_{\mathbf{M}}$ for each $S \in \mathbf{\Gamma}$, is a model (in the sense of classical Henkin style higher-order logic) of the theory T. Vice versa, for each classical model $\mathbf{M}$ of T there is an $\mathbf{L}$-model $\mathbf{N}$ of* $\mathrm{FCT}(T)$ *such that $\mathbf{N}^c$ is isomorphic to $\mathbf{M}$.*

Therefore, $T \vdash \varphi$ iff $\mathrm{FCT}(T) \vdash \varphi'$*, for any $\mathbf{\Gamma}$-formula φ (where the first provability is in classical Henkin style higher-order logic and the second in* FCT*).*

Example 4.3 *Let τ be a constant for a class of classes and T the theory with the axioms:*

- $\mathrm{HCrisp}(\tau)$
- $\emptyset \in \tau$
- $\mathrm{V} \in \tau$
- $(\forall \mathcal{X})(\mathrm{HCrisp}(\mathcal{X}) \rightarrow (\mathcal{X} \subseteq \tau \rightarrow \bigcup \mathcal{X} \in \tau))$
- $(\forall X, Y)(\mathrm{HCrisp}(X)\ \&\ \mathrm{HCrisp}(Y) \rightarrow (X \in \tau\ \&\ Y \in \tau \rightarrow X \cap Y \in \tau))$

Then in each $\mathbf{L}$-model of the theory T, the constant τ is represented by a classical topology on the universe of objects.

A Fuzzy Class Theory

In this section, we present an overview of Fuzzy Class Theory (FCT) in order to provide the reader with the necessary background. Note that this is only a brief introduction to the most basic concepts of FCT with the aim to keep the paper self-contained. Readers who want to understand all details should not expect to find all necessary material in this paper. Instead, they are referred to the freely available primer [2].

In the first paper [1], FCT was based on the logic ŁΠ [6]. In this paper, we use the logic $\mathrm{MTL}_\triangle$; obviously all definitions and basic results of [1] can be transferred from ŁΠ to $\mathrm{MTL}_\triangle$. For an introduction to $\mathrm{MTL}_\triangle$, see [5]; a more detailed treatment on first-order $\mathrm{MTL}_\triangle$ with crisp equality can be found in [7]).

Definition A.1 *Fuzzy Class Theory* (over $\mathrm{MTL}_\triangle$) is a theory over multi-sorted first-order logic $\mathrm{MTL}_\triangle$ with crisp equality. We have sorts for individuals of the zeroth order (i.e., atomic objects) denoted by lowercase variables $a, b, c, x, y, z, \ldots$; individuals of the first order (i.e., fuzzy classes) denoted by uppercase variables $A, B, X, Y, \ldots$; individuals of the second order (i.e., fuzzy classes of fuzzy classes) denoted by calligraphic variables $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y}, \ldots$; the individuals of the n-th order denoted by $X^{(n)}$. Individuals $\xi_1, \ldots, \xi_k$ of each order can form k-tuples (for any $k \geq 0$), denoted by $\langle \xi_1, \ldots, \xi_k \rangle$; tuples are governed by the usual axioms known from classical mathematics (e.g., tuples equal if and only if their respective constituents equal). Furthermore, for each variable x of any order n and for each formula φ there is a class term $\{x \mid \varphi\}$ of order $n+1$.

Besides the logical predicate of identity, the only primitive predicate is the membership predicate $\in$ between successive sorts (i.e., between individuals of the n-th order and individuals of the $(n+1)$-st order, for any n). The axioms for $\in$ are the following (for variables of all orders):

($\in$1) $y \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(y)$, for each formula φ (comprehension axioms)

($\in$2) $(\forall x)\triangle(x \in A \leftrightarrow x \in B) \to A = B$ (extensionality)

Moreover, we use all axioms and deduction rules of first-order $\mathrm{MTL}_\triangle$. Theorems, theories, proofs, etc., can be defined completely analogously.

Observation A.2 Since the language of FCT is the same at each order, defined symbols of any order can be shifted to all higher orders as well. Since furthermore the axioms of FCT have the same form at each order, all theorems on FCT-definable notions are preserved by uniform upward order-shifts.

Convention A.3 For better readability, let us make the following conventions:

- We use $(\forall x \in A)\varphi$, $(\exists x \in A)\varphi$ as abbreviations for $(\forall x)(x \in A \to \varphi)$ and $(\exists x)(x \in A \,\&\, \varphi)$, respectively.
- We use the notation $\{x \in A \mid \varphi\}$ as abbreviation for $\{x \mid x \in A \,\&\, \varphi\}$.
- We use $\{\langle x_1, \ldots, x_k\rangle \mid \varphi\}$ as abbreviation for $\{x \mid (\exists x_1)\ldots(\exists x_k)(x = \langle x_1, \ldots, x_k\rangle \,\&\, \varphi)\}$.
- The formulae $\varphi \,\&\, \ldots \,\&\, \varphi$ (n times) are abbreviated φ^n; instead of $(x \in A)^n$, we can write $x \in^n A$ (analogously for other predicates).
- Furthermore, $x \notin A$ is shorthand for $\neg(x \in A)$; analogously for other binary predicates.
- We use Ax and $Rx_1 \ldots x_n$ synonymously for $x \in A$ and $\langle x_1, \ldots, x_n\rangle \in R$, respectively.
- A chain of implications of the form $\varphi_1 \to \varphi_2$, $\varphi_2 \to \varphi_3, \ldots, \varphi_{n-1} \to \varphi_n$ will for short be written as $\varphi_1 \longrightarrow \varphi_2 \longrightarrow \cdots \longrightarrow \varphi_n$; analogously for the equivalence connective.

Definition A.4 In FCT, we define the following elementary fuzzy set operations:

$$\begin{array}{rll} \emptyset & =_{\mathrm{df}} & \{x \mid 0\} \\ \mathrm{V} & =_{\mathrm{df}} & \{x \mid 1\} \\ \{a\} & =_{\mathrm{df}} & \{x \mid x = a\} \\ \mathrm{Ker}(A) & =_{\mathrm{df}} & \{x \mid \triangle(x \in A)\} \\ A \cap B & =_{\mathrm{df}} & \{x \mid x \in A \,\&\, x \in B\} \\ A \sqcap B & =_{\mathrm{df}} & \{x \mid x \in A \wedge x \in B\} \\ A \sqcup B & =_{\mathrm{df}} & \{x \mid x \in A \vee x \in B\} \end{array}$$

Definition A.5 In FCT, we define basic properties of fuzzy relations as follows:

$$\begin{array}{rll} \mathrm{Refl}(R) & \equiv_{\mathrm{df}} & (\forall x)Rxx \\ \mathrm{Sym}(R) & \equiv_{\mathrm{df}} & (\forall x, y)(Rxy \to Ryx) \\ \mathrm{Trans}(R) & \equiv_{\mathrm{df}} & (\forall x, y, z)(Rxy \,\&\, Ryz \to Rxz) \end{array}$$

Definition A.6 Further we define in FCT the following elementary relations between fuzzy sets:

$$\begin{array}{rll} \mathrm{Crisp}(A) & \equiv_{\mathrm{df}} & (\forall x)\triangle(x \in A \vee x \notin A) \\ A \subseteq B & \equiv_{\mathrm{df}} & (\forall x)(x \in A \to x \in B) \\ A \approx B & \equiv_{\mathrm{df}} & (\forall x)(x \in A \leftrightarrow x \in B) \\ A \approx^{m,n} B & \equiv_{\mathrm{df}} & (A \subseteq^n B) \,\&\, (B \subseteq^m A) \end{array}$$

Let us recall the standard conventions:

$$A \approx^m B \equiv_{\mathrm{df}} A \approx^{m,m} B \qquad A \approx B \equiv_{\mathrm{df}} A \approx^1 B$$

Definition A.7 The union and intersection of a class of classes are functions defined as

$$\begin{array}{rll} \bigcup \mathcal{A} & =_{\mathrm{df}} & \{x \mid (\exists A \in \mathcal{A})(x \in A)\} \\ \bigcap \mathcal{A} & =_{\mathrm{df}} & \{x \mid (\forall A \in \mathcal{A})(x \in A)\} \end{array}$$

The models of FCT are systems (closed under definable operations) of fuzzy sets (and fuzzy relations) of all orders over some crisp universe U, where the membership functions of fuzzy subsets take values in some $\mathrm{MTL}_\triangle$-chain. Intended models are those which contain *all* fuzzy subsets and fuzzy relations over U (of all orders); we call such models *full*. Models in which moreover the $\mathrm{MTL}_\triangle$-chain is standard (i.e., given by a left-continuous t-norm on the unit interval $[0, 1]$) correspond to Zadeh's [10] original notion of fuzzy set; therefore we call them *Zadeh models*.

FCT is sound with respect to Zadeh (or full) models; thus, whatever we prove in FCT is true about real-valued (or $\mathbf{L}$-valued for any $\mathrm{MTL}_\triangle$-chain $\mathbf{L}$) fuzzy sets and relations. Although the theory of Zadeh models is not *completely* axiomatizable, the axiomatic system of FCT approximates it very well: the comprehension axioms ensure the existence of (at least) all fuzzy sets which are *definable* (by a formula of FCT), and the axioms of extensionality ensure that fuzzy sets are determined by their membership functions. This axiomatization is sufficient for almost all practical purposes; it can be characterized as *simple type theory over fuzzy logic* (cf. [9]) or *Henkin-style higher-order fuzzy logic.*

Acknowledgement

The work of both authors was partly supported by the grant No. B100300502 of the Grant Agency of the Academy of Sciences of the Czech Republic and partly by the Institutional Research Plan AV0Z10300504.

References

[1] Libor Běhounek and Petr Cintula. Fuzzy class theory. *Fuzzy Sets and Systems*, 154(1):34–55, 2005.

[2] Libor Běhounek and Petr Cintula. Fuzzy Class Theory: A primer v1.0. Technical Report V-939, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 2006. Available at www.cs.cas.cz/research/library/reports_900.shtml.

[3] Libor Běhounek and Petr Cintula. Features of mathematical theories in formal fuzzy logic. To appear in proceeding of IFSA 2007, available at www.cs.cas.cz/hp, 2007.

[4] Radim Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*, volume 20 of *IFSR International Series on Systems Science and Engineering*. Kluwer Academic/Plenum Press, New York, 2002.

[5] Francesc Esteva and Lluís Godo. Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124(3):271–288, 2001.

[6] Francesc Esteva, Lluís Godo, and Franco Montagna. The ŁΠ and ŁΠ$\frac{1}{2}$ logics: Two complete fuzzy systems joining Łukasiewicz and product logics. *Archive for Mathematical Logic*, 40(1):39–67, 2001.

[7] Petr Hájek and Petr Cintula. Triangular norm predicate fuzzy logics. To appear in Proceedings of Linz Seminar 2005, 2007.

[8] Franco Montagna. Storage operators and multiplicative quantifiers in many-valued logics. *Journal of Logic and Computation*, 14(2):299–322, 2004.

[9] Vilém Novák. On fuzzy type theory. *Fuzzy Sets and Systems*, 149(2):235–273, 2004.

[10] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

Interior-Based Topology in Fuzzy Class Theory

Libor Běhounek*
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07 Prague,
Czech Republic
behounek@cs.cas.cz

Tomáš Kroupa†
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Prague,
Czech Republic
kroupa@utia.cas.cz

Abstract

Fuzzy topology based on interior operators is studied in the fully graded framework of Fuzzy Class Theory. Its relation to graded notions of fuzzy topology given by open sets and neighborhoods is shown.

Keywords: Fuzzy topology, Fuzzy Class Theory, interior operator, neighborhood.

1 Introduction

Fuzzy topology is a discipline of fuzzy mathematics developed since the beginning of the theory of fuzzy sets [13, 16, 21, 20, 22, 19]. Besides established approaches to fuzzy topology (categorial, lattice-valued, etc.), recent advances in metamathematics of fuzzy logic have enabled an approach to fuzzy topology based on formal fuzzy logic. The framework of higher-order fuzzy logic, also known as Fuzzy Class Theory [4], is especially suitable for fuzzy topology, as it easily accommodates fuzzy sets of fuzzy sets (of arbitrary orders), which are constantly encountered in fuzzy topology.

In classical mathematics, topology can be defined in several equivalent ways: by a system of open (closed) sets, by a system of neighborhoods, or by an interior (closure) operator. These definitions, however, are no longer equivalent in fuzzy logic. Notions of fuzzy topology given by open sets and neighborhoods have been investigated in the framework of Fuzzy Class Theory in [9]. In the present paper we focus on fuzzy topology given by interior operators. Unlike the authors of previous studies of interior and closure operators (e.g., [15, 10, 11]), we work in the fully graded and formal framework of Fuzzy Class Theory, following the methodology of [6]. This approach yields a specific kind of results [8], incomparable to those obtained in traditional fuzzy mathematics: they are on the one hand more general (namely fully graded, i.e., admitting partially valid assumptions), while on the other hand limited to the scope of applicability of deductive fuzzy logic [2].

*The work of the first author was supported by grant No. B100300502 of GA AV ČR and the Institutional Research Plan AV0Z10300504.

†The work of the second author was supported by grant No. B100300502 of GA AV ČR and grant No. 1M0572 of MŠMT ČR.

2 Preliminaries

Fuzzy Class Theory FCT, introduced in [4], is an axiomatization of Zadeh's notion of fuzzy set in formal fuzzy logic. We use its variant defined over $\mathrm{MTL}_\triangle$ [14], the logic of all left-continuous t-norms, which is arguably [2] the weakest fuzzy logic with good inferential properties for fully graded fuzzy mathematics in the framework of formal fuzzy logic [6].

We assume the reader's familiarity with $\mathrm{MTL}_\triangle$; for details on this logic see [14]. Here we only recapitulate its standard real-valued semantics:

$\&$	...	a left-continuous t-norm $*$
$\to$	...	the residuum $\Rightarrow$ of $*$, defined as $x \Rightarrow y =_{\mathrm{df}} \sup\{z \mid z * x \le y\}$
$\wedge, \vee$	...	min, max
$\neg$	...	$x \Rightarrow 0$
$\leftrightarrow$	...	the bi-residuum: $\min(x \Rightarrow y, y \Rightarrow x)$
$\triangle$	...	$\triangle x = 1 - \mathrm{sgn}(1 - x)$
$\forall, \exists$	...	inf, sup

Definition 2.1 Fuzzy Class Theory FCT *is a formal theory over multi-sorted first-order fuzzy logic (in this paper,* $\mathrm{MTL}_\triangle$*), with the sorts of variables for*

- *Atomic objects (lowercase letters* $x, y, \dots$*)*
- *Fuzzy classes of atomic objects (uppercase letters* $A, B, \dots$*)*
- *Fuzzy classes of fuzzy classes of atomic objects (Greek letters* $\tau, \sigma, \dots$*)*

- *Fuzzy classes of the third order (in this paper denoted by sans serif letters* $\mathsf{A}, \mathsf{B}, \mathsf{a}, \mathsf{b}, \dots$*)*
- *Etc., in general for fuzzy classes of the n-th order* $(X^{(n)}, Y^{(n)}, \dots)$

Besides the crisp identity predicate $=$*, the language of* FCT *contains:*

- *The membership predicate* $\in$ *between objects of successive sorts*
- *Class terms* $\{x \mid \varphi\}$ *of order* $n+1$*, for any variable* x *of any order* n *and any formula* φ
- *Symbols* $\langle x_1, \dots, x_k \rangle$ *for* k*-tuples of individuals* $x_1, \dots, x_k$ *of any order*

FCT *has the following axioms (for all formulae* φ *and variables of all orders):*

- *The logical axioms of multi-sorted first-order logic* $\mathrm{MTL}_\triangle$
- *The axioms of crisp identity:*

$$x = x$$
$$x = y \to (\varphi(x) \to \varphi(y))$$
$$\langle x_1, \dots, x_k \rangle = \langle y_1, \dots, y_k \rangle \to x_i = y_i$$

- *The comprehension axioms:*

$$y \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(y)$$

- *The extensionality axioms:*

$$(\forall x) \triangle (x \in A \leftrightarrow x \in B) \to A = B$$

Note that in FCT, fuzzy sets are rendered as a *primitive notion* rather than modeled by membership functions. In order to capture this distinction, fuzzy sets are in FCT called *fuzzy classes;* the name *fuzzy set* is reserved for membership functions in the models of the theory.

The models of FCT are systems of fuzzy sets (and fuzzy relations) of all orders over a crisp universe of discourse, with truth degrees taking values in an $\mathrm{MTL}_\triangle$-chain $\mathbf{L}$ (e.g., the interval $[0,1]$ equipped with a left-continuous t-norm); thus all theorems on fuzzy classes provable in FCT are true statements about $\mathbf{L}$-valued fuzzy sets. Notice however that the theorems of FCT have to be derived from its axioms by the rules of the fuzzy logic $\mathrm{MTL}_\triangle$ rather than classical Boolean logic. For details on proving theorems of FCT see [7] or [5].

In formulae of FCT we employ usual abbreviations known from classical mathematics or traditional fuzzy mathematics, including those listed in Table 1. Usual rules of precedence apply to the connectives of $\mathrm{MTL}_\triangle$. Furthermore we define standard derived notions of FCT, summarized in Table 2, for all orders of fuzzy classes.

Table 1: Abbreviations used in the formulae of FCT

Ax	$\equiv_{\mathrm{df}}$	$x \in A$
$x_1 \dots x_k$	$=_{\mathrm{df}}$	$\langle x_1, \dots, x_k \rangle$
$x \notin A$	$\equiv_{\mathrm{df}}$	$\neg(x \in A)$
$(\forall x \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\forall x)(x \in A \to \varphi)$
$(\exists x \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\exists x)(x \in A \,\&\, \varphi)$
$(\forall x_1, \dots, x_k \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\forall x_1 \in A) \dots (\forall x_k \in A)\varphi$
$(\exists x_1, \dots, x_k \in A)\varphi$	$\equiv_{\mathrm{df}}$	$(\exists x_1 \in A) \dots (\exists x_k \in A)\varphi$
$\{x \in A \mid \varphi\}$	$=_{\mathrm{df}}$	$\{x \mid x \in A \,\&\, \varphi\}$
$\{t(x_1, \dots, x_k) \mid \varphi\}$	$=_{\mathrm{df}}$	$\{z \mid z = t(x_1, \dots, x_k) \,\&\, \varphi\}$
$y = F(x)$	$\equiv_{\mathrm{df}}$	Fxy, if $\triangle \operatorname{Fnc} F$ (see Tab. 2) is proved or assumed
φ^n	$\equiv_{\mathrm{df}}$	$\varphi \,\&\, \dots \,\&\, \varphi$ (n times)
$\varphi^\triangle$	$\equiv_{\mathrm{df}}$	$\triangle\varphi$

Fuzzy counterparts of classical mathematical notions are in the present paper defined following the methodology sketched in [18, §5] and further elaborated in [4, §7], namely by choosing a suitable formula that expresses the classical definitions and re-interpreting it in fuzzy logic.

A distinguishing feature of FCT is that not only the membership predicate $\in$, but all defined notions are in general fuzzy (unless they are defined as provably crisp). FCT thus presents a fully graded approach to fuzzy mathematics. The importance of full gradedness in fuzzy mathematics is explained in [7, 3, 1]: its main merit lies in that it allows inferring relevant information even when a property of fuzzy sets is not fully satisfied. Fuzzy topology has a long tradition of attempting full gradedness, cf. graded definitions and theorems in [19, 22].

3 Open Fuzzy Topology

In classical mathematics, topology introduced by means of open sets is given by a crisp system τ of crisp subsets of a ground set V, where τ is required to satisfy certain conditions (closedness under $\bigcup, \cap, \emptyset, \mathrm{V}$, and possibly further properties, e.g., separation axioms). Generalization by admitting fuzzy subsets leads in FCT to regarding open fuzzy topology as a (possibly fuzzy) class of (possibly fuzzy) subclasses of the ground class V, i.e., a fuzzy class τ of the second order.[1]

[1] We keep the ground class crisp to avoid problems with quantification relativized to a fuzzy domain; generalization to fuzzy topological spaces with fuzzy universes is a topic

Table 2: Defined notions of FCT

$\emptyset$	$=_{\mathrm{df}}$	$\{x \mid 0\}$ … empty class
V	$=_{\mathrm{df}}$	$\{x \mid 1\}$ … universal class
$\mathrm{Ker}\, A$	$=_{\mathrm{df}}$	$\{x \mid \triangle Ax\}$ … kernel
αA	$=_{\mathrm{df}}$	$\{x \mid \alpha \,\&\, Ax\}$ … α-resize
$-A$	$=_{\mathrm{df}}$	$\{x \mid \neg Ax\}$ … complement
$A \cap B$	$=_{\mathrm{df}}$	$\{x \mid Ax \,\&\, Bx\}$ … (strong) intersection
$A \sqcap B$	$=_{\mathrm{df}}$	$\{x \mid Ax \wedge Bx\}$ … min-intersection
$A \cup B$	$=_{\mathrm{df}}$	$\{x \mid Ax \veebar Bx\}$ … (strong) union
$A \times B$	$=_{\mathrm{df}}$	$\{xy \mid Ax \,\&\, By\}$ … Cartesian product
$\mathrm{Rng}\, R$	$=_{\mathrm{df}}$	$\{y \mid (\exists x)Rxy\}$ … range
$\bigcup \tau$	$=_{\mathrm{df}}$	$\{x \mid (\exists A \in \tau)(x \in A)\}$ … class union
$\bigcap \tau$	$=_{\mathrm{df}}$	$\{x \mid (\forall A \in \tau)(x \in A)\}$ … class intersec.
$\mathrm{Pow}\, A$	$=_{\mathrm{df}}$	$\{X \mid X \subseteq A\}$ … power class
$\mathrm{Hgt}\, A$	$\equiv_{\mathrm{df}}$	$(\exists x)Ax$ … height
$\mathrm{Plt}\, A$	$\equiv_{\mathrm{df}}$	$(\forall x)Ax$ … plinth
$\mathrm{Crisp}\, A$	$\equiv_{\mathrm{df}}$	$(\forall x)\triangle(Ax \vee \neg Ax)$ … crispness
$\mathrm{Refl}\, R$	$\equiv_{\mathrm{df}}$	$(\forall x)Rxx$ … reflexivity
$\mathrm{Trans}\, R$	$\equiv_{\mathrm{df}}$	$(\forall x, y, z)(Rxy \,\&\, Ryz \to Rxz)$ … transitivity
$\mathrm{Preord}\, R$	$\equiv_{\mathrm{df}}$	$\mathrm{Refl}\, R \,\&\, \mathrm{Trans}\, R$ … preorder
$\mathrm{Fnc}\, R$	$\equiv_{\mathrm{df}}$	$(\forall x, y, y')(Rxy \,\&\, Rxy' \to y = y')$ … functionality
$A \subseteq B$	$\equiv_{\mathrm{df}}$	$(\forall x)(Ax \to Bx)$ … inclusion
$A \approx B$	$\equiv_{\mathrm{df}}$	$(A \subseteq B) \wedge (B \subseteq A)$ … weak bi-incl.
$A \cong B$	$\equiv_{\mathrm{df}}$	$(A \subseteq B) \,\&\, (B \subseteq A)$ … strong bi-incl.

When investigating open fuzzy topologies, we are interested in such τ that satisfy analogous (but fuzzified) closure conditions as in classical topology. These are given by the following predicates that express the (degree of) closedness of τ under $\bigcup$ and $\cap$:

$$\mathsf{ic}(\tau) \equiv_{\mathrm{df}} (\forall A, B \in \tau)(A \cap B \in \tau)$$
$$\mathsf{Uc}(\tau) \equiv_{\mathrm{df}} (\forall \sigma \subseteq \tau)\big(\bigcup \sigma \in \tau\big)$$

These conditions (plus $\emptyset \in \tau$ and $\mathrm{V} \in \tau$) can be regarded as characteristic of open fuzzy topology. However, when studying open fuzzy topologies, we do *not* in general require that these axioms be satisfied as in classical topology. This is because they are (like all formulae of FCT) interpreted in many-valued logic; thus they need not be simply true or false, but are always true *to some degree.* By restricting our attention just to the systems that fully satisfy the above axioms, we would completely disregard systems that satisfy them to a degree of, e.g., 0.9999, even though graded theorems of FCT can provide us with useful information about such systems. Therefore we study *all* systems τ, no matter to which degree they satisfy the above axioms. Similarly we proceed also in fuzzification of other definitions of fuzzy topology in the following sections.

It turns out [9] that besides the predicate Uc, also predicates of the following forms are often met in the study of fuzzy topology (for $m, n \geq 1$):

$$\mathsf{Uc}_{m,n}(\tau) \equiv_{\mathrm{df}} (\forall \sigma)\big(\sigma \subseteq^m \tau \to \bigcup(\sigma \cap .\overset{n}{.}. \cap \sigma) \in \tau\big)$$

Note that because $\varphi \,\&\, \varphi$ is not generally equivalent to φ in $\mathrm{MTL}_\triangle$ (nor in stronger fuzzy logics except for Gödel fuzzy logic or stronger), $\sigma \cap .\overset{n}{.}. \cap \sigma$ does not generally equal σ (only $\sigma \cap .\overset{n}{.}. \cap \sigma \subseteq \sigma$ holds for all σ). Similarly $(\sigma \subseteq \tau)^m$ is in general stronger than simple $\sigma \subseteq \tau$ if $m > 1$. Recall that the larger m, the stronger φ^m; informally φ^m can be understood as m-times stressed φ (consult, e.g., [7] for the role of multiple conjunction in formal proofs). Thus, like Uc, the predicate $\mathsf{Uc}_{m,n}$ expresses the closedness of τ under a certain operation similar to the union of subsystems, only the condition of what counts as a subsystem is strengthened by m and the union itself is strengthened by n.

By convention, we also admit the value "$\triangle$" for either m or n or both (cf. the last line of Table 1). Then, e.g., $\mathsf{Uc}_{\triangle,1}(\tau)$ expresses the closedness of τ under the unions of *crisp* subsystems of τ, while $\mathsf{Uc}_{1,\triangle}(\tau)$ expresses the closedness of τ under the unions of *kernels* of subsystems of τ (i.e., only full members of the subsystem enter the union).

For convenience, we define a predicate that puts the properties monitored in open fuzzy topologies together. Since each of the properties can appear with varying multiplicity in theorems, we have to add further indices that parameterize the multiplicity of each of the conditions:

Definition 3.1 *We define the predicate indicating the degree to which τ is an (e, v, i, u, m, n)–*open fuzzy topology *as*

$$\mathsf{OTop}^{e,v,i,u}_{m,n}(\tau) \equiv_{\mathrm{df}} (\emptyset \in \tau)^e \,\&\, (\mathrm{V} \in \tau)^v \,\&\, \mathsf{ic}^i(\tau) \,\&\, \mathsf{Uc}^u_{m,n}(\tau)$$

For the sake of brevity, we drop the subscripts if both equal 1*, and similarly for the superscripts.*

The properties of open fuzzy topologies have been studied in [9]. Since in this paper we are mainly interested in the interior operator, we repeat here the definition of the interior operator induced by an open fuzzy topology and list its basic properties.

Definition 3.2 *Given a class of classes τ, we define the* interior *of a class A in τ as*

$$\mathrm{Int}_\tau(A) =_{\mathrm{df}} \bigcup\{B \in \tau \mid B \subseteq A\}$$

for future work. Since in this paper we always work within a single topological space, we can identify the ground class with the universal class V.

Proposition 3.3 *It is provable in* FCT*:*

(i) $\mathrm{Int}_\tau(A) \subseteq A$

(ii) $A \in \tau \to \mathrm{Int}_\tau(A) \approxeq A$

(iii) $A \subseteq B \to \mathrm{Int}_\tau(A) \subseteq \mathrm{Int}_\tau(B)$

(iv) $\mathrm{Int}_\tau(A \sqcap B) \subseteq \mathrm{Int}_\tau(A) \sqcap \mathrm{Int}_\tau(B)$

Proposition 3.4 *It is provable in* FCT*:*

(i) $\mathrm{V} \in \tau \to \mathrm{Int}_\tau(\mathrm{V}) \approxeq \mathrm{V}$

(ii) $\mathsf{Uc}(\tau) \to \mathrm{Int}_\tau(\mathrm{Int}_\tau(A)) \approxeq \mathrm{Int}_\tau(A)$

(iii) $\mathsf{ic}(\tau) \to \mathrm{Int}_\tau(A) \cap \mathrm{Int}_\tau(B) \subseteq \mathrm{Int}_\tau(A \cap B)$

Propositions 3.3 and 3.4 show that the interior operator generated by an open fuzzy topology τ satisfies properties expected from an interior operator—unconditionally in Proposition 3.3, and to a guaranteed degree (depending on the degree to which τ satisfies the conditions required from open fuzzy topologies) in Proposition 3.4.

If the antecedent conditions in Propositions 3.3 and 3.4 are fulfilled to the full degree, so are the conclusions. In particular, we have the following corollary:

Corollary 3.5 FCT *proves:*

(i) $\triangle(A \in \tau) \to \mathrm{Int}_\tau(A) = A$

(ii) $\triangle\, \mathsf{Uc}(\tau) \to \mathrm{Int}_\tau(\mathrm{Int}_\tau(A)) = \mathrm{Int}_\tau(A)$

In words, whenever a fuzzy class A is fully in τ, it equals its interior (no matter what conditions τ does or does not satisfy to which degree). Similarly, if τ is fully closed under fuzzy unions, interiors are stable in τ.

It will further be seen in Section 5 that an open fuzzy topology can vice versa be recovered from a primitive interior operator, under conditions similar to those above.

4 Neighborhood Fuzzy Topology

In classical mathematics, topology can also be introduced by assigning a system of neighborhoods to each point of a ground set V. Such a neighborhood system can be represented by a relation Nb between elements and subsets of V, where $\mathrm{Nb}(x, A)$ represents the fact that $A \subseteq \mathrm{V}$ is a neighborhood of $x \in \mathrm{V}$. The notion of neighborhood-based fuzzy topology, obtained by fuzzification of the classical notion in FCT, just allows the relation Nb and the class A in $\mathrm{Nb}(x, A)$ to be fuzzy.[2] Thus in FCT, neighborhood fuzzy topologies will be second-order relations between atomic objects and first-order classes, i.e., classes Nb such that $\triangle(\mathrm{Nb} \subseteq \mathrm{V} \times \mathrm{Ker}\,\mathrm{Pow}\,\mathrm{V})$.

Neighborhood systems are in classical topology required to satisfy certain conditions. Fuzzified versions of these conditions will be of interest in neighborhood-based fuzzy topology, too:

Definition 4.1 *Let* Nb *be a second-order class such that* $\triangle(\mathrm{Nb} \subseteq \mathrm{V} \times \mathrm{Ker}\,\mathrm{Pow}(\mathrm{V}))$*. Then we define the following predicates:*

$$\begin{aligned}
&\mathsf{N}_1(\mathrm{Nb}) \equiv_{\mathrm{df}} (\forall x)\,\mathrm{Nb}(x, \mathrm{V})\\
&\mathsf{N}_2(\mathrm{Nb}) \equiv_{\mathrm{df}} (\forall x, A)(\mathrm{Nb}(x, A) \to x \in A)\\
&\mathsf{N}_3(\mathrm{Nb}) \equiv_{\mathrm{df}} (\forall x, A, B)(\mathrm{Nb}(x, A)\ \&\ A \subseteq B \to \mathrm{Nb}(x, B))\\
&\mathsf{N}_4(\mathrm{Nb}) \equiv_{\mathrm{df}} (\forall x, A, B)(\mathrm{Nb}(x, A)\ \&\ \mathrm{Nb}(x, B) \to \mathrm{Nb}(x, A \cap B))\\
&\mathsf{N}_5(\mathrm{Nb}) \equiv_{\mathrm{df}} (\forall x, A)(\mathrm{Nb}(x, A) \to (\exists B)(B \subseteq A\ \&\ \mathrm{Nb}(x, B)\ \&\ (\forall y \in B)\,\mathrm{Nb}(y, B))
\end{aligned}$$

For convenience, we aggregate them in the following defined predicate:

Definition 4.2 *We define the predicate indicating the degree to which* Nb *is a* $(k_1, \ldots, k_5)$–neighborhood fuzzy topology *as follows:*

$$\mathsf{NTop}^{k_1,\ldots,k_5}(\mathrm{Nb}) \equiv_{\mathrm{df}} \mathrm{Nb} \subseteq^{\triangle} \mathrm{V} \times \mathrm{Ker}\,\mathrm{Pow}(\mathrm{V})\ \&\ \mathop{\&}_{i=1}^{5} \mathsf{N}_i^{k_i}(\mathrm{Nb})$$

Basic properties of neighborhood fuzzy topologies and their relation to open fuzzy topologies have been summarized in [9]. Here we restrict our attention to their relationship to interior-based topologies. The following definition internalizes in FCT the classical definition of the interior of a class A:

Definition 4.3 *Given a binary predicate* Nb *between elements and classes, we define*

$$\mathrm{Int}_{\mathrm{Nb}}(A) =_{\mathrm{df}} \{x \mid \mathrm{Nb}(x, A)\}$$

The behavior of $\mathrm{Int}_{\mathrm{Nb}}$ w.r.t. Kuratowski's (fuzzified) axioms of interior operators is studied in the following section.

5 Interior Fuzzy Topology

In classical topology, an interior operator is a function Int that assigns to each subset A of a ground set V

[2] We again keep the ground set V crisp for simplicity and identify it with the universal class; see footnote 1.

a set $\mathrm{Int}(A) \subseteq \mathrm{V}$. In FCT we allow both the argument A and the output $\mathrm{Int}(A)$ of the function to be fuzzy.[3] Fuzzy interior operators are thus construed as crisp second-order functions, i.e., classes Int such that $\mathrm{Int} \subseteq^{\triangle} \mathrm{Ker\,Pow(V)} \times \mathrm{Ker\,Pow(V)} \,\&\, \triangle\,\mathrm{Fnc(Int)}$.

The degrees to which Int satisfies (fuzzy versions of) Kuratowski's axioms for interior operators are given by the following predicates:

Definition 5.1 *For second-order classes* Int *such that* $\mathrm{Int} \subseteq^{\triangle} \mathrm{Ker\,Pow(V)} \times \mathrm{Ker\,Pow(V)} \,\&\, \triangle\,\mathrm{Fnc(Int)}$ *we define the following predicates:*

$$\mathsf{K}_1(\mathrm{Int}) \equiv_{\mathrm{df}} \mathrm{Int(V)} \approxeq \mathrm{V}$$
$$\mathsf{K}_2(\mathrm{Int}) \equiv_{\mathrm{df}} (\forall A)(\mathrm{Int}(A) \subseteq A)$$
$$\mathsf{K}_3(\mathrm{Int}) \equiv_{\mathrm{df}} (\forall A)(\mathrm{Int}(\mathrm{Int}(A)) \approxeq \mathrm{Int}(A))$$
$$\mathsf{K}_4(\mathrm{Int}) \equiv_{\mathrm{df}} (\forall A, B)(\mathrm{Int}(A) \cap \mathrm{Int}(B) \subseteq \mathrm{Int}(A \cap B))$$

Unlike in classical topology, in $\mathrm{MTL}_{\triangle}$ these conditions do not imply the monotonicity of Int. Therefore we define also the following predicates:

$$\mathsf{Mon}(\mathrm{Int}) \equiv_{\mathrm{df}} (\forall A, B)(A \subseteq B \to \mathrm{Int}(A) \subseteq \mathrm{Int}(B))$$
$$\mathsf{K}_5(\mathrm{Int}) \equiv_{\mathrm{df}} (\forall A, B)(\mathrm{Int}(A \sqcap B) \subseteq \mathrm{Int}(A) \sqcap \mathrm{Int}(B))$$

Although Mon and K_5 are not equivalent, the following relationships between them hold:

Proposition 5.2 *It is provable in* FCT*:*

1. $\mathsf{K}_5(\mathrm{Int}) \to \mathsf{Mon}(\mathrm{Int})$
2. $\mathsf{Mon}^2(\mathrm{Int}) \to \mathsf{K}_5(\mathrm{Int})$
3. $\triangle\,\mathsf{K}_5(\mathrm{Int}) \leftrightarrow \triangle\,\mathsf{Mon}(\mathrm{Int})$

For convenience, we gather the conditions K_1–K_5 into one predicate ITop:[4]

Definition 5.3 *We define the notion of* $(k_1, \ldots, k_5)$–interior fuzzy topology *by the predicate*

$$\mathsf{ITop}^{k_1,\ldots,k_5}(\mathrm{Int}) \equiv_{\mathrm{df}}$$
$$\mathrm{Int} \subseteq^{\triangle} \mathrm{Ker\,Pow(V)} \times \mathrm{Ker\,Pow(V)} \,\&\, \triangle\,\mathrm{Fnc(Int)} \,\&\, \mathop{\&}_{i=1}^{5} \mathsf{K}_i^{k_i}(\mathrm{Int})$$

[3] Again we keep V crisp and identify it with the universal class as in footnote 1. The function Int itself is conceived as crisp as well, to keep the correspondence to logical functions of [17]; if needed, it can be fuzzified by a similarity relation as in [1].

[4] It is not much important whether we take K_5 or Mon in the definition of ITop, as Proposition 5.2 "translates" between the two variants.

Open classes can be defined by means of the interior operator as usual:

$$\tau_{\mathrm{Int}} =_{\mathrm{df}} \{A \mid A \subseteq \mathrm{Int}(A)\}$$

The following graded theorem shows that if Int satisfies Kuratowski's axioms to a large degree, then τ_{Int} satisfies the properties of open fuzzy topologies to a large degree, and the interior operator generated by τ_{Int} equals Int to a large degree. Notice, however, that we have only got $\mathsf{OTop}_{2,1}(\tau_{\mathrm{Int}})$ rather than $\mathsf{OTop}(\tau_{\mathrm{Int}})$; in other words, we can only prove that the system of classes open w.r.t. a fuzzy Kuratowski interior operator is closed under unions of families "doubly included" in the system.

Theorem 5.4 FCT *proves:*

$$\mathsf{ITop}^{1,1,1,1,2}(\mathrm{Int}) \to \mathsf{OTop}_{2,1}(\tau_{\mathrm{Int}}) \,\&\, (\forall A)(\mathrm{Int}(A) \approxeq \mathrm{Int}_{\tau_{\mathrm{Int}}}(A))$$

Corollary 5.5 FCT *proves:*

$$\triangle\,\mathsf{ITop}(\mathrm{Int}) \to \triangle\,\mathsf{OTop}_{2,1}(\tau_{\mathrm{Int}}) \,\&\, \mathrm{Int} = \mathrm{Int}_{\tau_{\mathrm{Int}}}$$

Vice versa, interiors in well-behaved open fuzzy topologies are well-behaved fuzzy interior operators:

Theorem 5.6 FCT *proves:*

$$\mathsf{OTop}^{0,1,1,1}(\tau) \to \mathsf{ITop}(\mathrm{Int}_\tau) \,\&\, (\forall A)(A \in \tau \leftrightarrow A \subseteq \mathrm{Int}_\tau(A))$$

Corollary 5.7 FCT *proves:*

$$\triangle\,\mathsf{OTop}(\tau) \to \triangle\,\mathsf{ITop}(\mathrm{Int}_\tau) \,\&\, \tau = \tau_{\mathrm{Int}_\tau}$$

Neighborhoods can also be defined by means of the interior operator as usual:

$$\mathrm{Nb}_{\mathrm{Int}}(x, A) \equiv_{\mathrm{df}} x \in \mathrm{Int}(A)$$

It is immediate that $\mathrm{Nb}_{\mathrm{Int}}$ and $\mathrm{Int}_{\mathrm{Nb}}$ of Definition 4.3 are mutually inverse, i.e.,

$$\mathrm{Int} = \mathrm{Int}_{\mathrm{Nb}_{\mathrm{Int}}}$$
$$\mathrm{Nb} = \mathrm{Nb}_{\mathrm{Int}_{\mathrm{Nb}}}$$

Moreover we have the following correspondence between the predicates ITop and NTop:

Theorem 5.8 FCT *proves:*

1. $\mathsf{ITop}^{1,2,2,1,1}(\mathrm{Int}) \to \mathsf{NTop}(\mathrm{Nb}_{\mathrm{Int}})$
2. $\mathsf{NTop}^{1,3,3,2,1}(\mathrm{Nb}) \to \mathsf{ITop}(\mathrm{Int}_{\mathrm{Nb}})$

As a corollary, we get the perfect match between the conditions ITop and NTop when true to degree 1:

Corollary 5.9 FCT *proves:*

$$\triangle\,\mathsf{ITop}(\mathrm{Int}) \leftrightarrow \triangle\,\mathsf{NTop}(\mathrm{Nb}_{\mathrm{Int}}), \quad \mathrm{Int} = \mathrm{Int}_{\mathrm{Nb}_{\mathrm{Int}}}$$
$$\triangle\,\mathsf{NTop}(\mathrm{Nb}) \leftrightarrow \triangle\,\mathsf{ITop}(\mathrm{Int}_{\mathrm{Nb}}), \quad \mathrm{Nb} = \mathrm{Nb}_{\mathrm{Int}_{\mathrm{Nb}}}$$

We conclude by giving three examples of interior-based fuzzy topology.

Example 5.10 The operation sending a fuzzy class to its kernel is an interior operator that fully satisfies all of Kuratowski's axioms, as FCT proves

$$\begin{aligned}\operatorname{Ker}\mathrm{V} &= \mathrm{V}\\ \operatorname{Ker}A &\subseteq A\\ \operatorname{Ker}\operatorname{Ker}A &= \operatorname{Ker}A\\ \operatorname{Ker}A \cap \operatorname{Ker}B &= \operatorname{Ker}(A \cap B)\\ \operatorname{Ker}(A \sqcap B) &= \operatorname{Ker}A \sqcap \operatorname{Ker}B\end{aligned}$$

by [4, §3.4]. Thus $\triangle\,\mathsf{ITop}(\mathrm{Ker})$; we call it the *kernel fuzzy topology.*

In the kernel fuzzy topology, a class is fully open iff it is crisp: $\triangle(A \in \tau_{\mathrm{Ker}}) \leftrightarrow \operatorname{Crisp}A$. Partially open classes are those whose fuzzy elements only have low membership degrees. Since all crisp classes (including singletons) are open in the kernel fuzzy topology, it is a generalization of the notion of *discrete* crisp topology, with which it coincides in 2-valued models.

Example 5.11 Define the interior of A as $(\operatorname{Plt}A)\mathrm{V}$ (see Table 2 for the definitions of plinth and resize); i.e., $x \in \operatorname{Int}A \equiv_{\mathrm{df}} (\forall y)Ay$. In other words, the membership function of $\operatorname{Int}A$ is constant and all elements belong to $\operatorname{Int}A$ to the degree which is the infimum of the membership function of A. Then it is provable in FCT that $\triangle\,\mathsf{ITop}(\mathrm{Int})$; we call it the *plinth fuzzy topology.*

A class is fully open in the plinth topology iff it is a resize of the universal class. Thus, the plinth fuzzy topology is *stratified* (stratified topologies are defined as those in which all classes $\alpha\mathrm{V}$ are open [21, 19]). Partially open in the plinth topology are such classes whose membership functions have small amplitudes (i.e., the differences between their suprema and infima), as $\tau_{\mathrm{Int}} = \{A \mid \operatorname{Hgt}A \to \operatorname{Plt}A\}$. Since the only crisp open classes in the plinth topology are $\emptyset$ and V, it is a generalization of the notion of *anti-discrete* crisp topology (with which it coincides in 2-valued models).

Example 5.12 In [12], an operation of the *opening* of a fuzzy set under a fuzzy relation has been introduced. In [3] the definition has been generalized to the graded framework of FCT and its graded properties have been investigated. The definition can be rephrased as follows:

$$\operatorname{Int}_R(A) =_{\mathrm{df}} \{y \mid (\exists x)(Rxy \,\&\, (\forall z)(Rxz \to Az))\}$$

From results proved in [3] it follows that for any relation R, the operator fully satisfies the conditions K_2, K_3, and K_5. If R is a crisp preorder, then furthermore Int_R fully satisfies K_4. Since $\mathsf{K}_1(\operatorname{Int}_R)$ is equivalent to $\mathrm{V} \subseteq \operatorname{Rng}R$, we get

$$\triangle\operatorname{Preord}R \,\&\, \operatorname{Crisp}R \to \triangle\,\mathsf{ITop}(\operatorname{Int}_R)$$

This result can be generalized to a larger class of fuzzy relations: e.g., instead of crispness, $R = R \cap R$ is sufficient for $\triangle\,\mathsf{K}_4(\operatorname{Int}_R)$ if $\triangle\operatorname{Preord}R$; both conditions can further be relaxed if ITop is not required to degree 1. Furthermore it is shown in [3] that for any R we have $\operatorname{Int}_R = \operatorname{Int}_{\tau_{\operatorname{Int}_R}}$.

References

[1] L. Běhounek. Extensionality in graded properties of fuzzy relations. In *Proceedings of 11th IPMU Conference*, pages 1604–1611, Paris, 2006. Edition EDK.

[2] L. Běhounek. On the difference between traditional and formal fuzzy logic. Submitted, preprint available at www.cs.cas.cz/hp, 2007.

[3] L. Běhounek, U. Bodenhofer, and P. Cintula. Relations in fuzzy class theory: Initial steps. Submitted, preprint available at www.cs.cas.cz/hp, 2007.

[4] L. Běhounek and P. Cintula. Fuzzy class theory. *Fuzzy Sets and Systems*, 154(1):34–55, 2005.

[5] L. Běhounek and P. Cintula. Fuzzy class theory as foundations for fuzzy mathematics. In *Fuzzy Logic, Soft Computing and Computational Intelligence: 11th IFSA World Congress*, volume 2, pages 1233–1238, Beijing, 2005. Tsinghua University Press/Springer.

[6] L. Běhounek and P. Cintula. From fuzzy logic to fuzzy mathematics: A methodological manifesto. *Fuzzy Sets and Systems*, 157(5):642–646, 2006.

[7] L. Běhounek and P. Cintula. Fuzzy Class Theory: A primer v1.0. Technical Report V-939, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 2006. Available at www.cs.cas.cz/research/library/reports_900.shtml.

[8] L. Běhounek and P. Cintula. Features of mathematical theories in formal fuzzy logic. To appear in the proceedings of IFSA 2007, preprint available at www.cs.cas.cz/hp, 2007.

[9] L. Běhounek and T. Kroupa. Topology in Fuzzy Class Theory: Basic notions. To appear in the proceedings of IFSA 2007, preprint available at www.cs.cas.cz/hp.

[10] R. Bělohlávek and T. Funioková. Fuzzy interior operators. *International Journal of General Systems*, 33(4):415–430, 2004.

[11] R. Bělohlávek and T. Funioková and Vilém Vychodil. Fuzzy closure operators with truth stressers. *Logic Journal of the IGPL*, 13:503–513, 2005.

[12] U. Bodenhofer. A unified framework of opening and closure operators with respect to arbitrary fuzzy relations. *Soft Computing*, 7:220–227, 2003.

[13] C. Chang. Fuzzy topological spaces. *Journal of Mathematical Analysis and Applications*, 24:182–190, 1968.

[14] F. Esteva and L. Godo. Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124(3):271–288, 2001.

[15] W. Gähler. The general fuzzy filter approach to fuzzy topology I, II. *Fuzzy Sets and Systems*, 76:205–224, 225–246, 1995.

[16] J. Goguen. The fuzzy Tychonoff theorem. *Journal of Mathematical Analysis and Applications*, 43:734–742, 1973.

[17] P. Hájek. Function symbols in fuzzy logic. In *Proceedings of the East–West Fuzzy Colloquium*, pages 2–8, Zittau/Görlitz, 2000. IPM.

[18] U. Höhle. Fuzzy real numbers as Dedekind cuts with respect to a multiple-valued logic. *Fuzzy Sets and Systems*, 24(3):263–278, 1987.

[19] U. Höhle. *Many Valued Topology and Its Applications*. Kluwer, Boston, MA, 2001.

[20] U. Höhle and A. Šostak. A general theory of fuzzy topological spaces. *Fuzzy Sets and Systems*, 73:131–149, 1995.

[21] R. Lowen. Fuzzy topological spaces and fuzzy compactness. *Journal of Mathematical Analysis and Applications*, 56:623–633, 1976.

[22] M. Ying. Fuzzy topology based on residuated lattice-valued logic. *Acta Mathematica Sinica (English Series)*, 17:89–102, 2001.

Valverde-Style Representation Results in a Graded Framework

Libor Běhounek
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
187 02 Prague 8, Czech Republic
behounek@cs.cas.cz

Ulrich Bodenhofer
Institute of Bioinformatics
Johannes Kepler University Linz
Altenberger Str. 69
A-4040 Linz, Austria
bodenhofer@bioinf.jku.at

Petr Cintula
Institute of Computer Science
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2
187 02 Prague 8, Czech Republic
cintula@cs.cas.cz

Abstract

This paper generalizes the well-known representations of fuzzy preorders and similarities according to Valverde to the graded framework of Fuzzy Class Theory (FCT). The results demonstrate that FCT is a powerful tool and that new results and interesting constructions can be obtained by considering fuzzy relations in the graded framework of FCT.

Keywords: Fuzzy Class Theory, Fuzzy Relations, Graded Properties, MTL Logic.

1 Introduction

This paper aims at generalizing two of the most important and influential theorems in the theory of fuzzy relations—Valverde's representation theorems for fuzzy preorders and similarities [23]:

Consider a fuzzy relation $R\colon U \times U \to [0,1]$. R is a fuzzy preorder with respect to some left-continuous triangular norm $*$ if and only if there exists a family $(f_i)_{i\in I}$ of score functions $U \to [0,1]$ such that R can be represented as (for all $x, y \in U$)

$$R(x,y) = \inf_{i\in I}(f_i(x) \Rightarrow f_i(y)),$$

where $\Rightarrow$ is the residual implication of $*$. Moreover, R is a similarity with respect to $*$ if and only if there exists a family $(f_i)_{i\in I}$ of score functions $U \to [0,1]$ such that R can be represented as (for all $x, y \in U$)

$$R(x,y) = \inf_{i\in I}(f_i(x) \Leftrightarrow f_i(y)),$$

where $\Leftrightarrow$ is the residual bi-implication of $*$. These two results hold for fuzzy preorders and similarities, respectively, but clearly they do not provide us with any insight if the relations fail to fulfill those requirements.

In the early 1990's, Gottwald has introduced what he calls *graded properties of fuzzy relations* [17–19], a framework in which it is possible to deal with partial—*graded*—fulfillment of properties like reflexivity, transitivity, etc. In this framework, it is not only possible to define properties in a graded way, but also to generalize theorems on fuzzy relations in the sense that an assertion about a certain sub-class of fuzzy relations holds to a degree that depends on the degree to which the relations fulfill the necessary properties. Even though these ideas seem obvious and meaningful, Gottwald's approach unfortunately found only little resonance (exceptions are, for instance, [6,21]), mainly because it is not a full-fledged axiomatic framework and is not strictly separating syntax from semantics. For this reason, proofs are complicated and difficult.

With the advent of Fuzzy Class Theory (FCT) [3], a formal axiomatic framework is available in which it is just natural to consider properties of fuzzy relations in a graded manner. Notions are inspired by (and derived from) the corresponding notions of classical mathematics [4]; the syntax of FCT is close to the syntax of classical mathematical theories and the proofs in FCT resemble the proofs of the corresponding classical theorems. Therefore, it is technically easier to handle graded properties of fuzzy relations than in Gottwald's previous works and it is possible to access deeper results than in Gottwald's framework.

This paper is devoted to this advancement, concentrating on Valverde's famous representation theorems for fuzzy preorders and similarities. In the tradition of Cantor [9], Valverde uses score functions to represent relations. As these score functions map to the unit interval, they can also be considered as fuzzy sets, which facilitates a reformulation of these results in FCT.

The paper is organized as follows. After some preliminaries concerning FCT in Section 2, we introduce graded properties in Section 3. Section 4 is devoted to the generalization of Valverde's representation theorem for fuzzy preorders, while Section 5 deals with the

corresponding results for similarities. Note that this paper is an excerpt of a larger manuscript that has been submitted for publication [2]. For background information and proof details, readers are referred to this upcoming article.

2 Preliminaries

We aim this paper at researchers in the theory and applications of fuzzy relations to attract their interest in graded theories of fuzzy relations. In the traditional theory of fuzzy relations, it is not usual to separate formal syntax from semantics as it is the case in FCT. So it may be difficult for some readers who are new to FCT to follow the results. Therefore, we would like to provide the readers basically with a dictionary that improves understanding of the results in this paper and that demonstrates how the results would translate to the traditional language of fuzzy relations. For a more formal introduction to FCT, readers are referred to the appendix of this paper and the freely available primer [5].

FCT strictly distinguishes between its syntax and semantics; that is, we distinguish between a formal syntax of formulae and the fuzzy relations modeling them. This feature has two important consequences: (i) To make this distinction clear, the objects of the formal theory are called *fuzzy classes* and not *fuzzy sets*. The name *fuzzy set* is reserved for membership functions in the *models* of the theory (see Appendix). Nevertheless, the theorems of FCT about fuzzy classes are always valid for fuzzy *sets* and fuzzy relations. Thus, whenever we speak of classes, the reader can always safely substitute usual fuzzy sets for "classes". (ii) FCT screens off direct references to truth values; truth degrees belong to the *semantics* of FCT, rather than to its syntax. Thus, there are *no variables for truth degrees* in the language of FCT. The degree to which an element x belongs to a fuzzy class A is expressed simply by the atomic formula $x \in A$ (which can alternatively be written in a more traditional way as Ax).

The algebraic structure of truth degrees in the semantics of FCT is that of $\mathrm{MTL}_\triangle$-chains [13,20].[1] If the domain of truth values is the unit interval $[0,1]$, $\mathrm{MTL}_\triangle$-chains are characterized as algebras

$$([0,1], *, \Rightarrow, \min, \max, 0, 1, \triangle),$$

where $*$ is a left-continuous t-norm, $\Rightarrow$ is its residual implication, and $\triangle$ is a unary operation defined as

$$\triangle x = \begin{cases} 1 & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

[1] Note that even though in the first paper [3] FCT was based on the logic ŁΠ [14], the logic $\mathrm{MTL}_\triangle$ is sufficient for our present needs.

This means that we can translate the results to the language of fuzzy relations in the following way, where we may specify an arbitrary universe of discourse U and a left-continuous t-norm $*$ (with the residuum $\Rightarrow$):

FCT	**Fuzzy relations**
object variable x	element $x \in U$
(fuzzy) class A	fuzzy set $A \in \mathcal{F}(U)$
2nd-order fuzzy class $\mathcal{A}$	fuzzy set $\mathcal{A} \in \mathcal{F}(\mathcal{F}(U))$
unary predicate	fuzzy subset of U, $\mathcal{F}(U)$, etc.
n-ary predicate	n-ary f. rel. on U^n, $(\mathcal{F}(U))^n$, etc.
strong conjunction &	left-continuous t-norm $*$
implication $\to$	residual implication $\Rightarrow$
weak conjunction $\wedge$	minimum
weak disjunction $\vee$	maximum
negation $\neg$	the function $\neg x = (x \Rightarrow 0)$
equivalence $\leftrightarrow$	bi-residuum: $\min(x \Rightarrow y, y \Rightarrow x)$
universal quantifier $\forall$	infimum
existential quantifier $\exists$	supremum
predicate $=$	crisp identity
predicate $\in$	evaluation of membership function
class term $\{x \mid \varphi(x)\}$	f. set def. as $Ax = \varphi(x)$, for $x \in U$

For details on the syntax of FCT and defined notions see Appendix A and [3, 5].

Let us now shortly consider two examples. For instance, the truth degree of $A \subseteq B$, defined by the formula $(\forall x)(x \in A \to x \in B)$ in FCT (see Definition A.3) is in an $\mathrm{MTL}_\triangle$-chain computed as

$$\inf_{x \in U} (Ax \Rightarrow Bx),$$

which is a well-known concept of fuzzy inclusion (see [1, 6, 7, 18] and many more). The degree of reflexivity $\mathrm{Refl}(R)$, defined as $(\forall x)Rxx$ in Section 3, is nothing else but

$$\inf_{x \in U} Rxx.$$

We shall now proceed to how the theorems in the following sections should be read in a graded way (although they do not necessarily look graded at first glance). In traditional (fuzzy) logic, a theorem is read as follows:

> *If* a (non-graded) assumption is true
> (i.e., fully true, since non-graded),
> *then* a (non-graded) conclusion is (fully) true.

If an implication is provable in FCT, by soundness, it always holds to degree 1. Now take into account that, in all $\mathrm{MTL}_\triangle$-chains (comprising all standard $\mathrm{MTL}_\triangle$-chains), the following correspondence holds:

$$(x \Rightarrow y) = 1 \text{ if and only if } x \leq y.$$

So an implication that we can prove in FCT can be read as follows:

> *The more* a (graded) assumption is true
> (even if partially),
> *the more* a (graded) conclusion is true
> (i.e., at least as true as the assumption).

In other words, the truth degree of an assumption is a lower bound for the truth degree of the conclusion in provable implications.

Remark: To motivate and illustrate the results in this paper, we will use several examples. In order to make them compact and readable, we will, *in examples,* deviate from our principle to keep formulae separate from their semantics. Instead of mentioning models over some logics, we will simply say that we use some standard logic, for instance, standard Łukasiewicz logic (standing for the standard $\mathrm{MTL}_\triangle$-chain induced by the Łukasiewicz t-norm; analogously for other logics). In examples, we shall furthermore not distinguish between predicate symbols and the fuzzy sets or relations that model them. Instead of saying that a certain model of a fuzzy predicate R fulfills reflexivity to a degree of 0.8, we will simply write $\mathrm{Refl}(R) = 0.8$. This is not the cleanest way of writing it, but it is short and expressive, and it should always be clear to the reader what is meant.

3 Basic Graded Properties of Fuzzy Relations

As an important prerequisite, we first define graded variants of well-known properties of fuzzy relations in the framework of FCT.

Definition 3.1 In FCT, we define basic properties of fuzzy relations as follows:

$$\begin{array}{rll} \mathrm{Refl}(R) & \equiv_{\mathrm{df}} & (\forall x) Rxx \\ \mathrm{Sym}(R) & \equiv_{\mathrm{df}} & (\forall x, y)(Rxy \to Ryx) \\ \mathrm{Trans}(R) & \equiv_{\mathrm{df}} & (\forall x, y, z)(Rxy \,\&\, Ryz \to Rxz) \end{array}$$

Example 3.2 Let us shortly provide a simple example to illustrate the concepts introduced in Definition 3.1. Consider the domain $U = \{1, \ldots, 6\}$ and the following fuzzy relation (for convenience, in matrix notation):

$$P_1 = \begin{pmatrix} 1.0 & 1.0 & 0.5 & 0.4 & 0.3 & 0.0 \\ 0.8 & 1.0 & 0.4 & 0.4 & 0.3 & 0.0 \\ 0.7 & 0.9 & 1.0 & 0.8 & 0.7 & 0.4 \\ 0.9 & 1.0 & 0.7 & 1.0 & 0.9 & 0.6 \\ 0.6 & 0.8 & 0.8 & 0.7 & 1.0 & 0.7 \\ 0.3 & 0.5 & 0.6 & 0.4 & 0.7 & 1.0 \end{pmatrix}$$

It is easy to check that P_1 is a fuzzy preorder with respect to the Łukasiewicz t-norm $\max(x + y - 1, 0)$, hence, taking standard Łukasiewicz logic, we obtain $\mathrm{Refl}(P_1) = 1$ and $\mathrm{Trans}(P_1) = 1$. In this setting, one can easily compute $\mathrm{Sym}(P_1) = 0.4$ (note that, for a finite fuzzy relation R, in standard Łukasiewicz logic, $\neg\,\mathrm{Sym}(R)$ is nothing else but the largest difference between the two values Rxy and Ryx).

Now let us see what happens if we add some disturbances to P_1. Consider the following fuzzy relation:

$$P_2 = \begin{pmatrix} 1.00 & 1.00 & 0.56 & 0.40 & 0.30 & 0.00 \\ 0.87 & 1.00 & 0.33 & 0.44 & 0.26 & 0.02 \\ 0.67 & 0.92 & 0.93 & 0.87 & 0.70 & 0.39 \\ 0.93 & 1.00 & 0.64 & 1.00 & 0.97 & 0.67 \\ 0.52 & 0.79 & 0.82 & 0.71 & 1.00 & 0.59 \\ 0.27 & 0.50 & 0.61 & 0.41 & 0.72 & 1.00 \end{pmatrix}$$

Simple computations give the following results: $\mathrm{Refl}(P_2) = 0.93$, $\mathrm{Sym}(P_2) = 0.41$, $\mathrm{Trans}(P_2) = 0.85$ (for standard Łukasiewicz logic again).

Example 3.3 Consider $U = \mathbb{R}$ and let us define the following parameterized class of fuzzy relations (with $a, c > 0$):

$$E_{a,c}xy = \min(1, \max(0, a - \tfrac{1}{c}|x - y|))$$

It is well known that, for $a = 1$, we obtain fuzzy equivalence relations with respect to the Łukasiewicz t-norm [10, 11, 22, 23]; hence, using standard Łukasiewicz logic again, $\mathrm{Refl}(E_{1,c}) = 1$, $\mathrm{Sym}(E_{1,c}) = 1$, and $\mathrm{Trans}(E_{1,c}) = 1$ for all $c > 0$. On the contrary, it is obvious that reflexivity in the non-graded manner cannot be maintained for $a < 1$. Actually, we obtain

$$\mathrm{Refl}(E_{a,c}) = \min(1, a).$$

for all $a, c > 0$. Similarly, it is a well-known fact that, for $a > 1$, transitivity in the non-graded sense is violated [12]. Regarding graded transitivity, we obtain the following:

$$\mathrm{Trans}(E_{a,c}) = \min(1, \max(0, 2 - a))$$

None of these results depends on the parameter c, as c only corresponds to a re-scaling of the domain. We can conclude that the larger a, the more reflexive, but less transitive, $E_{a,c}$ is. Figure 1 shows two examples.

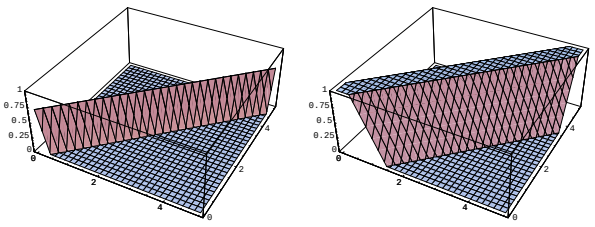


Figure 1: The fuzzy relations $E_{0.7,2}$ (left) and $E_{1.4,1}$ (right). From Example 3.3, we can infer that $\mathrm{Refl}(E_{0.7,2}) = 0.7$, $\mathrm{Trans}(E_{0.7,2}) = \mathrm{Refl}(E_{1.4,1}) = 1$, and $\mathrm{Trans}(E_{1.4,1}) = 0.6$.

In classical mathematics, special properties of relations are rarely studied completely independently of

each other. Instead, these properties most often occur in some combinations in the definitions of special classes of relations—with (pre)orders and equivalence relations being two most fundamental examples. The same is true in the theory of fuzzy relations, where fuzzy (pre)orders and fuzzy equivalence relations are the most important classes. Compound properties of this kind are defined as conjunctions of some of the simple properties of Definition 3.1. In the non-graded case, the properties are crisp, so the conjunction we need is the classical Boolean conjunction. In FCT, however, all properties are graded, so it indeed matters which conjunction we take. Thus, besides the (more usual) combinations by strong conjunction & (corresponding to the t-norm in the standard case), we also define their weak variants combined by weak conjunction (corresponding to the minimum). In this paper, we restrict to the investigation of fuzzy preorders and similarities.[2]

Definition 3.4 In FCT, we define the following compound properties of fuzzy relations:

$$\begin{aligned} \mathrm{Preord}(R) &\equiv_{\mathrm{df}} \mathrm{Refl}(R)\ \&\ \mathrm{Trans}(R) \\ \mathrm{wPreord}(R) &\equiv_{\mathrm{df}} \mathrm{Refl}(R) \wedge \mathrm{Trans}(R) \\ \mathrm{Sim}(R) &\equiv_{\mathrm{df}} \mathrm{Refl}(R)\ \&\ \mathrm{Sym}(R)\ \&\ \mathrm{Trans}(R) \\ \mathrm{wSim}(R) &\equiv_{\mathrm{df}} \mathrm{Refl}(R) \wedge \mathrm{Sym}(R) \wedge \mathrm{Trans}(R) \end{aligned}$$

Example 3.5 Let us shortly revisit Example 3.2. We can conclude the following:

$$\begin{aligned} \mathrm{Preord}(P_1) &= 1 & \mathrm{Preord}(P_2) &= 0.78 \\ \mathrm{wPreord}(P_1) &= 1 & \mathrm{wPreord}(P_2) &= 0.85 \\ \mathrm{Sim}(P_1) &= 0.4 & \mathrm{Sim}(P_2) &= 0.19 \\ \mathrm{wSim}(P_1) &= 0.4 & \mathrm{wSim}(P_2) &= 0.41 \end{aligned}$$

The values in the second column once more demonstrate why it is justified to speak of strong and weak properties—the properties with strong conjunction get smaller truth degrees and thus are harder fulfil. Obviously, the implications $\mathrm{Preord}(R) \to \mathrm{wPreord}(R)$ and $\mathrm{Sim}(R) \to \mathrm{wSim}(R)$ hold.

Example 3.6 For the family of fuzzy relations defined in Example 3.3, we obtain the interesting result

$$\mathrm{Preord}(E_{a,c}) = \mathrm{wPreord}(E_{a,c}) = \max(0, 1 - |1 - a|),$$

from which we can infer that $\mathrm{Preord}(E_{a,c}) = \mathrm{wPreord}(E_{a,c}) = 1$ if and only if $a = 1$. Note that $\mathrm{Sym}(E_{a,c}) = 1$, so $\mathrm{Sim}(E_{a,c}) = \mathrm{Preord}(E_{a,c})$ and $\mathrm{wSim}(E_{a,c}) = \mathrm{wPreord}(E_{a,c})$, which implies that $\mathrm{Sim}(E_{a,c}) = \mathrm{wSim}(E_{a,c}) = 1$ if and only if $a = 1$.

[2] In line with Zadeh's original work [25], we use the term *similarity (relation)* synonymously for fuzzy equivalence (relation).

4 Representation of Fuzzy Preorders

This section aims at generalizing Valverde's representation theorem for fuzzy preorders. We will proceed as follows: we first generalize Fodor's characterization by means of traces and then use this characterization to prove the generalization of Valverde's theorem. Note that Valverde's original proof [23] implicitly follows the same lines.

So given a fuzzy relation R, let us first consider the fuzzy relation R^ℓ defined as

$$R^\ell xy \equiv_{\mathrm{df}} (\forall z)(Rzx \to Rzy)$$

This is called the *left trace* of R [15, 16]. Analogously, we can define the *right trace* as

$$R^r xy \equiv_{\mathrm{df}} (\forall z)(Ryz \to Rxz).$$

Observe the meaning of the following expressions:

$$\begin{aligned} R^\ell \subseteq R &\quad\leftrightarrow\quad (\forall x, y)[(\forall z)(Rzx \to Rzy) \to Rxy] \\ R \subseteq R^\ell &\quad\leftrightarrow\quad (\forall x, y)[Rxy \to (\forall z)(Rzx \to Rzy)] \\ R \approx R^\ell &\quad\leftrightarrow\quad (\forall x, y)[Rxy \leftrightarrow (\forall z)(Rzx \to Rzy)] \end{aligned}$$

Now we can formulate characterizations of graded reflexivity and transitivity, which are not difficult to prove in FCT.

Theorem 4.1 *The following properties hold in* FCT*:*

$$\begin{aligned} \mathrm{Refl}(R) &\quad\leftrightarrow\quad R^\ell \subseteq R \\ \mathrm{Trans}(R) &\quad\leftrightarrow\quad R \subseteq R^\ell \end{aligned}$$

As a corollary we obtain graded versions of Fodor's characterizations [15, Theorems 4.1, 4.3, and Corollary 4.4]. For the two notions of fuzzy equality $\approx$ and $\cong$, see Definition A.3.

Corollary 4.2 *The following is provable in* FCT*:*

$$\begin{aligned} \mathrm{wPreord}(R) &\leftrightarrow R \approx R^\ell \\ \mathrm{Preord}(R) &\leftrightarrow R \cong R^\ell \\ R \approx^2 R^\ell &\longrightarrow \mathrm{Preord}(R) \longrightarrow R \approx R^\ell \end{aligned}$$

Note that, regardless of the symmetry of R, we can replace R^ℓ in the above characterizations by the right trace as well.

Now we have all prerequisites for formulating and proving a graded version of Valverde's representation theorem for preorders. In order to make notations more compact, let us define two graded notions of *Valverde preorder representation* (a strong one and a

weak one), for a given fuzzy relation R and a fuzzy class of fuzzy classes $\mathcal{A}$:

$$\begin{aligned}\mathrm{ValP}(R,\mathcal{A}) &\equiv_{\mathrm{df}} R \cong \{\langle x,y\rangle \mid (\forall A\in\mathcal{A})(Ax\to Ay)\}\\ \mathrm{wValP}(R,\mathcal{A}) &\equiv_{\mathrm{df}} R \approx \{\langle x,y\rangle \mid (\forall A\in\mathcal{A})(Ax\to Ay)\}\end{aligned}$$

The predicates ValP and wValP express the degree to which the fuzzy class $\mathcal{A}$ represents the relation R.

Then we can prove the following essential result for preorders and weak preorders.

Theorem 4.3 FCT *proves the following:*

$$\begin{aligned}&(\mathcal{A}\subseteq\mathcal{A}\cap\mathcal{A})\ \&\ \mathrm{ValP}^2(R,\mathcal{A})\longrightarrow \mathrm{Preord}(R)\\ &\qquad\longrightarrow(\exists\mathcal{A})(\mathrm{Crisp}(\mathcal{A})\ \&\ \mathrm{ValP}(R,\mathcal{A}))\\ &(\mathcal{A}\subseteq\mathcal{A}\cap\mathcal{A})\ \&\ \mathrm{wValP}^3(R,\mathcal{A})\longrightarrow \mathrm{wPreord}(R)\\ &\qquad\longrightarrow(\exists\mathcal{A})(\mathrm{Crisp}(\mathcal{A})\ \&\ \mathrm{wValP}(R,\mathcal{A}))\end{aligned}$$

It can be shown that the exponents in this theorem cannot be lowered, see [2] for a counterexample. Obviously, this theorem is more complicated than Valverde's original result; it is an example where the graded framework does not provide us with just a plain copy of the non-graded (or crisp) result. The following corollary gives us a result that is comparable with Valverde's original theorem.

Corollary 4.4 *The following equivalences are provable in* FCT*:*

$$\triangle\,\mathrm{Preord}(R)\longleftrightarrow R=R^{\ell}\longleftrightarrow \tag{1}$$

$$(\exists\mathcal{A})(\triangle(\mathcal{A}\subseteq\mathcal{A}\cap\mathcal{A})\ \&\ \triangle\,\mathrm{ValP}(R,\mathcal{A})) \tag{2}$$

$$\longleftrightarrow(\exists\mathcal{A})(\mathrm{Crisp}(\mathcal{A})\ \&\ \triangle\,\mathrm{ValP}(R,\mathcal{A})) \tag{3}$$

Observe that also the analogous formulas with wPreord and wValP are equivalent to those in this corollary.

Example 4.5 Let us shortly revisit Example 3.2 (in which we use standard Łukasiewicz logic). The fuzzy relation P_1 was actually constructed from the following crisp family of three fuzzy sets $\mathcal{A}=\{A_1,A_2,A_3\}$ that are defined as follows (for convenience, in vector notation):

$$\begin{aligned}A_1&=(0.7,0.8,0.2,0.5,0.4,0.6)\\ A_2&=(0.3,0.5,0.6,0.4,0.7,1.0)\\ A_3&=(1.0,1.0,0.6,0.4,0.3,0.0)\end{aligned}$$

Although the formula (3) is a perfect copy of Valverde's non-graded representation, the corollary still has graded elements—note that in (2), the class $\mathcal{A}$ may still be a fuzzy class of fuzzy classes, if only it satisfies $\triangle(\mathcal{A}\subseteq\mathcal{A}\cap\mathcal{A})$. Recall that in Gödel logic, this condition is fulfilled by *all* fuzzy classes $\mathcal{A}$, and that in any logic it is satisfied by a system $\mathcal{A}$ in a model if all degrees of membership in $\mathcal{A}$ are idempotent with respect to conjunction.

The degree of $A\in\mathcal{A}$ may be considered as a weighting factor that controls the influence of a specific A on the final result. Corollary 4.4 requires all membership degrees in $\mathcal{A}$ to be idempotent to ensure that the relation represented by $\mathcal{A}$ is a fuzzy preorder, but its graded version in Theorem 4.3 also shows that (loosely speaking) it will *almost* be a fuzzy preorder if $\mathcal{A}$ *almost* satisfies $\mathcal{A}\subseteq\mathcal{A}\cap\mathcal{A}$ (e.g., in standard Łukasiewicz logic if it is close to crispness).

Example 4.6 Let us consider a $[0,1]$-valued fuzzy logic with the triangular norm

$$x*y=\begin{cases}\max(x+y-\frac{1}{2},0) & \text{if } x\in[0,\frac{1}{2}]^2,\\ \min(x,y) & \text{otherwise,}\end{cases}$$

i.e. a simple ordinal sum with a scaled Łukasiewicz t-norm in $[0,\frac{1}{2}]^2$ and the Gödel t-norm anywhere else. It is clear that the set of idempotent elements of this t-norm is $\{0\}\cup[\frac{1}{2},1]$ and that the corresponding residual implication is given as

$$x\Rightarrow y=\begin{cases}1 & \text{if } x\le y,\\ \max(y,\frac{1}{2}-x+y) & \text{otherwise.}\end{cases}$$

Now reconsider $U=\{1,\dots,6\}$ and the three fuzzy sets A_1, A_2 and A_3 from Example 4.5 and define a fuzzy class of fuzzy classes $\mathcal{A}$ such that $\mathcal{A}A_1=0.9$, $\mathcal{A}A_1=1.0$, and $\mathcal{A}A_3=0.8$. Since all three values are idempotent elements of $*$, we can be sure by Theorem 4.3 that the construction $R_1=_{\mathrm{df}}\{\langle x,y\rangle \mid (\forall A\in\mathcal{A})(Ax\to Ay)\}$ always gives us a fuzzy preorder in the given logic. In this particular example, we obtain the following:

$$R_1=\begin{pmatrix}1.0&1.0&0.2&0.4&0.3&0.0\\0.3&1.0&0.2&0.4&0.3&0.0\\0.3&0.5&1.0&0.4&0.3&0.0\\0.4&1.0&0.2&1.0&0.4&0.1\\0.3&0.5&0.3&0.4&1.0&0.2\\0.3&0.5&0.2&0.4&0.4&1.0\end{pmatrix}$$

If we repeat this construction and define a fuzzy relation $R_2=_{\mathrm{df}}\{\langle x,y\rangle \mid (\forall A\in\mathcal{A})(Ax\to Ay)\}$ with $\mathcal{A}$ defined as above, but the connectives interpreted in standard Łukasiewicz logic, we obtain the following:

$$R_2=\begin{pmatrix}1.0&1.0&0.6&0.6&0.5&0.2\\0.8&1.0&0.5&0.6&0.5&0.2\\0.7&0.9&1.0&0.8&0.9&0.6\\0.9&1.0&0.8&1.0&1.0&0.8\\0.6&0.8&0.9&0.7&1.0&0.9\\0.3&0.5&0.6&0.4&0.7&1.0\end{pmatrix}.$$

Straightforward calculations show that $\mathrm{Refl}(R_2) = 1$ and $\mathrm{Trans}(R_2) = \mathrm{Preord}(R_2) = \mathrm{wPreord}(R_2) = 0.8$. This is not at all contradicting to Theorem 4.3, as $\mathcal{A} \subseteq \mathcal{A} \cap \mathcal{A}$ holds only to a degree of 0.6 in standard Łukasiewicz logic.

5 Representations of Similarities

In his landmark paper [23], Valverde not only considers fuzzy preorders, but also similarities (as obvious from the title of this paper). So the question naturally arises how we can modify the above results in the presence of symmetry. As will be seen next, the modifications are not as straightforward as in the non-graded case. We first define a fuzzy relation $R^{\ell s}$ as

$$R^{\ell s}xy \equiv_{\mathrm{df}} (\forall z)(Rzx \leftrightarrow Rzy)$$

(for a given fuzzy relation R). There is no particular name for this fuzzy relation in literature. In analogy to Section 4, let us call it *left symmetric trace* of R.

The following lemma demonstrates how this notion is related to the defining properties of similarities.

Theorem 5.1 *The following are theorems of* FCT*:*

$$R^{\ell s} \subseteq R \leftrightarrow \mathrm{Refl}(R) \tag{4}$$
$$R \subseteq R^{\ell s} \rightarrow \mathrm{Trans}(R) \tag{5}$$
$$R \cong R^{\ell s} \rightarrow \mathrm{Sym}(R) \tag{6}$$
$$\mathrm{Sym}(R)\ \&\ \mathrm{Trans}(R) \rightarrow R \subseteq R^{\ell s} \tag{7}$$

The following theorem provides us with an analogue of Corollary 4.2, unfortunately, with looser bounds on the left-hand side.

Corollary 5.2 FCT *proves:*

$$R \approx^4 R^{\ell s} \longrightarrow R \cong^2 R^{\ell s} \longrightarrow \mathrm{Sim}(R)$$
$$\mathrm{Sim}(R) \longrightarrow R \cong R^{\ell s} \longrightarrow R \approx R^{\ell s}$$
$$R \approx^2 R^{\ell s} \longrightarrow R \cong R^{\ell s} \longrightarrow \mathrm{wSim}(R)$$
$$\mathrm{wSim}^2(R) \longrightarrow R \approx R^{\ell s}$$

The question arises whether it is really necessary to require $\cong$ rather than $\approx$ in (6). The following example tells us that this is indeed the case. It also implies that $R \approx R^{\ell s} \rightarrow \mathrm{wSim}(R)$ does *not* hold in general.

Example 5.3 Consider $U = \{1, 2\}$, standard Łukasiewicz logic, and the following fuzzy relation:

$$R = \begin{pmatrix} 0.5 & 1.0 \\ 0.0 & 0.5 \end{pmatrix}$$

It is obvious that $\mathrm{Refl}(R) = 0.5$ and $\mathrm{Sym}(R) = 0$. Moreover, routine calculations show $\mathrm{Trans}(R) = 1$ and that $R^{\ell s}$ is given as follows:

$$R^{\ell s} = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}$$

So, we finally obtain that $(R \approx R^{\ell s}) = 0.5$, while $(R \cong R^{\ell s}) = 0$.

Now we can formulate a graded version of Valverde's representation theorem for similarities. Analogously to the above considerations, let us define the graded notion of *Valverde similarity representation* (strong one and weak one) for a given fuzzy relation R and a fuzzy class $\mathcal{A}$:

$$\mathrm{ValS}(R, \mathcal{A}) \equiv_{\mathrm{df}} R \cong \{\langle x, y\rangle \mid (\forall A \in \mathcal{A})(Ax \leftrightarrow Ay)\}$$
$$\mathrm{wValS}(R, \mathcal{A}) \equiv_{\mathrm{df}} R \approx \{\langle x, y\rangle \mid (\forall A \in \mathcal{A})(Ax \leftrightarrow Ay)\}$$

In the same way as for preorders, we can prove Valverde's representation theorem of similarities and weak similarities.

Theorem 5.4 FCT *proves the following:*

$$(\mathcal{A} \subseteq \mathcal{A} \cap \mathcal{A})\ \&\ \mathrm{ValS}^3(R, \mathcal{A}) \rightarrow \mathrm{Sim}(R) \tag{8}$$
$$\mathrm{Sim}(R) \rightarrow (\exists \mathcal{A})(\mathrm{Crisp}(\mathcal{A})\ \&\ \mathrm{ValS}(R, \mathcal{A})) \tag{9}$$
$$(\mathcal{A} \subseteq \mathcal{A} \cap \mathcal{A})\ \&\ \mathrm{wValS}^3(R, \mathcal{A}) \rightarrow \mathrm{wSim}(R) \tag{10}$$
$$\mathrm{wSim}^2(R) \rightarrow (\exists \mathcal{A})(\mathrm{Crisp}(\mathcal{A})\ \&\ \mathrm{wValS}(R, \mathcal{A})) \tag{11}$$

Again, this theorem is more complicated than Valverde's original representation of similarities. In the following corollary, analogously to preorders, we can infer a result very similar to Valverde's original theorem in case that the corresponding properties are fulfilled to degree 1.

Corollary 5.5 *The following equivalences are provable in* FCT*:*

$$\triangle\, \mathrm{Sim}(R) \longleftrightarrow R = R^{\ell s} \longleftrightarrow \tag{12}$$
$$(\exists \mathcal{A})(\triangle(\mathcal{A} \subseteq \mathcal{A} \cap \mathcal{A})\ \&\ \triangle\, \mathrm{ValS}(R, \mathcal{A})) \tag{13}$$
$$\longleftrightarrow (\exists \mathcal{A})(\mathrm{Crisp}(\mathcal{A})\ \&\ \triangle\, \mathrm{ValS}(R, \mathcal{A})) \tag{14}$$

Again, like in the case of preorders, we can add equivalent lines with wValS instead of ValS, and (13) has a graded ingredient—the class $\mathcal{A}$ may be a fuzzy class of fuzzy classes.

6 Concluding Remarks

The present paper has generalized Valverde's famous representation theorems for fuzzy preorders and similarities to the fully graded framework of Fuzzy Class

Theory (FCT). In the formal setting of FCT, this generalization can be done relatively easily compared to Gottwald's semi-formal framework of graded properties of fuzzy relations. At the same time, we have seen that the results are not just obtained by simply rewriting known theorems. Indeed we obtain new results that even give rise to interesting new constructions (as demonstrated by Example 4.6).

A Appendix: Fuzzy Class Theory

In this section, we present a self-contained list of definitions related to Fuzzy Class Theory (FCT). For a complete and detailed introduction to FCT, the reader is referred to the freely available primer [5].

Definition A.1 *Fuzzy Class Theory* (over $\mathrm{MTL}_\triangle$) is a theory over multi-sorted first-order logic $\mathrm{MTL}_\triangle$ with crisp equality. There are sorts for individuals of the zeroth order (i.e., atomic objects), denoted by lowercase variables $a, b, c, x, y, z, \ldots$; individuals of the first order (i.e., fuzzy classes), denoted by uppercase variables $A, B, X, Y, \ldots$; individuals of the second order (i.e., fuzzy classes of fuzzy classes), denoted by calligraphic variables $\mathcal{A}, \mathcal{B}, \mathcal{X}, \mathcal{Y}, \ldots$; etc. Individuals $\xi_1, \ldots, \xi_k$ of each order can form k-tuples (for any $k \geq 0$), denoted by $\langle \xi_1, \ldots, \xi_k \rangle$; tuples are governed by the usual axioms known from classical mathematics (e.g., that tuples equal if and only if their respective constituents equal). Furthermore, for each variable x of any order n and for each formula φ there is a class term $\{x \mid \varphi\}$ of order $n+1$.

Besides the logical predicate of identity, the only primitive predicate is the membership predicate $\in$ between successive sorts (i.e., between individuals of the n-th order and individuals of the $(n+1)$-st order, for any n). The axioms for $\in$ are the following (for variables of all orders):

$$(\in 1) \quad y \in \{x \mid \varphi(x)\} \leftrightarrow \varphi(y), \text{ for each formula } \varphi \qquad \text{(comprehension axioms)}$$

$$(\in 2) \quad (\forall x)\triangle(x \in A \leftrightarrow x \in B) \rightarrow A = B \qquad \text{(extensionality)}$$

Moreover, we use all axioms and deduction rules of multi-sorted first-order logic $\mathrm{MTL}_\triangle$ with crisp identity. Theorems, proofs, etc., are defined completely analogously as in classical logic.

Convention A.2 For better readability, we make the following conventions:

- We use $(\forall x \in A)\varphi$, $(\exists x \in A)\varphi$ as abbreviations for $(\forall x)(x \in A \rightarrow \varphi)$ and $(\exists x)(x \in A \,\&\, \varphi)$, respectively.
- $\{x \in A \mid \varphi\}$ is shorthand for $\{x \mid x \in A \,\&\, \varphi\}$.
- We use $\{\langle x_1, \ldots, x_k \rangle \mid \varphi\}$ as abbreviation for $\{x \mid (\exists x_1) \ldots (\exists x_k)(x = \langle x_1, \ldots, x_k \rangle \,\&\, \varphi)\}$.
- The formulae $\varphi \,\&\, \ldots \,\&\, \varphi$ (n times) are abbreviated φ^n; instead of $(x \in A)^n$, we can write $x \in^n A$ (analogously for other predicates).
- We use Ax and $Rx_1 \ldots x_n$ as synonyms for $x \in A$ and $\langle x_1, \ldots, x_n \rangle \in R$, respectively.
- A chain of implications

  $$\varphi_1 \rightarrow \varphi_2,\ \varphi_2 \rightarrow \varphi_3,\ \ldots,\ \varphi_{n-1} \rightarrow \varphi_n$$

  is, for short, written as $\varphi_1 \longrightarrow \varphi_2 \longrightarrow \cdots \longrightarrow \varphi_n$ (and analogously for the equivalence connective).

Definition A.3 We define the following elementary relations between fuzzy sets in FCT:

$$\begin{array}{rcl} \mathrm{Crisp}(A) & \equiv_{\mathrm{df}} & (\forall x)\triangle(x \in A \vee x \notin A) \\ A \subseteq B & \equiv_{\mathrm{df}} & (\forall x)(x \in A \rightarrow x \in B) \\ A \cong B & \equiv_{\mathrm{df}} & (A \subseteq B) \,\&\, (B \subseteq A) \\ A \approx B & \equiv_{\mathrm{df}} & (\forall x)(x \in A \leftrightarrow x \in B) \end{array}$$

The *models* of FCT are systems (closed under definable operations) of fuzzy sets (and fuzzy relations) of all orders over some crisp universe U, where the membership functions of fuzzy subsets take values in some $\mathrm{MTL}_\triangle$-chain. *Intended* models are those which contain *all* fuzzy subsets and fuzzy relations over U (of all orders). Models in which moreover the $\mathrm{MTL}_\triangle$-chain is standard (i.e., given by a left-continuous t-norm on the unit interval $[0, 1]$) correspond to Zadeh's [24] original notion of fuzzy set; therefore we call them *Zadeh models*. FCT is sound with respect to Zadeh models, therefore all theorems provable in FCT are true statements about fuzzy sets and relations in the traditional sense.

Acknowledgement

The work of Libor Běhounek was supported by the Program "Information Society", Project 1ET100300517. The work of Petr Cintula was partly supported by the grant No. B100300502 of the Grant Agency of the Academy of Sciences of the Czech Republic and partly by by the Institutional Research Plan AV0Z10300504.

References

[1] Wyllis Bandler and Ladislav J. Kohout. Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, 4:183–190, 1980.

[2] Libor Běhounek, Ulrich Bodenhofer, and Petr Cintula. Relations in fuzzy class theory: initial steps. Submitted to Fuzzy Sets and Systems.

[3] Libor Běhounek and Petr Cintula. Fuzzy class theory. *Fuzzy Sets and Systems*, 154(1):34–55, 2005.

[4] Libor Běhounek and Petr Cintula. From fuzzy logic to fuzzy mathematics: A methodological manifesto. *Fuzzy Sets and Systems*, 157(5):642–646, 2006.

[5] Libor Běhounek and Petr Cintula. Fuzzy Class Theory: A primer v1.0. Technical Report V-939, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 2006. Available at www.cs.cas.cz/research/library/reports_900.shtml.

[6] Radim Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*, volume 20 of *IFSR International Series on Systems Science and Engineering*. Kluwer Academic/Plenum Press, New York, 2002.

[7] Ulrich Bodenhofer. A similarity-based generalization of fuzzy orderings preserving the classical axioms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(5):593–610, 2000.

[8] Ulrich Bodenhofer, Bernard De Baets, and János Fodor. A compendium of fuzzy weak orders: Representations and constructions. To appear in Fuzzy Sets and Systems, 2007.

[9] Georg Cantor. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen*, 46:481–512, 1895.

[10] Bernard De Baets and Radko Mesiar. Pseudometrics and T-equivalences. *Journal of Fuzzy Mathematics*, 5(2):471–481, 1997.

[11] Bernard De Baets and Radko Mesiar. Metrics and T-equalities. *Journal of Mathematical Analysis and Applications*, 267:331–347, 2002.

[12] Martine De Cock and Etienne E. Kerre. On (un)suitable fuzzy relations to model approximate equality. *Fuzzy Sets and Systems*, 133(2):137–153, 2003.

[13] Francesc Esteva and Lluís Godo. Monoidal t-norm based logic: Towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124(3):271–288, 2001.

[14] Francesc Esteva, Lluís Godo, and Franco Montagna. The ŁΠ and ŁΠ$\frac{1}{2}$ logics: Two complete fuzzy systems joining Łukasiewicz and product logics. *Archive for Mathematical Logic*, 40(1):39–67, 2001.

[15] János Fodor. Traces of fuzzy binary relations. *Fuzzy Sets and Systems*, 50(3):331–341, 1992.

[16] János Fodor and Marc Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer Academic Publishers, Dordrecht, 1994.

[17] Siegfried Gottwald. Fuzzified fuzzy relations. In Robert Lowen and Marc Roubens, editors, *Proceedings of 4th IFSA Congress*, volume Mathematics (ed. P. Wuyts), pages 82–86, Brussels, 1991.

[18] Siegfried Gottwald. *Fuzzy Sets and Fuzzy Logic: Foundations of Application—from a Mathematical Point of View*. Vieweg, Wiesbaden, 1993.

[19] Siegfried Gottwald. *A Treatise on Many-Valued Logics*, volume 9 of *Studies in Logic and Computation*. Research Studies Press, Baldock, 2001.

[20] Petr Hájek and Petr Cintula. Triangular norm predicate fuzzy logics. To appear in Proceedings of Linz Seminar 2005, 2006.

[21] Joan Jacas and Jordi Recasens. Fuzzified properties of fuzzy relations. In *Proceedings of the 9th IPMU Conference*, volume 1, pages 157–161, Annecy, 2002.

[22] Enric Trillas and Llorenç Valverde. An inquiry into indistinguishability operators. In Heinz J. Skala, Settimo Termini, and Enric Trillas, editors, *Aspects of Vagueness*, pages 231–256. Reidel, Dordrecht, 1984.

[23] Llorenç Valverde. On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328, 1985.

[24] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[25] Lotfi A. Zadeh. Similarity relations and fuzzy orderings. *Information Sciences*, 3:177–200, 1971.

Enriched Generic Algebras of Fuzzy Relations

Ladislav J. Kohout
Dept of Computer Science, Florida State University,
Tallahassee, Florida 32306-4530, U.S.A.
kohout@cs.fsu.edu

Abstract

The paper presents an overview of a computation friendly calculus of fuzzy relations. It is presented within the framework of an enriched generic algebra of relations that we have been developing since 1979. It is enriched with BK-non-associative products of relations of three kinds: triangle subproduct $\lhd$, superproduct $\rhd$ and the square product $\Box$. The BK-products have been useful in concise formulation of new theorems in relational mathematics as well as in a number of practical applications in medicine, engineering, information retrieval and elsewhere. The concise algebraic manipulation is also advantageous in symbolic computing. Currently we are engaged in developing an equational theorem checker in which the BK-products play a substantial role. In this endeavour, Tarski style relational calculi play an essential role.

Keywords: Fuzzy relations, non-associative compositions, Fuzzy BK-products, relational calculus, BL fuzzy logic.

1 Enriched Calculus of Fuzzy Relations

1.1 Enhancing Expressive Power of Calculus of Relations

There are six distinguishing features of the BK-product systems of relations that facilitate the unification of different many-valued systems of fuzzy relations and enhance their practical applicability.

1 Non-associative BK-products are introduced and used in definitions of relational properties and in computations. These products are defined for both homogeneous and heterogeneous relations.

2 Homomorphisms between relations are extended from mappings used in the literature to general relations. This yields *generalized morphisms* important for practical solving of relational inequalities and equations.

3 Relational properties are not only global but also local (important for applications).

4 The unified treatment of computational algorithms by means of matrix notation is used.

5 The theory unifying crisp and fuzzy relations in some distinguished logics makes it possible to represent a whole *finite nested family* of crisp relations with special properties as a *single* cutworthy [11] fuzzy relation for the purpose of computation. After completing the computations, the resulting fuzzy relation is again converted by α-cuts to a nested family of crisp relations, thus increasing the computing performance considerably.

6 Relations in their predicate forms are distinguished from their satisfaction sets; foresets and aftersets of relations are used in addition to relational predicates. This makes it possible to introduce interpretable linguistic labels (semiotic descriptors) that have a clearly defined meaning within the domains of their applications. Then one can develop an algebra of meaning defined by equations and inequalities that provides a computational basis for forming ontologies in knowledge engineering applications as well as in *computing with words*.

1.2 Basic Notions

Propositional Form. A *binary relation from A to B* is given by an open predicate ____P____ with two empty slots; when the first is filled with the name a of an element of A and the second with the name b of an element of B, there results a proposition. If aPb is true, we write aR_Pb and say that *a is* R_P*-related to*

b[1]. The lattice of all binary (two-place, 2-argument) relations from A to B is denoted by $\mathcal{R}(A \rightsquigarrow B)$. Relations of this kind are called *heterogeneous*. When the set B happens to be the same as A, we speak of relations **within a set** or *in a set*, or *on a set*, and call these *homogeneous*.

The **Satisfaction Set** The *satisfaction set* or *extension set* of a relation $R \in \mathcal{R}(A \rightsquigarrow B)$ is the set of all those pairs $(a, b) \in A \times B$ for which it holds:

$$R_S = \{(a, b) \in A \times B \mid aRb\}$$

Clearly R_S is a subset of the Cartesian product $A \times B$. We have $R_P = R'_P \Longrightarrow R_S = R'_S$. That means, knowing R_P, we know R_S; knowing R_S, we know everything about R_P except the wording of its "name" ____P____. Because in symbolic computing with relations we need to deal with names of predicates R_P, it is essential distinguish notationally by R_P the cases when we do not deal just with satisfaction sets R_S(i.e. extensions) of relations.

The **Extensionality Convention** says that, regardless of their propositional wordings, two relations should be regarded as the same if they hold, or fail to hold between exactly the same pairs: $R_S = R'_S \Longrightarrow R_P = R'_P$. Once the extensionality convention has been adopted, then there is a one-to-one correspondence between the subsets R_S of $A \times B$ and the (distinguishable) relations R_P in $\mathcal{R}(A \rightsquigarrow B)$. Since R_S and R_P now uniquely determine each other, the current fashion for set-theoretical parsimony suggests that they be identified. We, however, maintain the distinction in principle.

1.3 Crisp and Fuzzy Nonassociative Compositions of Relations

In 1977 Bandler and Kohout [2] introduced nonassociative relational compositions $\triangleleft, \triangleright, \square$ that further extended the crisp relational calculus [26],[27],[23],[12],[13],[14]. The fuzzy version of these products was first published in [3], for succinct surveys see [9], [23].

The family $\{\triangleleft, \triangleright, \circ\}$ is based on the positive fragment of logic, avoiding carefully the negation. Hence many algebraic properties of this triad of relational products carry over from the crisp relational calculus to fuzzy calculi for residuated systems, e.g. up to the monoidal logics based relational calculi[2].

[1]When it is unnecessary to emphasize the propositional form the subscript is dropped in R_P, writing: R, aRb.

[2]Other calculi are investigated (distinct from the monoidal logic based ones) which are based on noncommutative and/or.

Table 1: DEFINITIONS OF RELATIONAL PRODUCTS

PRODUCT TYPE	SET-BASED DEFINITION
Zadeh's Circle product:	$x(R \circ S)z \Leftrightarrow$ xR intersects Sz
BK-Triangle Subproduct:	$x(R \triangleleft S)z \Leftrightarrow xR \,\tilde{\subseteq}\, Sz$
BK-Triangle Superprod.:	$x(R \triangleright S)z \Leftrightarrow xR \,\tilde{\supseteq}\, Sz$
BK-Square product:	$x(R \square S)z \Leftrightarrow xR \cong Sz$

MANY-VALUED LOGIC DEF.	TENSOR NOTATION
$(R \circ S)_{ik} = \bigvee_j (R_{ij} \& S_{jk})$	$(R \circ S)_{ik} = R_{ij} \circ S^{jk}$
$(R \triangleleft S)_{ik} = \bigwedge_j (R_{ij} \rightarrow S_{jk})$	$(R \triangleleft S)_{ik} = R_{ij} \triangleleft S^{jk}$
$(R \triangleright S)_{ik} = \bigwedge_j (R_{ij} \leftarrow S_{jk})$	$(R \triangleright S)_{ik} = R_{ij} \triangleright S^{jk}$
$(R \square S)_{ik} = \bigwedge_j (R_{ij} \equiv S_{jk})$	$(R \square S)_{ik} = R_{ij} \square S^{jk}$

Table 1 of definitions three different notational forms for BK-products:

1. the notation using the concept of set inclusion and equality [4, 5].
2. many-valued logic (MVL) based notation[3], which uses the quantifier $\bigwedge$ and a connective $\rightarrow$, or $\leftarrow$, or $\equiv$.
3. the tensor notation[4].

1.4 Granular Form of BK-Products

The set based-definition uses fuzzy granules: subsets of the elements of the relations composed. Conceptually, and also for computational reasons, it is essential to distinguish two different kinds of granules: foresets and aftersets.

The fuzzy **afterset** of $x \in X$ is the *fuzzy subset of* Y consisting of the elements $y \in Y$ to which x is related by R (where $\mu_A x = \delta\{xRy\}$, the degree to which x and y are R-related):

$$xR = \{y \mid y \in Y \text{ and } \delta\{xRy\} > 0\}.$$

The fuzzy **foreset** of $y/\delta\{xRy\} \in Y$ is the *fuzzy subset of* X consisting of all the elements $x \in X$ which are related by R to y (where $\mu_A y = \delta\{xRy\}$, the degree to which x and y are R-related):

$$Ry = \{x/\delta\{xRy\} \mid x \in X \text{ and } \delta\{xRy\} > 0\}.$$

[3]With appropriately defined fuzzy power set and fuzzy set inclusion, the notational forms (1) and (2) of these relational compositions are algebraically equivalent.

[4]The tensor notation preserves in addition the inner structure of the composition when the right hand side of the form (3) is used in the formulas.

The notions of afterset and foreset of an element can be extended to afterset and foreset of a set in (at least) two distinct but equally important ways: an *inclusive* or *exclusive* afterset / foreset (see [8]).

The **inclusive** after- and foresets are given by

$$A'R = A'_R \circ R, \quad RB' = R \circ B'_R$$

The **exclusive** after- and foresets are given by

$$[A']R = A'_R \lhd R, \quad R[B'] = R \rhd B'_R$$

More explicitly, its component-wise definition involves indexed elements.

$$C_k = (A \circ R) = \bigvee_i (A_i \ \& \ R_{ik})$$

$$C_k = (A \lhd R) = \bigwedge_i (A_i \rightarrow R_{ik})$$

$$C_k = (A \rhd R) = \bigwedge_i (A_i \leftarrow R_{ik})$$

where A_i is the membership (characteristic function) giving the degree to which the predicate $a_i \in A$ is TRUE; and R_{ij} is the degree to which the predicate $R_{ij} \in R$ is TRUE, where R_{ij} is an element of R.

1.5 N-ary Relations

Operations can also be defined on n-ary relations for $n > 2$.

1.5.1 Intersections, Unions, Inclusions

Intersection, union, complementation and inclusion all have their definitions, and endow the set $\mathcal{R}_F(X_1, ..., X_n)$, of all fuzzy $n-$ary relations on $X_1, ..., X_n$ with the usual lattice, structure within which the crisp relations constitute a Boolean sublattice.

1.5.2 Products:

$\mathcal{R}_F(X_1, ..., X_r) \times \mathcal{R}_F(Y_1, ..., Y_s)$ yields
$\mathcal{R}_F(X_1, ..., \hat{X}_m, ..., X_r, Y_1, ..., \hat{Y}_p, ..., Y_s)$.

Products play an equally important part with $n-$ary relations [7],[9] as they do with binary. There are a number of different kinds, which can be distinguished as follows. As usual, the matrix formulation will be the clearest.

We are given an $r-$ary relation R on $(X_1, ..., X_r)$ and an $s-$ary relation on $(Y_1, ..., Y_s)$ such that (at least) one of the X sets, say X_m is the same as one of the Y sets, say Y_p. There are also given two operations on the numbers in [0, 1], which we symbolize by $\oplus$ and $\odot$. We indicate by circumflex or hat above a set or an index its omission from the list in which it is shown. Then we have the following

Definition 1 *The $(\oplus, m, p, \odot)$-product of R and S is the $(r + s - 2)$-ary relation T on the sets $(X_1, ..., \hat{X}_m, ..., X_r, Y_1, ..., \hat{Y}_p, ..., Y_s)$ given by*

$$T_{i_1 ... \hat{i}_m ... i_r, j_1 ... \hat{j}_p ... j_s} =$$

$$\bigoplus_{i_m = i_p} (R_{i_1 ... i_m ... i_r} \odot S_{j_1 ... j_p ... j_s}).$$

Where $\vee$ indicates maximum or, where appropriate, supremum, and $\wedge$ indicates minimum or infimum, and & a t-norm. The usual Zadeh's circle product is given by $\circ = (sup, 2, 1, \wedge)$. The harsh and mean subtriangular products are respectively $\lhd_h = (inf,\ 2, 1, \rightarrow)$ and $\lhd_m = (1/ \mid x_2 \mid \sum, 2, 1, \rightarrow)$ and correspondingly for supertriangular and square products. These same kinds of products play equally important role with $n-$ary relations. The only enrichment is the possible substitutions for the 2 and the 1. It may be worth noting that the ordinary matrix product of conventional linear algebra is $(\sum, 2, 1, \cdot)$; the influence of this on the notation of the definition should be clear.

2 Tarski's Calculus of Crisp Relations

The axioms defining the theory of crisp binary homogeneous relations were given by Tarski in 1941, in his now classical paper [29]. Using the first order predicate logic, he gave twelve axioms (1)–(12). Let us take for example the following axioms:

4 $(\forall x)(\forall y)(\forall z)[(xRy \wedge yEz) \rightarrow xRz]$

11 $(\forall x)(\forall z)[x(R \circ S)z \leftrightarrow (\exists y)(xRy \wedge ySz)]$

12 $R = S \leftrightarrow (\forall x)(\forall y)[xRy \leftrightarrow xSy)]$

Axiom (4) determines the algebraic behaviour of E; namely, it tells us that E is the two-sided identity. Axiom (11) defines the conventional relation composition $\circ$ (i.e. *circle product*). Finally, (12) defines the *equality* of two relations.

Using the predicate logic, Tarski has constructed the calculus of relations as a part of more comprehensive logical theory consisting of the following components:

1. Individual variables *x, y, z, ...*;
2. relation variables *R, S, P, Q, ...*
3. Logical constants of a predicate calculus: e.g. connectives $\wedge, \vee, \&, \rightarrow, \equiv, \neg; \forall, \exists$.
4. Rules of inference.

From variables and constants various expressions are formed, namely:

- Elementary sentences, xRy (x is related to y).
- Compound sentences: well-formed formulas composed in the usual way of sentences by means connectives and quantifiers.

This basic system is extended by introducing further constants that are specific to the calculus of relations. Tarski [29] uses the following eleven constants:

1. Relations: U.. universal, O ... null , E identity, D ... diversity.
2. Unary operations: the complement, the converse (transpose).
3. Binary operations: $\sqcup$, $\sqcap$, $\circ$ .... relational composition (product) and its dual $\bullet$.
4. The identity predicate: $=$

Relational designations are the expressions formed from relation variables, relation constants, and operation signs.

Elementary sentences in this extension are expressions of the form 'xRy' and '$R = S$', where 'x' and 'y' stand for any individual variables and 'R' and 'S' for any relational designation.

After eliminating individual variables and logical constants Tarski obtained axioms of pure relational calculus that contain only *relational constants*, relational *unary operations*, relational *binary operations* and the *identity predicate*. In this way Tarski obtained fifteen axioms (I)-(XV) of crisp (Boolean, non-fuzzy) relational calculus [29].

As an illustration, let us look at

Axiom X. $(R \circ (S \circ Q)) = ((R \circ S) \circ Q)$

Axiom XI. $R \circ E = R$

Axiom XIII. $((R \circ S) \sqcap Q^T = O) \rightarrow ((S \circ Q) \sqcap R^T = O)$

Axioms (I) – (XV) of relational calculus can be translated back into the first order logic expressions by means of the first order logic relational axioms (1) – (12). This translation yields axioms (I-C) – (XV-C). For illustration, the axioms IX, X, XI, XIII are translated as follows:

X-C. $(\forall x)(\forall v)((\exists y)(xRy \wedge (\exists z)(ySz \wedge zQv)) \equiv (\exists z)((\exists y)(xRy \wedge ySz) \wedge zQv))$

XI-C. $(\forall x)(\forall y)((xRy \wedge yEy) \equiv xRy)$

XIII-C $(\forall x)(\forall z)(((x(R \circ S)z \wedge xQ^T z) \equiv xOz) \rightarrow ((\forall y)(\forall x)((y(S \circ Q)x \wedge yR^T x) \equiv yOx))$

Note that whereas axioms (I) – (XV) do not have quantifiers and do not refer to the individual elements of relations, the axioms (I-C) – (XV-C) do. In fact, the latter quantify over the variables denoting individual elements.

3 BK-Calculus in Tarski Style: Crisp Relations

BK-products and all our crisp work was strongly motivated by the work of Otakar Boruvka. The style of mathematical proofs was essentially similar to what we have learned from works of Otakar Boruvka, Eduard Čech, and van Der Waerden[5] Only when the work was substantially developed, and we started looking at strict formalization in predicate calculus we understood the significance of Tarski's paper for our work.

3.1 Characterization of Special Properties of Relations Between Two Sets

Self-inverse circle product is very useful in characterization of *special properties of relations* between two sets. Using the product, one can characterize these properties in purely relational way, without directly referring to individual elements of the relations involved.

Theorem 2 *Special properties of a heterogeneous relation* $R \in \mathcal{R}(X \rightsquigarrow Y)$

1. R *is covering* $\Leftrightarrow E_X \sqsubseteq R \circ R^{-1}$.
2. R *is univalent* $\Leftrightarrow R^{-1} \circ R \sqsubseteq E_Y$.
3. R *is onto* $\Leftrightarrow E_Y \sqsubseteq R^{-1} \circ R$.
4. R *is separating* $\Leftrightarrow R \circ R^{-1} \sqsubseteq E_X$.

where E_X *and* E_Y *are the left and right identities, respectively. Note that the relational inclusion* $\sqsubseteq$ *is distinguished from the set inclusion* $\subseteq$.

3.2 Characterization of Special Properties of Relations On a Set

Defining a relation R on a set (i.e. a homogeneous relation, $R \in \mathcal{R}(X \rightsquigarrow X)$) allows one to work with

[5] Wyllis Bandler's Ph.D. supervisor was van Der Waerden. Indeed, without WB's familiarity with van Der Waerden's methods and my knowledge of Boruvka's and Čech's work our work on crisp relations [2] would not have developed that way. It was a lucky coincidence that it is translatable into the style of Tarski [29]. This provided a bridge to using BL logic of Hájek for rigorous fuzzification.

additional properties. Relational properties of homogeneous relations are well covered in the literature. *Local reflexivity* is an exception. It appeared in [2] and was generalized to fuzzy relations in [6], leading to new computational algorithms for identification of relational properties of data-sets both crisp and fuzzy relations [6],[10].

1. *Covering* $\Leftrightarrow$ every x_i is related by R to something $\Leftrightarrow \forall i \in I, \quad \exists j \in I, \quad \text{s.t.} \quad R_{ij} = 1.$
2. Locally reflexive $\Leftrightarrow$ if x_i is related to anything, or if anything is related to x_i, then x_i is related to itself $\Leftrightarrow \forall i \in I, \quad R_{ii} = max_j(R_{ij}, R_{ji})$
3. *Reflexive* $\Leftrightarrow$ covering and locally reflexive $\Leftrightarrow \forall i \in I, \quad R_{ii} = 1$

Unfortunately, it is absent from the textbooks, yet it is extremely important in applications of relational methods to analysis of the real life data (see the notion of participant in the next two subsections).

3.2.1 Partitions IN and ON a Set

A *partition* **on** a set X is a division of X into non-overlapping (and nonempty) subsets called blocks. A partition **in** a set X is a partition on the subset of X called the *subset of participants*.

There is a one-to-one correspondence between partitions **in** X and *local equivalences* (i.e. locally reflexive, symmetric relations and transitive relations) in $\mathcal{R}(X \rightsquigarrow B)$. The partitions in X (so also the local equivalences in $\mathcal{R}(X \rightsquigarrow B)$) form a lattice with `"___is-finer-than___"` as its ordering relation.

To obtain a global equivalence[6] one needs to add the covering property to the properties of local equivalence, so that local reflexivity turns into a (total) reflexivity. Equivalences have the following universal representation:

Theorem 3 *[8],[2]*
$R = R\square R^{-1}$ *if and only if* R *is an equivalence.*

3.2.2 Tolerances and Overlapping Classes

• $R \circ R^T$ is always symmetric and locally reflexive
• $R \circ R^T$ is a tolerance iff R is covering.
• $R\square R^T$ is always a (local) tolerance
• $R\square R^T \sqsubseteq R$ iff R is reflexive
• $E \sqsubseteq R \sqsubseteq R\square R^T$ iff R is an equivalence
• $R\square R^T = R$ iff R is an equivalence
• $R\square R^T \sqsubseteq R \circ R^T$ iff R is covering.

Tests for tolerance and equivalence.

[6] usually called just "equivalence".

It is not always the case that one manages, or even attempts, to classify participants into non-overlapping blocks. Local *tolerance* (i.e. locally reflexive and symmetric) relations lead to classes which may well overlap, where one participant may belong to more than one class.

4 Fuzzification

Many of the theorems and formulas proven in standard predicate logic discussed in the previous section generalize to BL logics when the negation is not involved. The purpose of this section is to demonstrate that the relational calculus considerably simplifies the proofs and relational computations in BL fuzzy logics.

4.1 Residuated bootstrap of BK-products

Important formulas of the BK-enriched relational calculus are given by the next theorem. This theorem which describes the interrelationship of $\circ, \lhd, \rhd$ plays a substantial role in further development of the theory of crisp and fuzzy relations. Because these formulas depend on residuation, they carry over into relational theories based on t-norms and corresponding residuated implication operators[7] [25],[20].

Theorem 4 *Residuated bootstrap of BK-products [21],[25] For arbitrary* $V \in \mathcal{B}(A \rightsquigarrow C)$,

$$(R \circ S \sqsubseteq V) \quad \equiv \quad (R \sqsubseteq V \rhd S^T) \quad \equiv \quad (S \sqsubseteq R^T \lhd V)$$

Proof: [22],[25]
$BL \vdash (\forall x)(\varphi \to \nu) \equiv (\exists x)\varphi \to \nu$ *(by [15], T 5.1.4(2)).*
The substitutions $\varphi := xRy\&ySz, \nu := xVz$ *yield*
$\vdash (\forall y)((xRy\&ySz) \to xVz) \equiv ((\exists y)(xRy\&ySz) \to xVz)$
$\vdash (\forall y)((xRy\&ySz) \to xVz) \to ((\exists y)(xRy\&ySz) \to xVz)$ *by [15], L 2.2.16(25).*
$\vdash (\forall x)(\forall z)(\forall y)((xRy\&ySz) \to xVz) \to (\forall x)(\forall z)((\exists y)(xRy\&ySz) \to xVz))$ *by [15], T 5.1.16(5) and MP.*
$\vdash (\forall x)(\forall y)(xRy \to (\forall z)(xVz \leftarrow zS^Ty)) \to (\forall x)(\forall z)((\exists y)(xRy\&ySz) \to xVz))$ *By [15], T 15.1.14(1).*
$(A) \vdash (\forall x)(\forall y)(xRy \to x(V \rhd S^T)y) \to (\forall x)(\forall z)(x(R \circ S)z \to xVz)$
Similarly, we derive (B)
$(B) \vdash (\forall x)(\forall y)(xRy \to x(V \rhd S^Ty) \leftarrow (\forall x)(\forall z)(x(R \circ S)z \to xVz)$
Substituting $\varphi := (A), \psi := (B)$ *into* $\vdash \varphi \to (\psi \to \varphi\&\psi)$,
applying MP twice and using the definition of $\equiv$ *yields*
$(B) \vdash (\forall x)(\forall y)(xRy \to x(V \rhd S^T)y) \equiv$

[7] In fact, it generalizes even further, to fuzzy Monoidal logics [24].

$(\forall x)(\forall z)(x(R \circ S)z \rightarrow xVz)$
which is $\vdash (R \sqsubseteq V \rhd S^T) \equiv (R \circ S \sqsubseteq V)$

Similar proofs can be given for other equivalences of Th. 4.

Theorem 4 has been proven in BL logic. It is an important equality of BK-relational calculus that can be used directly for proofs of other theorems. We shall give some examples in the sections that follow. For further details and discussion of the importance of this theorem as the core statement in axiomatization of relations see [25].

4.1.1 Fuzzy Generalised Morphisms

Very important for distributed knowledge networking [19],[28] is a generalization of conventional homomorphisms defined constructively by BK-products [21].:

Definition 5 *Let F, R, G, S be the relations between the sets A, B, C, D such that $R \in \mathcal{R}(A \rightsquigarrow B)$, $S \in \mathcal{R}(C \rightsquigarrow D)$, $F \in \mathcal{R}(A \rightsquigarrow C)$, $G \in \mathcal{R}(B \rightsquigarrow D)$. The conditions that (for all $a \in A, b \in B, c \in C, d \in D$) aFc and aRb and bGd imply cSd, can be expressed in any of the following ways:*
(i) $FRG : S$ are **forward compatible**
(ii) F, G are **generalized homomorphisms** *from R to S.*

Let $FRG : S$ denote relations that satisfy the forward compatibility criterion $(F^{-1} \circ R \circ G \sqsubseteq S)$. This criterion is fulfilled iff $(R \sqsubseteq F \lhd S \rhd G^{-1})$. Furthermore, we have the following theorem:

Theorem 6

$$(F^{-1} \circ R \circ G \sqsubseteq S) \equiv (R \sqsubseteq F \lhd S \rhd G^{-1})$$

The direct proof in BL logic is more involved than the proof of Theorem 4 (bootstrap). It gets simplified if done in the equalities of Th. 4 [25],[20].
Proof:
Substituting $T := F^{-1}$, $U := R \circ G$, $V := S$ we obtain $(F^{-1} \circ R \circ G \sqsubseteq S) \equiv (R \circ G \sqsubseteq F \lhd S)$.
Substituting $T := R$, $U := G$, $V := F \lhd S$ we obtain $(R \circ G \sqsubseteq F \lhd S) \equiv (R \sqsubseteq F \lhd S \rhd G^{-1})$.
Transitivity of equivalences yields $(F^{-1} \circ R \circ G \sqsubseteq S) \equiv (R \sqsubseteq F \lhd S \rhd G^{-1})$. This completes the proof.

4.2 Characterization of Fuzzy Classivalent Relations

Another important relational property is *classivalence* (partial difunctionality).

Definition 7 *A fuzzy relation $R \in \mathcal{R}(X \rightarrow X)$ in a t-norm residuated logic (BL) is called classivalent (partial difunctional) when it satisfies the following condition: $(\forall a)(\forall b)(\forall a')(\forall b')(aRb \& a'Rb \& a'Rb' \rightarrow aRb')$ where $\&$ is a t-norm and $\rightarrow$ is the residuated implication operator associated with $\&$.*

Theorem 8 *It is provable in BL that a fuzzy relation is classivalent (partial bifunctional) iff $R \circ R^T \circ R \sqsubseteq R$*

Theorem 9 *Let $R \in \mathcal{R}(X \rightsquigarrow X)$ be a classivalent (partial functional) relation a t-norm residuated logic BL. Then the following equivalence holds:*

$$(R \circ R^T \circ R \sqsubseteq R) \equiv (R \circ R^T \sqsubseteq R \rhd (R^T \circ R))$$
$$\equiv (R^T \circ R \sqsubseteq R^T \lhd R)$$

Proof:
Substituting $R := R$, and also $S := R^T \circ R$ into the equivalences of Theorem 4 (Residuated Bootstrap) yields $(R \circ R^T \circ R \sqsubseteq R) \equiv (R \sqsubseteq R \rhd (R^T \circ R)) \equiv (R^T \circ R \sqsubseteq R^T \lhd R)$ ¶

In a similar way, the *residuated bootstrap of BK-products* is used to prove (in BL) other theorems (cf. [25]) concerned with the theory of Generalized Morphisms. A software tool GMorph is described in [17].

4.3 Fuzzification of Tarski's Relational Systems

Axioms I, II, III, V, VII, VIII, IX, X, XIII of the original system of Tarski [29] described in Sec. 2 above carry over to BL completely. The remaining axioms IV, VI, XII, XIV, XV have more alternatives.

We have seen that in his 1941 paper, Tarski introduced two different axiomatizations for the fragment of the calculus of relations. The axioms introduced by the first method based on the predicate logic are axioms (1) – (12) mentioned in Sec.2 above[8]. The second method is exemplified by axioms I – XV listed above. The second method led subsequently to the theory of relational algebras.

The first axiomatization of relational algebras appeared in [16] by Jónsson and Tarski.

They added to all basic axioms of Boolean algebra the following formulas:

JT1: $(R^T)^T = R$
JT2: $(R \circ S) \circ T = R \circ (S \circ T)$
JT3: $(R \sqcup S) \circ T = R \circ T \sqcup S \circ T$
JT4: $R \circ (S \sqcup T) = (R \circ S) \sqcup (R \circ T)$
JT5: $(R \sqcup S)^T = R^T \sqcup S^T$
JT6: $R \circ E = R$
JT7: $R^T \circ \overline{(R \circ S)} \sqsubseteq \overline{S}$

[8] For the full list see [29].

The most interesting is JT7. It contains 3 operations: 1-ary complement and transpose; and 2-ary circle product $\circ$.

It is more difficult to prove that the last axiom, namely

$$R^T \circ \overline{R \circ S} \sqsubseteq \overline{S}$$

caries over to BL fuzzification. We shall now prove the theorem to that effect.

The substitutions $R := R^T$; $S := \overline{R \circ S}$; $V := \overline{S}$ into the formula of Theorem 4 yield

Theorem 10

$$R^T \circ \overline{R \circ S} \sqsubseteq \overline{S} \quad \equiv \quad \overline{R \circ S} \sqsubseteq (R^T)^T \lhd \overline{S}$$

Theorem 11

$$\overline{R \circ S} \sqsubseteq (R^T)^T \lhd \overline{S}$$

is a 1-tautology in BL.

5 Conclusions

Taking the Program of Tarski, extending the mathematics of relations – this can be considered as the goal of our programme, This is compatible with the General System theory. Indeed, Klir in his GS publications [18] emphasises the importance of relations in GST. This, of course, should be compatible with the programme of fuzzy mathematics. Already, our theory of extended relations defines relations as systems with foresets and aftersets; relations are presents as predicates, as well as satisfaction sets. Link with map algebras is also important. This has also has to be considered.

Now, the question is, how this link of relational algebra/calculus with fuzzy set theory is to be provided. L-categories of Wyllis Bandler give the lead [1]. There are two different approaches: one is meta–mathematics and the other is eso–mathematical use of category theory[9]

Esomathematical use of category theory already indicates that to distinguish different roles, different contexts[10] we need to compose morphism bunches of different colours, and composition of such bunches is not necessarily associative [1].

[9]This is connected with external and internal – when we try import from the external to the internal language, the connectives. ...This is also connected to formal fuzzy logic, inference rules and implication, when we import the deduction theorem into the system.

[10]the associativity of category theory is harmful there

References

[1] W. Bandler. Some esomathematical uses of category theory. In G. Klir, editor, *Applied General Systems Research: Recent Developments and Trends*, NATO Conference Series II, Vol. 5, pages 243–255. Plenum Press, New York, 1978.

[2] W. Bandler and L. Kohout. *Mathematical Relations, their Products and Generalized Morphisms.* Tech. report, Man-Machine Systems Laboratory, EES-MMS-REL 77-3, Dept. of Electrical Eng., University of Essex, Colchester, Essex, U.K., 1977.

[3] W. Bandler and L. Kohout. Fuzzy relational products and fuzzy implication operators. In *International Workshop on Fuzzy Reasoning Theory and Applications*, London, September 1978. Queen Mary College, University of London. Also available as FRP-1 report.

[4] W. Bandler and L. Kohout. Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, 4:13–30, 1980. Reprinted in: *Readings in Fuzzy Sets for Intelligent Systems*, D. Dubois, H. Prade and R. Yager (eds.), Morgan Kaufmann Publishers, San Mateo, Calif., 1993, pages 88-96.

[5] W. Bandler and L. Kohout. Fuzzy relational products as a tool for analysis and synthesis of the behaviour of complex natural and artificial systems. In P. Wang and S. Chang, editors, *Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems*, pages 341–367. Plenum Press, New York and London, 1980.

[6] W. Bandler and L. Kohout. Fast fuzzy relational algorithms. In A. Ballester, D. Cardús, and E. Trillas, editors, *Proc. of the Second Internat. Conference on Mathematics at the Service of Man*, pages 123–131, Las Palmas, 1982. (Las Palmas, Canary Islands, Spain, 28 June - 3 July), Universidad Politechnica de las Palmas.

[7] W. Bandler and L. Kohout. Mathematical relations. In *Fuzzy Research Project*, University of Essex, Colchester, U.K., August 1982. Dept. of Mathematics. FRP-23 report.

[8] W. Bandler and L. Kohout. On the general theory of relational morphisms. *International Journal of General Systems*, 13:47–66, 1986.

[9] W. Bandler and L. Kohout. Relations, mathematical. In M. Singh, editor, *Systems and Control Encyclopedia*, pages 4000 – 4008. Pergamon Press, Oxford, 1987.

[10] W. Bandler and L. Kohout. Special properties, closures and interiors of crisp and fuzzy relations. *Fuzzy Sets and Systems*, 26(3):317–332, June 1988.

[11] W. Bandler and L. Kohout. Cuts commute with closures. In B. Lowen and M. Roubens, editors, *Fuzzy Logic: State of the Art*, pages 161–167. Kluwer Academic, Boston and Dordrecht, 1993.

[12] R. Bělohlávek. Similarity relations and BK-relational products. *Information Sciences*, 126(1-4):287–295, 2000.

[13] R. Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer Academic/Plenum Press, New York, 2002.

[14] P. Hájek. A remark on Bandler-Kohout products of relations. *Internat. Journal of General Systems*, 25(2):165–166, 1996.

[15] P. Hájek. *Metamathematics of Fuzzy Logics*. Kluwer, Dordrecht, 1998.

[16] B. Jónsson and A. Tarski. Representation problems for relational algebras. *Bulletin of Amer. Math. Soc.*, 54:80 and 1192, 1948.

[17] E. Kim and L. Kohout. Design of GMORPH by means of Activity Structures. In R. M. et al., editor, *Proc. of IFSA 1997 (The world Congress of International Fuzzy Systems Association, Prague), vol. I.*, pages 84–89. IFSA, June 1997.

[18] G. Klir. *Architecture of Systems Problem Solving*. Plenum Press, New York, 1985.

[19] L. Kohout. Dynamic sharing of cognitive structures by means of generalized morphisms. In P. Wang, editor, *Proc. of the 2nd Joint Conf. on Information Sciences (Sept. 28 - Oct. 1, Wrightsville Beach, NC, USA)*, pages 597–600. Duke University, 1995.

[20] L. Kohout. Generalized morphisms in BL-logics. In *Logic Colloquium 1998 (The 1998 ASL European Summer Meeting, Prague)*. Association for Symbolic Logic, August 9-15 1998. (Extended abstract presenting the main mathematical theorems published the Bulletin of ASL No. 1, March 1999, pp. 116-117.).

[21] L. Kohout. Generalized morphisms in BL-logics. *The Bulletin of Symbolic Logic*, 5(1):116–117, 1999.

[22] L. Kohout. Basic algebra of BK products of relations in t-norm fuzzy logics. In R. Langari and J. Yen, editors, *Proc. of FUZZ-IEEE2000*, pages 599–604, Piscataway, NJ, 2000. IEEE Neural Network Council, IEEE.

[23] L. Kohout. Boolean and fuzzy relations. In P. Pardalos and C. Floudas, editors, *The Encyclopedia of Optimization*, pages 189–202. Kluwer, Boston, 2001. vol.I, A-D.

[24] L. Kohout. Fuzzy relational equations in monoidal logics. In *Proc. 2004 IEEE Internat Conference on Fuzzy Systems*, pages 1555–1559, New York, July 2004. IEEE Neural Network Society, IEEE.

[25] L. Kohout. Theory of fuzzy generalized morphisms and relational inequalities. *Internat. J. of General Systems*, 33(4):339–360, 2004.

[26] L. Kohout and W. Bandler. Fuzzy relational products in knowledge engineering. In V. Novák et al., editor, *Fuzzy Approach to Reasoning and Decision Making*, pages 51–66. Academia and Kluwer, Prague and Dordrecht, 1992.

[27] L. Kohout and W. Bandler. Use of fuzzy relations in knowledge representation, acquisition and processing. In L. Zadeh and J. Kacprzyk, editors, *Fuzzy Logic for the Management of Uncertainty*, pages 415–435. John Wiley, New York, 1992.

[28] J. Muhammad and L. Kohout. The role of generalized morphisms in co-operative activities of agents. In P. Wang, editor, *Joint Conference on Information Sciences – JCIS'98 Proceedings*, pages 56–59, Research Triangle Park, N.C., October 1998. Association for Intelligent Machinery. In vol.1.

[29] A. Tarski. Calculus of relations. *Journal of Symbolic Logic*, 6(3):73–89, 1941.

Uncertainty as a Modality over T-norm Based Logics

Enrico Marchioni
Department of Information and Communication Sciences
Universitat Oberta de Catalunya
Rambla del Poblenou 156, 08018 Barcelona (Spain)
enrico@iiia.csic.es

Abstract

In this work we propose a general approach for representing uncertainty measures in the framework of t-norm based logics. This approach is extended also to classes of measures like probability, possibility, necessity, lower and upper probability. We show that, under certain conditions, the logical consistency of a theory of uncertainty is tantamount to the coherence of a related assessment of rational values. Finally, we characterize the basic requirements that guarantee the compactness of coherent assessments.

Keywords: T-norm based logics, Fuzzy measures, Coherence, Compactness.

1 Introduction

Measures of uncertainty aim at formalizing the strength of our beliefs in the occurrence of some events by assigning to those events a degree of uncertainty. From the mathematical point of view a measure of uncertainty is a real-valued function that gives an event a value from the real unit interval $[0, 1]$. A well-known example is given by probability measures which try to capture our degree of confidence in the occurrence of events by real-valued assessments. Esteva, Hájek, and Godo proposed in [16, 12] a new interpretation of measures of uncertainty in the framework of t-norm based logics. Given a sentence as "The proposition φ is plausible (probable, believable)", its degrees of truth can be interpreted as the degree of uncertainty of the proposition φ. Indeed, the higher is our degree of confidence in φ, the higher the degree of truth of the above sentence will result. In some sense, the predicate "is plausible (believable, probable)" can be regarded as a modal operator over the proposition φ. Then, given a measure of uncertainty μ, we can define modal many-valued formulas $\kappa(\varphi)$, whose interpretation is given by a real number corresponding to the degree of uncertainty assigned to φ under μ. Furthermore, we can translate the peculiar axioms governing the behavior of an uncertainty measure into formulas of a certain t-norm based logic, depending on the operations we need to represent.

Previous particular results concerning the representation of measures of uncertainty were presented in several works. We can mention the treatment of probability measures, necessity measures and belief functions proposed by Esteva, Hájek, and Godo in [16, 15, 12], [16], and [13], respectively; the treatment of conditional probability proposed by the present author and Godo in [14]; the treatment of (generalized) conditional possibility and necessity given by the present author in [22]; and finally the treatment of simple and conditional non-standard probability given by Flaminio and Montagna in [10].

Here, our aim will consist in giving a general and comprehensive treatment of the representation of measures of uncertainty. In particular, we will show how it is possible to represent classes of measures such as probabilities, lower and upper probabilities, possibilities and necessities. Important properties of the functions of t-norm based logics will then be useful in order to prove relevant features of the classes of measures represented.

Recall that, given a set W of possible situations, a *fuzzy measure* [23] is a mapping μ from the Boolean algebra of subsets of W into the real unit interval $[0, 1]$ satisfying the following properties: (i) $\mu(\bot) = 0$, (ii) $\mu(\top) = 1$, (iii) if $\vdash \varphi \to \psi$ then $\mu(\varphi) \leq \mu(\psi)$.

Besides such common properties, each kind of fuzzy measure differs from the others in how the measure associated to compound propositions is computed from the marginal ones. In other words, what specifies the behavior of a fuzzy measure is the way how from assessments of uncertainty concerning separated events we can determine the degree associated to their com-

bination. In a certain sense we can say that classes of fuzzy measures are characterized by the satisfaction of some compositional properties. However, it is well known that fuzzy measures cannot be fully compositional. This means that the degree of confidence in any compound proposition φ cannot be always computed from the degree assigned to its subformulas (see [6]).

Now, we briefly recall the definitions of some classes of measures we will deal with in the following sections.

Probability measures [21] are fuzzy measures defined over a *σ-algebra*[1]. Probabilities measures over σ-algebras are fuzzy measures which satisfy the law of countable additivity, i.e.: if $\bigwedge_{i=1}^{n} \varphi_i \leftrightarrow \bot$ then $\mu\left(\bigvee_{i=1}^{n} \varphi_i\right) = \sum_{i=1}^{n} \varphi_i$, for all $n \in \mathbb{N}$. Those measures are also called *countably additive probabilities*. If we do not require the algebra to be closed under countable unions, we define the class of probability measures, called *finitely additive probabilities*, as the class of all those fuzzy measures (over Boolean algebras) which satisfy the law of finite additivity: if $\vdash \varphi \wedge \psi \leftrightarrow \bot$ then $\mu(\varphi \vee \psi) = \mu(\varphi) + \mu(\psi)$. Here we deal with finitely additive probabilities only. We denote the class of probability measures by $\mathcal{P}$, and each measure in $\mathcal{P}$ by P.

Possibility measures (see [24, 5]) are a class of fuzzy measures satisfying the following law of composition w.r.t. the maximum t-conorm: $\mu(\varphi \vee \psi) = \max(\mu(\varphi), \mu(\psi))$. We denote the class of possibility measures by $\mathcal{P}i$, and each measure in $\mathcal{P}i$ by Π. Similarly, *Necessity measures* [5] are fuzzy measures satisfying the following law of composition w.r.t. the minimum t-norm: $\mu(\varphi \wedge \psi) = \min(\mu(\varphi), \mu(\psi))$. We denote the class of necessity measures by $\mathcal{N}$, and each measure in $\mathcal{N}$ by N. Possibility and Necessity measures are dual in the sense that, given a possibility measure Π (a necessity measure N), one can derive its dual necessity measure (possibility measure) from it by means of the standard involutive negation $n_s(x) = 1 - x$. Indeed, $N(\varphi) = 1 - \Pi(\neg\varphi)$ and $\Pi(\varphi) = 1 - N(\neg\varphi)$. Probability measures are, on the contrary, self-dual, since the dual measure of a probability measure still is a probability measure: $P(\varphi) = 1 - P(\neg\varphi)$.

Given a set of probability measures P_i over the same Boolean algebra, the *upper probability* $P^{\Delta}(\varphi)$ is defined as $\sup\{P_i(\varphi)\}$ and the *lower probability* $P_{\nabla}(\varphi)$ is defined as $\inf\{P_i(\varphi)\}$ (see [17]). Upper and lower probabilities are dual, since from an upper probability we can define a lower probability as $P_{\nabla}(\varphi) = 1 - P^{\Delta}(\neg\varphi)$, and viceversa.

Upper and lower probabilities can be also seen as classes of fuzzy measures. Indeed, as shown by Anger and Lembcke in [1], any upper probability is a fuzzy measure μ such that for all natural numbers m, n, k, and all $\varphi_1, \ldots, \varphi_m$, if $\{\{\varphi_1, \ldots, \varphi_m\}\}$ is an (n, k)-cover[2] of $(\varphi, \top)$, then

$$(\sharp) \qquad k + n\mu(\varphi) \leq \sum_{i=1}^{m} \mu(\varphi_i).$$

Notice that Halpern and Pucella proved in [18] that when the sample space is finite there are only finitely many instances of the the above property. Indeed, there exists constants $k_0, k_1, \ldots$ such that if W is a finite set, for all natural numbers $m, n, k \leq k_{|W|}$, and all $\varphi_1, \ldots, \varphi_m$, if $\{\{\varphi_1, \ldots, \varphi_m\}\}$ is an (n, k)-cover of $(\varphi, \top)$, then $(\sharp)$ holds.

Similarly we can see any lower probability as a fuzzy measure μ such that for all natural numbers m, n, k, and all $\varphi_1, \ldots, \varphi_m$, if $\{\{\varphi_1, \ldots, \varphi_m\}\}$ is an (n, k)-cover of $(\varphi, \top)$, then

$$(\sharp\sharp) \qquad k + n\mu(\varphi) \geq \sum_{i=1}^{m} \mu(\varphi_i).$$

This paper is organized as follows. In the next section we introduce the uncertainty logic $\mathcal{M}(\mathcal{L})$ based on a t-norm based logic $\mathcal{L}$ in order to reason about the class $\mathcal{M}$ of fuzzy measures. We give a completeness result, and show how this logic can be easily extended to represent classes of measures like probabilities, possibilities, necessities, lower and upper probabilities. Furthermore, we briefly mention expressive power issues related to the possibility of defining the notion of conditioning and having rational truth-constants in the language. In Section 3 we deal with the notions of coherence and compactness of rational assessments and show their connections with suitable theories defined in the logical framework. We end with some final remarks.

Notice that, due to space reasons, we do not provide any background notion concerning t-norm based logics. The interested reader can find all the required concepts in the papers cited throughout this work.

[1] Recall that, given a set W, a σ-algebra is a collection of subsets of W closed under complementation and countable unions.

[2] A proposition φ is said to be *covered* n times by a multiset $\{\{\varphi_1, \ldots, \varphi_m\}\}$ of propositions if every situation in which φ is true makes true at least n propositions from $\varphi_1, \ldots, \varphi_m$ as well. An (n, k)-*cover* of $(\varphi, \top)$ is a multiset $\{\{\varphi_1, \ldots, \varphi_m\}\}$ that covers $\top$ k times and covers φ $n + k$ times.

2 Logics of Uncertainty

2.1 The base logic

Let $\mathcal{L}$ be any t-norm based logic [15, 7], or any of its expansions, and let $\mathcal{M}$ be any class of fuzzy measures. $\mathcal{M}(\mathcal{L})$ is built up over $\mathcal{L}$ extending its language by including modal formulas which represent the uncertainty given by a fuzzy measure $\mu \in \mathcal{M}$. We define the language in two steps. First, we have classical Boolean formulas φ, ψ, etc., defined in the usual way from the classical connectives $(\wedge, \neg)$ and from a countable set V of propositional variables $p, q, \ldots$, etc. The set of Boolean formulas is denoted by L. Moreover given any set $D \subseteq L$, we denote by $Con(D)$ the set of sentences which logically follow from D in classical logic. Moreover, $Taut(L)$ will denote the set of classical tautologies.

Elementary modal sentences are formulas of the form $\kappa(\varphi)$, where κ is a unary operator taking as arguments Boolean sentences. Compound modal formulas are built by means of the $\mathcal{L}$-connectives. Nested modalities are not allowed.

Definition 2.1 The axioms of the logic $\mathcal{M}(\mathcal{L})$ are the following:

(i) The set $Taut(L)$ of classical Boolean tautologies

(ii) Axioms of $\mathcal{L}$ for modal formulas

(iii) The following axiom:
$(\mathcal{M}1) \qquad \neg\kappa(\bot)$

Deduction rules of $\mathcal{M}(\mathcal{L})$ are those of $\mathcal{L}$, plus:

(iv) *modalization*: from $\vdash \varphi$ (i.e. φ is derivable in Classical Logic) derive $\kappa(\varphi)$

(v) *monotonicity*: from $\vdash \varphi \to \psi$ derive $\kappa(\varphi) \to \kappa(\psi)$.

The language we have defined clearly is a hybrid language. Indeed, any theory (set of formulas) we will deal with will be of the form $\Gamma = D \cup T$, where D contains only non-modal formulas and T contains only modal formulas. In the following D will always refer to a recursive non-modal theory. Notice that there is no direct interaction between non-modal and modal formulas, with the exception of the application of the above rules of inference. The role of modalization and monotonicity only consists in generating new modal formulas which can then be used in the deduction. Therefore, we are led to define in $\mathcal{M}(\mathcal{L})$ the notion of proof from a theory, written $\vdash_{\mathcal{M}(\mathcal{L})}$, in a non-standard way, at least when the theory contains non-modal formulas.

Definition 2.2 The proof relation $\vdash_{\mathcal{M}(\mathcal{L})}$ between sets of formulas and formulas is defined by:

1. $D \cup T \vdash_{\mathcal{M}(\mathcal{L})} \varphi$ if $\varphi \in Con(D)$;
2. $T \vdash_{\mathcal{M}(\mathcal{L})} \Phi$ if Φ follows from T in the usual way from the above axioms and rules;
3. $D \cup T \vdash_{\mathcal{M}(\mathcal{L})} \Phi$ if $T \cup D^{\mathcal{M}} \vdash_{\mathcal{M}(\mathcal{L})} \Phi$;

where $D^{\mathcal{M}} = \{\kappa(\varphi) : \varphi \in Con(D)\}$.

We now define the semantics for $\mathcal{M}(\mathcal{L})$ by introducing *$\mathcal{M}$-Kripke structures.*

Definition 2.3 A $\mathcal{M}$-Kripke model is a structure $K = \langle W, \mathcal{U}, e, \mu \rangle$, where:

- W is a non-empty set of possible worlds.
- $\mathcal{U}$ is a Boolean algebra of subsets of W.
- $e : V \times W \to \{0, 1\}$ is a *Boolean* evaluation of the propositional variables, that is, $e(p, w) \in \{0, 1\}$ for each propositional variable $p \in V$ and each world $w \in W$. Any given truth-evaluation $e(\cdot, w)$ is extended to Boolean propositions as usual. For a Boolean formula φ, we will denote by $[\varphi]_W$ the set of worlds in which φ is true, i.e. $[\varphi]_W = \{w \in W \mid e(\varphi, w) = 1\}$.
- $\mu : \mathcal{U} \to [0, 1]$ is a fuzzy measure over $\mathcal{U}$, such that $[\varphi]_W$ is μ-measurable for any non-modal φ.
- $e(\cdot, w)$ is extended to elementary modal formulas by defining
$$e(\kappa(\varphi), w) = \mu([\varphi]_W),$$
and to arbitrary modal formulas according to the $\mathcal{M}(\mathcal{L})$-semantics.

A structure K is a model for Φ, written $K \models \Phi$, if $e^K(\Phi) = 1$. If T is a set of formulas, we say that K is a model of T if $K \models \Phi$ for all $\Phi \in T$. The notion of logical entailment relative to a class of structures $\mathcal{K}$, written $\models_{\mathcal{K}}$, is then defined as follows:

$$\Gamma \models_{\mathcal{K}} \Phi \text{ iff } K \models \Phi \text{ for each } K \in \mathcal{K} \text{ model of } \Gamma.$$

If $\mathcal{K}$ denotes the whole class of $\mathcal{M}$-Kripke structures we shall write $\Gamma \models_{\mathcal{M}(\mathcal{L})} \Phi$. When $\models_{\mathcal{K}} \Phi$ holds we will say that Φ is *valid* in $\mathcal{K}$, i.e. when Φ gets value 1 in all structures $K \in \mathcal{K}$.

Proposition 2.4 (Soundness) *The logic $\mathcal{M}(\mathcal{L})$ is sound with respect to the class of $\mathcal{M}$-Kripke structures.*

Let $D \subset L$ be any given non-modal (propositional) theory (possibly empty). For any $\varphi, \psi \in L$, define $\varphi \sim_D \psi$ iff $\varphi \leftrightarrow \psi$ follows from D in classical propositional logic, i.e. if $\varphi \leftrightarrow \psi \in Con(D)$. The relation $\sim_D$ is an equivalence relation in L and $[\varphi]_D$ will denote the equivalence class of φ. Obviously, the quotient set $L/_{\sim_D}$ forms a Boolean algebra which is isomorphic to a subalgebra $\mathcal{B}(\Omega_D)$ of the power set of the set Ω_D of Boolean interpretations of the crisp language L which are model of D. For each $\varphi \in L$, we shall identify the equivalence class $[\varphi]_D$ with the set $\{\omega \in \Omega_D \mid \omega(\varphi) = 1\} \in \mathcal{B}(\Omega_D)$ of models of D that make φ true. We shall denote by $\mathcal{M}(D)$ the set of fuzzy measures defined over $L/_{\sim_D}$ or, equivalently, on $\mathcal{B}(\Omega_D)$.

Notice that each fuzzy measure $\mu \in \mathcal{M}(D)$ induces a $\mathcal{M}$-Kripke structure $\langle \Omega_D, \mathcal{B}(\Omega_D), e_\mu, \mu \rangle$ where $e_\mu(p, \omega) = \omega(p) \in \{0, 1\}$ for each $\omega \in \Omega_D$ and each propositional variable p. We shall denote by $\mathcal{K}_D$ the class of $\mathcal{M}$-Kripke structures which are models of D, and by $\mathcal{MS}(D)$ the class of $\mathcal{M}$-Kripke models $\{(\Omega_D, \mathcal{B}(\Omega_D), e_\mu, \mu) \mid \mu \in \mathcal{M}(D)\}$. Obviously, $\mathcal{MS}(D) \subset \mathcal{K}_D$.

Abusing the language, we will say that a fuzzy measure $\mu \in \mathcal{M}(D)$ is a *model* of a modal theory T whenever the induced Kripke structure $\langle \Omega_D, \mathcal{B}(\Omega_D), e_\mu, \mu \rangle$ is a model of T (obviously $\langle \Omega_D, \mathcal{B}(\Omega_D), e_\mu, \mu \rangle$ is a model of D as well).

Given the above notions, we now prove a completeness result for $\mathcal{M}(\mathcal{L})$.

Theorem 2.5 ((Finite) Strong completeness) *Let $\mathcal{L}$ be any t-norm based logic (or any of its expansions). If $\mathcal{L}$ is finitely strongly standard complete, then let T be a finite modal theory over $\mathcal{M}(\mathcal{L})$, D a finite non-modal theory and Φ a modal formula. Then*

$$T \cup D \vdash_{\mathcal{M}(\mathcal{L})} \Phi \text{ iff } e_\mu(\Phi) = 1$$

for each fuzzy measure $\mu \in \mathcal{M}(D)$ model of T.

Moreover, if $\mathcal{L}$ is strongly standard complete the same holds for infinite theories.

2.2 Classes of measures

We now see how we can easily extend the above logic in order to treat particular classes of measures of uncertainty. Let $\mathcal{L}$ be any t-norm based logic (or any of its expansions), and let $\mathcal{M}'$ be a class of fuzzy measures. We say that $\mathcal{L}$ is *compatible* with $\mathcal{M}'$ if the real valued operations needed to compute values of compound propositions are definable in $\mathcal{L}$. Clearly, a careful analysis of which functions are definable in certain t-norm based logics can help us know which among those logics are suitable for representing certain classes of measures.

Probability. As for probabilities, we need the sum and the standard involutive negation, which are only available in expansions of the Łukasiewicz logic [15], that are then the only t-norm based logics compatible with probability measures.

Let $\mathcal{L}$ be a t-norm based logic compatible with $\mathcal{P}$. Then, the logic $\mathcal{P}(\mathcal{L})$ is obtained from $\mathcal{M}(\mathcal{L})$ by adding the following axioms:

$(\mathcal{M}2)$ $\kappa(\varphi \to \psi) \leftrightarrow (\kappa(\varphi) \to \kappa(\varphi))$,

$(\mathcal{M}3)$ $\kappa(\varphi \vee \psi) \leftrightarrow (\kappa(\varphi) \to \kappa(\varphi \wedge \psi)) \to \kappa(\psi)$,

$(\mathcal{M}4)$ $\kappa(\neg\varphi) \leftrightarrow \neg\kappa(\varphi)$.

Notice that in presence of axiom $(\mathcal{M}2)$ the monotonicity rule is derivable.

$\mathcal{P}$-Kripke models are $\mathcal{M}$-Kripke models where μ is a finitely additive probability measure.

Possibility and Necessity. As for possibility measures we only need the minimum t-norm, hence every t-norm based logic is compatible with $\mathcal{P}i$.

Let $\mathcal{L}$ be a t-norm based logic compatible with $\mathcal{P}i$. Then, the logic $\mathcal{P}i(\mathcal{L})$ is obtained from $\mathcal{M}(\mathcal{L})$ by adding the following axiom:

$(\mathcal{M}5)$ $\kappa(\varphi \vee \psi) \leftrightarrow \kappa(\varphi) \vee \kappa(\psi)$.

$\mathcal{P}i$-Kripke models are $\mathcal{M}$-Kripke models where μ is a possibility measure.

As for necessity measures we only need the maximum t-conorm [20], hence every t-norm based logic is compatible with $\mathcal{N}$.

Let $\mathcal{L}$ be a t-norm based logic compatible with $\mathcal{N}$. Then, the logic $\mathcal{N}(\mathcal{L})$ is obtained from $\mathcal{M}(\mathcal{L})$ by adding the following axiom:

$(\mathcal{M}6)$ $\kappa(\varphi \wedge \psi) \leftrightarrow \kappa(\varphi) \wedge \kappa(\psi)$.

$\mathcal{N}$-Kripke models are $\mathcal{M}$-Kripke models where μ is a necessity measure.

Lower and Upper Probability. As for upper probabilities, notice that the condition $(\sharp)$ is equivalent to $\frac{k}{m} + \frac{n}{m}\mu(\varphi) \leq \sum_{i=1}^{m} \frac{\mu(\varphi_i)}{m}$, given that $n, k \leq m$. It is then clear that $\sum_{i=1}^{m} \frac{\mu(\varphi_i)}{m} \leq 1$, and so it makes sense to rely on t-norm based logics. It is evident that a logic is compatible with the class $\mathcal{P}^\Delta$ only if it allows the

representation rational numbers, the product of rationals and formulas, and addition. Thus, any extension (or expansion) of RŁ [11], or RPPŁ$'_\Delta$ [19] represents a suitable choice. Furthermore, the presence of the standard involutive negation makes possible to define also lower probabilities.

Let $\mathcal{L}$ be a t-norm based logic compatible with $\mathcal{P}^\Delta$. The logic $\mathcal{P}^\Delta(\mathcal{L})$ is obtained from $\mathcal{M}(\mathcal{L})$ by adding the rule (UP): if

$$\varphi \to \bigvee_{\binom{J\subseteq\{1,\dots,m\}}{|J|=k+n}} \bigwedge_{j\in J} \varphi_j, \quad \text{and} \quad \bigvee_{\binom{J\subseteq\{1,\dots,m\}}{|J|=k}} \bigwedge_{j\in J} \varphi_j$$

are propositional tautologies, then derive

$$k.\delta_m \oplus n.\delta_m\kappa(\varphi) \to \bigoplus_{j=1}^{m} \delta_m\kappa(\varphi_j),$$

if $\mathcal{L}$ is an extension (expansion of RŁ), or derive

$$\overline{\tfrac{k}{m}} \oplus \left(\overline{\tfrac{n}{m}} *_\pi \kappa(\varphi)\right) \to \bigoplus_{j=1}^{m} \left(\overline{\tfrac{1}{m}} *_\pi \kappa(\varphi_j)\right),$$

if $\mathcal{L}$ is an extension (expansion) of RPPŁ$'_\Delta$.

The semantics for $\mathcal{P}^\Delta$ is given by $\mathcal{P}^\Delta$-Kripke models, i.e. $\mathcal{M}$-Kripke models where μ is an upper probability measure.

As for the class of lower probability measures, given that they are fuzzy measures characterized by (♯♯), it is obvious that the logics compatible with $\mathcal{P}^\Delta$ are the same that are compatible with $\mathcal{P}_\nabla$. Furthermore, notice that (♯♯) is equivalent to $\sum_{i=1}^{m} \frac{\mu(\varphi_i)}{m} \ominus \frac{k}{m} \leq \frac{n}{m}\mu(\varphi)$. It is then clear that $\frac{n}{m}\mu(\varphi) \leq 1$, and so, again, it makes sense to rely on t-norm based logics.

Let $\mathcal{L}$ be a t-norm based logic compatible with $\mathcal{P}_\nabla$. The logic $\mathcal{P}_\nabla(\mathcal{L})$ is obtained from $\mathcal{M}(\mathcal{L})$ by adding the rule (LP): if

$$\varphi \to \bigvee_{\binom{J\subseteq\{1,\dots,m\}}{|J|=k+n}} \bigwedge_{j\in J} \varphi_j, \quad \text{and} \quad \bigvee_{\binom{J\subseteq\{1,\dots,m\}}{|J|=k}} \bigwedge_{j\in J} \varphi_j$$

are propositional tautologies, then derive

$$\bigoplus_{j=1}^{m} \delta_m\kappa(\varphi_j) \ominus k.\delta_m \to n.\delta_m\kappa(\varphi),$$

if $\mathcal{L}$ is an extension (expansion of RŁ), or derive

$$\bigoplus_{j=1}^{m} \left(\overline{\tfrac{1}{m}} *_\pi \kappa(\varphi_j)\right) \ominus \overline{\tfrac{k}{m}} \to \left(\overline{\tfrac{n}{m}} *_\pi \kappa(\varphi)\right),$$

if $\mathcal{L}$ is an extension (expansion) of RPPŁ$'_\Delta$.

The semantics for $\mathcal{P}_\nabla$ is given by $\mathcal{P}_\nabla$-Kripke models, i.e. $\mathcal{M}$-Kripke models where μ is a lower probability measure.

We can now prove the following completeness theorem.

Theorem 2.6 *Let $\mathcal{L}$ be any t-norm based logic, and let $\mathcal{M}'$ be any class of measures among $\mathcal{P}$, $\mathcal{P}i$, $\mathcal{N}$, $\mathcal{P}^\Delta$, and $\mathcal{P}_\nabla$. If the following conditions are satisfied:*

1. *$\mathcal{L}$ is compatible with $\mathcal{M}'$,*
2. *$\mathcal{L}$ is (finitely) strongly standard complete,*

then $\mathcal{M}'(\mathcal{L})$ is (finitely) strongly standard complete.

2.3 Expansions: truth-constants and definability of conditional measures

The expressive power of the above logics can be significantly increased if we aim at representing other features of certain fuzzy measures. First of all, notice that relying on a t-norm based logic including rational truth-constants (see [9]) would allow to represent several statements concerning assessment of rational values like

- "the degree of uncertainty of φ is 0.8" as $\kappa(\varphi) \leftrightarrow \overline{0.8}$,
- "the degree of uncertainty of φ is at least 0.8" as $\overline{0.8} \to \kappa(\varphi)$,
- "the degree of uncertainty of φ is at most 0.8" as $\kappa(\varphi) \to \overline{0.8}$.

Not having truth constants would yield a purely qualitative treatment in which only comparative statements can be expressed. The advantage of the presence of truth constants will be made even clearer in next section.

Another increase of expressive power would consist in allowing the definition of conditional measures from the simple marginal measures represented in the logic.

The degree of confidence in the occurrence of an event might have to be changed when new information comes at hand. This results in an update of the sample space which is commonly treated in theories of uncertainty by the concept of conditioning.

Conditional probability. The updated probability measure $P(\cdot|\chi)$ (i.e. the probability of an event given the occurrence of χ) called *conditional probability*, is defined as $P(\varphi|\chi) = \frac{P(\varphi\wedge\chi)}{P(\chi)}$, provided that $P(\chi) > 0$. If $P(\chi) = 0$ the conditional probability remains then undefined.

In order to define conditioning from marginal probabilities we need division. Clearly, the only possibility to express division in t-norm based logics is given by logics containing the product implication. Hence, any extension (or expansion) of Łukasiewicz logic having

the product implication, like ŁΠ and ŁΠ$\frac{1}{2}$ (see [8]), represents a suitable choice.

Conditional possibility. In general, the conditional possibility $\Pi(\varphi|\chi)$ can be viewed as the solution to the equation $\Pi(\varphi \wedge \chi) = x * \Pi(\chi)$, where $*$ is a continuous t-norm (continuity guarantees the existence of a solution), and $\Pi(\varphi|\chi)$ is defined as the greatest solution. This would then be equivalent to the following equation: $\Pi(\varphi|\chi) = \Pi(\chi) \Rightarrow_* \Pi(\varphi \wedge \chi)$, where $\Rightarrow_*$ is the residuum of the t-norm $*$.

A classical treatment, proposed by Dubois and Prade [4], consists in taking the minimum t-norm, to obtain a qualitative definition. This, however, yields some technical problems when the sample space is infinite, given that Gödel implication is not continuous. This can be avoided by defining a probability-like conditioning by means of the product t-norm. Indeed, as shown in [3], not any t-norm can be used if we want the mapping $\Pi(\cdot|\chi)$ to be a possibility measure. If we rely on an arbitrary space, $*$ must be a strict t-norm, i.e. continuous, Archimedean and without zero-divisors [20]. If the universe is finite, $*$ needs not be Archimedean, and then we can recover the minimum t-norm.

Hence, two natural definitions for conditional possibility are obtained from the above equation by taking $\Rightarrow_*$ as the Gödel implication or the product implication. Hence, such types of derived conditioning can be framed in any extension (or expansion) of Gödel logic, the former, and in any extension (or expansion) of Product logic, the latter [15].

Notice that conditioning for necessity measure is not in general defined from marginal necessities, but it is derived from conditional possibilities (see [5], for the details). Furthermore, there is no clear notion of conditional lower or upper probability as derived from a single measure. Conditional lower and upper probabilities are rather defined from a set of probabilities (see [17]).

3 Rational Assessments: Compactness, Coherence and Consistency

In this section we lay out a link between consistency of modal theories and the coherence of rational assessments of fuzzy measures and conditional measures. In order to do so, we need some previous notions and results concerning satisfiability, compactness and consistency.

A detailed investigation of compactness of many logics based on continuous t-norms was presented in [2]. The notion of *satisfiability* proposed there generalizes the classical one by admitting various degrees of simultaneous satisfiability.

Definition 3.1 [2] For a set Γ of formulas in a t-norm based logic and $K \subseteq [0,1]$, we say that Γ is *K-satisfiable* if there exists an evaluation e such that $e(\varphi) \in K$ for all $\varphi \in \Gamma$. The set Γ is said to be *finitely K-satisfiable* if each finite subset of Γ is K-satisfiable. We say that a logic is *K-compact* if K-satisfiability is equivalent to finite K-satisfiability. A logic satisfies the *compactness property* if it is K-compact for each closed subset of $[0,1]$.

In particular we should mention that t-norm based logics only having continuous truth-functions, like Łukasiewicz Logic, do enjoy the compactness property.

Theorem 3.2 ([2]) *Let $\mathcal{L}$ be a given t-norm based logic whose connectives only have continuous truth-functions. Then $\mathcal{L}$ has the compactness property.*

The above result clearly still holds when we deal with theories in which the interpretations of all connectives correspond to continuous truth functions.

In many real-life situations assessments of uncertainty are not precisely made over a set of events with a specific algebraic structure. Still, such assessments must be required to be coherent, that is: they must satisfy the axioms of a fuzzy measure whenever they are extended over the whole Boolean algebra generated by those events.

Definition 3.3 Let $\mathcal{M}$ be a class of fuzzy measures, $\mathcal{C}$ be a countable set of events, and μ be a real-valued assessment defined on $\mathcal{C}$. We call μ a *$\mathcal{M}$-coherent fuzzy measure* if there is a fuzzy measure $\mu' \in \mathcal{M}$ over the Boolean algebra generated by $\mathcal{C}$ such that $\mu(\varphi) = \mu'(\varphi)$ for all $\varphi \in \mathcal{C}$.

It is clear that by relying on a t-norm based logic $\mathcal{L}$ in which rational truth constants are definable we can represent rational assessments w.r.t. to a fuzzy measure. This will allow us to show that checking the coherence of a rational assessment over a countable set of events is tantamount to checking consistency of a suitably defined theory in $\mathcal{M}(\mathcal{L})$.

First of all, a clarification has to be made. Here, given a class of fuzzy measures $\mathcal{M}'$, $\mathcal{M}'(\mathcal{L})$ will denote an extension of $\mathcal{M}(\mathcal{L})$ over a t-norm based logic with rational truth constants compatible with $\mathcal{M}'$ being complete w.r.t. to $\mathcal{M}'$-Kripke models. For instance, $\mathcal{M}'(\mathcal{L})$ might correspond to either $\mathcal{P}(\mathcal{L})$, $\mathcal{P}i(\mathcal{L})$, or $\mathcal{N}(\mathcal{L})$. Now, we need theories of the form $\Gamma = \{\kappa(\varphi_i) \leftrightarrow \overline{\alpha_i}\}$ in order to have models in which assessments of fuzzy measures are not only 1-valued. Of course, we cannot take into account real-valued assessments, since we only have rational truth-constants in

our language. Then we obtain that for any rational assessment its coherence is equivalent to the consistency of the respective theory in $\mathcal{M}'(\mathcal{L})$, given that its extension induces a $\mathcal{M}'$-Kripke structure which is a model of such a theory. However, there is an important restriction. Indeed, to obtain the mentioned result we need the logic $\mathcal{L}$ to have the compactness property, or the connective $\leftrightarrow$ to be continuous, since we need to exploit the above compactness results (see [22]). Since $\varphi \leftrightarrow \psi$ is defined as $(\varphi \leftrightarrow \psi)\&(\psi \leftrightarrow \varphi)$, it is obvious that both & and $\rightarrow$ must have continuous truth functions. Up to isomorphism, the only continuous t-norm having a continuous residuum is the Łukasiewicz t-norm. This implies, in this case, that $\mathcal{L}$ must be an extension (or expansion) of RPL [15].

Theorem 3.4 *Let $\theta = \{\mu^*(\varphi_i) = \alpha_i\}$ be a rational assessment, and let $\mathcal{M}'$ be a class of fuzzy measures. Suppose that the following conditions are satisfied:*

i. $\mathcal{L}$ is a t-norm based logic with rational truth constants

ii. $\mathcal{L}$ either has the compactness property or is an extension (or expansion) of RPL

iii. $\mathcal{L}$ has a finitary notion of derivability

iv. $\mathcal{L}$ is compatible with $\mathcal{M}'$

v. $\mathcal{M}'(\mathcal{L})$ is (finitely) strongly complete w.r.t. $\mathcal{M}'$-Kripke structures.

Then θ is $\mathcal{M}'$-coherent iff the theory $\Gamma_\theta = \{\kappa(\varphi_i) \leftrightarrow \overline{\alpha}_i\}$ is consistent in $\mathcal{M}'(\mathcal{L})$, i.e. $\Gamma_\theta \nvdash_{\mathcal{M}'(\mathcal{L})} \bar{0}$.

Given the above theorems, it is now easy to prove a compactness result for coherent assessments. This means that when we have a rational assessment to a countable set of events, such an assessment is coherent if and only if its restriction to each finite subset of that set also is coherent. Indeed, since any of such coherent restrictions can be translated into a theory which is consistent by Theorem 3.4, the whole corresponding theory is consistent, and consequently, again by Theorem 3.4, the corresponding assessment is coherent. Notice that this result concerns rational assessments of fuzzy measures only, and it is proved by purely logical means.

Theorem 3.5 *Let $\mathcal{C} = \{\varphi_i\}$ be a countable family of events, let $\theta = \{\mu^*(\varphi_i) = \alpha_i\}$ be a rational assessment over $\mathcal{C}$, and let $\mathcal{M}'$ be a class of fuzzy measures. Suppose that the following conditions are satisfied:*

i. $\mathcal{L}$ is a t-norm based logic with rational truth constants

ii. $\mathcal{L}$ either has the compactness property or is an extension (or expansion) of RPL

iii. $\mathcal{L}$ has a finitary notion of derivability

iv. $\mathcal{L}$ is compatible with $\mathcal{M}'$

v. $\mathcal{M}'(\mathcal{L})$ is (finitely) strongly complete w.r.t. $\mathcal{M}'$-Kripke structures.

Let $\theta_{\downarrow_\mathcal{I}}$ be the restriction of θ to each finite $\mathcal{I}$, such that $\mathcal{I} \subset \mathcal{C}$. Then:

θ is $\mathcal{M}'$-coherent iff $\theta_{\downarrow_\mathcal{I}}$ is $\mathcal{M}'$-coherent for every $\mathcal{I}$.

4 Final Remarks

As far as we know, the only comprehensive logical treatment of uncertainty measures is the one proposed by Halpern [17]. In such a work, a modal operator ℓ, standing for likelihood, is applied over Boolean formulas, so that $\ell(\varphi)$ is a likelihood term interpreted as "the uncertainty of φ". A basic likelihood formula is an expression of the form $a_1\ell(\varphi_1) + \cdots + a_k\ell(\varphi_k) > b$, where $a_1, \ldots, a_k, b$ are real numbers and $k \geq 1$. Likelihood formulas are Boolean combinations of basic likelihood formulas. The language resulting from the foregoing description is called $\mathcal{L}^{QU}$, where QU stands for *quantitative uncertainty*. From $\mathcal{L}^{QU}$ we can then build up a logic for a class of measures, by introducing the adequate axioms. Given that likelihood formulas are linear inequalities, we also have to introduce all substitution instances of valid linear inequality formulas as axioms. The semantics for $\mathcal{L}^{QU}$ is given by Kripke models $\langle W, \mathcal{U}, e, \mu\rangle$ equipped with a measure belonging to a certain class. Halpern showed how to treat probabilities, possibilities, belief functions and upper probabilities obtaining sound and complete axiomatizations.

This approach is strongly based on the presence of axioms of linear inequalities which allow to represent basic operations between formulas. Our approach exploits the advantage given by the fact that in t-norm based logics the operations associated to the evaluation of the connectives are functions defined over $[0, 1]$, which correspond, directly or up to some combinations, to operations used to compute degrees of uncertainty. Then such algebraic operations can be embedded in the connectives of the many-valued logical framework, resulting in clear and elegant formalizations. Given that there is a whole family of t-norm based logics, the choice of the logic to exploit to represent a specific class of measures will clearly depend on the operations we need to represent. This permits to avoid the introduction of instances of linear inequalities, since they are directly given by the functions associated to the connectives of some logics. For instance,

Łukasiewicz logic and its expansions allow the representation of piecewise linear functions, and hence are the most suitable choice for the representation of linear equalities and inequalities. Moreover, in the case of possibility and necessity measures, for instance, we might not even need to use linear inequalities. What we need are just the minimum and the maximum operators plus the possibility of expressing comparative statements which is immediately given by the implication connective.

Therefore, in our treatment we do not need to add axioms for having peculiar operations, since the possible presence of those operations just relies on an adequate choice of the base logic. Having functions embedded in our logics also implies that some properties of the chosen logic might be inherited by the kind of measures we define in it. Indeed, once proven the connection between the consistency of a suitably defined theory in our logic and the coherence of the related assessment, properties like compactness for those assessments can be easily studied by purely logical means.

To conclude, we would like to point out that we do not deem that the t-norm based approach is better than the others. The study carried out in this work might be just an overt example of the advantages t-norm based logics can provide.

Acknowledgement

The author acknowledges partial support of the Spanish project MULOG TIN2004-07933-C03-01.

References

[1] ANGER B., LEMBCKE J. Infinitely subadditive capacities as upper envelopes of measures. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 68, 403–414, 1985.

[2] CINTULA P. AND NAVARA M. Compactness of fuzzy logics. *Fuzzy Sets and Systems* 143, 59-73, 2004.

[3] DE BAETS B., TSIPORKOVA E., AND MESIAR R. Conditioning in possibility theory with strict order norms. *Fuzzy Sets and Systems* 106, 121–129, 1999.

[4] DUBOIS D. AND PRADE H. The logical view of conditioning and its application to possibility and evidence theories. *Int. J. of Approximate Reasoning*, 4, 23–46, 1990.

[5] DUBOIS, D. AND PRADE H. Possibility theory: Qualitative and quantitative aspects. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, D. M.Gabbay and Ph. Smets (eds), Vol. 1, 169–226, Kluwer Academic Publisher, Dordrecht, The Netherlands, 1998.

[6] DUBOIS, D. AND PRADE H. Possibility Theory, probability theory and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32, 35–66, 2001.

[7] ESTEVA F. AND GODO L. Monoidal t-norm based logic: towards a logic for left-continuous t-norms. *Fuzzy Sets and Systems*, 124, 271–288, 2001.

[8] ESTEVA F., GODO L. AND MONTAGNA F. The ŁΠ and ŁΠ$\frac{1}{2}$ logics: two complete fuzzy logics joining Łukasiewicz and product logic. *Archive for Mathematical Logic*, 40, 39–67, 2001.

[9] ESTEVA F., GODO L. AND NOGUERA C. Survey on expansions of t-norm based logics with truth-constants In *Proceedings of the Linz Symposium 2005*, to appear.

[10] FLAMINIO T. AND MONTAGNA F. A logical and algebraic treatment of conditional probability. *Archive for Mathematical logic*, 44, 245–262, 2005.

[11] GERLA B. *Many-valued Logics of Continuous T-norms and their Functional Representation*. PhD Thesis, Università di Milano, Italy, 2001.

[12] GODO L., ESTEVA F. AND HÁJEK P. Reasoning about probability using fuzzy logic. *Neural Network World*, 10, No. 5, 811–824, 2000.

[13] GODO L., ESTEVA F. AND HÁJEK P. A fuzzy modal logic for belief functions. *Fundamenta Informaticae*, Vol. 57, Numbers 2-4, 127–146, 2003.

[14] GODO L., MARCHIONI E. Coherent conditional probability in a fuzzy logic setting. *Logic Journal of the IGPL*, Vol. 14, Number 3, 457–481, 2006.

[15] HÁJEK P. *Metamathematics of fuzzy logic,* Kluwer Academic Publisher, Dordrecht, The Netherlands, 1998.

[16] HÁJEK P., GODO L. AND ESTEVA F. Fuzzy Logic and Probability. In *Proceedings of the 11 th. Conference Uncertainty in Artificial Intelligence 95 (UAI'95)*, 237–244, 1995.

[17] HALPERN J.Y. *Reasoning about Uncertainty*. The MIT Press, Cambridge Massachusetts, 2003.

[18] HALPERN J.Y., PUCELLA R. A logic for reasoning about upper probabilities. *Journal of Artificial Intelligence Research*, 17, 57–81, 2002.

[19] HORČÍK R. AND CINTULA P. Product Łukasiewicz logic. *Archive for Mathematical Logic*, 43(4), 477–503, 2004.

[20] KLEMENT E. P., MESIAR R. AND PAP E. *Triangular Norms*. Kluwer Academic Publisher, Dordrecht, The Netherlands, 2000.

[21] KOLMOGOROV A.N. *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York, 1960.

[22] MARCHIONI E. Possibilistic conditioning framed in fuzzy logics. *International Journal of Approximate Reasoning*, Vol. 43, Issue 2, 133–165, 2006.

[23] SUGENO M. *Theory of fuzzy integrals and its applications*. Phd. Dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[24] ZADEH L. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3–28, 1978.

Modal systems based on many-valued logics

F. Bou
IIIA - CSIC
Campus UAB s/n
08193, Bellaterra, Spain
fbou@iiia.csic.es

F. Esteva
IIIA - CSIC
Campus UAB s/n
08193, Bellaterra, Spain
esteva@iiia.csic.es

L. Godo
IIIA - CSIC
Campus UAB s/n
08193, Bellaterra, Spain
godo@iiia.csic.es

Abstract

We propose a general semantic notion of modal many-valued logic. Then, we explore the difficulties to characterize this notation in a syntactic way and analyze the existing literature with respect to this framework.

Keywords: modal many-valued logic, modal fuzzy logic.

1 Introduction

The purpose of this research is the search of a syntactic notion of *modal many-valued logic*[1] that generalizes the notion of (normal) modal logic [5, 1]. Our search is motivated by semantic issues. That is, we understand modal many-valued logics as logics defined by Kripke frames (possibly with many-valued accessibility relations) where every world follows the rules of a many-valued logic, this many-valued logic being the same for every world. The reader will find the details of this approach in Section 2.

Throughout the paper we will show the reader the difficulties of this search and we will try to specify which conditions should satisfy this syntactic notion. We will also review the works in the literature that fits inside our framework.

Unfortunately, we have not been able to found a syntactic characterization of the notion of modal many-valued logic and so the question remains open.

2 A semantic approach

In this section we start giving the definition of the modal many valued logic $Log_{\Box}(\mathbf{A}, \mathsf{F})$ associated with an algebra $\mathbf{A}$ and a class of $\mathbf{A}$-valued Kripke frames F.

[1] In particular, modal fuzzy logics will be inside this class.

The language of this new logic is, by definition, the propositional language generated by a set Var of propositional variables[2] together with the connectives given by the algebraic signature of $\mathbf{A}$ expanded with a new unary connective: the necessity[3] operator $\Box$. The set of formulas of the resulting language will be denoted by $Fm_{\Box}$.

We point out that the intended meaning of the universe A is a set of truth-values. The only requirements in our definition will be that the algebra $\mathbf{A}$ is a complete lattice, and that the algebraic language of $\mathbf{A}$ contains, besides meet $\wedge$ and join $\vee$, a constant 1 and an implication $\rightarrow$.

We stress that these conditions are quite weak and a lot of well-known algebras satisfy them, for instance, complete FL-algebras [20] and complete BL-algebras [13]. Hence, in particular we can consider that $\mathbf{A}$ is any of the three basic continuous t-norm algebras: *Lukasiewicz algebra* $[\mathbf{0}, \mathbf{1}]_{\mathbf{L}}$, *product algebra* $[\mathbf{0}, \mathbf{1}]_{\mathbf{\Pi}}$ and *Gödel algebra* $[\mathbf{0}, \mathbf{1}]_{\mathbf{G}}$. We also note that due to the fact that the free algebra with countable generators (i.e., the Lindenbaum-Tarski algebra) of any of the previous varieties of algebras is not complete it is not included in our framework.

An $\mathbf{A}$-*valued Kripke frame* is a pair $\mathfrak{F} = \langle W, R \rangle$ where W is a set (of *worlds*) and R is a binary relation valued in A (i.e., $R : W \times W \longrightarrow A$) called *accessibility relation*. It is said that the Kripke frame is *classical* in case that the range of R is included in $\{0, 1\}$[4]. Whenever $\mathbf{A}$ is fixed, we will denote by Fr and CFr the classes of all $\mathbf{A}$-valued Kripke frames and all $\mathbf{A}$-valued classical Kripke frames. For the rest of the paper we will mainly focus on these two classes since they provide in some sense minimal logics.

Before introducing $Log_{\Box}(\mathbf{A}, \mathsf{F})$ we need to define what

[2] In most cases it is assumed that $Var = \{p_0, p_1, p_2, \ldots\}$.

[3] Later on we will give some ideas about how to develop these ideas with the possibility operator $\Diamond$.

[4] Here 0 means the minimum of $\mathbf{A}$ and 1 its maximum.

is an **A**-valued Kripke model. An **A**-*valued Kripke model* is a triple $\mathfrak{M} = \langle W, R, e\rangle$ where $\langle W, R\rangle$ is an **A**-valued Kripke frame and e is a map, called *valuation*, assigning to each variable in Var and each world in W an element of A. The map e can be uniquely extended to a map $\bar{e} : Fm_{\Box} \times W \longrightarrow A$ satisfying that:

- $\bar{e}$ is an algebraic homomorphism, in its first component, for the connectives in the algebraic signature of **A**, and
- $\bar{e}(\Box\varphi, w) = \bigwedge\{R(w, w') \to \bar{e}(\varphi, w') : w' \in W\}$.

Although the functions e and $\bar{e}$ are different there will be no confusion between them, and so sometimes we will use the same notation e for both.

Following the same definitions than in the Boolean modal case [5, 1] it is clear how to define *validity* of a $Fm_{\Box}$-formula in an **A**-valued Kripke model and in an **A**-valued Kripke frame.

Now we are ready to introduce the *modal many-valued logic* $Log_{\Box}(\mathbf{A}, \mathsf{F})$. It is defined as the set of formulas $\varphi \in Fm_{\Box}$ satisfying that for every **A**-valued Kripke model $\langle W, R, e\rangle$ over a frame $\langle W, R\rangle$ in F and for every world w in W, it holds that $e(\varphi, w) = 1$.

Remark 1 *For the sake of simplicity in this paper we restrict ourselves to adding the necessity operator $\Box$, but analogously we could have considered a possibility operator ruled by the condition*[5]

$$e(\Diamond\varphi, w) = \bigvee\{R(w, w') \odot e(\varphi, w') : w' \in W\}.$$

Note 2 *We stress that for the case that* **A** *is the Boolean algebra of two elements all previous definitions correspond to the standard terminology in the field of modal logic (cf. [5, 1]). As far as the authors know the first one to talk about this way of extending the valuation e into the modal many-valued realm was* M. FITTING *in [9].*

Up to now we have considered a logic as a set of formulas. Besides this way to consider logics, it is also common to consider them as consequence relations, e.g., [2]. Following this approach next we define two different consequence relations.

The *modal many-valued local consequence* $\models_{l(\mathbf{A},\mathsf{F})}$ associated with an algebra **A** and a class of **A**-valued Kripke frames F is defined by the following equivalence:

$$\Gamma \models_{l(\mathbf{A},\mathsf{F})} \varphi \quad \text{iff}$$

for every **A**-valued Kripke model $\langle W, R, e\rangle$ over a frame $\langle W, R\rangle$ in F and for every world w in W, it holds that if $e(\gamma, w) = 1$ for every $\gamma \in \Gamma$, then $e(\varphi, w) = 1$.

The *modal many-valued global consequence* $\models_{g(\mathbf{A},\mathsf{F})}$ associated with an algebra **A** and a class of **A**-valued Kripke frames F is given by the following definition:

$$\Gamma \models_{g(\mathbf{A},\mathsf{F})} \varphi \quad \text{iff}$$

for every **A**-valued Kripke model $\langle W, R, e\rangle$ over a frame $\langle W, R\rangle$ in F, it holds that if $e(\gamma, w) = 1$ for every $\gamma \in \Gamma$ and every world w in W, then $e(\varphi, w) = 1$ for every world w in W.

We point out that the set of theorems of both consequence relations is precisely the set $Log_{\Box}(\mathbf{A}, \mathsf{F})$.

3 Differences with the modal Boolean case

General Considerations. Let us assume that we have fixed an algebra **A** satisfying the previous requirements. In order to simplify our discussion we will also assume that it is a FL_{ew}-algebra [20][6], i.e., a residuated lattice. This hypothesis will allow us to use the residuation law.

In order to find a successful syntactic definition of the notions introduced in the previous section[7] first of all we would need to settle a completeness theorem for the logics introduced in the previous section. In particular, we should know how to axiomatize the minimal one $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$. What formulas must we add to an axiomatization of the many-valued logic defined by **A** in order to obtain a complete axiomatization of $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$?

The fact that the famous modal axiom (K) (sometimes called *normality* axiom)

$$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi) \qquad (\mathrm{K})$$

does not in general belong to $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$ is what makes difficult even to suggest an axiomatization of $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$. As a simple counterexample we can consider the logic $Log_{\Box}([\mathbf{0}, \mathbf{1}]_{\mathbf{L}}, \mathsf{Fr})$ and the Kripke model given by $W = \{a, b\}$, $R(a, a) = 1$, $R(a, b) = 1/2$, $e(p_0, a) = 1$, $e(p_0, b) = 1/2$, $e(p_1, a) = 1$ and $e(p_1, b) = 0$. Then, $e(\Box(p_0 \to p_1) \to (\Box p_0 \to \Box p_1), a) = 1/2$.

[5] The connective $\odot$ is what sometimes is called in the literature fusion, multiplicative conjunction, etc. (see [13, 20]).

[6] We remind the reader that in particular all BL-algebras [13] are FL_{ew}-algebras.

[7] In the modal Boolean case it is well-known the existence of modal logics that are Kripke frame incomplete. Hence, the searched definition of modal many-valued logic will have to include more logics that the ones introduced in Section 2.

It is easy to check that the *necessity rule*, from φ follows $\Box\varphi$, holds for $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$. Another property that holds for $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$ is the monotonicity of the necessity operator, i.e., if $\varphi \to \psi$ is in the logic then also $\Box\varphi \to \Box\psi$ is in the logic. Moreover, it is possible to see that

$$(\Box\varphi \wedge \Box\psi) \leftrightarrow \Box(\varphi \wedge \psi)$$

is valid under our semantics. However, in general

$$(\Box\varphi \odot \Box\psi) \to \Box(\varphi \odot \psi)$$

and

$$(\Box(\varphi \to \psi) \odot \Box\varphi) \to \Box\psi$$

fail. The reader can notice that the last formula is equivalent to (K) thanks to the residuation law.

Although in general (K) does not belong to $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$, let us remark two particular cases where (K) holds. For these two cases the difficulties disappear quite a lot as we will see in examples of Section 4. The first one is when $\odot$ and $\wedge$ coincide in the algebra $\mathbf{A}$. In particular this means that (K) belongs to $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{Fr})$. And the second one is when F is the class of classical Kripke frames CFr, i.e., for any algebra $\mathbf{A}$ all $\mathbf{A}$-valued classical Kripke frames satisfy the normality condition. In particular this means that (K) belongs to $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{L}}, \mathsf{CFr})$ and $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{\Pi}}, \mathsf{CFr})$.

Transfer Properties. We are going to show with three counterexamples that in general metalogical properties are lost when we move from the modal Boolean case to the modal many-valued one. This implies that in order to attack future problems for modal many-valued logics we will need to introduce new machinery, what makes this new field a really exciting and appealing one.

First of all we point out that the fact that two algebras $\mathbf{A}$ and $\mathbf{B}$ generate the same variety does not imply that $Log_{\Box}(\mathbf{A}, \mathsf{CFr}) = Log_{\Box}(\mathbf{B}, \mathsf{CFr})$. As a counterexample we can consider $\mathbf{A}$ as the standard Gödel algebra $[\mathbf{0},\mathbf{1}]_{\mathbf{G}}$ and $\mathbf{B}$ as its subalgebra of universe $\{0\} \cup [1/2, 1]$. It is not hard to see that $\Box\neg\neg p \to \neg\neg\Box p$ belongs to $Log_{\Box}(\mathbf{B}, \mathsf{CFr})$ while fails in $Log_{\Box}(\mathbf{A}, \mathsf{CFr})$.

Secondly we notice that it can happen that two classes F_1 and F_2 of classical Kripke frames have different modal many-valued logics for an algebra $\mathbf{A}$ while for the case of the Boolean algebra of two elements they share the same logic. Why? It is well-known that the modal logic $S4$ is generated both by the class F_1 of finite quasi-orders (perhaps fails the antisymmetric property) and the class F_2 of infinite partial orders. However, $\Box\neg\neg p \to \neg\neg\Box p$ belongs to $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{F}_1)$ while fails in $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{F}_2)$.

Table 1: Fitting systems ($A = \{t_1, \ldots, t_n\}$)

$$\varphi \Rightarrow \varphi$$

$$\frac{\Gamma \Rightarrow \Delta}{\Gamma, \Gamma' \Rightarrow \Delta, \Delta'}$$

$$\frac{\Gamma \Rightarrow \Delta, \varphi \qquad \Gamma, \varphi \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

$$\varphi_1 \supset \varphi_2, \varphi_2 \supset \varphi_3 \Rightarrow \varphi_1 \supset \varphi_3$$

$$\frac{\Gamma, t_i \supset \varphi \Rightarrow \Delta, t_i \supset \psi \quad \text{for every } i \in \{1, \ldots, n\}}{\Gamma \Rightarrow \Delta, \varphi \supset \psi}$$

$$\frac{\Gamma, \psi \supset t_i \Rightarrow \Delta, \varphi \supset t_i \quad \text{for every } i \in \{1, \ldots, n\}}{\Gamma \Rightarrow \Delta, \varphi \supset \psi}$$

$$\Rightarrow \varphi \wedge \psi \supset \varphi$$
$$\Rightarrow \varphi \wedge \psi \supset \psi$$
$$\varphi_3 \supset \varphi_1, \varphi_3 \supset \varphi_2 \Rightarrow \varphi_3 \supset (\varphi_1 \wedge \varphi_2)$$

$$\Rightarrow \varphi \supset \varphi \vee \psi$$
$$\Rightarrow \psi \supset \varphi \vee \psi$$
$$\varphi_1 \supset \varphi_3, \varphi_2 \supset \varphi_3 \Rightarrow (\varphi_1 \vee \varphi_2) \supset \varphi_3$$

$$\Rightarrow t_i \supset t_j \quad \text{, if } t_i \leq t_j$$
$$t_i \supset t_j \Rightarrow \quad \text{, if } t_i \not\leq t_j$$

$$(\varphi_1 \wedge \varphi_2) \supset \varphi_3 \Rightarrow \varphi_1 \supset (\varphi_2 \supset \varphi_3)$$
$$\varphi_1 \supset (\varphi_2 \supset \varphi_3) \Rightarrow (\varphi_1 \wedge \varphi_2) \supset \varphi_3$$

$$\varphi \Rightarrow (\mathit{true} \supset \varphi)$$
$$(\mathit{true} \supset \varphi) \Rightarrow \varphi$$

$$\frac{\Rightarrow \varphi}{\Rightarrow \Box\varphi}$$

$$\Rightarrow \Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$$
$$\Rightarrow (t_i \supset \Box\varphi) \equiv \Box(t_i \supset \varphi), \text{ if } i \in \{1, \ldots, n\}$$

Lastly we remark that it is possible to have that $Log_{\Box}(\mathbf{2}, \mathsf{F})$ enjoys the finite Kripke frame property while $Log_{\Box}(\mathbf{A}, \mathsf{F})$ does not. A counterexample is given by the standard Gödel algebra $[\mathbf{0},\mathbf{1}]_{\mathbf{G}}$ and the class F of classical quasi-orders. The failure of the finite Kripke frame property of $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{F})$ is witnessed, for instance, by the formula $\Box\neg\neg p \to \neg\neg\Box p$.

4 Examples in the literature

In the last years there has been a growing number of papers about combining modal and many-valued logics. Some approaches differ from ours, like [6, 19], but others stay as particular cases of our framework. Among the ones that fit in our framework we can cite [9, 10, 14, 11, 12, 15, 4].

Next we will discuss the known axiomatizations in the literature of logics of the form $Log_{\Box}(\mathbf{A}, \mathsf{F})$ where $\mathbf{A}$ is non Boolean and F is the class of all Kripke frames or

the class of all classical Kripke frames[8]. Indeed, the only known ones are for logics satisfying axiom (K), i.e., the authors are unaware of any axiomatization for a case where axioms (K) fails. This remains as a challenge.

A is a finite Heyting algebra. This case was considered by M. FITTING in [10, Section6]. The language includes constants for every element of the fixed algebra $\mathbf{A}$ (i.e., for every truth value), what simplifies the proofs and allows to give a unified presentation of the calculus to axiomatize $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$. The last statement refers to the fact that all these calculus share the same schemes without constants. The calculus is given using sequents and can be found in Table 1. Completeness of this calculus means that $Log_{\Box}(\mathbf{A}, \mathsf{Fr})$ coincides with the set of formulas $\varphi \in Fm_{\Box}$ such that the sequent $\Rightarrow \varphi$ is derivable using the calculus in Table 1. We notice that using the constants it is very easy to see that $Log_{\Box}(\mathbf{A}, \mathsf{Fr}) \neq Log_{\Box}(\mathbf{A}, \mathsf{CFr})$. Other papers that study these cases are [17, 18, 16].

$\mathbf{Log}_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{Fr})$. This case has been studied by X. CAICEDO and R. RODRÍGUEZ in [4]. They have proved that this logic is axiomatized by the calculus given in Table 2. The proof is based on the construction of a canonical model[9], which indeed is classical. From here it follows that for every class of Kripke frames F, it holds that $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{Fr}) = Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{CFr})$[10]. Therefore, for the case of $[\mathbf{0},\mathbf{1}]_{\mathbf{G}}$ we already know how to introduce the notion of *modal many-valued logic*: it is any set of $Fm_{\Box}$-formulas that contains the formulas in Table 2 and is closed under the rules in Table 2.

$\mathbf{Log}_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{\L}}, \mathsf{CFr})$. The recent paper [15] by G. HANSOUL and B. TEHEUX axiomatizes the normal modal logic $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{\L}}, \mathsf{CFr})$ with the infinite calculus given in Table 3. The proof is based on the construction of a classical canonical model. Surprisingly this proof does not need the presence in the language of constants for every truth value. The trick to avoid the introduction of constants is based on a result of [21] (see [15, Definition 5.3]).

A slightly different approach. As we have claimed before it is unknown how to manage the resulting non-normal logics. One possibility to avoid this difficulty is to introduce graded modalities $\Box_t$ (where $t \in \mathbf{A}$) corresponding to the cuts of the many-valued accessibility relation, i.e., using

$$e(\Box_t \varphi, w) = \bigwedge \{e(\varphi, w') : R(w, w') \geq t\}$$

to extend the valuation. Then, it is easy to see that all modalities $\Box_t$ are normal. We notice that in some particular cases, axiomatizations for these graded modalities has been found in the literature (see for instance [8, 22, 3]). The case considered in [3] corresponds to consider the n-valued Łukasiewicz chain algebra $\mathbf{L_{n-1}}$ (see [7]) and having constants in the language for every element in the n-valued Łukasiewicz algebra. The axiomatization given in [3] is shown in Table 4. An interesting fact about this case is that $\Box$ is definable in the new language because

$$(\Box\varphi) \leftrightarrow \bigwedge \{t \to \Box_t \varphi : t \in \mathbf{L}_{n-1}\}$$

is valid under our semantics.

Table 2: Inference Rules of $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{Fr})$

$$(\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))$$
$$\varphi \to (\psi \to \varphi)$$
$$(\varphi \wedge \psi) \to (\psi \wedge \varphi)$$
$$(\varphi \wedge (\psi \wedge \chi)) \to ((\psi \wedge \varphi) \wedge \chi)$$
$$(\varphi \to (\psi \to \chi)) \leftrightarrow ((\varphi \wedge \psi) \to \chi)$$
$$((\varphi \to \psi) \to \chi) \to (((\psi \to \varphi) \to \chi) \to \chi)$$
$$0 \to \varphi$$
$$\varphi \to (\varphi \wedge \varphi)$$
$$\neg\varphi \leftrightarrow (\varphi \to 0)$$
$$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$$
$$\neg\neg\Box\varphi \to \Box\neg\neg\varphi$$
From φ infer $\Box\varphi$
From φ and $\varphi \to \psi$, infer ψ

Table 3: Inference Rules of $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{\L}}, \mathsf{CFr})$

$$(\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))$$
$$\varphi \to (\psi \to \varphi)$$
$$((\varphi \to \psi) \to \psi) \to ((\psi \to \varphi) \to \varphi)$$
$$(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$$
$$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$$
$$\Box(\varphi \odot \varphi) \leftrightarrow (\Box\varphi \odot \Box\varphi)$$
$$\Box(\varphi \oplus \varphi) \leftrightarrow (\Box\varphi \oplus \Box\varphi)$$
$$\Box(\varphi \oplus \varphi^m) \leftrightarrow (\Box\varphi \oplus (\Box\varphi)^m) \text{ for every } m \geq 0$$
From φ infer $\Box\varphi$
From φ and $\varphi \to \psi$, infer ψ

5 Main Open Problems

In opinion of the authors the main open problems in this field are the search of axiomatizations for $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{\L}}, \mathsf{Fr})$ and $Log_{\Box}([\mathbf{0},\mathbf{1}]_{\mathbf{\Pi}}, \mathsf{Fr})$ in case they are recursively axiomatizable. The main difficulties here are the lack of normality of these logics.

[8] Hence, we do not consider cases where the class of frames satisfy some extra conditions, e.g., reflexive and $\odot$-transitive frames. The reason why we do not talk about them is because we want to consider the minimal logics.

[9] This technique also gives strong completeness in the sense that the caculus in Table 2 axiomatizes $\models_{l([\mathbf{0},\mathbf{1}]_{\mathbf{G}}, \mathsf{Fr})}$.

[10] We stress that this does not contradict facts like that $\Box p \to p$ and $\Box\neg p \to \neg p$ define, over $[\mathbf{0},\mathbf{1}]_{\mathbf{G}}$, the same class of classical Kripke frames while they define different classes of Kripke frames.

Table 4: Inference Rules for $\Box_0, \ldots, \Box_1$ ($t \in L_{n-1}$)

$$(\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))$$
$$\varphi \to (\psi \to \varphi)$$
$$((\varphi \to \psi) \to \psi) \to ((\psi \to \varphi) \to \varphi)$$
$$(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$$
$$(\neg\varphi) \leftrightarrow (\varphi \to 0)$$
$$(\varphi \odot \psi) \leftrightarrow \neg(\varphi \to \neg\psi)$$
$$n.\varphi \to (n-1).\varphi$$
$$(m.\varphi^{m-1})^n \leftrightarrow (n.\varphi^m), \quad 2 \le m \le n-2 \text{ and } m \not| (n-1)$$
$$(t_i \to t_j) \leftrightarrow t_k, \quad \text{if } t_k = t_i \to t_j$$
$$\Box_t(\varphi \to \psi) \to (\Box_t\varphi \to \Box_t\psi)$$
$$(\Box_t\varphi \odot \Box_t\psi) \to \Box_t(\varphi \odot \psi)$$
$$\Box_{t_i}\varphi \to \Box_{t_j}\varphi, \quad \text{if } t_i \le t_j$$
$$\Box_0\varphi \to \varphi$$
$$\neg\Box_0\neg\varphi \to \Box_0\neg\Box_0\neg\varphi$$
$$\Box_{t_i}(t_j \to \varphi) \leftrightarrow (t_j \to \Box_{t_i}\varphi)$$
From φ infer $\Box_t\varphi$
From φ and $\varphi \to \psi$, infer ψ

Once there is an axiomatization for them (if any) it seems easy to find the right definition of modal many-valued logic. And once we know the definition of the class of modal many-valued logics the next step will be their study with all possible techniques: algebras[11], Kripke frames, Kripke models, sequent calculus, etc

References

[1] P. Blackburn, M. de Rijke, and Y. Venema. *Modal logic*. Number 53 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, 2001.

[2] W. J. Blok and D. Pigozzi. *Algebraizable logics*, volume 396 of *Mem. Amer. Math. Soc.* A.M.S., Providence, January 1989.

[3] F. Bou, F. Esteva, L. Godo, and R. Rodríguez. n-Łukasiewicz modal logic. Manuscript, 2007.

[4] X. Caicedo and R. Rodríguez. A Gödel similarity-based modal logic. Manuscript. A shortened version was published as "A Gödel modal logic", in: Proc. of Logic, Computability and Randomness 2004. Cordoba, Argentina, 2007.

[5] A. Chagrov and M. Zakharyaschev. *Modal Logic*, volume 35 of *Oxford Logic Guides*. Oxford University Press, 1997.

[6] A. Ciabattoni, G. Metcalfe, and F. Montagna. Adding modalities to MTL and its extensions. Manuscript. To appear in the Proceedings of the Linz Symposium 2005, 2005.

[11] Remember that extensions of algebraizable logics are algebraizable [2].

[7] R. Cignoli, I. M. L. D'Ottaviano, and D. Mundici. *Algebraic foundations of many-valued reasoning*, volume 7 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 2000.

[8] F. Esteva, P. Garcia, L. Godo, and R. Rodríguez. A modal account of similarity-based reasoning. *International Journal of Approximate Reasoning*, 16(3-4):235–260, 1997.

[9] M. Fitting. Many-valued modal logics. *Fundamenta Informaticae*, 15:235–254, 1992.

[10] M. Fitting. Many-valued modal logics, II. *Fundamenta Informaticae*, 17:55–73, 1992.

[11] S. Frankowski. Definable classes of many valued Kripke frames. *Bulletin of the Section of Logic*, 35(1):27–36, 2006.

[12] S. Frankowski. General approach to many valued Kripke models. *Bulletin of the Section of Logic*, 35(1):11–26, 2006.

[13] P. Hájek. *Metamathematics of fuzzy logic*, volume 4 of *Trends in Logic—Studia Logica Library*. Kluwer Academic Publishers, Dordrecht, 1998.

[14] P. Hájek and D. Harmancová. A many-valued modal logic. In *Proceedings IPMU'96. Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1021–1024, Granada, 1996. Universidad de Granada, The IEEE Computer Society Press.

[15] G. Hansoul and B. Teheux. Completeness results for many-valued Łukasiewicz modal systems and relational semantics, 2006. Available at http://arxiv.org/abs/math/0612542.

[16] C. D. Koutras. A catalog of weak many-valued modal axioms and their corresponding frame classes. *Journal of Applied Non-Classical Logics*, 13(1):47–72, 2003.

[17] C. D. Koutras and C. Nomikos. The computational complexity of satisfiability in many-valued modal logic. In *Proceedings of the 3rd Panhellenic Logic Symposium*, Anogia, Greece, July 2001.

[18] C. D. Koutras, C. Nomikos, and P. Peppas. Canonicity and completeness results for many-valued modal logics. *Journal of Applied Non-Classical Logics*, 12(1):7–41, 2002.

[19] A. M. Mironov. Fuzzy modal logics. *Journal of Mathematical Sciences*, 128(6):3641–3483, 2005.

[20] H. Ono. Substructural logics and residuated lattices - an introduction. In V. F. Hendricks and J. Malinowski, editors, *50 Years of Studia Logica*, volume 21 of *Trends in Logic—Studia Logica Library*, pages 193–228. Dordrecht, 2003.

[21] P. Ostermann. Many-valued modal propositional calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 34(4):343–354, 1988.

[22] N. Y. Suzuki. Kripke frame with graded accessibility and fuzzy possible world semantics. *Studia Logica*, 59(2):249–269, 1997.

Session 3

Copulas – E. P. Klement and F. Durante

Different types of convexity and concavity for copulas

E. Alvoni
Dipartimento MATEMATES
Università di Bologna
Viale Filopanti 5
I-40126 Bologna, Italy
elisabetta.alvoni@unibo.it

F. Durante
Department of Knowledge-Based
Mathematical Systems
Johannes Kepler University
A-4040 Linz, Austria
fabrizio.durante@jku.at

P.L. Papini
Dipartimento di Matematica
Università di Bologna
Piazza Porta S.Donato 5
I-40126 Bologna, Italy
papini@dm.unibo.it

C. Sempi
Dipartimento di Matematica
Ennio De Giorgi
Università del Salento
I-73100 Lecce, Italy
carlo.sempi@unile.it

Abstract

We present different notions of convexity and concavity for copulas and we study the relationships among them.

Keywords: Copulas, Schur-concavity, Quasi-concavity.

1 Introduction

A function $C\colon [0,1]^2 \to [0,1]$ is called *copula* if, for every x, x', y, y' in $[0,1]$, for $x \le x', y \le y'$,

$$C(x,0) = C(0,x) = 0,\ C(x,1) = C(1,x) = x, \quad (1)$$

$$C(x',y') - C(x,y') - C(x',y) + C(x,y) \ge 0. \quad (2)$$

In other words, a copula is a binary aggregation operator that has neutral element 1 and which satisfies inequality (2), called the *2–increasing property* [16]. In 1959, Abe Sklar introduced this concept in order to link a bivariate distribution function to its marginals [18]. Since then, copulas have played an important rôle not only in probability theory and statistics, but also in multi–criteria decision making and in fuzzy set theory. Classical examples of copulas are $\Pi(x,y) = xy$, $M(x,y) = \min(x,y)$ and $W(x,y) = \max(x+y-1, 0)$. In particular, for each copula C, we have $W \le C \le M$ pointwise on $[0,1]^2$. The class of copulas shall be denoted by $\mathcal{C}$, which is a convex and compact (with respect to the L^∞ norm) subset in the class of all continuous functions from $[0,1]^2$ into $[0,1]$.

The following two construction of copulas will be considered in the sequel: ordinal sum and ϕ–transformations; they are recalled here.

Let $(C_i)_{i\in\mathcal{I}}$ be a family of copulas indexed by the (at most) countable set $\mathcal{I}$. Let $(]a_i, b_i[)_{i\in\mathcal{I}}$ be a family of pairwise disjoint subintervals of $[0,1]$ indexed by the same set $\mathcal{I}$. The *ordinal sum* of $(C_i)_{i\in\mathcal{I}}$ with respect to $(]a_i, b_i[)_{i\in\mathcal{I}}$ is the copula $C\colon [0,1]^2 \to [0,1]$ given by

$$C(x,y) = \begin{cases} a_i + (b_i - a_i) C_i\left(\frac{x-a_i}{b_i-a_i}, \frac{y-a_i}{b_i-a_i}\right) & \text{on }]a_i, b_i[^2, \\ M(x,y) & \text{otherwise.} \end{cases}$$

Usually, we denote $C = (\langle a_i, b_i, C_i\rangle)_{i\in\mathcal{I}}$.

Let C be a copula and let $\phi\colon [0,1] \to [0,1]$ be a continuous and concave bijection with $\phi(0) = 0$ and $\phi(1) = 1$. Then the *ϕ–transform* [9, 13] of C is the copula C_ϕ given by

$$C_\phi(x,y) = \phi^{-1}(C(\phi(x), \phi(y))). \quad (3)$$

A copula C is symmetric if $C(x,y) = C(y,x)$ for all x, y in $[0,1]$, and, in particular, the copulas W, Π and M are symmetric, but not every copula is symmetric. As showed in [11, 17], we can assign to each copula C a degree of non–symmetry, given by

$$\sigma_C = \sup\{|C(x,y) - C(y,x)|, x, y \in [0,1]\}.$$

It was proved in [11, 17] that $\sigma_C \le 1/3$ for every copula C, and the value $1/3$ is attained.

Recently, investigations on various notions of convexity for copulas have received much attention because of their potential applications: see, for example, [1, 16] and the recent papers [2, 4, 7, 8]. Here, we revisit different types of convexity and study their relationships through several examples. In particular, we shall note that, if a copula C satisfies some conditions of convexity or concavity, then the bounds for the degree of non–symmetry σ_C can be improved.

2 Global and directional convexity

We start with the classical notion of convexity.

Definition 1 *A copula C is called (globally) convex if, for all x_1, x_2, y_1, y_2 and λ in $[0,1]$,*

$$C(\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2) \le \lambda C(x_1, x_2) + (1-\lambda) C(y_1, y_2). \quad (4)$$

A copula C is called (globally) concave if (4) *holds with the reverse inequality sign.*

In the class of copulas, convexity and concavity are strong properties in view of the following result.

Proposition 1 ([6]) *Let C be a copula.*

(a) *C is convex if, and only if, $C = W$;*

(b) *C is concave if, and only if, $C = M$.*

Weak versions of these properties are given by the following definitions.

Definition 2 *A copula C is called directionally convex if, for every z_0 in $[0,1]$, the functions $x \mapsto C(x, z_0)$ and $y \mapsto C(z_0, y)$ are convex, viz. for all x, y and λ in $[0,1]$,*

$$C(\lambda x + (1-\lambda)y, z_0) \leq \lambda C(x, z_0) + (1-\lambda)C(y, z_0),$$

$$C(z_0, \lambda x + (1-\lambda)y) \leq \lambda C(z_0, x) + (1-\lambda)C(z_0, y).$$

A copula C is called directionally concave if, for every z_0 in $[0,1]$, the functions $x \mapsto C(x, z_0)$ and $y \mapsto C(z_0, y)$ are concave.

Notice that W is directionally convex, M is directionally concave, Π is both directionally convex and concave. Such properties are useful in view of the following statistical interpretation.

Proposition 2 ([16]) *Let C be the copula associated with a pair (X, Y) of continuous random variables. Then:*

(a) *C is directionally convex if, and only if, Y is stochastically decreasing in X and X is stochastically decreasing in Y;*

(b) *C is directionally concave if, and only if, Y is stochastically increasing in X and X is stochastically increasing in Y.*

Example 1 *Let C_α be a member of the Farlie-Gumbel-Morgenstern family of copulas defined, for all α in $[-1,1]$ by*

$$C_\alpha(x, y) = xy(1 + \alpha(1-x)(1-y)).$$

Then C_α is directionally convex if $\alpha \in [-1, 0]$ and it is directionally concave if $\alpha \in [0, 1]$.

Directionally convex copulas are also called *P–increasing copulas* [5, 10] and they are used in order to ensure that the pointwise composition of two 2–increasing aggregation operators is also 2–increasing.

Proposition 3 *The class of all directionally concave (resp. convex) copulas is a convex and compact (with respect to the to the L^∞ norm) subset of $\mathcal{C}$.*

Note that an ordinal sum of directionally concave copula is also directionally concave, but a non-trivial ordinal sum of directionally convex copula is not directionally convex (because M is directionally concave). Moreover, the ϕ–transform of directionally concave (resp. convex) copula may not be directionally concave (resp. convex).

For directionally concave and convex copulas, we have the following bounds for the degree of non–symmetry.

Proposition 4 *Let C be a copula.*

(a) *If C is directionally concave, then $\sigma_C \leq 1/9$.*

(b) *If C is directionally convex, then $\sigma_C \leq 3 - \sqrt{2}$.*

Both bounds are sharp.

3 Schur–concave copulas

The notion of Schur–concavity was introduced in the context of majorization ordering [14] and it is here reformulated in the class of copulas.

Definition 3 *A copula C is called Schur–concave if, for all x, y and λ in $[0,1]$,*

$$C(x, y) \leq C(\lambda x + (1-\lambda)y, (1-\lambda)x + \lambda y). \qquad (5)$$

A copula C is called Schur–convex if (5) *holds with the reverse inequality sign.*

The following result allows to investigate only Schur–concave copulas.

Proposition 5 ([8]) *W is the only Schur–convex copula.*

Every Schur–concave copula C is symmetric (and hence $\sigma_C = 0$), but the converse implication is not true.

Example 2 *Let C be the copula defined by*

$$C(x, y) = \begin{cases} \frac{xy}{2} & \text{on } \left[0, \frac{1}{2}\right]^2, \\ \frac{x\,(3y-1)}{2} & \text{on } \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right], \\ \frac{y\,(3x-1)}{2} & \text{on } \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right], \\ \frac{xy+x+y-1}{2} & \text{on } \left[\frac{1}{2}, 1\right]^2. \end{cases}$$

Then C is symmetric, but

$$C\left(\frac{6}{10}, \frac{4}{10}\right) = \frac{32}{200} < \frac{33}{200} = C\left(\frac{7}{10}, \frac{3}{10}\right),$$

and, hence, C is not Schur–concave.

Remark 1 *If $z = C(s,t)$ is the surface associated with a Schur–concave copula C, then the intersections of the surface with all the vertical planes of the form $s+t=2x$, for all $x \in [0,1]$ and $s \in [0,x]$, are curves that are decreasing from (x,x) to $(2x,0)$, if $x \leq 1/2$, and from (x,x) to $(2x-1,1)$, otherwise.*

For a copula, the notion of Schur–concavity can be expressed in terms of its derivatives.

Proposition 6 ([8]) *Let C be a continuously differentiable copula. Then C is Schur–concave on $[0,1]^2$ if, and only if,*

(a) *C is symmetric;*

(b) *for all $x, y \in [0,1]$, $x \geq y$, $\frac{\partial C(x,y)}{\partial x} \leq \frac{\partial C(x,y)}{\partial y}$.*

As a consequence, it is easily proved that the copula Π is Schur–concave. Moreover, the following copulas are Schur–concave:

- associative copulas (in particular, M, W and Π);
- Fréchet copulas, $C_{\alpha,\beta} = \alpha M + (1-\alpha-\beta)\Pi + \beta W$ for α and β in $[0,1]$, $\alpha+\beta=1$;
- Farlie–Gumbel–Morgenstern copulas, $C_\alpha(x,y) = xy + \alpha xy\,(1-x)\,(1-y)$ for α in $[-1,1]$.

We denote by $\mathcal{C}_{SC}$ the class of all Schur–concave copulas.

Proposition 7 ([8]) *The set $\mathcal{C}_{SC}$ is a convex and compact (with respect to the to the L^∞ norm) subset of $\mathcal{C}$.*

Moreover, the class $\mathcal{C}_{SC}$ is closed with respect to ordinal sums and bijective concave transformations.

Proposition 8 ([8]) *The ordinal sum of Schur–concave copulas is Schur–concave.*

Proposition 9 *Let $\phi\colon [0,1] \to [0,1]$ be a continuous and concave bijection with $\phi(0)=0$ and $\phi(1)=1$. If C is Schur–concave, then the ϕ–transform C_ϕ given by (3) is also Schur–concave.*

Proof. Let x, y and λ be in $[0,1]$. Because ϕ is concave, we have

$$\begin{aligned} \phi(\lambda x + (1-\lambda)y) &\geq \lambda\phi(x) + (1-\lambda)\phi(y), \\ \phi((1-\lambda)x + \lambda y) &\geq (1-\lambda)\phi(x) + \lambda\phi(y). \end{aligned}$$

Moreover, since C is Schur–concave, we have

$$C(\lambda\phi(x) + (1-\lambda)\phi(y), (1-\lambda)\phi(x) + \lambda\phi(y)) \geq C(\phi(x), \phi(y)).$$

But C is increasing in each variable so that

$$C_\phi(\lambda x + (1-\lambda)y, (1-\lambda)x + \lambda y) \geq C_\phi(x,y),$$

which is the desired assertion. □

4 Weak Schur–concave copulas

Recently, E.P. Klement, R. Mesiar and E. Pap [12] raised some problems on binary aggregation operators. Specifically, in Problem 5 they suggested to study the inequality

$$C\left(\max(x-a,0), \min(x+a,1)\right) \leq C(x,x), \qquad (6)$$

for all $x \in [0,1]$ and for all $a \in]0,1/2[$. In [4], this problem was investigated for the class of copulas and the connection between inequality (6) and the notion of Schur–concavity was stressed.

In the spirit of these investigations, in [7] the following weakened form of Schur–concavity is introduced.

Definition 4 *A copula C is called weakly Schur-concave (WSC, for short) if, for all $x \in [0,1]$ and for all $a \in]0,1/2[$, both*

$$C\left(\max(x-a,0), \min(x+a,1)\right) \leq C(x,x) \qquad (7)$$

and

$$C\left(\min(x+a,1), \max(x-a,0)\right) \leq C(x,x), \qquad (8)$$

hold.

If (7) and (8) are satisfied by C with reverse inequality sign, then C is said to be weakly Schur-convex.

Proposition 10 ([7]) *If a copula C is weakly Schur–convex, then $C(t,t) = \max(2t-1,0)$ for every t in $[0,1]$.*

Therefore, by using [3, 15], if C is weakly Schur–convex, then C has the following representation

$$C(x,y) = \begin{cases} \frac{C_1(2x,2y-1)}{2} & \text{on } \left[0,\frac{1}{2}\right] \times \left[\frac{1}{2},1\right], \\ \frac{C_2(2x-1,2y)}{2} & \text{on } \left[\frac{1}{2},1\right] \times \left[0,\frac{1}{2}\right], \\ W(x,y) & \text{otherwise,} \end{cases}$$

for suitable copulas C_1 and C_2.

Remark 2 *Geometrically speaking, a copula C is weakly Schur–concave if the maximum of C along the line*

$$d_k = \{(x,y) \in [0,1]^2 \mid x+y = 2k\},$$

for every $k \in [0,1]$, is attained at the point (k,k). Moreover, if a copula C is differentiable, then the property WSC implies that the partial derivatives of C at the points of the main diagonal of $[0,1]^2$ are equal.

Important examples of WSC copulas are W, Π and M. If C is a Schur–concave copula, then it is easily proved that C is also WSC. But, on the other hand, if a copula C is WSC, then it does not need to be Schur–concave.

Example 3 *Let C be the copula defined by*

$$C(x,y) = \begin{cases} \frac{M(3x,3y-2)}{3} & \text{on } \left[0,\frac{1}{3}\right]\times\left[\frac{2}{3},1\right], \\ \frac{M(3x-1,3y-1)}{3} & \text{on } \left[\frac{1}{3},\frac{2}{3}\right]\times\left[\frac{1}{3},\frac{2}{3}\right], \\ \frac{M(3x-2,3y)}{3} & \text{on } \left[\frac{2}{3},1\right]\times\left[0,\frac{1}{3}\right], \\ W(x,y) & \text{otherwise.} \end{cases}$$

Then C is WSC, but C is not Schur–concave. In fact, given the points $\left(\frac{2}{10},\frac{7}{10}\right)$ and $\left(\frac{3}{10},\frac{6}{10}\right)$, we have

$$C\left(\frac{3}{10},\frac{6}{10}\right) = 0 < \frac{1}{30} = C\left(\frac{2}{10},\frac{7}{10}\right),$$

which implies that C is not Schur–concave.

We denote by $\mathcal{C}_{WSC}$ the class of all WSC copulas.

Proposition 11 ([7]) *The set $\mathcal{C}_{WSC}$ is a convex and compact (with respect to the to the L^∞ norm) subset of $\mathcal{C}$.*

Moreover, the class $\mathcal{C}_{WSC}$ is closed with respect to ordinal sums and bijective concave transformations.

Proposition 12 ([7]) *The ordinal sum of WSC copulas is WSC.*

Proposition 13 ([7]) *Let $\phi\colon [0,1]\to[0,1]$ be a continuous and concave bijection with $\phi(0)=0$ and $\phi(1)=1$. If C is WSC, then the ϕ–transform of C given by (3) is also WSC.*

A WSC copula does not need to be symmetric. For instance, consider the copula

$$C(x,y) = \begin{cases} xy, & x \le y; \\ \frac{(x+y)^2}{4}, & y < x \le 2\sqrt{y} - y; \\ y, & x \ge 2\sqrt{y} - y. \end{cases}$$

For WSC copulas, we have the following bound for the degree of non–symmetry.

Proposition 14 ([7]) *Let C be a WSC copula. Then $\sigma_C \le 1/4$.*

Notice that a WSC copula C such that $\sigma_C = 1/4$ is defined in the following way:

$$C(x,y) = \begin{cases} \max(x+y-1,0), & x \ge \frac{3}{4} \text{ or } x+y \ge \frac{3}{2}, \\ \max\left(y-\frac{1}{4},0\right), & y \le x \le \frac{3}{4}, \\ x, & x \le y - \frac{1}{2}, \\ \max\left(\frac{x+y}{2}-\frac{1}{4},0\right), & \text{otherwise.} \end{cases}$$

5 Quasi–concave copulas

The notion of quasi–concavity was introduced in the context of optimization theory and it is here reformulated in the class of copulas.

Definition 5 *A copula C is called quasi–concave if, for all x_1, x_2, y_1, y_2 and λ in $[0,1]$,*

$$C(\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2) \ge \min(C(x_1,x_2), C(y_1,y_2)). \quad (9)$$

A copula C is said to be quasi–convex if, for all x_1, x_2, y_1, y_2 and λ in $[0,1]$,

$$C(\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2) \le \max(C(x_1,x_2), C(y_1,y_2)). \quad (10)$$

The following result allows us to investigate only quasi–concave copulas.

Proposition 15 *W is the only quasi–convex copula.*

Notice that every associative copula is quasi–concave, and in particular W, Π and M are quasi–concave.

Remark 3 *Geometrically speaking, a copula C is quasi–concave if, given a segment $\overline{PQ}$ in $[0,1]^2$, the value of C at any internal point of this segment is not smaller than either the value of C at P or the value of C at Q.*

We denote by $\mathcal{C}_{QC}$ the class of all quasi–concave copulas.

Proposition 16 ([2]) *The set $\mathcal{C}_{QC}$ is a convex and compact (with respect to the to the L^∞ norm) subset of $\mathcal{C}$.*

Moreover, the class $\mathcal{C}_{QC}$ is closed with respect to ordinal sums and bijective concave transformations.

Proposition 17 *The ordinal sum of quasi–concave copulas is quasi–concave.*

Proof. The proof can be easily reproduced by using the geometrical interpretation of quasi–concavity given above. □

Proposition 18 *Let $\phi\colon [0,1]\to[0,1]$ be a continuous and concave bijection with $\phi(0)=0$ and $\phi(1)=1$. If C is quasi–concave, then the ϕ–transform of C given by (3) is also quasi–concave.*

Proof. Let x_1, x_2, y_1, y_2 and λ be in $[0,1]$. Because ϕ is concave, we have

$$\begin{aligned}\phi(\lambda x_1 + (1-\lambda)y_1) &\geq \lambda\phi(x_1) + (1-\lambda)\phi(y_1),\\ \phi(\lambda x_2 + (1-\lambda)y_2) &\geq \lambda\phi(x_2) + (1-\lambda)\phi(y_2).\end{aligned}$$

Moreover, since C is increasing in each variable and quasi–concave, we have

$$\begin{aligned}&C(\phi(\lambda x_1 + (1-\lambda)y_1), \phi(\lambda x_2 + (1-\lambda)y_2))\\ \geq\ &C(\lambda\phi(x_1) + (1-\lambda)\phi(y_1), \lambda\phi(x_2) + (1-\lambda)\phi(y_2))\\ \geq\ &\min(C(\phi(x_1), \phi(x_2)), C(\phi(y_1), \phi(y_2))).\end{aligned}$$

But ϕ is strictly increasing so that

$$\begin{aligned}&C_\phi(\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2)\\ &\qquad\geq \min(C_\phi(x_1, x_2), C_\phi(y_1, y_2)),\end{aligned}$$

which is the desired assertion. □

A quasi–concave copula does not need to be symmetric (so it is not necessarily Schur–concave): see, for instance, [16, section 3.2.1]. Moreover, we can improve the estimation of the degree of non–symmetry in the class of quasi–concave copulas.

Proposition 19 ([2]) *Let C be a quasi–concave copula. Then $\sigma_C \leq 1/5$, and the value $1/5$ is attained.*

In the symmetric case, we obtain the following result.

Proposition 20 ([2]) *For a quasi–concave copula C, the following statements are equivalent:*

(a) *C is symmetric,*

(b) *C is WSC,*

(c) *C is Schur–concave.*

Notice that a symmetric and Schur–concave copula does not need to be quasi–concave ([16, Example 3.2.8]). Example 2 in [4], instead, describes a symmetric and WSC copula that is neither Schur–concave nor, as a consequence, quasi–concave.

Proposition 21 ([2]) *A directionally concave copula is quasi–concave.*

The converse implication of the above proposition is not true, even when C is symmetric.

References

[1] C. Alsina, M.J. Frank and B. Schweizer, Associative Functions: Triangular Norms and Copulas, World Scientific, Hackensack, 2006.

[2] E. Alvoni, P.L. Papini, Quasi-concave copulas, asymmetry, transformations, Comment. Math. Univ. Carolinae, in press.

[3] B. De Baets and H. De Meyer, Copulas and the pairwise probabilistic comparison of ordered lists, in: Proceedings of the 10th International Conference IPMU, Perugia, 2004, pp. 1091–1098.

[4] F. Durante, Solution of an open problem for associative copulas, Fuzzy Sets and Systems 152 (2005) 411–415.

[5] F. Durante, R. Mesiar, P.L. Papini, C. Sempi, 2–increasing binary aggregation operators, Inform. Sci. 177 (2007) 111–129.

[6] F. Durante, J.J. Quesada-Molina, C. Sempi, Semicopulas: characterizations and applicability, Kybernetika 42 (2006) 287–302.

[7] F. Durante, P.L. Papini, A weakening of Schur–concavity for copulas, Fuzzy Sets and Systems, in press.

[8] F. Durante, C. Sempi, Copulæ and Schur–concavity, Int. Math. J. 3 (2003) 893–905.

[9] F. Durante and C. Sempi, Copula and semicopula transforms, Int. J. Math. Math. Sci. 2005 (2005), 645–655.

[10] F. Durante, C. Sempi, On the characterization of a class of binary operations on bivariate distribution functions, Publ. Math. Debrecen 69 (2006) 47–63.

[11] E.P. Klement, R. Mesiar, How non-symmetric can a copula be?, Comment. Math. Univ. Carolinae 47 (2006) 141–148.

[12] E.P. Klement, R. Mesiar, E. Pap, Problems on triangular norms and related operators, Fuzzy Sets and Systems 145 (2004) 471–479.

[13] E.P. Klement, R. Mesiar and E. Pap, Archimax copulas and invariance under transformations, C.R. Math. Acad. Sci. Paris 340 (2005) 755–758.

[14] A. Marshall, I. Olkin, Inequalities: Theory of Majorization and its Applications, Academic Press, New York, 1979.

[15] R. Mesiar, J. Szolgay, W–ordinal sums of copulas and quasi–copulas, in: Proceedings of the MAGIA 2004 Conference, Kočovce, Slovak Republic, October 2004, pp. 78-83.

[16] R. B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.

[17] R. B. Nelsen, Extremes of nonexchangeability, Statist. Papers 48 (2007) 329–336.

[18] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

A method for constructing multivariate copulas

F. Durante
Department of Knowledge-Based Mathematical Systems
Johannes Kepler University
A-4040 Linz, Austria
fabrizio.durante@jku.at

J.J. Quesada-Molina
Departamento de Matemática Aplicada
Universidad de Granada
E-18071 Granada, Spain
jquesada@ugr.es

M. Úbeda-Flores
Departamento de Estadística y Matemática Aplicada
Universidad de Almería
E-04120 Almería, Spain
mubeda@ual.es

Abstract

We provide a method for constructing a class of multivariate copulas depending on a univariate function. We study some properties of this class and present several examples. The same circle of ideas is used in a similar construction of quasi–copulas.

Keywords: Copula, Quasi–copula, Concordance, Kendall's tau.

1 Introduction

An n-dimensional *copula* (briefly n-copula) is a function $C\colon \mathbb{I}^n \to \mathbb{I}$ ($\mathbb{I} = [0,1]$) which satisfies:

(C1) for every $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ in $\mathbb{I}^n$, $C(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0, and $C(\mathbf{u}) = u_k$ whenever all coordinates of $\mathbf{u}$ are 1 except u_k;

(C2) for every $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $\mathbb{I}^n$ such that $a_k \leq b_k$ for all $k = 1, 2, \ldots, n$, $V_C([\mathbf{a}, \mathbf{b}]) = \sum \mathrm{sgn}(\mathbf{c}) C(\mathbf{c}) \geq 0$, where $[\mathbf{a}, \mathbf{b}]$ denotes the *n-box* $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$, the sum is taken over all the *vertices* $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ of $[\mathbf{a}, \mathbf{b}]$, $c_k \in \{a_k, b_k\}$ $(1 \leq k \leq n)$, and $\mathrm{sgn}(\mathbf{c}) = 1$ if $c_k = a_k$ for an even number of indices k's, and $\mathrm{sgn}(\mathbf{c}) = -1$ if $c_k = a_k$ for an odd number of indices k's.

In view of *Sklar's Theorem* [15], the joint distribution function H of the random vector $(X_1, X_2, \ldots, X_n)$ with univariate marginals $F_1, F_2, \ldots, F_n$ can be expressed, for every $\mathbf{x} \in \mathbb{R}^n$, by

$$H(\mathbf{x}) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)), \qquad (1)$$

where C is an n-copula that is uniquely determined on $\mathrm{Range}\, F_1 \times \mathrm{Range}\, F_2 \times \cdots \times \mathrm{Range}\, F_n$.

Conversely, given n univariate distribution functions $F_1, F_2, \ldots, F_n$, and an n–copula C, the function H given by (1) is an n–dimensional distribution function.

In particular, Π_n and M_n, defined for all $\mathbf{u}$ in $\mathbb{I}^n$ by $\Pi_n(\mathbf{u}) = \prod_{i=1}^{n} u_i$ and $M_n(\mathbf{u}) = \min\{u_1, u_2, \ldots, u_n\}$, are the n-copula of independent and comonotone random variables, respectively.

Thus, copulas are useful tools in the construction of multivariate distributions with given marginals: it suffices to construct a multivariate copula and, hence, attach to it some univariate marginals.

In [8, 12], several methods for constructing copulas are given; however, most of them concern the bivariate case and no simple extension to the n–dimensional case $(n \geq 3)$ is provided.

Recently [4], F. Durante introduced a new family of bivariate copulas depending on a univariate function. Specifically, under some assumptions on a function $f\colon \mathbb{I} \to \mathbb{I}$, the family of copulas given by

$$C_f(x, y) = \min\{x, y\} f(\max\{x, y\})$$

is studied (for more details about this family see [5]). In [6], it was proved that this family arises in a natural way when we impose that the horizontal and vertical sections of a bivariate copula are linear on some segments of the unit square.

In this note, we extend the above family of copulas to the n–dimensional case $(n \geq 3)$ and we study its properties. More details can be found in [7].

2 The new family of n-copulas

Given a continuous function $f\colon \mathbb{I} \to \mathbb{I}$, we define the function $C_f : \mathbb{I}^n \to \mathbb{I}$ given by

$$C_f(u_1, u_2, \ldots, u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}) \qquad (2)$$

where $u_{[1]}, \ldots, u_{[n]}$ denote the components of $(u_1, u_2, \ldots, u_n) \in \mathbb{I}^n$ rearranged in increasing order. We aim at studying the conditions under which C_f is a copula.

Theorem 1. *Let $f\colon \mathbb{I} \longrightarrow \mathbb{I}$ be a continuous function and let C_f be the function defined by (2). Then C_f is an n–copula if, and only if,*

(a) *$f(1) = 1$;*

(b) *f is increasing;*

(c) *the function $t \to f(t)/t$ is decreasing on $(0, 1]$.*

A function f satisfying the assumptions of Theorem 1 is called a *generator*. Every generator f is the restriction to $\mathbb{I}$ of a univariate distribution function. In particular, if $(U_1, U_2, \ldots, U_n)$ is a vector of n random variables uniformly distributed on $\mathbb{I}$ with n-copula C_f, then

$$P(\max\{U_1, U_2, \ldots, U_n\} \leq t | U_1 \leq t) = f^{n-1}(t).$$

Note that if $f_1(t) = t$ for all $t \in \mathbb{I}$, then $C_{f_1} = \Pi_n$; and, if $f_2(t) = 1$ for all $t \in \mathbb{I}$, then $C_{f_2} = M_n$. In general, if f is a generator, then $t \leq f(t) \leq 1$. Moreover, $f(t)/t$ is decreasing on $(0, 1]$ if, and only if, f is star–shaped, viz. $f(\alpha t) \geq \alpha f(t)$ for all α, t in $\mathbb{I}$; in particular, if f is concave, then $f(t)/t$ is decreasing on $(0, 1]$.

In the sequel, we will denote by Φ the class of all generators. When building n-copulas of type (2), the class Φ plays a major rôle. Some properties of this class are presented in the following result.

Proposition 1. *Let f and g be two continuous functions from $\mathbb{I}$ onto $\mathbb{I}$. The following statements hold:*

(a) *if f and g are in Φ, then $\alpha f + (1-\alpha)g$ is in Φ for every α in $\mathbb{I}$;*

(b) *if f and g are in Φ, then the functions $\min\{f, g\}$ and $\max\{f, g\}$ are in Φ;*

(c) *if f and g are in Φ, then the composition $f \circ g$ is in Φ.*

Example 1. For any α in $\mathbb{I}$, consider the function $f\colon \mathbb{I} \to \mathbb{I}$ given by $f_\alpha(t) = \alpha t + \overline{\alpha}$, with $\overline{\alpha} := 1 - \alpha$. Then, f_α is in Φ, and the n-copula $C_{f_\alpha} = C_\alpha$ defined by (2) is given by

$$C_\alpha(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} (\alpha u_{[i]} + \overline{\alpha}).$$

In particular, in the bivariate case, we obtain the well–known Fréchet family of copulas $C_\alpha(u_1, u_2) = \alpha u_1 u_2 + (1-\alpha)\min\{u_1, u_2\}$ (see [8, Family B11] and [12]).

Example 2. For any $\alpha \geq 1$, consider the function $f\colon \mathbb{I} \to \mathbb{I}$ given by $f_\alpha(t) := \min\{\alpha t, 1\}$. Then, f_α is in Φ, and the n-copula $C_{f_\alpha} = C_\alpha$ defined by (2) is given by

$$C_\alpha(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} \min\{\alpha u_{[i]}, 1\}.$$

In particular, in the bivariate case, C_α is the *ordinal sum* of the 2-copulas $\{\Pi_2, M_2\}$ with respect to the partition $\{[0, 1/\alpha], [1/\alpha, 1]\}$ (see [12] for more details).

Example 3. For any α in $\mathbb{I}$, consider the function $f\colon \mathbb{I} \to \mathbb{I}$ given by $f_\alpha(t) = t^\alpha$. Then, f_α is in Φ, and the n-copula $C_{f_\alpha} = C_\alpha$ defined by (2) is given by

$$C_\alpha(\mathbf{u}) = u_{[1]}^{1-\alpha} \prod_{i=1}^{n} u_i^\alpha,$$

i.e., $C_\alpha = (M_n)^{1-\alpha}(\Pi_n)^\alpha$. Note that $C_0 = M_n$ and $C_1 = \Pi_n$. Moreover, C_α can be considered as a generalization of the Cuadras-Augé family of 2-copulas [2]. In this case, every copula C_α is a *multivariate extreme copula*, viz. $C_\alpha(u_1^t, u_2^t, \ldots, u_n^t) = (C_\alpha(u_1, u_2, \ldots, u_n))^t$ for every $t > 0$, which is a useful property in multivariate extreme value theory, as showed in [8].

Notice that n–copula C_f is *symmetric*, viz. the value of C_f does not change by permuting its arguments.

The mixed derivative of order n of an n–copula C, $\frac{\partial^n C}{\partial u_1 \ldots \partial u_n}(\mathbf{u})$, exists almost everywhere on $\mathbb{I}^n$. In particular, an n–copula C is said to be *absolutely continuous* if

$$C(\mathbf{u}) = \int_0^{u_1} \cdots \int_0^{u_n} \frac{\partial^n C(\mathbf{t})}{\partial t_1 \ldots \partial t_n}\, d\mathbf{t},$$

otherwise, C has a *singular component*. Every n–copula C_f of type (2), except Π_n, has a singular component. In fact, as illustrated in [8, pages 14–15], it suffices to note that, for every $i = 1, 2, \ldots, n$, the mapping $t \mapsto \frac{\partial C}{\partial u_i}(u_1, \ldots, u_{i-1}, t, u_{i+1} \ldots, u_n)$ has a jump discontinuity. For instance, in the bivariate case, the first derivative of C_f is given by

$$\frac{\partial C_f}{\partial u}(u, v) = \begin{cases} f(v), & \text{if } u < v, \\ v\, f'(u), & \text{otherwise.} \end{cases}$$

For a fixed v_0, the mapping $t \mapsto \frac{\partial C_f}{\partial u}(t, v_0)$ has a jump discontinuity in v_0, and, thus, C_f has a singular component along the main diagonal of the unit square. By using [8, Theorem 1.1], the mass of this singular component is given by

$$m = \int_0^1 (f(t) - t f'(t))\, dt = 2 \int_0^1 f(t)\, dt\, - 1.$$

This m has a graphical interpretation when f admits an inverse; in fact, m is the area of the region of the unit square between the graph of f and the graph of f^{-1}.

3 Statistical properties

Now, we give a statistical interpretation of the new family. Let $W_1, W_2, \ldots, W_n$, Z be $n+1$ independent random variables such that W_i has distribution function f satisfying parts (a), (b) and (c) in Theorem 1, for all $i = 1, 2, \ldots, n$, and Z has distribution function $g(t) = t/f(t)$. Note that $g(1) = 1$ and g is increasing, since $f(t)/t$ is decreasing. Consider the random variables $U_i = \max\{W_i, Z\}$, for all $i = 1, 2, \ldots, n$. Then, for every $(u_1, u_2, \ldots, u_n)$, the distribution function of the random vector $(U_1, U_2, \ldots, U_n)$ is given by

$$P(U_1 \leq u_1, \ldots, U_n \leq u_n) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]}),$$

and, hence, it is a copula of type (2).

From a statistical point of view, the study of concordance in a family of multivariate distributions has also a great interest: this is the topic of the following result. We recall that, given two n-copulas C_1 and C_2, C_1 is said to be *more concordant than* C_2 (written $C_1 \succ C_2$) if both $C_1 \geq C_2$ and $\overline{C}_1 \geq \overline{C}_2$ hold, where, if $\mathbf{U}$ is a random vector with joint d.f. given by the n-copula C, then $\overline{C}$ is the *survival function* associated with C defined by $\overline{C}(\mathbf{u}) = P[\mathbf{U} > \mathbf{u}]$. For more details see [8, 12].

Theorem 2. *Let f and g be two generators and let C_f and C_g be two n–copulas of type* (2) *defined by $C_f(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} f(u_{[i]})$ and $C_g(\mathbf{u}) = u_{[1]} \prod_{i=2}^{n} g(u_{[i]})$, respectively, for every $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{I}^n$. Then, we have that $C_f \prec C_g$ if, and only if, both the following conditions are satisfied:*

(i) $\prod_{i=2}^{n} f(u_{[i]}) \leq \prod_{i=2}^{n} g(u_{[i]})$ *for all* $\mathbf{u}$ *in* $\mathbb{I}^n$, *and*

(ii) *for all* $\mathbf{u}$ *in* $\mathbb{I}^n$

$$\sum_{i=2}^{n} (f(u_{[i]}) - g(u_{[i]})) + (1 - u_{[1]}) \cdot \left(\prod_{i=2}^{n} (1 - f(u_{[i]})) - \prod_{i=2}^{n} (1 - g(u_{[i]})) \right) \leq 0. \quad (3)$$

In particular, for $n = 2$, we have that $C_f \prec C_g$ if, and only if, $f \leq g$.

Another way to summarize the information about the concordance for copulas is represented by the so–called *multivariate measures of concordance* ([16]), which are non-parametric measures of multivariate association for a continuous random vector with associated n-copula C. The most common measure is Kendall's tau (see [10, 11]), which is given by

$$\tau_n(C) = \frac{1}{2^{n-1} - 1} \left(2^n \int_{\mathbb{I}^n} C(\mathbf{u}) \, \mathrm{d}C(\mathbf{u}) - 1 \right).$$

For a copula of type (2), it has the following expression.

Theorem 3. *Let C_f be the n–copula given by* (2) *via Theorem* 1. *Then, Kendall's tau associated with C_f is given by:*

$$\tau_n(C_f) = \frac{1}{2^{n-1} - 1} \left(2^{n-1} - 2^n \sum_{k=2}^{n} \frac{k}{2^{k-1}} \int_{\mathbb{I}} t(1 - f^2(t))^{k-1} \, \mathrm{d}t - 1 \right).$$

Example 4. Consider the copula C_α of Example 3. The value of Kendall's tau for C_α is given by:

$$\tau_n(C) = \frac{1}{2^{n-1} - 1} \left(2^{n-1} - \frac{2^n}{\alpha} \sum_{k=2}^{n} \frac{k}{2^k} \operatorname{Beta}\left(\frac{1}{\alpha}, k \right) - 1 \right).$$

4 A new family of n-quasi–copulas

The notion of *quasi–copula* was introduced by Alsina, Nelsen and Schweizer [1] in order to show that a certain class of operations on univariate distribution functions is not derivable from corresponding operations on random variables defined on the same probability space (see also [13] for the multivariate case). Cuculescu and Theodorescu [3] characterized an n-dimensional quasi–copula (or n-quasi–copula) as a function Q from $\mathbb{I}^n$ onto $\mathbb{I}$, which satisfies (C1), and, in place of (C2), both the weaker conditions:

(Q1) Q is *non-decreasing in each variable*;

(Q2) Q satisfies the *Lipschitz condition*,

$$|Q(\mathbf{u}) - Q(\mathbf{v})| \leq \sum_{i=1}^{n} |u_i - v_i|$$

for all $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{I}^n$.

While every n-copula is an n-quasi–copula, there exist *proper* n-quasi–copulas, i.e., n-quasi–copulas which are not n-copulas. For example,

$$W_n(\mathbf{u}) = \max\left\{ 0, \sum_{i=1}^{n} u_i - n + 1 \right\}$$

is an n–copula if, and only if, $n = 2$, and a proper n-quasi–copula for all $n \geq 3$. Recently, n-quasi–copulas have been used to express the pointwise best-possible bounds on nonempty sets of distribution functions, n-copulas or n-quasi–copulas (see [14]). Here we present the characterization of quasi–copulas of type (2).

Theorem 4. *Let $f : \mathbb{I} \to \mathbb{I}$ be a continuous function, and let C_f be the function defined by (2). Then, C_f is an n–quasi–copula if, and only if, the following statements are satisfied:*

(i) $f(1) = 1$,

(ii) f *is increasing,*

(iii) $u_1(f(u_2) - f(u_1)) \leq u_2 - u_1$ *for every* $u_1, u_2 \in \mathbb{I}$, *with* $u_1 < u_2$.

Proof. If C_f is an n-quasi–copula, then it is easy to check that (i) and (ii) hold.

To prove (iii), consider $u_1, v_1 \in \mathbb{I}$ with $u_1 < v_1$. From (Q2) we have that

$$C_f(v_1, u_1, 1, \ldots, 1) - C_f(u_1, u_1, 1, \ldots, 1) \leq v_1 - u_1,$$

that is $u_1(f(v_1) - f(u_1)) \leq v_1 - u_1$, and we obtain (iii).

Conversely, given a continuous $f : \mathbb{I} \to \mathbb{I}$ satisfying (i), (ii) and (iii), let C_f be the function defined by (2). Then it is easily proved that C_f satisfies (C1) and (Q1). In order to prove that C_f satisfies (Q2), let $u_1, v_1, \ldots, v_n$ be $(n+1)$ points in $\mathbb{I}$ such that $u_1 < v_1$. Notice that, because C_f is symmetric, it is enough to show that

$$\lambda = C_f(v_1, v_2, \ldots, v_n) - C_f(u_1, v_2, \ldots, v_n) \leq v_1 - u_1. \quad (4)$$

Three cases will be considered.

1. If $v_{[1]} = v_1$, we obtain that

$$\lambda = (v_1 - u_1) \prod_{i=2}^{n} f(v_i),$$

and therefore (4) holds.

2. If $v_{[1]} = v_j \leq u_1$ for some $j \in \{2, 3, \ldots, n\}$, then we have that

$$\lambda = v_j(f(v_1) - f(u_1)) \prod_{\substack{i=2 \\ i \neq j}}^{n} f(v_i),$$

and (4) follows from (iii) and the fact that $f(t) \leq 1$ for all $t \in \mathbb{I}$.

3. Finally, if $u_1 < v_{[1]} = v_j < v_1$ for some j belongs to $\{2, 3, \ldots, n\}$, then

$$\lambda = [C_f(v_j, v_2, \ldots, v_n) - C_f(u_1, v_2, \ldots, v_n)] + [C_f(v_1, v_2, \ldots, v_n) - C_f(v_j, v_2, \ldots, v_n)].$$

Inequality (4) is, hence, a consequence of the preceding two cases.

Hence, the proof is completed. □

Corollary 1. *Let $f : \mathbb{I} \longrightarrow \mathbb{I}$ be a differentiable function, and let C_f be the function defined by (2). Then, C_f is an n-quasi–copula if, and only if, the following statements are satisfied:*

(i) $f(1) = 1$;

(ii) f *is increasing;*

(iii) $uf'(u) \leq 1$ *for every* $u \in \mathbb{I}$.

We now provide an example of a proper n-quasi–copula of type (2).

Example 5. Consider the function $f(u) = u + u^2 - u^3$ for every $u \in \mathbb{I}$, and let C_f be the function defined by (2). Then, it is easy to check that $f(1) = 1$, f is increasing on $[0, 1]$, but C_f is not an n-copula since the function $f(u)/u$ is increasing on $[0, 1/2]$. However, it is easy to see that $uf'(u) = u(1 + 2u - 3u^2)$ satisfies part (iii) in Corollary 1, and thus, C_f is a proper n-quasi–copula.

5 Concluding remarks

We have introduced a new family of multivariate copulas depending on a univariate function. Now, we will show how to construct many other copulas starting from our method and a result from [9].

Consider a continuous and increasing bijection $\phi : \mathbb{I} \to \mathbb{I}$, and suppose that ϕ^{-1} is absolutely monotonic of order n on $\mathbb{I}$, viz. ϕ^{-1} admits derivatives up to order n on $\mathbb{I}$ and, for $i = 1, 2, \ldots, n$,

$$\frac{\mathrm{d}^i(\phi^{-1})(t)}{\mathrm{d}t^i} \geq 0.$$

From [9, Theorem 4.7], for every n–copula C, we have that the mapping $C_\phi : \mathbb{I}^n \to \mathbb{I}$ defined by

$$C_\phi(u_1, \ldots, u_n) = \phi^{-1}(C(\phi(u_1), \ldots, \phi(u_n))) \quad (5)$$

is also an n–copula.

In particular, if $C = C_f$ is a copula of type (2) generated by f, then the function $C_{f,\phi}$ defined by

$$C_{f,\phi}(u_1, u_2, \ldots, u_n) = \phi^{-1}\left[\phi(u_{[1]}) \prod_{i=2}^{n} f(\phi(u_{[i]}))\right]$$

is also an n–copula.

For $g_1 = -\ln \phi$ and $g_2 = -\ln(f \circ \phi)$, equation (5) can be written into the form

$$C_{g_1,g_2}(u_1, u_2, \ldots, u_n) = g_1^{-1}\left[g_1(u_{[1]}) + \sum_{i=2}^{n} g_2(u_{[i]})\right].$$

The reader will recognize that the last expression is a direct generalization of the Archimedean family of multivariate copulas [12].

Acknowledgement

The second and third authors acknowledge the support by the Ministerio de Educación y Ciencia (Spain) and FEDER, under research project MTM2006-12218.

References

[1] C. Alsina, R.B. Nelsen, B. Schweizer, On the characterization of a class of binary operations on distribution functions, Statist. Probab. Lett. 17 (1993) 85–89.

[2] C.M. Cuadras, J. Augé, A continuous general multivariate distribution and its properties, Comm. Statist. - Theory Methods 10 (1981) 339–353.

[3] I. Cuculescu, R. Theodorescu, Copulas: diagonals and tracks, Rev. Roumaine Math. Pures Appl. 46 (2001) 731–742.

[4] F. Durante, A new family of symmetric bivariate copulas, C.R. Acad. Sci. Paris, Sér. I, 344 (2007) 195–198.

[5] F. Durante, A new class of symmetric bivariate copulas, J. Nonparametr. Stat., in press.

[6] F. Durante, R. Mesiar, A. Kolesárová, C. Sempi, Semilinear copulas, submitted.

[7] F. Durante, J.J. Quesada-Molina, M. Úbeda-Flores, On a family of multivariate copulas for aggregation processes, submitted.

[8] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, New York, 1997.

[9] P.M. Morillas, A method to obtain new copulas from a given one, Metrika 61 (2005) 169–184.

[10] R.B. Nelsen, Nonparametric measures of multivariate association, in: L. Rüschendorf, B. Schweizer, M.D. Taylor (Eds.) Distribution Functions with Fixed Marginals and Related Topics, Institute of Mathematical Statistics (Lecture Notes – Monograph Series Volume 28), Hayward CA, 1996, pp. 223–232.

[11] R.B. Nelsen, Concordance and copulas: a survey, in: C. Cuadras, J. Fortiana, J.A. Rodríguez-Lallena (Eds.) Distributions with Given Marginals and Statistical Modelling, Kluwer Academic Publishers, Dordrecht, 2002, pp. 169–178.

[12] R.B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.

[13] R.B. Nelsen, J.J. Quesada-Molina, B. Schweizer, C. Sempi, Derivability of some operations on distribution functions, in: L. Rüschendorf, B. Schweizer and M.D. Taylor (Eds.), Distributions with Fixed Marginals and Related Topics, IMS Lecture Notes-Monograph Series, Hayward, CA, 1996, pp. 233–243.

[14] J.A. Rodríguez-Lallena, M. Úbeda-Flores, Best-possible bounds on sets of multivariate distribution functions, Comm. Statist. - Theory Methods 33 (2004) 805–820.

[15] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

[16] M.D. Taylor, Multivariate measures of concordance, Ann. Inst. Statist. Math., in press.

Semigroups of semicopulas and evolution of dependence at increase of age

Rachele Foschi
Department of Mathematics
University "La Sapienza", Rome
foschi@mat.uniroma1.it

Fabio Spizzichino
Department of Mathematics
University "La Sapienza", Rome
fabio.spizzichino@uniroma1.it

Abstract

We consider a pair of exchangeable lifetimes X, Y and the families of the conditional survival functions $\overline{F}_t(x,y)$ of $(X-t, Y-t)$ given $(X>t, Y>t)$. We analyze some properties of dependence and of ageing for $\overline{F}_t(x,y)$ and some relations among them.

Keywords: copulas, semicopulas, evolution of dependence under truncation, level curves of survival functions

1 Introduction

Let X, Y be exchangeable non-negative random variables and denote by $\overline{F}(x,y), \overline{G}(x), \hat{C}(u,v)$ the corresponding joint survival function, marginal univariate survival function and survival copula respectively; namely

$$\overline{F}(x,y) = P(X>x, Y>y) \tag{1}$$

$$\overline{G}(x) = \overline{F}(x,0) = P(X>x)$$

$$\hat{C}(u,v) = \overline{F}\{\overline{G}^{-1}(u), \overline{G}^{-1}(v)\}. \tag{2}$$

We assume that $\overline{F}(x,y)$ is a continuous survival function which is strictly decreasing on $\mathbb{R}_+$ in each variable. This in particular implies that $\overline{G}(x)$ is a continuous, strictly decreasing survival function; we also assume $\overline{G}(0)=1$.
Let us consider the function $B:[0,1]\times[0,1]\to[0,1]$ defined by

$$B(u,v) = \exp\{-\overline{G}^{-1}\left(\overline{F}(-\log u, -\log v)\right)\}. \tag{3}$$

It is immediate to see that B satisfies boundary conditions for a copula, is increasing in each variable and continuous, but it does not satisfy, generally, the rectangular inequality. For this reason we say that B is generally a semicopula ([2, 5]). However, it turns out to be a copula in several cases of interest.

The function B can be used to describe certain "bivariate ageing" properties of the pair (X,Y) and has been called "bivariate aging function".

By imposing appropriate dependence conditions on B, it is possible to characterize some conditions of bivariate ageing for X,Y; this can be used to analyze some relations existing among univariate ageing, bivariate ageing, and stochastic dependence (see [2]).

From a more technical point of view, relevant features of B are that it describes the family of the level curves of $\overline{F}$ and it permits to give a representation of $\overline{F}$ in terms of the pair $\left(\overline{G}, B\right)$, see Eq. (5) below.

An item of general interest is the conditional survival function:

$$\overline{F}_t(x,y) = P(X>t+x, Y>t+y | X>t, Y>t), \tag{4}$$

for $t>0$. In fact the study of the evolution in time of the survival functions $\overline{F}_t(x,y)$ can be interesting in several fields (see e.g. [1, 3, 4, 11]).

As a natural consequence of the introduction of the family $\{\overline{F}_t\}_{t\geq 0}$, it is of interest to study the evolution of the families denoted by $\{B_t\}_{t\geq 0}$, $\{\hat{C}_t\}_{t\geq 0}$, with obvious use of the notation (see also Section 2).

In this paper we aim to point out both analogies and structural differences between $\{B_t\}_{t\geq 0}$ and $\{\hat{C}_t\}_{t\geq 0}$. In this frame we obtain some further results along the lines indicated in [1, 2]. In particular we give some examples of how information about $\{\hat{C}_t\}_{t\geq 0}$ can be derived from the analysis of $\{B_t\}_{t\geq 0}$ and vice-versa.

Section 2 is devoted to recall some basic facts and notation from [1, 2] and to point out some further basic properties of $\{B_t\}_{t\geq 0}$, $\{\hat{C}_t\}_{t\geq 0}$. Specific results related with evolution of dependence and bivariate ageing will be presented in Section 3.

2 Some basic facts

First we briefly recall some of the arguments contained in [1, 2]. As immediate consequences of (2) and (3) respectively, we obtain

$$\begin{aligned} \overline{F}\ (x,y) &= \hat{C}\{\overline{G}\,(x)\,,\overline{G}\,(y)\},\\ \overline{F}\ (x,y) &= \overline{G}\{-\log B\left(e^{-x},e^{-y}\right)\}. \end{aligned} \tag{5}$$

Furthermore we can easily obtain, still as a consequence of Eqs. (2) and (3), the following relations between B and $\hat{C}$:

$$B(u,v)=\exp\left[-\overline{G}^{-1}\{\hat{C}\left(\overline{G}(-\log u),\overline{G}\,(-\log v)\right)\}\right]; \tag{6}$$

$$\hat{C}(u,v)=\overline{G}\left\{-\log B\left(e^{-\overline{G}^{-1}(u)},e^{-\overline{G}^{-1}(u)}\right)\right\}. \tag{7}$$

Remark 1. *Notice that, if $\overline{G}(x)=e^{-x}$, i.e. if $\overline{G}(-\log u)=u$, then $B=\hat{C}$ and B is thus certainly a copula. More generally, we also observe that, if $\overline{G}(-\log u)$ is concave, then B is copula. This fact follows by the general method of transforming copulas by means of*

$$C_{\phi}(u,v)=\phi^{-1}\left(C(\phi(u),\phi(v))\right),$$

with $\phi:[0,1]\to[0,1]$, ϕ bijective and concave (see e.g. [6, 8, 9, 10]).

Let us consider now the joint law of the residual lifetimes $(X-t,Y-t)$, conditional on the observation of the survival data $\{X>t,Y>t\}$. For $t>0$ we put

$$\begin{aligned} \overline{F}_t\,(x,y) &= P\left(X>t+x,Y>t+y|X>t,Y>t\right),\\ \overline{G}_t\,(x) &= P\left(X>t+x|X>t,Y>t\right), \end{aligned}$$

so that we can write

$$\overline{F}_t\,(x,y)=\frac{\overline{F}\ (x+t,y+t)}{\overline{F}\ (t,t)}, \tag{8}$$

$$\overline{G}_t\,(x)=\frac{\overline{F}\ (x+t,t)}{\overline{F}\ (t,t)}. \tag{9}$$

We are interested in studying the survival copula and the ageing function of $\overline{F}_t$. For this reason we need that $\overline{G}_t\,(x)$ is continuous and strictly decreasing on $\mathbb{R}_+$ in each variable. This is guaranteed in view of the assumption that $\overline{F}(x,y)$ is a continuous and strictly decreasing on $\mathbb{R}_+$. This assumption is also equivalent to $\hat{C}(u,v)$, $B(u,v)$ being strictly increasing in u, $\forall\ 0<v\leq 1$.
For $0\leq u\leq 1$, $0\leq v\leq 1$, we then put

$$\hat{C}_t(u,v)\equiv\overline{F}_t\{\overline{G}_t^{-1}\,(u)\,,\overline{G}_t^{-1}\,(v)\} \tag{10}$$

and

$$B_t(u,v)\equiv\exp[-\overline{G}_t^{-1}\{\overline{F}_t\,(-\log u,-\log v)\}]. \tag{11}$$

Remark 2. *For $t=0$, B_t coincides with B as given in formula (3).*

In view of (8) and (9), the relation between $\hat{C}_t$ and $\hat{C}$ has an explicit form given by

$$\hat{C}_t(u,v)=\frac{\hat{C}[\overline{G}(\overline{G}_t^{-1}(u)+t),\overline{G}(\overline{G}_t^{-1}(v)+t)]}{\hat{C}(\overline{G}(t),\overline{G}(t))}. \tag{12}$$

We notice that the function $\hat{C}_t$ in (12) is actually a copula for any $t>0$.

Remark 3. *Notice that Eq. (12) contains the term $\overline{G}$. However, by adopting a different parametrization for $\hat{C}_t$, we realize that the family $\{\hat{C}_t\}_{t\geq 0}$ only depends on $\hat{C}$. In fact, by letting $\hat{C}_{(z)}:=\hat{C}_{\overline{G}^{-1}(z)}$ we can write*

$$\hat{C}_{(z)}(u,v)=\frac{\hat{C}\left(R(z,u\hat{C}(z,z)),R(z,v\hat{C}(z,z))\right)}{\hat{C}(z,z)}, \tag{13}$$

where

$$R(z,w):=\sup\{u|\hat{C}(u,z)\leq w\}$$

is the residuum *of the copula $\hat{C}$ (see e.g. [7]).*

The structure of the relation between B_t and B is radically different from the one binding $\hat{C}_t$ and $\hat{C}$. In fact it can only be given in an implicit form, as follows by Lemma 12 of [1]: $B_t\ (u,v)$ is such that

$$B\left(ue^{-t},ve^{-t}\right)=B\left(B_t(u,v)e^{-t},e^{-t}\right); \tag{14}$$

actually $B_t\ (u,v)$ is the unique solution σ of the equation

$$B\left(ue^{-t},ve^{-t}\right)=B\left(\sigma e^{-t},e^{-t}\right).$$

Eq. (14) will have in the following a basic role in proving some properties of $\{B_t\}_{t\geq 0}$.

Remark 4. *For any $t>0$, B_t only depends on B.*

Other similarities between $\{B_t\}_{t\geq 0}$ and $\{\hat{C}_t\}_{t\geq 0}$ are shown by the following propositions, that will be used in Section 3.

Proposition 1. *$\{B_t\}_{t\geq 0}$ is a semigroup, i.e.*

$$(B_s)_r=(B_r)_s=B_{r+s}\qquad\forall t,s\geq 0.$$

Proposition 2. *$\{\hat{C}_t\}_{t\geq 0}$ is a semigroup.*

Proposition 1 and Proposition 2 can be proved independently one of another. However, e.g., Proposition 1 can be immediately obtained from Proposition 2 in view of the formula 6 applied to B_t and $\hat{C}_t$. For this reason we proceed with the proof of Proposition 2.

Proof. For fixed $r>0$ we consider the survival model with joint survival function $\overline{F}_r$ and marginal $\overline{G}_r$.

We now notice the semigroup property of the families $\{\overline{F}_t\}_{t\geq 0}$ and $\{\overline{G}_t\}_{t\geq 0}$. We can associate to the new model the families $\{(\overline{F}_r)_s\}_{s\geq 0}$ and $\{(\overline{G}_r)_s\}_{s\geq 0}$, such that, for any $t, s \geq 0$,

$$(\overline{F}_r)_s = \overline{F}_{r+s} \text{ and } (\overline{G}_r)_s = \overline{G}_{r+s}.$$

By definition, the survival copula of the model $\overline{F}_r$ is

$$\hat{C}^{(r)}(u,v) \equiv \overline{F}_r(\overline{G}_r^{-1}(u), \overline{G}_r^{-1}(v)) = \hat{C}_r(u,v).$$

We have to prove that

$$(\hat{C}^{(r)})_s = \hat{C}_{r+s} \qquad \forall r, s \geq 0.$$

By applying (12) to $(\hat{C}^{(r)})_s$, we obtain

$$\hat{C}_s^{(r)}(u,v) = \frac{\hat{C}^{(r)}[\overline{G}_r(\overline{G}_{r+s}^{-1}(u)+s), \overline{G}_r(\overline{G}_{r+s}^{-1}(v)+s)]}{\hat{C}^{(r)}(\overline{G}_r(s), \overline{G}_r(s))}. \tag{15}$$

By applying again (12) to $\hat{C}^{(r)}$ in Eq. (15),

$$\hat{C}_s^{(r)}(u,v) = \frac{\hat{C}[\overline{G}(\overline{G}_{r+s}^{-1}(u)+s+r), \overline{G}(\overline{G}_{r+s}^{-1}(v)+s+r)]}{\hat{C}(\overline{G}(r+s), \overline{G}(r+s))}. \tag{16}$$

The thesis follows by pointing out that the right hand side of the last equation effectively coincides with $\hat{C}_{r+s}(u,v)$. □

Remark 5. *$\hat{C}$ is strictly increasing (in each variable) if and only if $\hat{C}_t$ is strictly increasing.*
$\hat{C}$ is strictly increasing if and only if B is strictly increasing.
Hence B is strictly increasing if and only if B_t is strictly increasing.

One purpose of ours is to analyze increase or decrease of dependence between residual lifetimes. In this respect, we recall the following definitions.

Definition 1. *Let S_1, S_2 be two semicopulas. We write*

$$S_1 \preceq S_2$$

iff

$$S_1(u,v) \leq S_2(u,v) \qquad \forall\ u, v \in [0,1].$$

Definition 2. *(X_2, Y_2) is said* more concordant *than (X_1, Y_1) iff*

$$\hat{C}_1 \preceq \hat{C}_2.$$

Considering $X_2 = X_1 - t$ and $Y_2 = Y_1 - t$, monotonicity of the mapping $t \mapsto \hat{C}_t$ has the following meaning: if $t \mapsto \hat{C}_t$ is increasing in t, the residual lifetimes will be more and more dependent as age increases; on the contrary, if it is decreasing, the residual lifetimes will be less and less dependent.

We are then interested in describing analytical conditions for such monotonicity properties. In our parallel study of $\{\hat{C}_t\}_{t\geq 0}$ and $\{B_t\}_{t\geq 0}$, we are also interested in analytical conditions for monotonicity properties of the mapping $t \mapsto B_t$.

We denote, as usual, the common scalar product by $\cdot$ and the gradient operator by ∇, i.e. $\nabla \equiv \left(\frac{\partial}{\partial u}, \frac{\partial}{\partial v}\right)$. Furthermore, since we are dealing with exchangeable variables and hence with symmetric survival and ageing functions and survival copulas,

$$\nabla \hat{C}(z,z) \cdot (1,1) = 2\frac{\partial}{\partial \xi}\hat{C}(z,z),$$

where $\frac{\partial}{\partial \xi}$ denote the derivative in one of the variables.

As we can expect in view of Remark 3, the monotonicity properties of $t \mapsto \hat{C}_t$ can be characterized in terms of $\hat{C}$ and its residuum. We have in fact

Proposition 3. *The mapping $t \mapsto \hat{C}_t$ is increasing iff*

$$2\frac{\partial \hat{C}}{\partial \xi}(1,1)\left[\hat{C}(u,v) - \nabla \hat{C}(u,v)\cdot\left(u\frac{\partial R}{\partial w}(1,u), v\frac{\partial R}{\partial w}(1,v)\right)\right]$$
$$\geq \nabla \hat{C}(u,v)\cdot\left(\frac{\partial R}{\partial z}(1,u), \frac{\partial R}{\partial z}(1,v)\right). \tag{17}$$

Proof. In view of the semigroup property of $\{\hat{C}_t\}_{t\geq 0}$, it is sufficient to study the sign of the derivative of $\hat{C}_t$ w.r.t. t for $t = 0$. Since the change of parameter given by $z = \overline{G}(t)$ (see Remark 3) is strictly decreasing, instead of differentiating Eq. (12) w.r.t. t, we can differentiate the simpler Eq. (13) w.r.t. z. Thus we need to check that

$$\frac{\partial}{\partial z}\hat{C}_{(z)}(u,v)\Big|_{z=1} \leq 0.$$

To this purpose, we now compute the partial derivative $\frac{\partial}{\partial z}\hat{C}_{(z)}(u,v)$.

$$\frac{\partial}{\partial z}\hat{C}_{(z)}(u,v) =$$

$$\frac{1}{\hat{C}(z,z)^2}\Bigg\{\hat{C}(z,z)\nabla\hat{C}(R(z,u\hat{C}(z,z)), R(z,v\hat{C}(z,z)))\cdot$$

$$\cdot\left(\frac{dR}{dz}(z,u\hat{C}(z,z)), \frac{dR}{dz}(z,v\hat{C}(z,z))\right)$$

$$-\hat{C}\left(R(z,u\hat{C}(z,z)), R(z,v\hat{C}(z,z))\right)\nabla\hat{C}(z,z)\cdot(1,1)\Bigg\},$$

where

$$\frac{dR}{dz}(z,u\hat{C}(z,z)) =$$
$$\frac{\partial R}{\partial z}(z,u\hat{C}(z,z))+u\frac{\partial R}{\partial w}(z,u\hat{C}(z,z))\left[\nabla\hat{C}(z,z)\cdot(1,1)\right].$$

Since $[\hat{C}(z,z)]^2$ is positive for any $z>0$,

$$\frac{\partial}{\partial z}\hat{C}_{(z)}(u,v)\leq 0$$

iff

$$\hat{C}(z,z)\nabla\hat{C}(R(z,u\hat{C}(z,z)),R(z,v\hat{C}(z,z)))\cdot$$
$$\cdot\left(\frac{\partial R}{\partial z}(z,u\hat{C}(z,z))+2u\frac{\partial R}{\partial w}(z,u\hat{C}(z,z))\frac{\partial\hat{C}}{\partial\xi}(z,z),\right.$$
$$\left.\frac{\partial R}{\partial z}(z,v\hat{C}(z,z))+2v\frac{\partial R}{\partial w}(z,v\hat{C}(z,z))\frac{\partial\hat{C}}{\partial\xi}(z,z)\right)$$
$$-2\hat{C}\left(R(z,u\hat{C}(z,z)),R(z,v\hat{C}(z,z))\right)\frac{\partial}{\partial\xi}\hat{C}(z,z)\leq 0. \tag{18}$$

By putting $z=\overline{G}(0)=1$ in Eq. (18) and recalling that $\hat{C}(1,1)=1$ by definition of copula and $R(1,w)=w$, we obtain

$$\nabla\hat{C}(u,v)\cdot\left(\frac{\partial R}{\partial z}(1,u)+2u\frac{\partial R}{\partial w}(1,u)\frac{\partial\hat{C}}{\partial\xi}(1,1),\right.$$
$$\left.\frac{\partial R}{\partial z}(1,v)+2v\frac{\partial R}{\partial w}(1,v)\frac{\partial\hat{C}}{\partial\xi}(1,1)\right)-2\hat{C}(u,v)\frac{\partial\hat{C}}{\partial\xi}(1,1)\leq 0.$$

□

Remark 6. *Since $\hat{C}(z,z)$ is increasing in z, a sufficient condition for $z\mapsto\hat{C}_{(z)}$ being decreasing is that the numerator is decreasing in z, i.e.*

$$\nabla\hat{C}(u,v)\cdot\left(\frac{\partial R}{\partial z}(1,u)+2u\frac{\partial R}{\partial w}(1,u)\frac{\partial\hat{C}}{\partial\xi}(1,1),\right.$$
$$\left.\frac{\partial R}{\partial z}(1,v)+2v\frac{\partial R}{\partial w}(1,v)\frac{\partial\hat{C}}{\partial\xi}(1,1)\right)\leq 0.$$

Concerning the family $\{B_t\}_{t\geq 0}$, we have instead

Proposition 4. *$t\mapsto B_t$ is increasing if*

$$(u,v)\cdot\nabla B(u,v)\leq(B(u,v),1)\cdot\nabla B(B(u,v),1) \tag{19}$$

The proof is omitted.

Remark 7. *Let (X_1,Y_1) and (X_2,Y_2) be two exchangeable random vectors, having ageing functions B_1 and B_2 respectively. If $\overline{G}_1\geq\overline{G}_2$, $B_1\leq B_2$ implies that $\overline{F}_1\leq\overline{F}_2$.*
By applying this result to the family $\{B_t\}_{t\geq 0}$, we can conclude that $t\mapsto B_t$ increasing and $t\mapsto\overline{G}_t$ decreasing implies $t\mapsto\hat{C}_t$ increasing.

Remark 8. *The semigroup property has interesting consequences in the analysis of $\{B_t\}_{t\geq 0}$ and $\{\hat{C}_t\}_{t\geq 0}$. Consider two different families of semicopulas $\mathcal{C}$ and $\mathcal{C}'$ such that $S\in\mathcal{C}\Rightarrow S\in\mathcal{C}'$, i.e. $\mathcal{C}\subseteq\mathcal{C}'$.*
Suppose furthermore that $\hat{C}\in\mathcal{C}$ is equivalent to $\hat{C}_t\in\mathcal{C}'\ \forall t>0$.
Then we can conclude that the condition $\hat{C}_t\in\mathcal{C}\ \forall t>0$ is equivalent to the apparently weaker condition $\hat{C}_t\in\mathcal{C}'\ \forall t>0$. In fact, for given $r>0$, we have

$$\hat{C}\in\mathcal{C}\ \Rightarrow\ \hat{C}_t\in\mathcal{C}'\ \forall t>0\ \Rightarrow\ \hat{C}_t\in\mathcal{C}'\ \forall t>r+s$$
$$\Rightarrow\ (\hat{C}_r)_s\in\mathcal{C}'\ \forall s>0\ \Rightarrow\ \hat{C}_r\in\mathcal{C}.$$

This Remark has substantially been used along the proof of Proposition 3 and can be similarly applied in proving Proposition 4. The same general fact can also turn out to be useful in the arguments of the next section.

3 Some properties of ageing and dependence and their relation

We start this Section by analyzing some relations between dependence and ageing properties along the same line of [2]. For the purpose of a better understanding of both analogies and differences between the functions $\widehat{C}$ and B, we introduce in the analysis here the notion of TP_2.

We recall that a function $K:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is said to be TP_2 (Totally Positive of order 2) if, for $x'\leq x''$, $y'\leq y''$, it is

$$K(x'',y'')K(x',y')\geq K(x',y'')K(x'',y')$$

(see e.g [10] and references therein). Analogously a function $K:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is said to be RR_2 if, for $x'\leq x''$, $y'\leq y''$, it is

$$K(x'',y'')K(x',y')\leq K(x',y'')K(x'',y').$$

The following Proposition is analogous to some consequences of Propositions 5.2, 5.3, 5.4 of [2]. We however deal here with the concept of TP_2, that was not considered there; our proof is direct and independent of the results of [2].

Proposition 5.

1. *$\hat{C}\ TP_2$, $\overline{G}\ IFR\ \Rightarrow\ B\ TP_2$.*
2. *$B\ TP_2$, $\hat{C}\ RR_2\ \Rightarrow\ \overline{G}\ IFR$.*
3. *$B\ TP_2$, $\overline{G}\ DFR\Rightarrow\ \hat{C}\ TP_2$*

Proof. For simplicity sake let

$$x=-\log u,\ x'=-\log u',\ y=-\log v,\ y'=-\log v'$$

and

$$\alpha_{11} = \hat{C}(\overline{G}(x'), \overline{G}(y')), \quad \alpha_{12} = \hat{C}(\overline{G}(x'), \overline{G}(y)),$$

$$\alpha_{21} = \hat{C}(\overline{G}(x), \overline{G}(y')), \quad \alpha_{22} = \hat{C}(\overline{G}(x), \overline{G}(y)),$$

where $x' < x$ and $y' < y$. Thus we have

$$\alpha_{22} < \alpha_{12},\ \alpha_{21} < \alpha_{11}$$

and

$$-\log \alpha_{22} > -\log \alpha_{12},\ -\log \alpha_{21} > -\log \alpha_{11}.$$

1. In view of the adopted notation, the assumption $\hat{C}$ TP_2 becomes

   $$\alpha_{11}\alpha_{22} \geq \alpha_{12}\alpha_{21}$$

   or, equivalently,

   $$\log \alpha_{11} - \log \alpha_{12} \geq \log \alpha_{21} - \log \alpha_{22}.$$

   Furthermore, since $\overline{G}$ is IFR,

   $$R^{-1}(x) = \overline{G}^{-1}(e^{-x})$$

   is concave and increasing in x.
   Thus, applying $R^{-1}(\cdot)$ to $-\log \alpha_{ij},\ i, j = 1, 2$, we obtain

   $$R^{-1}(-\log \alpha_{12}) - R^{-1}(-\log \alpha_{11}) \geq$$

   $$\geq R^{-1}(-\log \alpha_{22}) - R^{-1}(-\log \alpha_{21})$$

   and hence

   $$\overline{G}^{-1}(\alpha_{11}) + \overline{G}^{-1}(\alpha_{22}) \leq \overline{G}^{-1}(\alpha_{12}) + \overline{G}^{-1}(\alpha_{21}). \tag{20}$$

   This is equivalent to B TP_2, in fact we can rewrite (20) as

   $$-\overline{G}^{-1}(\alpha_{11}) - \overline{G}^{-1}(\alpha_{22}) \geq -\overline{G}^{-1}(\alpha_{12}) - \overline{G}^{-1}(\alpha_{21}).$$

   By applying the exponential to both the members, we obtain

   $$e^{-\overline{G}^{-1}(\alpha_{11})} e^{-\overline{G}^{-1}(\alpha_{22})} \geq e^{-\overline{G}^{-1}(\alpha_{12})} e^{-\overline{G}^{-1}(\alpha_{21})},$$

   that is

   $$B(u, v)B(u', v') \geq B(u, v')B(u', v).$$

2. By the assumption $\hat{C}$ RR_2

   $$\alpha_{11}\alpha_{22} \leq \alpha_{12}\alpha_{21}.$$

   Thus, by putting

   $$u_{ij} := -\log \alpha_{ij}, \quad i, j = 1, 2,$$

   $$u_{22} - u_{21} \geq u_{12} - u_{11}, \tag{21}$$

   with $u_{22} > u_{21},\ u_{12} > u_{11}$. Furthermore, since The assumption B is TP_2, Eq. (20) holds, or, equivalently,

   $$R^{-1}(u_{12}) - R^{-1}(u_{11}) \geq R^{-1}(u_{22}) - R^{-1}(u_{21}). \tag{22}$$

   By (21) and since $R^{-1}(x)$ is increasing in x, this last inequality holds only if $R^{-1}(x)$ is concave in x, that is $\overline{G}$ is IFR.

3. We have to prove now that

   $$\alpha_{11}\alpha_{22} \geq \alpha_{12}\alpha_{21},$$

   that is equivalent to

   $$u_{22} - u_{21} \leq u_{12} - u_{11}.$$

   Since $\overline{G}$ is DFR, $R^{-1}(x)$ is increasing and convex in x. Thus, if $u_{22} - u_{21} > u_{12} - u_{11}$, we should have

   $$R^{-1}(u_{12}) - R^{-1}(u_{11}) < R^{-1}(u_{22}) - R^{-1}(u_{21}),$$

   that contradicts Eq. (22) and therefore the hypothesis that B is TP_2.

□

From now on we expand on ideas contained in [1]. For different families $\mathcal{C}$ of semicopulas, we analyze and compare conditions, on a survival model $\overline{F}$, of the type

$$\widehat{C}_t \in \mathcal{C}, \forall t \geq 0,$$

$$B_t \in \mathcal{C}, \forall t \geq 0.$$

More precisely, the families that will be considered are the following:

the families of PQD and NQD exchangeable semicopulas; we recall that a semicopula S is PQD or NQD if it is $S(u, v) \geq u \cdot v$ or $S(u, v) \leq u \cdot v$, respectively;

the families of LTD and LTI exchangeable semicopulas; we recall that a semicopula S is LTD or LTI if $\frac{S(u,v)}{u}$ is non-increasing or non-decreasing in u, respectively;

the families of exchangeable semicopulas $\mathcal{P}_+^{(3)}$ ($\mathcal{P}_-^{(3)}$), that were considered in [1] and [2], defined by the inequality

$$S(us, v) \geq S(s, uv)\ (S(us, v) \leq S(s, uv)),$$

$0 \leq v \leq u \leq 1$;

the families of TP_2 and RR_2 exchangeable semicopulas, that were already mentioned.

Fix now a family $\mathcal{C}$ among those listed just above. We notice that, in view of Remark 3, $\overline{F}$ being such that $\widehat{C}_t \in \mathcal{C}$, $\forall t \geq 0$ is just a condition on $\widehat{C}$; since we interpret $\widehat{C} \in \mathcal{C}$ as a condition of dependence, we can also interpret $\widehat{C}_t \in \mathcal{C}$, $\forall t \geq 0$ as a condition of dependence, that is typically stronger than $\widehat{C} \in \mathcal{C}$.

Similarly, in the spirit of [2] and in view of Remark 4, we can interpret $B_t \in \mathcal{C}$, $\forall t \geq 0$ as a condition of bivariate ageing.

In this way we can say that we are introducing here some potentially new notions of dependence and of bivariate ageing and want to analyze the relations existing among them. In this respect we develop and extend the approach in [1] by means of a few new remarks and results.

Let us start by considering the notion of PQD.

Proposition 6. *The condition $\widehat{C}_t$ PQD, $\forall t \geq 0$, is equivalent to*

$$\widehat{C}(u,u)\widehat{C}(u'',v'') \geq \widehat{C}(u,v'')\widehat{C}(u,u''), \qquad (23)$$

$\forall 0 \leq u \leq u'' \leq 1, 0 \leq u \leq v'' \leq 1$.

Proof. It is well known (and immediate to check) that the survival copula of a bivariate survival function $\overline{M}$ is PQD if and only if it is

$$\overline{M}(x,y) \geq \overline{G}_{\overline{M}}(x) \cdot \overline{G}_{\overline{M}}(y).$$

By taking $\overline{M} = \overline{F}_t$ we then see, in view of (4), that $\widehat{C}_t$ PQD, $\forall t \geq 0$, means

$$\frac{\overline{F}\,(t+x,t+y)}{\overline{F}\,(t,t)} \geq \frac{\overline{F}\,(t+x,t)}{\overline{F}\,(t,t)}\frac{\overline{F}\,(t,t+y)}{\overline{F}\,(t,t)},$$

$\forall\ 0 \leq t,x,y,$

that can also be written in the form

$$\widehat{C}\left(\overline{G}(t),\overline{G}(t)\right)\hat{C}\left(\overline{G}(t+x),\overline{G}(t+y)\right)$$
$$\geq \hat{C}\left(\overline{G}(t+x),\overline{G}(t)\right)\hat{C}\left(\overline{G}(t),\overline{G}(t+y)\right).$$

By the arbitrariness in the choice of $t,x,y \geq 0$, the proof can be completed by letting

$$u = \overline{G}(t),\ u'' = \overline{G}(t+x),\ v'' = \overline{G}(t+y).$$

□

We notice that the condition in (23) is weaker than $\widehat{C}$ TP_2 and strictly implies $\widehat{C}$ PQD.

As to the family of LTD semicopulas we can state

Proposition 7. *The condition $\widehat{C}_t$ LTD, $\forall t \geq 0$, is equivalent to $\widehat{C}$ being TP_2.*

Proof. It is also well known (and, again, immediate to check) that the survival copula of a bivariate survival function $\overline{M}$ is LTD if and only if it is

$$\frac{\overline{M}(x,y)}{\overline{G}_{\overline{M}}(x)} \text{ non-decreasing in } x, \forall y \geq 0.$$

By taking again $\overline{M} = \overline{F}_t$ we then see, in view of (4), that $\widehat{C}_t$ LFD,$\forall t \geq 0$, means

$$\frac{\overline{F}\,(t+x'',t+y)}{\overline{F}\,(t+x'',t)} \geq \frac{\overline{F}\,(t+x',t+y)}{\overline{F}\,(t+x',t)},$$

$\forall 0 \leq t,x',x'',y,$ with $x'' > x'$.

By the arbitrariness of t,x',y and the condition $x'' > x'$, we can easily see that the above inequality is equivalent to the TP_2 property of $\overline{F}$. □

As to the condition $\widehat{C}_t$ TP_2, $\forall t \geq 0$, we have the following

Proposition 8. *$\widehat{C}_t$ TP_2,$\forall t \geq 0$ is equivalent to $\widehat{C}$ TP_2.*

Proof. It is known (see e.g. [10]) that, for any fixed $t \geq 0$,

$$\overline{F}_t\ TP_2 \ \Leftrightarrow\ \hat{C}_t\ TP_2.$$

But, since

$$\overline{F}\ TP_2 \ \Rightarrow\ \overline{F}_t\ TP_2 \ \ \forall t \geq 0,$$

as straightly follows by the definitions of $\overline{F}_t$ and TP_2, it is sufficient $\hat{C}$ TP_2 to conclude that

$$\hat{C}_t\ TP_2\ \forall t \geq 0.$$

□

Remark 9. *An alternative proof of Proposition 8 can also be easily obtained by taking into account Proposition 7 and Remark 8. Notice that, even if $\widehat{C}$ TP_2 is a stronger condition than $\widehat{C}$ LTD, we have*

$$\widehat{C}_t\ TP_2, \forall t \geq 0 \text{ if and only if } \widehat{C}_t\ LTD, \forall t \geq 0.$$

We now pass to compare bivariate ageing properties for B and B_t, $\forall\ t \geq 0$.
First, we point out the following facts:

Lemma 1. *(see [2]) The condition*

$$B \in \mathcal{P}_+^{(3)} \qquad (24)$$

is equivalent to $\overline{F}_t$ being Schur-concave.

Lemma 2. *(see [1]) The condition (24) is equivalent to B_t being PQD $\forall\ t \geq 0$.*

By applying the arguments in the Remark 8, we immediately obtain also

Corollary 1. *The condition (24) is equivalent to*

$$B_t \in \mathcal{P}_+^{(3)} \ \forall\ t \geq 0.$$

Remark 10. *The Corollary 1 can be seen as an analog of Proposition 8. We can notice that, even if (24) is a stronger condition than B PQD, we have the chain of equivalences*

$$\overline{F} \text{ Schur-concave } \Leftrightarrow$$

$$B \in \mathcal{P}_+^{(3)} \ \Leftrightarrow\ B_t \in \mathcal{P}_+^{(3)}\ \forall t > 0 \ \Leftrightarrow\ B_t\ PQD\ \ \forall t \geq 0.$$

The condition $\overline{F}$ Schur-concave can be interpreted (see e.g. [2]) as a notion of bivariate IFR (Increasing Failure Rate). We can notice then that the above chain can be seen as a bivariate analog of the following well known equivalence, holding for univariate ageing:

$$\overline{F}\ IFR \ \Leftrightarrow\ \overline{F}_t\ IFR \ \Leftrightarrow\ \overline{F}_t\ NBU.$$

Concerning the TP_2 property for B, we can see that $B_t\ TP_2\ \forall t \geq 0$ is not implied by $B\ TP_2$.

On the other hand, the property $B_t\ TP_2\ \forall t \geq 0$ is actually also stronger than (24). In fact, as an immediate consequence of Lemma 2 and of the fact that $TP_2 \Rightarrow\ PQD$, we have

Corollary 2. $B_t\ TP_2\ \forall t \geq 0 \Rightarrow\ B \in \mathcal{P}_+^{(3)}$.

Remark 11. *While $B_t\ PQD\ \forall t \geq 0 \Rightarrow\ B \in \mathcal{P}_+^{(3)}$, we saw that $\hat{C}_t\ PQD\ \forall t \geq 0$ is not enough to get $\hat{C}\ TP_2$. However, we can still express $\hat{C}\ TP_2$ as a PQD-condition on models of residual lifetimes. Consider the family $\{\overline{F}_{a,b}\}_{a,b\geq 0}$ of joint survival functions,*

$$\begin{aligned}\overline{F}_{a,b}(x,y): &= P(X-a>x, Y-b>y | X>a, Y>b)\\ &= \frac{\overline{F}(x+a,y+b)}{\overline{F}(a,b)}\end{aligned}$$

(so that, with this notation, $\overline{F}_t(x,y) = \overline{F}_{t,t}(x,y)$) and the corresponding families

$$\{\overline{G}_{a,b}^{(X)}\}_{a,b\geq 0},\ \{\overline{G}_{a,b}^{(Y)}\}_{a,b\geq 0},\ \{\hat{C}_{a,b}\}_{a,b\geq 0}.$$

We can easily check that the following equivalences holds

$$\hat{C}_{a,b}\ PQD \ \Leftrightarrow\ \overline{F}_{a,b}\ PQD \text{ and } \hat{C}_{a,b}\ TP_2 \ \Leftrightarrow\ \overline{F}_{a,b}\ TP_2;$$

moreover,

$$\overline{F}_{a,b}\ PQD\ \forall\ a,\ b \ \Leftrightarrow\ \overline{F}_{a,b}\ TP_2\ \forall\ a,\ b \ \Leftrightarrow\ \overline{F}\ TP_2.$$

Results analogous to those above can be easily formulated for the negative dependence properties NQD, LTI, $S \in \mathcal{P}_-^{(3)}$, RR_2, respectively corresponding to PQD, LTD, $S \in \mathcal{P}_+^{(3)}$, TP_2.
By combining the above arguments, some statements analogous to those in Proposition 15 of [1] could also be obtained.

References

[1] B. Bassan, F. Spizzichino, On some properties of dependence and ageing for residual lifetimes in the exchangeable case, in: B. H. Lindqvist, K. A. Doksum (Eds.), *Mathematical and Statistical Methods in Reliability*. World Scientific Publ., River Edge, NJ (Series on Quality, Reliability and Engineering Statistics, Vol. 7), 2003, pp. 235-249.

[2] B. Bassan, F. Spizzichino, Relations among univariate ageing, bivariate ageing and dependence for exchangeable lifetimes, Journal of Multivariate Analysis 93 (2005) 313-339.

[3] A. Charpentier, A. Juri, Limiting dependence structures for tail events, with applications to credit derivatives, Journal of Applied Probability 43 (2006) 563-586.

[4] D. G. Clayton, A model for association in bivariate life tables and its application in epidemiologic studies of familial tendency in chronic disease incidence, Biometrika 65 (1978) 141-151.

[5] F. Durante, C. Sempi, Semicopulae, Kybernetika 41 (2004) 315-328.

[6] F. Durante, C. Sempi, Copula and Semicopula transforms, International Journal of Mathematics and Mathematical Sciences 4 (2005) 645-655.

[7] E. P. Klement, R. Mesiar, E. Pap, Triangular norms, Kluwer Academic Publishers, Dordrecht, 2000.

[8] E. P. Klement, R. Mesiar, E. Pap, Archimax copulas and invariance under transformations, Comptes Rendus Mathmatique. Acadmie des Sciences, Paris 340 (2005) 755-758.

[9] P. Morillas, A method to obtain a new copula from a given one, Metrika 61 (2005) 169-184.

[10] R. B. Nelsen, An Introduction to Copulas, Springer, New York, 2006.

[11] D. Oakes, A model for association in bivariate survival data, Journal of the American Statistical Association 84 (1989) 487-493.

AN AGGREGATE CLAIMS MODEL BETWEEN INDEPENDENCE AND COMONOTONE DEPENDENCE

Werner Hürlimann
IRIS integrated risk management ag
Dufourstrasse 80
P.O. BOX, CH-8034 Zürich
e-mail: werner.huerlimann@iris.ch

Abstract

We introduce a simple aggregate claims model, which is able to take into account a continuous range of positive dependence between independence and comonotone dependence. It is based on a multivariate extension of the one-parameter bivariate Fréchet copula, which finds a justification as follows. The chosen model uses only one additional dependence parameter, which is chosen such that it yields the most conservative model of aggregate claims with respect to the concordance order for the bivariate margins of this model. A possible numerical implementation of the aggregate claims distribution of the constructed model is proposed.

Keywords: multivariate Fréchet copula, independence, comonotone dependence, aggregate claims model

1 Mean and variance in aggregate claims models

The principal goal of the present study is the construction of a simple aggregate claims model, which is able to take into account a continuous range of positive dependence between independence and comonotone dependence. This is of great practical importance because the gap between independence and comonotone dependence is known to be significant. As a first preliminary step, we determine the minimum and maximum standard deviation in the aggregate model of a risky business with different positively dependent subunits, which are each subdivided into several product categories.

As in [6] suppose an insurance risk business consists of n different *subunits* $U_i, \quad i = 1,...,n$, which are each subdivided in m_i *product categories* $C_{ij}, \quad j = 1,...,m_i$. The *insurance risks* during some insurance period, which are associated to all these different business lines, are measured by the random variables $S_i, \quad i = 1,...,n$, $S_{ij}, \quad j = 1,...,m_i, \quad i = 1,...,n$, which denote the aggregate claims random variables, which are associated to U_i, C_{ij}. The main risk characteristics of the aggregate claims risk process are described as follows. The aggregate claims random variables are random sums of the type

$$S_{ij} = \sum_{k=1}^{N_{ij}} Y_{ij}^{(k)}, \quad S_i = \sum_{k=1}^{N_i} Y_i^{(k)}, \tag{1.1}$$

where N_i, N_{ij} are claim number random variables describing the frequency in the planning units and product categories, and $Y_i^{(k)}, Y_{ij}^{(k)}$ are claim size random variables describing severity given the k-th claim has occurred in the planning units and product categories. It is assumed that the $Y_i^{(k)} \sim Y_i$ respectively $Y_{ij}^{(k)} \sim Y_{ij}$ are identically distributed random variables.

Given are the first two moments of the frequency and severity random variables associated to the product categories, that is

μN_{ij} : mean number of claims of the product categories C_{ij}

σN_{ij} : standard deviation of the number of claims of the product categories C_{ij}

μY_{ij} : mean claim size of the product categories C_{ij}

σY_{ij} : standard deviation of the claim size of the product categories C_{ij}

It follows that the mean and variance of the aggregate claims random variables of the product categories are given by

$$\begin{aligned} &\mu S_{ij} = \mu N_{ij} \cdot \mu Y_{ij}, \\ &(\sigma S_{ij})^2 = \mu N_{ij} \cdot (\sigma Y_{ij})^2 \\ &+ (\sigma N_{ij})^2 \cdot (\mu Y_{ij})^2 \end{aligned} \tag{1.2}$$

Since $S_i = \sum_{j=1}^{m_i} S_{ij}$ the mean associated to frequency, severity and aggregate claims random variables of the subunits are obtained from the formulas:

$$\begin{aligned} \mu N_i &= \sum_{j=1}^{m_i} \mu N_{ij}, \\ \mu Y_i &= \sum_{j=1}^{m_i} \left(\frac{\mu N_{ij}}{\mu N_i} \right) \mu Y_{ij}, \\ \mu S_i &= \sum_{j=1}^{m_i} \mu S_{ij} \end{aligned} \tag{1.3}$$

To obtain the standard deviation (or variance) of the aggregate claims random variables of the subunits for positively dependent product categories, one notes that there are 4 basic models related to the extremal situations of independence and comonotone dependence of the claim number and claim size random variables.

(V1) N_{ij} independent, Y_{ij} independent

(V2) N_{ij} independent, Y_{ij} comonotone dependent

(V3) N_{ij} comonotone dependent, Y_{ij} independent

(V4) N_{ij} comonotone dependent, Y_{ij} comonotone dependent

The standard deviation (or variance) in these 4 basic cases are calculated using the following well-known relationships:

N_{ij} independent =>

$$(\sigma N_i)^2 = \sum_{j=1}^{m_i} (\sigma N_{ij})^2 \tag{1.4}$$

N_{ij} comonotone dependent =>

$$\sigma N_i = \sum_{j=1}^{m_i} \sigma N_{ij} \tag{1.5}$$

Y_{ij} independent =>

$$(\sigma Y_i)^2 = \sum_{j=1}^{m_i} \left(\frac{\mu N_{ij}}{\mu N_i} \right)^2 (\sigma Y_{ij})^2 \tag{1.6}$$

Y_{ij} comonotone dependent =>

$$\sigma Y_i = \sum_{j=1}^{m_i} \left(\frac{\mu N_{ij}}{\mu N_i} \right) \sigma Y_{ij} \tag{1.7}$$

The variance of the aggregate claims in these basic models is given by

$$\begin{aligned} &(\sigma S_i)^2 = \\ &\mu N_i \cdot (\sigma Y_i)^2 + (\sigma N_i)^2 \cdot (\mu Y_i)^2 \end{aligned} \tag{1.8}$$

The main focus is now on the aggregate claims of the whole insurance risk business, which is described by the random variable

$$S = \sum_{i=1}^{n} S_i \,. \tag{1.9}$$

First of all it is clear that the means of frequency, severity and aggregate claims satisfy the relationships

$$\begin{aligned} \mu N &= \sum_{i=1}^{n} \mu N_i \,, \\ \mu Y &= \sum_{i=1}^{n} \left(\frac{\mu N_i}{\mu N} \right) \cdot \mu Y_i \,, \\ \mu S &= \sum_{i=1}^{n} \mu S_i \end{aligned} \tag{1.10}$$

where one uses that $N = \sum_{i=1}^{n} N_i$ and $Y = \sum_{i=1}^{n} \left(\frac{\mu N_i}{\mu N} \right) \cdot Y_i$. Second, if the random variables S_i are positively dependent between independence and comonotone dependence, the variance will be bounded by the two values

$$\begin{aligned} (\sigma S_{\min})^2 &= \mu N \cdot \left(\sum_{i=1}^{n} \left(\frac{\mu N_i}{\mu N} \right)^2 \cdot (\sigma Y_i)^2 \right) \\ &+ \left(\sum_{i=1}^{n} (\sigma N_i)^2 \right) \cdot (\mu Y)^2 \end{aligned} \tag{1.11}$$

$$\begin{aligned} (\sigma S_{\max})^2 &= \mu N \cdot \left(\sum_{i=1}^{n} \left(\frac{\mu N_i}{\mu N} \right) \cdot \sigma Y_i \right)^2 \\ &+ \left(\sum_{i=1}^{n} \sigma N_i \right)^2 \cdot (\mu Y)^2 \end{aligned} \tag{1.12}$$

2 A one-parameter multivariate Fréchet copula

The more detailed study of the aggregate claims model of a risky business requires the specification of a multivariate distribution for the risk of its subunits.

A natural framework for the construction of multivariate non-normal distributions is the method of copulas, justified by the theorem of Sklar in [12]. It permits a separate study and modeling of the marginal distributions and the dependence structure. According to [10], Section 4.1, a parametric family of distributions should satisfy four desirable properties:

a) There should exist an interpretation like a mixture or other stochastic representation.
b) The margins, at least the univariate and bivariate ones, should belong to the same parametric family and numerical evaluation should be possible.
c) The bivariate dependence between the margins should be described by a parameter and cover a wide range of dependence.
d) The multivariate distribution and density should preferably have a closed-form representation, at least numerical evaluation should be possible.

In general, these desirable properties cannot be fulfilled simultaneously. For example, multivariate normal distributions satisfy properties a), b) and c) but not d). The method of copulas satisfies property c) but implies only partial closedness under the taking of margins, and can lead to computational complexity as the dimension increases. In fact, parametric families of copulas that satisfy all of the desirable properties are seldom. In [7] such a parametric family, called *multivariate linear Spearman copula*, has been constructed (formula (4.9)). It is based on the method of mixtures of independent conditional distributions, a variant of [10], Section 4.5, which is described as follows.

To satisfy property b), let us focus on the n Fréchet classes $FC_i := FC_i(F_{ij},\ j \neq i), \quad i = 1,...,n$, of n-variate distributions for which the bivariate margins

$$\begin{aligned} F_{ij}(x_i, x_j) &= F_{(X_i, X_j)}(x_i, x_j) \\ &= C_{ij}\left[F_i(x_i), F_j(x_j)\right], \quad j \neq i, \end{aligned}$$

belong to a given parametric family of copulas $C_{ij}\left[u_i, u_j\right]$. Assume that the conditional distributions

$$F_{j|i}(x_j | x_i) = \frac{\partial C_{ij}}{\partial u_i}\left[F_i(x_i), F_j(x_j)\right] \tag{2.1}$$

are well-defined. The n-variate distribution such that the random variables $X_j,\ j \neq i$, are conditionally independent given X_i, is contained in FC_i and is defined by

$$F^{(i)}(x) = \int_{-\infty}^{x_i} \left[\prod_{j \neq i} F_{j|i}(x_j|t) \right] \cdot dF_i(t) \qquad (2.2)$$

Choosing appropriately the bivariate copulas $C_{ij}\lfloor u_i, u_j \rfloor$, it is possible to construct n-variate copulas $C^{(i)}(u_1,...,u_n), \quad i = 1,...,n,$ such that $F^{(i)}$ belongs to $C^{(i)}$ and the bivariate margins $F_{ij}, \ j \neq i,$ belong to C_{ij}. Moreover, any convex combination of the $C^{(i)}$'s, that is

$$\begin{aligned} &C(u_1,...,u_n) = \sum_{i=1}^{n} \lambda_i C^{(i)}(u_1,...,u_n), \\ &0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^{n} \lambda_i = 1 \end{aligned} \qquad (2.3)$$

is again a n-variate copula, which by appropriate choice may satisfy the desirable properties. In the following, the class (2.3) of all convex combinations of mixtures of independent conditional distributions with given bivariate margins is denoted by CIC.

In [7], the subclass of CIC, denoted here CICF, which is generated by the bivariate margins with Fréchet copulas

$$\begin{aligned} &C_{ij}(u_i, u_j) \\ &= (1-\theta_{ij}) u_i u_j + \theta_{ij} \min(u_i, u_j) \end{aligned} \qquad (2.4)$$

where $\theta_{ij} \in [0,1]$, has been studied in detail. It is well-known that the parameter θ_{ij} of the copula (2.4) coincides with the grade correlation coefficient introduced in [13], simply called Spearman coefficient.

Though presumably possible, the evaluation of the full distribution of aggregate claims using the multivariate distributions in CICF should be rather technical. Instead of this, we consider the simple *multivariate Fréchet copula*

$$\begin{aligned} &C(u_1,...,u_n) = (1-\theta) \cdot \left[\prod_{i=1}^{n} u_i \right] \\ &+ \theta \cdot \min_{1 \leq i \leq n}(u_i), \quad \theta \in [0,1]. \end{aligned} \qquad (2.5)$$

Intuitively, the copula (2.5) models the whole range of possible dependence structure between independence and comonotone dependence. It would also be possible to model similarly the possible dependence structure between independence and "minimal" dependence.

The practical application of this model is motivated as follows. Guided by the concern of "prudent" valuation, we require that the bivariate margins of (2.5) are at least as positively dependent as those obtained from the bivariate model (2.4) in the concordance ordering (e.g. [1], [14], [3]). Consider the bivariate copulas of the bivariate margins of (2.5), which are denoted and given by

$$\begin{aligned} &C^{ij}(u_i, u_j) \\ &= (1-\theta) u_i u_j + \theta \min(u_i, u_j) \end{aligned} \qquad (2.6)$$

Then the stated condition means that $C_{ij}(u_i, u_j) \leq C^{ij}(u_i, u_j)$, and this is equivalent to

$$\begin{aligned} &(\theta - \theta_{ij}) \cdot u_i u_j \\ &\leq (\theta - \theta_{ij}) \cdot \min(u_i, u_j) \end{aligned}. \qquad (2.7)$$

or

$$\theta \geq \theta_{ij}, \quad \text{for all } (i, j). \qquad (2.8)$$

Therefore, for prudent aggregate claims evaluation, it is reasonable to set the dependence parameter of the model (2.5) equal to

$$\theta = \max_{(i,j)} \{\theta_{ij}\}. \qquad (2.9)$$

This model choice yields the most conservative model for aggregate claims with respect to the concordance order for the bivariate margins of this model.

It should be pointed out that the considered very simple multivariate Fréchet copula, which models the whole range of dependence between independence and comonotone dependence, finds important practical applications in the fields of insurance and finance. The interested reader is invited to consult [8] and [9].

3 Multivariate Fréchet compound models of aggregate claims

Given that the volatility, as measured by the variance, varies between two bounds, what is a useful and practical way to get an aggregate claims distribution for the whole insurance risk business? To

model the positive dependence between independence and comonotone dependence, it is reasonable to consider a model based on the multivariate Fréchet copula, which has been introduced and motivated in Section 2. Besides notations and standard assumption introduced in Section 1, our *multivariate Fréchet compound model of aggregate claims* is based on the following assumptions:

(A1) The random vector $\boldsymbol{N} = (N_1,...,N_n)$ is multivariate Fréchet distributed with dependence parameter ρ_N.

(A2) The random vector $\boldsymbol{Y} = (Y_1,...,Y_n)$ is multivariate Fréchet distributed with dependence parameter ρ_Y.

(A3) The random variables N_i are independent from the Y_i's.

Let $\boldsymbol{N}^{\perp} = (N_1^{\perp},...,N_n^{\perp})$ be a version of $\boldsymbol{N}$ with independent components, $\boldsymbol{N}^{+} = (N_1^{+},...,N_n^{+})$ a version of $\boldsymbol{N}$ with comonotone dependent components, $\boldsymbol{Y}^{\perp} = (Y_1^{\perp},...,Y_n^{\perp})$ a version of $\boldsymbol{Y}$ with independent components, and $\boldsymbol{Y}^{+} = (Y_1^{+},...,Y_n^{+})$ a version of $\boldsymbol{Y}$ with comonotone dependent components. Under the multivariate Fréchet assumption (A1), the distribution of the aggregate claim number random variable $N = \sum_{i=1}^{n} N_i$ is then given by

$$F_N(k) = (1-\rho_N)\cdot F_{N^{\perp}}(k) + \rho_N \cdot F_{N^{+}}(k), \quad k = 0,1,2,... \tag{3.1}$$

where $N^{\perp} = \sum_{i=1}^{n} N_i^{\perp}$ and $N^{+} = \sum_{i=1}^{n} N_i^{+}$. In the following let us set $\lambda_i = \mu N_i$, $\lambda = \mu N = \sum_{i=1}^{n} \lambda_i$. Under the multivariate Fréchet assumption (A2) the distribution of $Y = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} \cdot Y_i$ is given by

$$F_Y(y) = (1-\rho_Y)\cdot F_{Y^{\perp}}(y) + \rho_Y \cdot F_{Y^{+}}(y), \quad y > 0 \tag{3.2}$$

where $Y^{\perp} = \sum_{i=1}^{n} Y_i^{\perp}$ and $Y^{+} = \sum_{i=1}^{n} Y_i^{+}$. The main interest lies in the distribution of the aggregate claims $S = \sum_{i=1}^{n} S_i = \sum_{i=1}^{N} Z_i$, where each Z_i has the distribution of Y. By assumption (A3) each Z_i is independent from N and the compound model of risk theory implies the well-known convolution representation for the aggregate claims distribution

$$F_S(s) = \sum_{k=0}^{\infty} p_N(k)\cdot F_Y^{*(k)}(s) \tag{3.3}$$

with the claim number probabilities

$$p_N(k) = F_N(k) - F_N(k-1), \quad k = 1,2,... \quad p_N(0) = F_N(0). \tag{3.4}$$

It is not difficult to see that the k-th convolution of Y has the representation

$$F_Y^{*(k)}(y) = \sum_{j=0}^{k} \binom{k}{j} \rho_Y^{j} (1-\rho_Y)^{k-j} \cdot \left[F_{Y^{+}}^{*(j)} * F_{Y^{\perp}}^{*(k-j)}\right](y) \tag{3.5}$$

It is clear that the extreme cases $\rho_N = \rho_Y = 0$ and $\rho_N = \rho_Y = 1$ yield the minimum and maximum variance of S as given above in (1.11) and (1.12). In the special cases $\rho_Y = 0$ and $\rho_Y = 1$, the aggregate claims distribution takes the simpler form:

$$F_{S^{\perp}}(s) = \sum_{k=0}^{\infty} p_N(k)\cdot F_{Y^{\perp}}^{*(k)}(s) \tag{3.6}$$

$$F_{S^{+}}(s) = \sum_{k=0}^{\infty} p_N(k)\cdot F_{Y^{+}}^{*(k)}(s) \tag{3.7}$$

Instead of (3.3) it is reasonable to consider the Fréchet model

$$F_S(s) = (1-\rho_S)\cdot F_{S^\perp}(s) + \rho_S \cdot F_{S^+}(s) \tag{3.8}$$

with some Spearman coefficient $\rho_S \in [0,1]$. This simpler model yields the minimum and maximum variance provided $\rho_N = \rho_S = 0$ respectively $\rho_N = \rho_S = 1$. Each pair $(\rho_N, \rho_S) \in [0,1]^2$ defines another *multivariate Fréchet compound model of aggregate claims*.

4 Numerical evaluation of a multivariate Fréchet compound model

For a numerical evaluation of our multivariate Fréchet compound model, we assume for simplicity that claim numbers are Poisson distributed, that is $N_i \sim Po(\lambda_i), \quad i = 1,...,n$, and that claim sizes are Gamma distributed, that is $Y_i \sim \Gamma(\alpha_i, \beta_i), \quad i = 1,...,n$. Note that the widely used compound Poisson Gamma aggregate claims model has been justified through a characterization result in Mathematical Statistics in [4]. The introduced claim size random variables $Y^\perp$, respectively Y^+, are the independent, respectively comonotone, sums of the gamma distributed random variables $\Gamma\left(\alpha_i, \frac{\lambda}{\lambda_i}\beta_i\right), \quad i = 1,...,n$, whose distribution functions can be determined using the methods presented in [5]. Through approximation it is possible to assume that $Y^\perp$ and Y^+ are also gamma distributed with mean

$$\mu_Y = \frac{1}{\lambda}\sum_{i=1}^{n} \lambda_i \frac{\alpha_i}{\beta_i} \tag{4.1}$$

and standard deviations

$$\sigma_{Y^\perp} = \frac{1}{\lambda}\sqrt{\sum_{i=1}^{n} \lambda_i^2 \frac{\alpha_i}{\beta_i^2}} \quad \text{(independent case)} \tag{4.2}$$

$$\sigma_{Y^+} = \frac{1}{\lambda}\sum_{i=1}^{n} \lambda_i \frac{\sqrt{\alpha_i}}{\beta_i} \quad \text{(comonotone case)} \tag{4.3}$$

In this situation the claim sizes are gamma distributed such that $Y^\perp \sim \Gamma(\alpha^\perp, \beta^\perp)$ and $Y^+ \sim \Gamma(\alpha^+, \beta^+)$, with

$$\alpha^\perp = k_{Y^\perp}^{-2},\ \beta^\perp = \alpha^\perp \cdot \mu_Y^{-1},$$
$$\alpha^+ = k_{Y^+}^{-2},\ \beta^+ = \alpha^+ \cdot \mu_Y^{-1},$$

where $k_{Y^\perp}, k_{Y^+}$ are the coefficients of variation of the claims size random variables $Y^\perp$ and Y^+. In the special case

$$\lambda_i = \frac{\lambda}{n},\ \alpha_i = \alpha,\ \beta_i = \beta,\ i = 1,...,n$$

the claim sizes are gamma distributed with parameters

$$\alpha^\perp = n\alpha,\ \beta^\perp = n\beta,\ \alpha^+ = \alpha,\ \beta^+ = \beta$$

In general, the *k*-th convolution of the claim sizes $Y^\perp$ and Y^+ are gamma distributed such that

$$(Y^\perp)^{*(k)} \sim \Gamma(k\alpha^\perp, \beta^\perp),$$
$$(Y^+)^{*(k)} \sim \Gamma(k\alpha^+, \beta^+)$$

The claim number random variables $N^\perp$ is Poisson distributed with parameter λ. An explicit analytical expression for the claim number distribution N^+ is not known unless $\lambda_i = \frac{\lambda}{n},\ i = 1,...,n$, in which case the probabilities are

$$\Pr(N^+ = k) = Po\left(\left[\frac{k}{n}\right]; \frac{\lambda}{n}\right),\ k = 0,1,2,...$$

The latter result suggests the following approximation in the general case. The claim number $N^\perp$ has the same distribution as the claim number random variable

$$\tilde{N}^\perp = \sum_{i=1}^{n} \tilde{N}_i^\perp,\ \tilde{N}_i \sim Po\left(\frac{\lambda}{n}\right)$$

This suggests to approximate the claim number N^+ by the random variable $\tilde{N}^+ = \sum_{i=1}^{n} \tilde{N}_i^+$, which implies that the probability $\Pr(N^+ = k)$ is approximately equal to

$$\Pr(\tilde{N}^+ = k) = Po\left(\left[\frac{k}{n}\right]; \frac{\lambda}{n}\right),\ k = 0,1,2,...$$

Proceeding this way, the distribution of the

multivariate Fréchet compound Poisson model of aggregate claims with gamma claim sizes can be evaluated using the following analytical expressions:

$$F_S(s) = (1-\rho_S)\cdot F_{S^\perp}(s) + \rho_S \cdot F_{S^+}(s) \quad (4.4)$$

$$F_{S^\perp}(s) = \sum_{k=0}^{\infty} p_N(k)\cdot G\left(\beta^\perp s; k\alpha^\perp\right) \quad (4.5)$$

$$F_{S^+}(s) = \sum_{k=0}^{\infty} p_N(k)\cdot G\left(\beta^+ s; k\alpha^+\right) \quad (4.6)$$

$$G(x;\alpha) = \frac{1}{\Gamma(\alpha)}\cdot\int_0^x t^{\alpha-1}e^{-t}dt \quad (4.7)$$

$$p_N(k) = (1-\rho_N)\cdot p_{N^\perp}(k) + \rho_N \cdot p_{N^+}(k), \quad k = 0,1,2,... \quad (4.8)$$

$$p_{N^\perp}(k) = Po(k;\lambda) - Po(k-1;\lambda), \quad k = 1,2,..., \quad p_{N^\perp}(0) = e^{-\lambda} \quad (4.9)$$

$$p_{N^+}(k) = Po\left(\left[\frac{k}{n}\right];\frac{\lambda}{n}\right) - Po\left(\left[\frac{k-1}{n}\right];\frac{\lambda}{n}\right) \quad (4.10)$$

$$k = 1,2,..., \quad p_{N^+}(0) = e^{-\frac{\lambda}{n}}$$

The following Table illustrates the exact calculation of the multivariate Fréchet compound Poisson gamma distribution in the special case

$$n = 3, \quad \lambda_i = \frac{1}{3}, \alpha_i = \frac{1}{4}, \beta_i = \frac{1}{4}, i = 1,2,3,$$

$$\rho_N = 0.37895, \rho_S = 0.2$$

It is immediately seen that the aggregate claims random variables are stochastically ordered in the dangerousness order, that is one has the inequality $S^\perp \le_D S \le_D S^+$ (e.g. [11] for a definition of dangerousness order). The interest of such inequalities is well-known because actuaries feel that positive dependence reveal a more dangerous situation compared to independence (e.g. [2]).

Table 4.1: Distribution of aggregate claims for a multivariate Fréchet compound model

Aggregate Claims s	$F_{S^\perp}(s)$	$F_S(s)$	$F_{S^+}(s)$
0	0.50000	0.50000	0.50000
1	0.70156	0.71676	0.77757
2	0.81401	0.82055	0.84673
3	0.88544	0.88623	0.88936
4	0.93003	0.92769	0.91835
5	0.95744	0.95374	0.93894
6	0.97415	0.97011	0.95393
7	0.98430	0.98044	0.96502
8	0.99045	0.98702	0.97331
9	0.99418	0.99126	0.97955
10	0.99645	0.99402	0.98429
11	0.99783	0.99585	0.98790
12	0.99868	0.99707	0.99066
13	0.99919	0.99791	0.99278
14	0.99951	0.99849	0.99441
15	0.99970	0.99889	0.99567
16	0.99982	0.99918	0.99664
17	0.99989	0.99939	0.99739
18	0.99993	0.99954	0.99797
19	0.99996	0.99965	0.99842
20	0.99997	0.99973	0.99877
21	0.99998	0.99980	0.99904
22	0.99999	0.99984	0.99925
23	0.99999	0.99988	0.99942
24	1.00000	0.99991	0.99954

References

[1] S. Cambanis, G. Simons and W. Stout, Inequalities for Ek(X,Y) when the marginals are fixed. Zeitschrift für Wahrscheinlichkeits-theorie und verw. Gebiete 36 (1976), 285-294.

[2] M. Denuit, J. Dhaene and C. Ribas, Does positive dependence between individual risks increase stop-loss premiums, Insurance: Mathematics and Economics 28 (2001) 305-08.

[3] J. Dhaene and M.J. Goovaerts. Dependency of risks and stop-loss premiums. ASTIN Bulletin 26 (1996), 201-212.

[4] W. Hürlimann, On mean scaled insurance risk models. Blätter der Deutschen Gesellschaft für Versicherungsmathematik XXIII(3) (1998), 251-265.

[5] W. Hürlimann, Analytical evaluation of economic risk capital for portfolios of Gamma risks. ASTIN Bulletin 31 (2001), 107-122.

[6] W. Hürlimann, Economic risk capital allocation from top down, Blätter der Deutschen Gesellschaft für Versicherungsmathematik XXV(4) (2002), 885-891.

[7] W. Hürlimann, Multivariate Fréchet copulas and conditional value-at-risk, International Journal of Mathematics and Mathematical Sciences 7 (2004), 345-364.

[8] W. Hürlimann, Analytical pricing and risk of best of warrant options (II) a multivariate extension. Appears in The Global Journal of Applied Mathematics and Mathematical Sciences 1(1) (2006).

[9] W. Hürlimann, A Gamma IBNR claims reserving model with dependent development periods, To be presented at the 37th International ASTIN Colloquium, Orlando, Florida, June 19-22 (2007).

[10] H. Joe, Multivariate Models and Dependence Concepts, Monographs on Statistics and Applied Probability 73, Chapman and Hall, London (1997).

[11] R. Kaas, A.E. Van Heerwaarden and M.J. Goovaerts, Ordering of Actuarial Risks, Caire Education Series, Amsterdam (1994).

[12] A. Sklar, Fonctions de répartition et leurs marges. Publications de l'Institut Statistique de l'Université de Paris 8 (1959), 229-31.

[13] C. Spearman, The proof and measurement of association between two things, American Journal of Psychology 15 (1904), 72-101.

[14] A.H. Tchen, Inequalities for distributions with given marginals. The Annals of Probability 8 (1980), 814-827.

Bounds for Value at Risk for Asymptotically Dependent Assets - the Copula Approach

Piotr Jaworski
Warsaw University, Institute of Mathematics, ul.Banacha 2, 02-097 Warszawa, Poland
email: jwptxa@mimuw.edu.pl

Abstract

The theory of copulas provides a useful tool for modeling dependence in risk management. In insurance and finance, as well as in other applications, dependence of extreme events is particularly important, hence there is a need for the detailed study of the tail behaviour of the multivariate copulas. In this paper we investigate the class of copulas having homogeneous lower tails. We show that having only such information on the structure of dependence of returns from assets is enough to get estimates on Value at Risk of the multiasset portfolio in terms of Value at Risk of one-asset portfolios.

Keywords: Copulas, Value at Risk, dependence of extreme events.

1 Introduction

In my presentation I shall deal with the advantages of modeling the dependence between the extremal events with the help of copulas. Let us consider the following case. An investor operating on an emerging market, has in his portfolio several currencies which are highly dependent. Let s_i, $i = 1, \ldots, d$ be the quotients of the currency rates at the end and at the beginning of the investment. Let w_i be the part of the capital invested in the i-th currency, $\sum w_i = 1$, $w_i \geq 0$. So the final value of the investment equals

$$W_1(w) = (w_1 s_1 + \ldots + w_d s_d) \cdot W_0.$$

For portfolio consisting of only one currency (say i -th) we put $w = e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$.

The crucial point is to estimate the risk of keeping the portfolio. As a measure of risk I shall consider "Value at Risk" (VaR), which last years became one of the most popular measures of risk in the "practical" quantitative finance (see for example [2, 20, 5, 6, 17, 16, 10, 19]). Roughly speaking the idea is to determine the biggest amount one can lose on a certain confidence level $1 - \alpha$

$$VaR_{1-\alpha}(w) = sup\{V : P(W_0 - W_1(w) \leq V) < 1 - \alpha\}.$$

Note that this quantity warns risk managers how much of "economic" capital (reserves) is needed to keep the solvency.
In order to determine VaR accurately one has to deal with the complexity of the problem. The extremes hardly follow the normal distribution law. Therefore the main challenge is to describe properly the interdependencies of risk factors (in our case the changes of currency rates). In this presentation, it will be based on copulas, which are scaleless dependency measures of random variables. I will show that sometimes it is enough to have only the partial information on the given copula.

The main result, I would like to present, is the following estimate of the Value at Risk of a given portfolio w in terms of Value at Risk of one-currency portfolios e_i (compare [14] for two dimensional case):

$$\sum w_i VaR_{1-\alpha}(e_i) \geq VaR_{1-\alpha}(w) \geq$$

$$\geq \sum w_i VaR_{1-\alpha'}(e_i),$$

where $\alpha' = \frac{\alpha^2}{C(\alpha,\ldots,\alpha)}$. The above estimate is valid for sufficiently small α under the mild assumptions:

- The lower tail part of the copula C of s_i's is nonzero and homogeneous of degree 1, i.e. for sufficiently small q

$$C(q) = L(q), \quad \forall t > 0 \quad L(tq) = tL(q).$$

- For $i = 1, \ldots, d$, for sufficiently small x, the function $G_i(x) = \frac{1}{F_i(x)}$, where F_i is the distribution function of s_i, is convex.

The first assumption is modelling the asymptotic dependence (compare [13] Th.2). For example it describes very well the behaviour of foreign exchange rates on an emerging market, where the *extremal* changes are usually due to the local factors (compare [14]).

The second one is fulfilled by a wide variety of probability laws. For example it is valid if the distributions of $-\ln s_i$ have the same upper tails as normal, Pareto or Gamma distribution (i.e. if their distribution functions coincide for enough big arguments). Moreover, it is easy to check that if $1 - F_i(-x)$ is a von Mises function, i.e

$$\exists z > 0 \quad \forall 0 < x < z \quad F_i(x) = c \exp\left(-\int_x^z \frac{1}{a(t)} dt\right),$$

where a is absolutely continuous and its density has limit 0 at the origin, then $G_i(x)$ is convex (for small x). Note that the von Mises functions played an important role in the Extreme Value Theory, they are classical examples of distribution functions belonging to the Maximum Domain of Attraction of the Gumbel Distribution (for details see [8] §3.3.3).

2 Notation

2.1 Copulas

We recall that a function

$$C : [0,1]^d \longrightarrow [0,1]$$

is called a copula (see [18] §2.10, [4] §4.1, [1] §4.4) if for every $u = (u_1, \ldots, u_d)$ and $v = (v_1, \ldots, v_d)$ $(u_i, v_i \in [0,1])$ and every $j \in \{1, \ldots, d\}$

$$\begin{aligned} &i) && u_j = 0 \Rightarrow C(u) = 0; \\ &ii) && (\forall i \neq j \;\; u_i = 1) \Rightarrow C(u) = u_j; \\ &iii) && u \preceq v \Rightarrow V_C(u,v) \geq 0, \end{aligned}$$

where $u \preceq v$ denotes the partial ordering on $\mathbb{R}^d$,

$$u \preceq v \Leftrightarrow \forall i \quad u_i \leq v_i,$$

and $V_C(u,v)$ is the C-volume of the rectangle $I(u,v)$, the one with lower vertex u and upper vertex v.

$$V_C(u,v) = \sum_{j_1=1}^{2} \ldots \sum_{j_d=1}^{2} (-1)^{j_1+\ldots+j_d} C(a_{1,j_1}, \ldots, a_{d,j_d}),$$

where $a_{i,1} = u_i$ and $a_{i,2} = v_i$ for $i = 1, \ldots, d$.
The functions with the last property are called n-nondecreasing. Those which fulfill the first one are called grounded.

Note that every copula is nondecreasing not only with respect to each variable but also with respect to the partial ordering $\preceq$. Moreover it is continuous and even Lipschitz ([18], Theorem 2.10.7, [4], Lemma 4.2)

$$|C(v) - C(u)| \leq \sum_{i=1}^{d} |v_i - u_i|.$$

REMARK **2.1** *(cf. [3], Th. 12.5) Every continuous, grounded, n-nondecreasing function*

$$H : [0,a]^d \longrightarrow \mathbb{R}$$

is a distribution function of a Borel measure μ_H on $[0,a]^d$

$$H(u) = \mu_H(I(0,u)),$$

$$\mu_H(I(u,v)) = \mu_H(int(I(u,v))) = V_H(u,v).$$

Due to the second condition every copula is a distribution function of a probability measure on the unit rectangle $[0,1]^d$ with uniform margins (compare [15], §1.6). Furthermore, the above remark remains true if H is defined on the whole multioctant $[0,+\infty)^d$.

Let $\mathcal{X}_i$, $i = 1, \ldots, d$ be random variables defined on the same probability space $(\Omega, \mathcal{M}, \mathbb{P})$. Their joint cumulative distribution $F_{\mathcal{X}}$ can be described using an appropriate copula $C_{\mathcal{X}}$ ("Sklar Theorem" see [18], Theorem 2.10.11, [4], Theorem 4.5):

$$F_{\mathcal{X}}(x) = C_{\mathcal{X}}(F_{\mathcal{X}_1}(x_1), \ldots, F_{\mathcal{X}_d}(x_d)),$$

where $F_{\mathcal{X}_i}$ are cumulative distributions of $\mathcal{X}_i$. Note that the strictly increasing transformations of random variables $\mathcal{X}_i$ do not affect the copula. Indeed, if

$$\mathcal{X}_i' = f_i(\mathcal{X}_i), \;\; i = 1, \ldots, d,$$

where f_i are strictly increasing (and so invertible), then

$$F_{\mathcal{X}'}(x) = F_{\mathcal{X}}(f_1^{-1}(x_1), \ldots, f_d^{-1}(x_d)) =$$

$$= C_{\mathcal{X}}(F_{\mathcal{X}_1}(f_1^{-1}(x)), \ldots, F_{\mathcal{X}_d}(f_d^{-1}(x_d))) =$$

$$= C_{\mathcal{X}}(F_{\mathcal{X}_1'}(x_1), \ldots, F_{\mathcal{X}_d'}(x_d)).$$

Therefore if one is interested in tail dependence of random variables rather than in their *individual* distribution, then the proper choice is to study the copula. The more so, since the copula is uniquely determined at every point u such, that the equations $F_{\mathcal{X}_i}(x_i) = u_i$ have solutions.

2.2 Model assumptions

We assume that for $t > 0$ the distribution function of each $s_i - F_i(t)$ is positive and the joint probability distribution of s_i's is continuous with respect to Lebesgue measure and is determined by a copula C

$$F_s(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

Furthermore there is a constant $\delta \in (0,1)$ such that:
A1. For $q = (q_1, \ldots q_d)$ and $0 \le q_i \le \delta$, $C(q) = L(q)$, where L is some nonzero positive homogeneous function of degree one $(\forall t > 0 \quad L(tq) = tL(q))$.
A2. For $i = 1, \ldots, d$ the function $G_i(t) = \frac{1}{F_i(t)}$ restricted to $t \in F_i^{-1}((0, \delta])$ is convex.

The second assumption implies that the preimage of δ consists of just one point and F_i restricted to $[0, F_i^{-1}(\delta)]$ is strictly increasing. Moreover we get a simpler formula for Value at Risk of one-asset portfolios.

COROLLARY **2.1** *For $\alpha \in (0, \delta]$,*

$$VaR_{1-\alpha}(e_i) = W_0 \cdot (1 - F_i^{-1}(\alpha)), \quad i = 1, \ldots, d.$$

In [11, 13] we showed that there is a large class of copulas which tails can be approximated by a homogeneous function L. We recall the basics about L's. Comparing [13] Theorem 3 and the construction from the proof of Proposition 6 (also [13]) one gets:

THEOREM **2.1** *For a homogeneous of degree 1 function L, $L : [0, +\infty)^d \to \mathbb{R}$, the following conditions are equivalent:*
1. L is equal to the lower tail of some copula C.
2. L is d-nondecreasing and

$$0 \le L(u) \le min(u_1, \ldots, u_d) \quad for \quad u \succeq 0.$$

3. L is continuous, grounded, d-nondecreasing and $\mu_L = m \times \mu_\Delta$, where m is the Lebesgue measure on the real halfline and μ_Δ is a measure on the unit simplex $\Delta = \{q \in \mathbb{R}_+^d : q_1 + \ldots + q_d = 1\}$ such that

$$\int_\Delta \frac{1}{q_i} d\mu_\Delta(q) \le 1 \quad for \quad i = 1, \ldots, d.$$

3 Upper estimate

We assume, that $\forall i \;\; w_i > 0$.

THEOREM **3.1** *For positive α such that*

$$\sum_{i=1}^d w_i F_i^{-1}(\alpha) \le \min\{w_j F_j^{-1}(\delta) : j = 1, \ldots, d\}$$

the following inequality holds

$$VaR_{1-\alpha}(w) \le w_1 VaR_{1-\alpha}(e_1) + \ldots + w_d VaR_{1-\alpha}(e_d).$$

For $\lambda = (\lambda_1, \ldots, \lambda_d)$, $\lambda_i > 0$, we put

$$Y_\lambda = \{q \in \mathbb{R}_+^d \;:\; \sum_{i=1}^d \frac{\lambda_i}{q_i} \ge 1\}.$$

LEMMA **3.1** .

$$\mu_L(Y_\lambda) \le \sum \lambda_i.$$

Proof.
We base on the fact, that the multioctant is the Cartesian product of a halfline and simplex

$$\mathbb{R}_+^d = \mathbb{R}_+ \times \Delta.$$

Since L is homogeneous,

$$\mu_L(Y_\lambda) = \int_\Delta m(\mathbb{R}_+ \xi \cap Y_\lambda) d\mu_\Delta(\xi).$$

The intersection of Y_λ and the halfline given by the vector ξ is a segment of length $\sum \frac{\lambda_i}{\xi_i}$,

$$\mathbb{R}_+ \xi \cap Y_\lambda = \{t : \sum \frac{\lambda_i}{t\xi_i} \ge 1\} = \{t : 0 \le t \le \sum \frac{\lambda_i}{\xi_i}\}.$$

Therefore

$$\mu_L(Y_\lambda) = \int_\Delta \sum \frac{\lambda_i}{\xi_i} d\mu_\Delta(\xi) =$$

$$= \sum \lambda_i \int_\Delta \sum \frac{1}{\xi_i} d\mu_\Delta(\xi) \le \sum \lambda_i.$$

□

For $r > 0$ we put

$$V_r = \{q \in \mathbb{R}_+^d \;:\; \sum_{i=1}^d w_i F_i^{-1}(q_i) \le r\}.$$

LEMMA **3.2** *For positive r and α such, that*

$$r = \sum_{i=1}^d w_i F_i^{-1}(\alpha) \le \min\{w_j F_j^{-1}(\delta) : j = 1, \ldots, d\}$$

the following inclusions hold

$$V_r \subset [0, \delta]^d, \quad V_r \subset Y_\lambda,$$

where

$$\lambda_i = \alpha \frac{w_i c_i^{-1}}{\sum w_j c_j^{-1}}; \quad c_j = F_j'(F_j^{-1}(\alpha)).$$

Proof.
If q belongs to V_r then

$$\sum_{i=1}^d w_i F_i^{-1}(q_i) \le r = \sum_{i=1}^d w_i F_i^{-1}(\alpha) \le \min\{w_j F_j^{-1}(\delta)\}.$$

Therefore

$$w_i F_i^{-1}(q_i) \le w_i F_i^{-1}(\delta),$$

and $q_i \le \delta$.

To proof the second inclusion, we use the convexity of $G_i = \frac{1}{F_i}$.

$$\frac{1}{q_i} - \frac{1}{\alpha} = \frac{1}{F_i(F_i^{-1}(q_i))} - \frac{1}{F_i(F_i^{-1}(\alpha))} =$$

$$= G_i(F_i^{-1}(q_i)) - G_i(F_i^{-1}(\alpha)) \ge$$

$$\ge G_i'(F_i^{-1}(\alpha)) \cdot \left(F_i^{-1}(q_i) - F_i^{-1}(\alpha)\right) =$$

$$= \frac{-F_i'(F_i^{-1}(\alpha))}{(F_i(F_i^{-1}(\alpha)))^2} \cdot \left(F_i^{-1}(q_i) - F_i^{-1}(\alpha)\right) =$$

$$= -\frac{c_i}{\alpha^2}\left(F_i^{-1}(q_i) - F_i^{-1}(\alpha)\right)$$

thus

$$F_i^{-1}(q_i) - F_i^{-1}(\alpha) \ge -\frac{\alpha^2}{c_i}\left(\frac{1}{q_i} - \frac{1}{\alpha}\right).$$

If q belongs to V_r then

$$0 \ge \sum_{i=1}^{d} w_i F_i^{-1}(q_i) - r =$$

$$= \sum_{i=1}^{d} w_i F_i^{-1}(q_i) - \sum_{i=1}^{d} w_i F_i^{-1}(\alpha) \ge$$

$$\ge -\sum_{i=1}^{d} \frac{w_i \alpha^2}{c_i}\left(\frac{1}{q_i} - \frac{1}{\alpha}\right) =$$

$$= -\alpha\left(\sum_{i=1}^{d} \frac{\lambda_i}{q_i} \sum_{j=1}^{d} \frac{w_j}{c_j} - \sum_{i=1}^{d} \frac{w_i}{c_i}\right) =$$

$$= -\alpha \sum_{j=1}^{d} \frac{w_j}{c_j}\left(\sum_{i=1}^{d} \frac{\lambda_i}{q_i} - 1\right).$$

So

$$0 \le \sum \frac{\lambda_i}{q_i} - 1,$$

and therefore q belongs to Y_λ.

□

Proof of theorem 3.1 .
In order to estimate $VaR_{1-\alpha}(w)$ we consider

$$1 - P\left(W_0 - W_1(w) \le \sum w_i VaR_{1-\alpha}(e_i)\right) =$$

$$= P\left(W_0 - W_1(w) \ge \sum w_i VaR_{1-\alpha}(e_i)\right) =$$

$$= P\left(1 - \sum w_i s_i \ge \sum w_i (1 - F_i^{-1}(\alpha))\right) =$$

$$= P\left(\sum w_i s_i \le \sum w_i F_i^{-1}(\alpha)\right) =$$

$$= P(\sum w_i s_i \le r) = \mu_C(V_r) =$$

$$= \mu_L(V_r) \le \mu_L(Y_\lambda) \le \sum \lambda_i = \alpha.$$

So

$$P\left(W_0 - W_1(w) \le \sum w_i VaR_{1-\alpha}(e_i)\right) \ge 1 - \alpha.$$

In such a way we obtain the estimate

$$VaR_{1-\alpha}(w) \le \sum w_i VaR_{1-\alpha}(e_i).$$

□

4 Lower estimate

LEMMA **4.1** *If $r = \sum_{i=1}^{d} w_i F_i^{-1}(\alpha)$ then the multicube $[0,\alpha]^d$ is contained in V_r.*

Proof. If $q_i \le \alpha$ then

$$F_i^{-1}(q_i) \le F_i^{-1}(\alpha).$$

Therefore

$$w_1 F_1^{-1}(q_1) + \ldots + w_d F_d^{-1}(q_d) \le$$
$$\le w_1 F_1^{-1}(\alpha) + \ldots w_d F_d^{-1}(\alpha) = r.$$

□

THEOREM **4.1** *For $\alpha < \delta$*

$$VaR_{1-L(1,\ldots,1)\alpha}(w) \ge \sum w_i VaR_{1-\alpha}(e_i).$$

Proof. (compare [7])
Due to the homogeneity we get

$$\mu_C([0,\alpha]^d) = C(\alpha,\ldots,\alpha) = L(1,\ldots,1)\alpha.$$

Let $r = \sum_{i=1}^{d} w_i F_i^{-1}(\alpha) = 1 - (w_1 VaR_{1-\alpha}(e_1) + \ldots + w_d VaR_{1-\alpha}(e_d))/W_0$. Since the cube $[0,\alpha]^d$ is contained in V_r, we have

$$\mu_C(V_r) \ge C(\alpha,\ldots,\alpha) = L(1,\ldots,1)\alpha.$$

Therefore the $L(1,\ldots,1)\alpha$ quantile of $\frac{W_1(w)}{W_0}$ is smaller than r. Thus

$$VaR_{1-L(1,\ldots,1)\alpha}(w) \ge (1-r)W_0 =$$
$$= w_1 VaR_{1-\alpha}(e_1) + \ldots + \ldots + w_d VaR_{1-\alpha}(e_d).$$

□

By substitution $\alpha := \alpha / L(\ldots)$ we get

COROLLARY **4.1** *For $\alpha < L(1,\ldots,1)\delta$*

$$VaR_{1-\alpha}(w) \ge \sum w_i VaR_{1-\alpha'}(e_i),$$

where $\alpha' = \frac{\alpha}{L(1,\ldots,1)} = \frac{\alpha^2}{C(\alpha,\ldots,\alpha)}$.

5 Final Remarks

The estimates obtained in theorems 3.1 and 4.1 are exact i.e. there is a copula fulfilling assumption A1 such that in both estimates we get equalities.

Lemma **5.1** *If* $C(q_1,\ldots,q_d) = min(q_1,\ldots,q_d)$ *then* $L(1,\ldots,1) = 1$ *and*

$$VaR_{1-\alpha}(w) = \sum w_i VaR_{1-\alpha}(e_i).$$

Proof.
Indeed, if $C(q) = min(q)$ then the measure μ_C is singular with mass uniformly distributed on the diagonal $\{q = (t,\ldots,t) : t \in [0,1]\}$. Therefore if $r = \sum_{i=1}^{d} w_i F_i^{-1}(\alpha)$ then

$$\mu_C(V_r) = \mu_C([0,\alpha]^d) = \alpha.$$

Hence

$$VaR_{1-\alpha}(w) = (1-r)W_0 =$$
$$= w_1 VaR_{1-\alpha}(e_1) + \ldots + \ldots + w_d VaR_{1-\alpha}(e_d).$$

□

References

[1] C.Alsina, M.Frank, B.Schweizer, *Associative Functions: Triangular Norms and Copulas,* World Scientific Publishing Co. 2006.

[2] Basle Committee on Banking Supervision, Amendment to the Capital Accord to Incorporate Market Risks, Basle 1996.

[3] P.Billingsley, *Probability and Measure,* John Wiley & Sons, Inc. 1979.

[4] U.Cherubini, E.Luciano, W.Vecchiato, *Copula Methods in Finance,* John Wiley & Sons Ltd 2004.

[5] CreditMetrics - Technical Document, 1997, J.P.Morgan & Co. Incorporated.

[6] P.J.Cumperayot, J.Danielsson, B.J.Jorgensen, C.G.de Vries, On the (Ir)Revelancy of Value-at-Risk Regulation, In: J.Franke, W.Härdle, G.Stahl, (editors), Measuring Risk in Complex Stochastic Systems, Lecture Notes in Statistics 147 Springer V. 2000.

[7] P.Embrechts, A.Höing, J.Alessandro, Using copulae to bound the value-at-risk for functions of dependent risks, Finance Stoch. 7 (2003) 145-167.

[8] P.Embrechts, C.Klüppelberg, T.Mikosch, *Modelling Extremal Events for Insurance and Finance,* Springer V. 1997.

[9] P.Embrechts, G.Puccetti, Bounds for functions of dependent risks, Finance Stoch. 10 (2006) 341-352.

[10] P.Jackson, D.J.Maude, W.Perraudin, Bank Capital and Value at Risk, The Journal of Derivatives, Spring 1997, 73-89.

[11] P.Jaworski, On uniform tail expansions of bivariate copulas, Applicationes Mathematicae 31.4 (2004) 397-415.

[12] P.Jaworski, Value at Risk in the presence of the power laws, Acta Physica Polonica B 36 (2005) 2575-2587.

[13] P.Jaworski, On uniform tail expansions of multivariate copulas and wide convergence of measures, Applicationes Mathematicae 33.2 (2006) 159-184.

[14] P.Jaworski, On Value at Risk for Foreign Exchange Rates - the Copula Approach, Acta Physica Polonica B 37 (2006) 3005-3016.

[15] H.Joe, *Multivariate Models and Dependence Concepts,* Chapman & Hall 1997.

[16] F.M.Longin, From Value-at-Risk to stress-testing: The extreme value approach, In: P.Embrechts, (editor), Extremes and Integrated Risk Management, Risk Waters Group Ltd 2000.

[17] A.J.McNeil, Extreme Value Theory for Risk Managers, In: P.Embrechts, (editor), Extremes and Integrated Risk Management, Risk Waters Group Ltd 2000.

[18] R.B.Nelsen, *An Introduction to Copulas,* Springer V. 1999.

[19] G.C.Pflug, Some remarks on the Value-at-Risk and the Conditional Value-at-Risk, In: S.Uryasev, (editor), *Probabilistic Constrained Optimization: Methodology and Applications,* Kluwer Academic Publishers 2000.

[20] Risk Metrics - Technical Document, 1996, Morgan Guaranty Trust Company of New York.

Session 4

Fuzzy Measures and Integrals: Theory – E. Pap and R. Mesiar

Pseudo-integral of set-valued functions

Tatjana Grbić
Faculty of Technical Sciences
University of Novi Sad
Trg D. Obradovica 6
Novi Sad, Serbia
tatjana@uns.ns.ac.yu

Ivana Štajner-Papuga
Department of Mathematics
and Informatics
University of Novi Sad
Trg D. Obradovica 4, Novi Sad, Serbia
stajner@im.ns.ac.yu

Ljubo Nedović
Faculty of Technical Sciences
University of Novi Sad
Trg D. Obradovica 6
Novi Sad, Serbia
nljubo@uns.ns.ac.yu

Abstract

An approach to the integration of set-valued functions based on the pseudo-integral has been proposed. Some basic properties of the pseudo-integral of set-valued functions have been shown.

Keywords: Pseudo-operations, Pseudo-integral, Set-valued functions.

1 Introduction

The main topic of this paper is integration of set-valued functions. Over the years, theory of set-valued functions, beside being an important mathematical theory, has become an important tool in several practical areas, specially in economic analysis (problems of individual demand, mean demand, competitive equilibrium, coalition production economies, etc.) (see [5]). Applications of integration of set-valued functions in economy analysis have roots in Aumann's research based on the classical Lebesgue integral ([1]). Some generalizations of this approach that relied on an extension of Lebesgue integral known as $\perp$-integral (see [18]) and on the Choquet integral had been investigated in [19] and [4], respectively.

Results presented in this paper belong to the theory of pseudo-analysis. As a rather new theory, pseudo-analysis has proved itself to be a vast source of powerful tools that are being successfully applied in many mathematical theories as well as in various practical problems (see [2, 6, 7, 9, 11, 12, 13, 14, 15, 16]). Having this in mind, integral proposed here is based on the pseudo-integral ([13]), i.e. pseudo-analysis' counterpart of the classical integral.

Section 2 contains preliminary notions, such as pseudo-operations, semiring, σ-$\oplus$-decomposable measure and pseudo-integral. The construction of the pseudo-integral of set-valued functions is given in the third section. Basic properties of this new type of integral are presented in the Section 4. Some concluding remarks are stated in the Section 5.

2 Preliminary notions

The basic preliminary notions needed in this paper are notions of pseudo-operations and semiring.

Let $[a, b]$ be closed subinterval of $[-\infty, +\infty]$ (in some cases semiclosed subintervals will be considered) and let $\preceq$ be total order on $[a, b]$. A semiring is structure $([a, b], \oplus, \odot)$ such that the following hold:

- $\oplus$ is *pseudo-addition*, i.e., a function $\oplus : [a, b] \times [a, b] \to [a, b]$ which is commutative, non-decreasing (with respect to $\preceq$), associative and with a zero element, denoted by $\mathbf{0}$;
- $\odot$ is *pseudo-multiplication*, i.e., a function $\odot : [a, b] \times [a, b] \to [a, b]$ which is commutative, positively non-decreasing ($x \preceq y$ implies $x \odot z \preceq y \odot z$, $z \in [a, b]_+ = \{x : x \in [a, b],\ \mathbf{0} \preceq x\}$), associative and for which exists a unit element denoted by $\mathbf{1}$;
- $\mathbf{0} \odot x = \mathbf{0}$;
- $x \odot (y \oplus z) = (x \odot y) \oplus (x \odot z)$.

There are three basic classes of semirings with continuous (up to some points) pseudo-operations. The first class contains semirings with idempotent pseudo-addition and non idempotent pseudo-multiplication. Semirings with strict pseudo-operations defined by monotone and continuous generator function $g : [a, b] \to [0, +\infty]$, i.e. g-semirings ([8, 10, 12, 13]), form the second class, and semirings with both idempotent operations belong to the third class. More on this structure can be found in [7, 9, 12, 13, 14].

Total order $\preceq$ is closely connected to the choice of the pseudo-addition. If $\oplus$ is an idempotent operation

(semirings of the first and the third class), total order is induced in the following way

$$x \preceq y \text{ if and only if } x \oplus y = y,$$

and if $([a,b], \oplus, \odot)$ is a semiring of the second class given by generator g, total order is given by

$$x \preceq y \text{ if and only if } g(x) \leq g(y).$$

In all three cases the strict order $\prec$ has the following form:

$$x \prec y \text{ if and only if } x \preceq y \text{ and } x \neq y.$$

Now, let $([a,b], \oplus, \odot)$ be a semiring and let $([a,b], \oplus)$ and $([a,b], \odot)$ be complete lattice ordered semigrups. Let us suppose that interval $[a,b]$ is endowed with metric d which is compatible with sup and inf and satisfies one of the following conditions:

i) $d(x_1 \oplus y_1, x_2 \oplus y_2) \leq d(x_1, x_2) + d(y_1, y_2)$,

ii) $d(x_1 \oplus y_1, x_2 \oplus y_2) \leq \max\{d(x_1, x_2), d(y_1, y_2)\}$.

Since construction of the pseudo-integral is similar to the construction of Lebesgue integral, necessary notion is also notion of the σ-$\oplus$-measure ([13]).

Let Σ be a σ-algebra of subset of a X. A set function $\mu : \Sigma \to [a,b]_+$ is the *σ-$\oplus$-measure* if

i) $\mu(\emptyset) = \mathbf{0}$,

ii) $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \bigoplus_{i=1}^{\infty} \mu(A_i) = \lim_{n\to\infty} \bigoplus_{i=1}^{n} \mu(A_i)$,

where $(A_i)_{i\in\mathbb{N}}$ is a sequence of pairwise disjoint sets from Σ.

If $\oplus$ is an idempotent operation, then disjointness of sets and condition i) can be omitted.

If $e : X \to [a,b]$ is an elementary function of the following representation

$$e = \bigoplus_{i=1}^{\infty} a_i \odot \chi_{A_i},$$

where $a_i \in [a,b]$, $A_i \in \Sigma$ and χ_A is *the pseudo-characteristic function* of a set A given by

$$\chi_A(x) = \begin{cases} \mathbf{1} & \text{for } x \in A, \\ \mathbf{0} & \text{for } x \notin A, \end{cases} \tag{1}$$

the pseudo-integral of e with respect to σ-$\oplus$-measure μ is

$$\int_X^{\oplus} e \odot d\mu = \bigoplus_{i=1}^{\infty} a_i \odot \mu(A_i).$$

The pseudo-integral of a bounded measurable function $f : X \to [a,b]$ is

$$\int_X^{\oplus} f \odot d\mu = \lim_{n\to\infty} \int_X^{\oplus} \varphi_n \odot d\mu, \tag{2}$$

where $(\varphi_n)_{n\in\mathbb{N}}$ is a sequence of elementary functions such that $d(\varphi_n(x), f(x)) \to 0$ uniformly while $n \to \infty$ and d is previously mentioned metric. Proof of existence of sequence $(\varphi_n)_{n\in\mathbb{N}}$ as well as construction of functions φ_n can be found in [13].

Remark 1 Construction of the pseudo-integral for non idempotent pseudo-addition additionally requires that for each $\varepsilon > 0$ exists a monotone ε-net in $f(X)$ (see [13]).

The pseudo-integral of function f on some arbitrary subset A of X is given by

$$\int_A^{\oplus} f \odot d\mu = \int_X^{\oplus} (f \odot \chi_A) \odot d\mu, \tag{3}$$

where χ_A is the pseudo-characteristic function (1).

3 Pseudo-integral of set-valued functions

The definition of set-valued pseudo-integral will be given in this section.

Let $([a,b], \oplus, \odot)$ be a semiring as described in the previous section. Let $f : X \to [a,b]_+$ be a measurable real-valued function, i.e., sets of the form $\{x | \alpha \prec f(x)\}$ are measurable for all $\alpha \in [a,b]_+$. Further more, let us suppose that function f is integrable with respect to pseudo-integral given by (2), that is, let $\int_X^{\oplus} f \odot d\mu$ exist as a finite value in the sense of semiring $([a,b], \oplus, \odot)$. The family of all measurable and integrable functions with respect to measure μ and pseudo-integral (2) will be denoted with $L^1_{\oplus}(\mu)$.

Let $\mathcal{F}$ be the class of all closed subsets of $[a,b]_+$. *A set-valued function* is a function from X to $\mathcal{F}\backslash\{\emptyset\}$. Further on, by a measurable set-valued function we shall consider set-valued functions with measurable graph.

Definition 2 *Let F be a set-valued function.* The pseudo-integral *of F on $A \in \Sigma$ is*

$$\int_A^{\oplus} F \odot d\mu = \left\{ \int_A^{\oplus} f \odot d\mu \,\middle|\, f \in S(F) \right\}, \tag{4}$$

where $S(F)$ is the family of μ-a.e. measurable selections of F, i.e.

$$S(F) = \left\{ f \in L^1_{\oplus}(\mu) \,|\, f(x) \in F(x) \;\; \mu - a.e. \text{ on } X \right\}.$$

Specially, when $\int^{\oplus}$ coincides with Lebesgue integral (see ([12, 13]), integral (4) is the classical Aumann's integral ([1]).

The first question that arises is the question of pseudo-integrability of set-valued functions.

Definition 3 *A set-valued function $F : X \to \mathcal{F}\backslash\{\emptyset\}$ is pseudo-integrable on $A \in \Sigma$ if*

$$\int_A^{\oplus} F \odot d\mu \neq \emptyset.$$

In order to address this issue, the following property of set-valued functions has to be defined:

Definition 4 *A set-valued function F is* pseudo-integrably bounded *if there is a function $h \in L^1_{\oplus}(\mu)$ such that:*

i) $\bigoplus_{\alpha \in F(x)} \alpha \preceq h(x)$, *for idempotent pseudo-addition,*

ii) $\sup_{\alpha \in F(x)} \alpha \preceq h(x)$, *for pseudo-addition given by increasing generator g,*

iii) $\inf_{\alpha \in F(x)} \alpha \preceq h(x)$, *for pseudo-addition given by decreasing generator g.*

Sufficient condition for pseudo-integrability of set-valued function F is given by the following proposition.

Proposition 5 *If F is a pseudo-integrably bounded set-valued function, then F is pseudo-integrable.*

Proof. Let F be a pseudo-integrably bounded set-valued function and let h be a function from Definition 4. Let f be a selection of F, i.e., $f(x) \in F(x)$ μ-a.e. on X. It can be easily shown that, under give assumptions, the following holds almost everywhere

$$f \preceq h.$$

Now, properties of the pseudo-integral will insure us

$$\int_X^{\oplus} f \odot d\mu \preceq \int_X^{\oplus} h \odot d\mu$$

(see [13]). Since we assumed that h is integrable function, due to the properties of semiring $([a,b], \oplus, \odot)$ function f is also integrable, i.e., set (4) is not empty. □

Example 6
(a) Let us consider semiring of the first class, e.g. $([a,b], \max, +)$. Then, σ-$\oplus$-measure μ is given by some function $l : \mathbb{R} \to [a,b]$ as $\mu(A) = \sup_{x \in A} l(x)$ (see [11, 12, 13]). In this case, pseudo-integral of some set-valued function F has the following form

$$\int_A^{\oplus} F \odot d\mu = \left\{ \sup_{x \in A} (f(x) + l(x)) \mid f \in S(F) \right\}.$$

(b) If $([a,b], \oplus, \odot)$ is a semiring of the second class pseudo-operations are given by generating function $g : [a,b] \to [0,\infty]$ as

$$x \oplus y = g^{-1}(g(x) + g(y)) \text{ and } x \odot y = g^{-1}(g(x)g(y))$$

(see [6, 8, 10, 12, 13, 15]), therefore pseudo-integral of some set-valued function F is

$$\int_{[c,d]}^{\oplus} F \odot d\mu = \left\{ g^{-1}\left(\int_{[c,d]} g \circ f \, d(g \circ \mu) \right) \Big| f \in S(F) \right\}.$$

(c) Semiring $([a,b], \min, \max)$ is a semiring of the third class. Now, σ-$\oplus$-measure μ is given by some function $l : \mathbb{R} \to [a,b]$ as $\mu(A) = \inf_{x \in A} l(x)$. In this case, pseudo-integral of some set-valued function F has the following form

$$\int_A^{\oplus} F \odot d\mu = \left\{ \inf_{x \in A} (\max(f(x), l(x))) \mid f \in S(F) \right\}.$$

Remark 7 Problem of pseudo-integral of set-valued function for semiring of the second class given by increasing continuous generator had been investigated in [3].

4 Basic properties of set-valued pseudo-integral

Some basic properties of integral given by Definition 4 will be presented in this section.

Proposition 8 *Let F be pseudo-integrable set-valued function, F_1 and F_2 pseudo-integrably bounded set-valued functions and let $A, B \in \Sigma$.*

i) If $A \subset B$ then $\int_A^{\oplus} F \odot d\mu \preceq \int_B^{\oplus} F \odot d\mu$.

ii) If $F_1 \preceq F_2$ then $\int_X^{\oplus} F_1 \odot d\mu \preceq \int_X^{\oplus} F_2 \odot d\mu$.

iii) If $\mu(A) = \mathbf{0}$ then $\int_A^{\oplus} F \odot d\mu = \{\mathbf{0}\}$.

iv) If $\mathbf{0} \prec \alpha$ then

$$\int_X^{\oplus} (\alpha \odot F) \odot d\mu = \alpha \odot \int_X^{\oplus} F \odot d\mu.$$

Remark 9 Relation "less or equal" applied on sets from $\mathcal{F}$ in Proposition 8 is given in the following sense: if $C, D \in \mathcal{F}$ and $C \preceq D$ then

$$\text{for all } x \in C \text{ there exists } y \in D \text{ such that } x \preceq y$$

and

$$\text{for all } y \in D \text{ there exists } x \in C \text{ such that } x \preceq y$$

Proof of Proposition 8.

i) Let us suppose that $x \in \int_A^{\oplus} F \odot d\mu$. Then, by Definition 4, there exists $f \in S(F)$ such that $x = \int_A^{\oplus} f \odot d\mu$. Now, (3) gives us following

$$x = \int_X^{\oplus} (f \odot \chi_A) \odot d\mu.$$

Since it can be easily shown that in our case $f \odot \chi_A \preceq f \odot \chi_B$, properties of pseudo-integral will ensure

$$\int_X^{\oplus} (f \odot \chi_A) \odot d\mu \preceq \int_X^{\oplus} (f \odot \chi_B) \odot d\mu.$$

Obviously,

$$\int_X^{\oplus} (f \odot \chi_B) \odot d\mu = \int_B^{\oplus} f \odot d\mu \in \int_B^{\oplus} F \odot d\mu,$$

therefor there exists

$$y = \int_B^{\oplus} f \odot d\mu \in \int_B^{\oplus} F \odot d\mu,$$

such that $x \preceq y$.

Proof that for $y \in \int_B^{\oplus} F \odot d\mu$ exists $x \in \int_A^{\oplus} F \odot d\mu$ such that $x \preceq y$ is analogous to the previously shown part "for all $x \in \int_A^{\oplus} F \odot d\mu$ exists $y \in \int_B^{\oplus} F \odot d\mu$ such that $x \preceq y$".

ii) As in i), it has to be proved that for all $x \in \int_X^{\oplus} F_1 \odot d\mu$ exists $y \in \int_X^{\oplus} F_2 \odot d\mu$ such that $x \preceq y$ and viceversa, i.e., that for all $y \in \int_X^{\oplus} F_2 \odot d\mu$ exists $x \in \int_X^{\oplus} F_1 \odot d\mu$ such that $x \preceq y$.

Since F_1 is pseudo-integrably bounded set-valued function, it is pseudo-integrable set-valued functions, i.e., $\int_X^{\oplus} F_1 \odot d\mu \neq \emptyset$, therefor let $x = \int_X^{\oplus} f \odot d\mu$ for some $f \in S(F_1)$. Let us suppose that $\oplus$ is max or strict pseudo-operation given by increasing generator. Now, having in mind that $F_1(u) \preceq F_2(u)$ for all $u \in X$, following function can be defined:

$$k(u) = \sup \{v \,|\, v \in F_2(u),\ f(u) \preceq v\}.$$

Obviously $f \preceq k$ and, sice $\mathcal{F}$ is a family of closed sets, $k(u) \in F_2(u)$. Based on the assumption that F_2 is also pseudo-integrably bounded set-valued function and on properties of pseudo-integral, it can be easily shown that $k \in L_{\oplus}^1$ and that required y is of the form

$$y = \int_X^{\oplus} k \odot d\mu \in \int_X^{\oplus} F_2 \odot d\mu.$$

If pseudo-addition in question is min or strict operation given by decreasing generator, for given $x = \int_X^{\oplus} f \odot d\mu$ required $y \in \int_X^{\oplus} F_2 \odot d\mu$ can be obtained through function

$$k(u) = \inf \{v \,|\, v \in F_2(u),\ f(u) \preceq v\}$$

as $y = \int_X^{\oplus} k \odot d\mu$.

Proof of part "for all $y \in \int_X^{\oplus} F_2 \odot d\mu$ there exists $x \in \int_X^{\oplus} F_1 \odot d\mu$ such that $x \preceq y$" if $\oplus$ is max or strict pseudo-operation given by increasing generator is based on function

$$k_2(u) = \inf \{v \,|\, v \in F_1(u),\ v \preceq f_2(u)\},$$

i.e., for $y = \int_X^{\oplus} f_2 \odot d\mu,\ f_2 \in S(F_2)$, required x is $\int_X^{\oplus} k_2 \odot d\mu$. Analogously, if $\oplus$ is min or strict pseudo-operation given by decreasing generator, then proof is based on function

$$k_2(u) = \sup \{v \,|\, v \in F_1(u),\ v \preceq f_2(u)\},$$

i.e., for $y = \int_X^{\oplus} f_2 \odot d\mu,\ f_2 \in S(F_2)$, required x is $\int_X^{\oplus} k_2 \odot d\mu$.

Statements iii) and iv) follow directly from Definition 4 and properties of the pseudo-integral (see [11, 13]). □

Also, for set-valued pseudo-integrals pseudo-convex property can be easily shown, where by pseudo-convex set following is considered:

$$\begin{gathered}\text{set } A \subseteq [a,b] \text{ is } \textit{pseudo-convex} \text{ if for} \\ x, y \in A \text{ and } \alpha, \beta \in [a,b]_+ \text{ where } \alpha \oplus \beta = \mathbf{1} \text{ holds} \\ \alpha \odot x \oplus \beta \odot y \in A.\end{gathered} \tag{5}$$

Proposition 10 *Let F be a measurable set-valued function such that $F(x)$ is pseudo-convex for μ-a.e. $x \in X$. Then, $\int_X^{\oplus} F \odot d\mu$ is a pseudo-convex subset of $[a,b]_+$.*

Proof. Proof follows directly from property of pseudo-convex sets (5) and Definition 4. □

5 Conclusion

Topic presented here belongs to a contemporary field of mathematics that has been successfully applied in economic analysis. Further research of this combination of set-valued functions theory and pseudo-analysis

will be directed to the problems of convergence for sequences of set-valued pseudo-integrals and possible applications.

Acknowledgement. This paper has been supported by the project MNZŽSS 144012. The second and the third authors are supported by the project "Mathematical Models for Decision Making under Uncertain Conditions and Their Applications" of the Academy of Sciences and Arts of Vojvodina supported by Provincial Secretariat for Science and Technological Development of Vojvodina.

References

[1] R. J. Aumann, Integrals of set-valued functions, Journal of Mathematical Analysis and Applictions 12 (1965) 1–12.

[2] P. Benvenuti, R. Mesiar and D. Vivona, Monotone Set Functions-Based Integrals, in: E. Pap (Ed.), *Handbook of Measure Theory, Volume II*. Elsevier, Amsterdam, 2002, pp. 1329–1379.

[3] C. Guo, D. Zhang, On set-valued fuzzy measures, Information Sciences 160 (2004) 13–25.

[4] L. C. Jang, B. M. Kil, Y. K. Kim, J. S. Kwon, Some properties of Choquet integrals of set-valued functions, Fuzzy Sets and Systems 91 (1997) 95–98.

[5] E. Klein, A C. Thompson, Theory of Correspondences, A Wi ley-Interscience Publication, New York, 1984.

[6] E.P. Klement, R. Mesiar and E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.

[7] V. N. Kolokoltsov and V.P. Maslov, Idempotent Analysis and Its Applications, Kluwer Academic Publishers, Dordrecht, 1997.

[8] A. Marková-Stupňanová, A note on g-derivative and g-integral, Tatra Mt. Math. Publ. 8 (1996) 71–76.

[9] V.P. Maslov, S.N. Samborskij, Idempotent Analysis, American Mathematical Society, Providence, 1992.

[10] R. Mesiar, Pseudo-linear integrals and derivatives based on a generator g, Tatra Mt. Math. Publ. 8 (1996) 67–70.

[11] E. Pap, An integral generated by decomposable measure, Univ. u Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 20 (1990) 135–144.

[12] E. Pap, g-calculus, Univ. u Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 23 (1993) 145–156.

[13] E.Pap, Null-Additive Set Functions, Kluwer Academic Publishers, Dordrecht-Boston-London, 1995.

[14] E. Pap,Pseudo-additive measures and their applications, in: E. Pap (Ed.), *Handbook of Measure Theory, Volume II*. Elsevier, Amsterdam, 2002, pp. 1403-1465.

[15] E. Pap, Applications of the generated pseudo-analysis to nonlinear partial differential equations, Contemporary Mathematics 377 (2005) 239–259.

[16] E. Pap, D. Vivona, Non-commutative and non-associative pseudo-analysis and its applications on nonlinear partial differential equations, J. Math. Anal. Appl. 246/2 (2000) 390–408.

[17] D. Zhang, C. Guo, Integrals of set-valued functions for $\perp$-decomposable measures, Fuzzy Sets and Systems 78 (1996) 341–346.

[18] S. Weber, Conditinal measures and their applications to fuzzy sets, Fuzzy Sets and Systems 42 (1991) 73–85.

[19] D. Zhang, C. Guo, Integrals of set-valued functions for $\perp$-decomposable measures, Fuzzy Sets and Systems 78 (1996) 341–346.

M-probability theory on IF-events

Beloslav Riečan
Faculty of Natural Sciences, Matej Bel University
Department of Mathematics
Tajovského 40
974 01 Banská Bystrica, Slovakia
riecan@fpv.umb.sk
and
Mathematical Institute of Slovak Acad. of Sciences
Štefánikova 49
SK-81473 Bratislava

Abstract

Following M. Krachounov ([5]), max and min operations with fuzzy sets are considered instead of Lukasiewicz ones ([6], [7], [8], [9]). Her the domain $\mathcal{F}$ of a probability $m : \mathcal{F} \to [0,1]$ consists on IF-events $A = (\mu_A, \nu_A)$ i.e. mappings from a measurable space $(\Omega, \mathcal{S})$ to the unit square ([1]). Local representation of sequences of M-observables by random variables is constructed and three kinds of convergences are characterized: convergence in distribution, convergence in measure, and almost everywhere convergence.

Keywords: IF-events, M-states, M-observables.

1 Introduction

Although the notion of an IF set is given uniquelly, operations on them present a large variety. In this paper we shall consider max and min operations. First let us repeat the basic definitions.

Let $(\Omega, \mathcal{S})$ be a measurable space. By an IF-event we mean any pair

$$A = (\mu_A, \nu_A)$$

of $\mathcal{S}$-measurable functions, such that $\mu_A \geq 0, \nu_A \geq 0$, and

$$\mu_A + \nu_A \leq 1.$$

An important notion is the ordering

$$A \leq B \Longleftrightarrow \mu_A \leq \mu_B, \nu_A \geq \nu_B.$$

We shall use the following connectives for $a, b \in R$:

$$a \vee b = \max(a, b),$$

$$a \wedge b = \min(a, b).$$

Hence consider a measurable space $(\Omega, \mathcal{S})$, where $\mathcal{S}$ is a σ-algebra of subsets of Ω, $\mathcal{F} = \{A = (\mu_A, \nu_A); \mu_A, \nu_A$ are non-negative, $\mathcal{S}$-measurable functions, $\mu_A + \nu_A \leq 1$ $\}$. According to [4] we shall define the probability on $\mathcal{F}$ using max-min connectives instead of the Lukasiewicz connectives,

$$A \vee B = (\mu_A \vee \mu_B, \nu_A \wedge \nu_B),$$

$$A \wedge B = (\mu_A \wedge \mu_B, \nu_A \vee \nu_B).$$

For distinguishing the two theories (max and min operations instead of Lukasiewicz operations) we shall speak about M-probability.

For the first time the probability of an IF-event $A = (\mu_A, \nu_A)$ was defined in [4] as the interval

$$\mathcal{P}(A) = [\int_\Omega \mu_A dP, 1 - \int_\Omega \nu_A dP].$$

As we have shown in [8] it is a special case of a general definition given axiomatically. Of course, also in the axiomatic approach, the probability $A \longmapsto \mathcal{P}(A) = [\mathcal{P}^\flat(A), \mathcal{P}^\sharp(A)]$ can be reduced to two states $\mathcal{P}^\flat, \mathcal{P}^\sharp : \mathcal{F} \to [0,1]$. On the other hand the study of states on IF-events cannot be reduced to two coordinate mappings on fuzzy sets (see Proposition 2.3 and Example 2.5).

Of course, from the classical point of view the probability theory has two basic notions: probability, and random variable, in the non-commutative case state and observable. This approach in the IF -probability case is very fruitful. The space of IF-events can be embedded in a suitable MV-algebra, hence all results of the MV-algebra probability theory ([11], [12]) can be applied to the IF-case ([6], [7], [8], [9], [10]).

On the other hand the Krachounov proposal [5] need probably another approach. Generally probability can be defined for any t-norm and t-conorm ([2]). Of course, in this moment we are able to prove the existence of the joint observable only in the case of Lukasiewicz connectives and the Zadeh max - min

connectives ([10]). The M-probability theory on IF-events cannot be reduced to the corresponding theory on fuzzy sets. On the other hand the obtained results working for IF-events are new also for M-probabilities on fuzzy events.

The existence of the joint observable is the key to the basic assertions of the probability theory. Our paper anobles to translate some known convergence theorems of the Kolmogorov theory to the M-probability case. We present as the main result the translation formulas (Theorem 3.3). ¿From the convergence of suitable sequence of random variables the corresponding convergence of a sequence of M-observables follows. In Section 2 some notions and preliminary useful results are presented.

2 States and observables

Definition 2.1 *A mapping $p : \mathcal{F} \to [0.1]$ is called M-state if the following properties are satisfied:*

(i) $m((1_\Omega, 0_\Omega)) = 1, m((0_\Omega, 1_\Omega)) = 0$;

(ii) $m(A) + m(B) = m(A \vee B) + m(A \wedge B)$ *for any* $A, B \in \mathcal{F}$;

(iii) $A_n \nearrow A, B_n \searrow B \Longrightarrow m(A_n) \nearrow m(A), m(B_n) \searrow m(B)$.

Definition 2.2 *A mapping $m : \mathcal{F} \to [0,1]$ is an IF-state, if the following properties are satisfied:*

(i) $m((1_\Omega, 0_\Omega)) = 1, m((0_\Omega, 1_\Omega)) = 0$;

(ii) $m(A) + m(B) = m(A \oplus B) + m(A \odot B))$ *for any* $A, B \in \mathcal{F}$;

(iii) $A_n \nearrow A, B_n \searrow B \Longrightarrow m(A_n) \nearrow m(A), m(B_n) \searrow m(B)$.

Here

$$A \oplus B = (\mu_A \oplus \mu_B, \nu_A \odot \nu_B)$$
$$A \odot B = (\mu_A \odot \mu_B, \nu_A \oplus \nu_B)$$
$$f \oplus g = \min(f + g, 1),$$
$$f \odot g = \max(f + g - 1, 0)$$

Proposition 2.3 *A mapping $m : \mathcal{F} \to [0,1]$ is an IF-state if and only if there exist a probability $P : \mathcal{S} \to [0,1]$ and $\alpha \in [0,1]$ such that*

$$m(A) = (1 - \alpha) \int_\Omega \mu_A dP + \alpha(1 - \int_\Omega \nu_A dP) \qquad (1)$$

for any $A = (\mu_A, \nu_A) \in \mathcal{F}$.

Proof. [8]

Proposition 2.4 *Any IF - state is an M-state.*

Proof. It follows by Prop.2.3.

Example 2.5 *Let $\mathcal{T}$ be the tribe of all $\mathcal{S}$-measurable functions $f : \Omega \to [0,1]$. Let m be an IF-state on $\mathcal{F}, m : \mathcal{F} \to [0,1]$. Evidently $(f, 1 - f) \in \mathcal{F}$, hence we can define the mapping $\overline{m} : \mathcal{T} \to [0,1]$ by the formula*

$$\overline{m}(f) = m((f, 1 - f)).$$

Evidently $\overline{m}$ is a state on $\mathcal{T}$, hence by the Butnariu - Klement theorem ([2]) there exists a probability $P : \mathcal{S} \to [0,1]$ such that

$$\overline{m}(f) = \int_\Omega f dP, \qquad (2)$$

for any $f \in \mathcal{T}$. Of course, the formula (2) does not imply (1), we see that the IF-approach cannot be coordinatwisely reduced to the fuzzy approach.

Example 2.6 *Fix $x_0 \in \Omega$ and put*

$$m(A) = \frac{1}{2}(\mu_A^2(x_0) + 1 - \nu_A^2(x_0)).$$

Since $(\mu_A \vee \mu_B)^2 + (\mu_A \wedge \mu_B)^2 = \mu_A^2 + \mu_B^2$, it is not difficult to see that m is an M-state. Put

$$\mu_A(x) = \mu_B(x) = \frac{1}{4}, \nu_A(x) = \nu_B(x) = \frac{1}{2}$$

for any $x \in \Omega$. Then

$$m(A) = m(B) = \frac{13}{32}.$$

On the other hand

$$A \oplus B = ((\frac{1}{2})_\Omega, 0_\Omega), A \odot B = (0_\Omega, 1_\Omega),$$

hence

$$m(A \oplus B) + m(A \odot B) = \frac{5}{8} + 0 \neq \frac{13}{32} + \frac{13}{32} = m(A) + m(B).$$

Although the probability theory on IF-events studied in [6, 7, 8, 9, 10] seems to be satisfactory, the previous facts lead us to an experience to create basic instruments for an alternative M-probability theory. Of course, the crucial notion is the notion of an M-observable.

Definition 2.7 *An M-observable is a mapping $x : \mathcal{B}(R) \to \mathcal{F}$ satisfying the following conditions:*

(i) $x(R) = (1_\Omega, 0_\Omega), x(\emptyset) = (0_\Omega, 1_\Omega)$;

(ii) $x(A \cup B) = x(A) \vee x(B), x(A \cap B) = x(A) \wedge x(B)$ *for any* $A, B \in \mathcal{B}(R)$;

(iii) $A_n \nearrow A, B_n \searrow B \Longrightarrow x(A_n) \nearrow x(A), x(B_n) \searrow x(B)$.

Proposition 2.8 *If $x : \mathcal{B}(R) \to \mathcal{F}$ is an M-observable, and $m : \mathcal{F} \to [0,1]$ is an M-state, then $m \circ x : \mathcal{B}(R) \to [0,1]$ is a probability measure.*

Proof. Evidently $m(x(R)) = m(1_\Omega) = 1$. Also continuity of $m \circ x$ is clear. Let $A \cap B = \emptyset$. Then $x(A) \wedge x(B) = x(\emptyset) = (0_\Omega, 1_\Omega)$. Therefore

$$\begin{aligned} m(x(A \cup B)) &= m(x(A) \vee x(B)) + m(x(A) \wedge x(B)) = \\ &= m(x(A)) + m(x(B)). \end{aligned}$$

Definition 2.9 *Let $x, y : \mathcal{B}(R) \to \mathcal{F}$ be M-observables. The joint M-observable of x and y is a mapping $h : \mathcal{B}(R^2) \to \mathcal{F}$ satisfying the following conditions:*

(i) $h(R^2) = (1_\Omega, 0_\Omega), h(\emptyset) = (0_\Omega, 1_\Omega)$;

(ii) $h(A \cup B) = h(A) \vee h(B), h(A \cap B) = h(A) \wedge h(B)$ *for any* $A, B \in \mathcal{B}(R^2)$;

(iii) $A_n \nearrow A, B_n \searrow B \Longrightarrow h(A_n \nearrow h(A), h(B_n) \searrow h(B)$;

(iv) $h(C \times D) = \min(x(C), y(D))$ *for any* $C, D \in \mathcal{B}(R)$.

Theorem 2.10 *For any M-observables there exists their joint M observable.*

Proof. [10] Theorem 2.2.2.

3 Convergence of M-observables

The aim of the paper is a characterization of sequences on M-observables by the convergence of random variables.

Definition 3.1 *Let $y_1, y_2, ...$ be a sequence of M-observables, $y_n : \mathcal{B}(R) \to \mathcal{F}, p : \mathcal{F} \to [0,1]$ be an M-state.*

(i) The sequence is said to converge in distribution to a function $F : R \to [0,1]$ if for each $t \in R$

$$\lim_{n\to\infty} p(y_n((-\infty, t))) = F(t).$$

(ii) The sequence is said to converge in measure to 0 if for each $\varepsilon > 0$

$$\lim_{n\to\infty} p(y_n((-\varepsilon, \varepsilon))) = 1.$$

(iii) The sequence converges to 0 almost everywhere, if

$$\lim_{p\to\infty} \lim_{k\to\infty} \lim_{i\to\infty} p(\bigvee_{n=k}^{k+i} y_n((-\frac{1}{p}, \frac{1}{p}))) = 1.$$

Definition 3.2 *Let $x_1, ..., x_n : \mathcal{B}(R) \to \mathcal{F}$ be M-observables, $h_n : \mathcal{B}(R^n) \to \mathcal{F}$ be their joint M-observable, $g_n : R^n \to R$ be a Borel function. Then we define $g_n(x_1, ..., x_n)$ by the formula*

$$g_n(x_1, ..., x_n) = h_n \circ g_n^{-1}.$$

It is easy to see that the mapping $y_n : \mathcal{B}(R) \to \mathcal{F}$ is an M-observable.

Theorem 3.3 *Let (x_n) be a sequence of M-observables, $x_n : \mathcal{B}(R) \to \mathcal{F}, h_n : \mathcal{B}(R^n) \to \mathcal{F}$ their joint M-observables $(n = 1, 2, ...), g_n : R^n \to R$ Borel functions, $y_n = g_n(x_1, ..., x_n), n = 1, 2, ...$ Then there exists a probability space $(X, \mathcal{S}, P)$ and a sequence (ξ_n) of random variables, $\xi_n : X \to R$ such that if $\eta_n = g_n(\xi_1, ..., \xi_n), n = 1, 2, ...)$, then*

(i) the sequence $y_1, y_2, ...$ converges in distribution to a function F if and only if so does the sequence $\eta_1, \eta_2, ...$;

(ii) $y_1, y_2, ...$ converges to 0 in measure p if and only if $\eta_1, \eta_2, ...$ converges to 0 in measure P;

(iii) if $\eta_1, \eta_2, ...$ converges P-almost everywhere to 0, then $y_1, y_2, ...$ converges p-almost evrywhere to 0.

Proof. Put $X = R^N . \mathcal{S} = \sigma(\mathcal{C})$, where $\mathcal{C}$ is the family of all cylinders in R^N. Put $p_n = p \circ h_n$. Then $\{p_n; n \in N\}$ form a consistent family of probability measures $p_n : \mathcal{B}(R^n) \to [0,1]$, i.e.

$$p_{n+1}(A \times R) = p_n(A), A \in \mathcal{B}(R^n), n = 1, 2, ...$$

By the Kolmogorov theorem there exists exactly one probability measure $P : \sigma(\mathcal{C}) \to [0,1]$ such that

$$P \circ \pi_n^{-1} = p_n, n = 1, 2, ...$$

where $\pi_n : R^N \to R^n$ is the projection. Put

$$\xi_n : R^N \to R, \xi_n((u_i)_{i=1}^\infty)) = u_n, n = 1, 2, ...$$

Then

$$P(\eta_n^{-1}(A)) = P((g_n(\xi_1, ..., \xi_n))^{-1}(A)) =$$

$$P(\pi_n^{-1}(g_n^{-1}(A))) = p(h_n(g_n^{-1}(A))) = p(y_n(A))$$

Therefore

$$p(y_n(-\infty, t)) = P(\eta_n^{-1}(-\infty, t)),$$

$$p(y_n((-\varepsilon,\varepsilon))) = P(\eta_n^{-1}((-\varepsilon,\varepsilon))),$$

what implies (i) and (ii). Let now η_n converges to 0 P-almost everywhere. We have

$$P(\bigcap_{n=k}^{k+i} \eta_n^{-1}((-\frac{1}{p},\frac{1}{p}))) =$$

$$p(h_{k+i}(\bigcap_{n=k}^{k+i}\{(t_1,.,,,t_{k+i}) : g_n(t_1,...,t_n) \in (-\frac{1}{p},\frac{1}{p})\})) \leq$$

$$\leq p(\bigwedge_{n=k}^{k+i} h_{k+i}(\{(t_1,...,t_{k+i}) : (t_1,...,t_n) \in g_n^{-1}((-\frac{1}{p},\frac{1}{p}))\})) =$$

$$= p(\bigwedge_{n=k}^{k+i} h_n \circ g_n^{-1}((-\frac{1}{p},\frac{1}{p}))) =$$

$$= p(\bigwedge_{n=1}^{k+i} y_n((-\frac{1}{p},\frac{1}{p}))).$$

Therefore

$$1 \leq \lim_{p\to\infty}\lim_{k\to\infty}\lim_{i\to\infty} p(\bigwedge_{n=k}^{k+i} y_n((-\frac{1}{p},\frac{1}{p}))) \leq 1,$$

hence $(y_n)_{n=1}^{\infty}$ converges to 0 p-almost everywhere.

Acknowledgement

The paper was supported by Grant VEGA 1/2002/05.

References

[1] K. Atanassov: Intuitionistic Fuzzy Sets: Theory and Applications, Physica Verlag, New York 1999.

[2] D. Butnariu, E. P. Klement: Triangular norm-based measures and their markov kernel representation, J. Math. Anal. Appl. 162, 1991, 11 - 143.

[3] R. L. O. Cignoli, I. M. L. D'Ottaviano, D. Mundici: Algebraic Foundations of Many-valued Reasoning, Kluwer, Dordrecht 2000.

[4] P. Grzegorzewski, E. Mrowka: Probability of intuitionistic fuzzy events, in: *Soft Methods in Probability, Statistics and Data Analysis (P. Grzegorzewski et al. eds.)*. Physica Verlag, New York 2002, 105 - 115.

[5] M. Krachounov: Intuitionistic probability and intuitionistic fuzzy sets, in: *First Intern. Workshop on IFS, Generalized Nets and Knowledge Engineering (E. El-Darzi, K. Atanassov, P. Chountas eds.)*. Univ. of Westminster, London 2006, 18 - 24. 2006, 714-717.

[6] K. Lendelová: Measure theory on multivalued logics and its applications, PhD thesis, M.Bel University Banská Bystrica 2005.

[7] M. Renčová, B. Riečan: Probability on IF-sets: an elementary approach, in: *First Intern. Workshop on IFS, Generalized Nets and Knowledge Engineering (E. El-Darzi, K. Atanassov, P. Chountas eds.)*. Univ. of Westminster, London 2006, 8 - 17.

[8] B. Riečan: On a problem of Radko Mesiar: general from of IF probabilities, Fuzzy Sets and Systems 152, 2006, 1485 - 1490.

[9] B. Riečan: On two concepts of probability on IF-sets, Notes on IFS 12, 2006, 69 - 72.

[10] B. Riečan: Probability theory on intuitionistic fuzzy events, in: *A volume in honour of Daniele Mundici's 60th birthday Lecture Notes in Computer Science*, Springer, Berlin 2007.

[11] B. Riečan, D. Mundici: Probability on MV-algebras, in: *Handbook of Measure Theory (E. Pap ed.)*. Elsevier, Amsterdam 2002, 869 -909.

[12] B. Riečan, T. Neubrunn: Integral, Measure, and Ordering, Kluwer, Dordrecht 1997.

Almost everywhere convergence in family of IF-events with product

Katarína Lendelová
Faculty of Natural Sciences, Matej Bel University
Department of Mathematics
Tajovského 40
SK-974 01 Banská Bystrica
lendelov@fpv.umb.sk

Abstract

The aim of this paper is to define the lower and upper limits on the family of IF-events with product. We compare two concepts of almost everywhere convergence and we show that they are equivalent, too.

Keywords: IF-event, Product, Lower limit, Upper limit, Almost everywhere convergence.

1 Introduction

In recent years the theory of IF-sets introduced by Atanassov ([1]) has been studied by many authors.

An *IF-set* A on a space Ω is a couple (μ_A, ν_A), where $\mu_A : \Omega \to [0,1]$, $\nu_A : \Omega \to [0,1]$ are functions such that $\mu_A(\omega) + \nu_A(\omega) \leq 1$ for each $\omega \in \Omega$ (see [1]). The function μ_A is called the membership function, the function ν_A is called the non membership function.

In [4] Grzegorzewski and Mrówka defined the probability on the family of IF-events

$$\mathcal{N} = \{(\mu_A, \nu_A)\ ; \mu_A + \nu_A \leq 1\},$$

where μ_A, ν_A are $\mathcal{S}$-measurable, as a mapping $\mathcal{P}$ from the family $\mathcal{N}$ to the set of all compact intervals in $\mathbf{R}$ by the formula

$$\mathcal{P}((\mu_A, \nu_A)) = \left[\int_\Omega \mu_A\ dP, 1 - \int_\Omega \nu_A\ dP\right],$$

where $(\Omega, \mathcal{S}, P)$ is probability space. This IF-probability was axiomatically characterized by B. Riečan (see[13]).

More general situation was studied in [16], where the author introduced the notion of IF-probability on the family

$$\mathcal{F} = \{(f,g)\ ;\ f,g \in \mathcal{T},\ f+g \leq 1\},$$

where $\mathcal{T}$ is a Lukasiewicz tribe, as a mapping $\mathcal{P}$ from the family $\mathcal{F}$ to the family $\mathcal{J}$ of all closed intervals $\langle a,b\rangle$ such that $0 \leq a \leq b \leq 1$. Variant of Central limit theorem and Weak law of large numbers were proved as an illustration of method applied on these IF-events. It can see in the papers [10, 11].

More general situation was used in [9]. The authors defined the IF-probability on the family $\mathcal{M} = \{(a,b) \in M, a+b \leq u\}$, where M is σ-complete MV-algebra, which can be identified with the unit interval of a unique ℓ-group G with strong unit u, in symbols,

$$M = \Gamma(G,u) = (\langle 0,u\rangle, 0, u, \neg, \oplus, \odot)$$

where $\langle 0,u\rangle = \{a \in G\ ;\ 0 \leq a \leq u\}$, $\neg a = u - a$, $a \oplus b = (a+b) \wedge u$, $a \odot b = (a+b-u) \vee 0$ (see [17]). We say that G is the ℓ-group (with strong unit u) corresponding to M.

By an ℓ-group we shall mean a lattice-ordered Abelian group. For any ℓ-group G, an element $u \in G$ is said to be a strong unit of G, if for all $a \in G$ there is an integer $n \geq 1$ such that $nu \geq a$.

The independence of IF-observables, the convergence of IF-observables and the Strong law of large numbers were studied on this family of IF-events, see [6, 7].

In paper [8] we defined the product operation on the family $\mathcal{F}$ of IF-events

$$\mathcal{F} = \{(f,g)\ ;\ f,g \in \mathcal{T},\ f+g \leq 1\},$$

where $\mathcal{T}$ is a Lukasiewicz tribe. We formulated the version of conditional IF-probability on this family, too. In this paper we introduce the notion of lower and upper limits and show two concepts of almost everywhere convergence. In *Section 2* we introduce the operations on $\mathcal{F}$ and $\mathcal{J}$, where $\mathcal{J}$ is the family of all closed intervals $\langle a,b\rangle$ such that $0 \leq a \leq b \leq 1$.

2 Basic notions

Now we introduce operations on $\mathcal{F}$. Let $A = (a_1, a_2)$, $B = (b_1, b_2)$. Then we define

$$A \oplus B = (a_1 \oplus b_1, a_2 \odot b_2) = \big((a_1 + b_1) \wedge 1, (a_2 + b_2 - 1) \vee 0\big),$$

$$A \odot B = (a_1 \odot b_1, a_2 \oplus b_2) = \big((a_1 + b_1 - 1) \vee 0, (a_2 + b_2) \wedge 1\big).$$

If $A_n = (a_{n1}, a_{n2})$, then we write

$$A_n \nearrow A \iff a_{n1} \nearrow a_1,\ a_{n2} \searrow a_2.$$

IF-probability $\mathcal{P}$ on $\mathcal{F}$ is a mapping from $\mathcal{F}$ to the family $\mathcal{J}$ of all closed intervals $\langle a, b\rangle$ such that $0 \le a \le b \le 1$. Here we define

$$\langle a, b\rangle + \langle c, d\rangle = \langle a + c, b + d\rangle,$$

$$\langle a_n, b_n\rangle \nearrow \langle a, b\rangle \iff a_n \nearrow a,\ b_n \nearrow b.$$

By an **IF-probability on** $\mathcal{F}$ we understand any function $\mathcal{P} : \mathcal{F} \to \mathcal{J}$ satisfying the following properties:

(i) $\mathcal{P}((1,0)) = \langle 1, 1\rangle = 1$; $\mathcal{P}((0,1)) = \langle 0, 0\rangle = 0$;

(ii) if $A \odot B = (0,1)$ and $A, B \in \mathcal{F}$, then $\mathcal{P}(A \oplus B) = \mathcal{P}(A) + \mathcal{P}(B)$;

(iii) if $A_n \nearrow A$, then $\mathcal{P}(A_n) \nearrow \mathcal{P}(A)$.

IF-probability $\mathcal{P}$ is called **separating**, if

$$\mathcal{P}\big((a_1, a_2)\big) = \langle \mathcal{P}^\flat(a_1), 1 - \mathcal{P}^\sharp(a_2)\rangle,$$

where the functions $\mathcal{P}^\flat, \mathcal{P}^\sharp : \mathcal{T} \to [0,1]$ are probabilities.

Example 2.1 *Let $\mathcal{T}$ be the tribe of all $\mathcal{S}$-measurable functions $f : \Omega \to [0,1]$. Let m be an IF-state on $\mathcal{F}$, $m : \mathcal{F} \to [0,1]$. Evidently $(f, 1-f) \in \mathcal{F}$, hence we can define the mapping $\overline{m} : \mathcal{T} \to [0,1]$ by the formula*

$$\overline{m}(f) = m((f, 1-f)).$$

We can see that $\overline{m}$ is a state on $\mathcal{T}$. Then by the Butnariu - Klement theorem ([2]) there exists a probability $P : \mathcal{S} \to [0,1]$ such that

$$\overline{m}(f) = \int_\Omega f dP$$

for any $f \in \mathcal{T}$. Of course, the previous formula does not imply the representation of IF-state given by formula

$$m(A) = (1-\alpha) \int_\Omega \mu_A dP + \alpha(1 - \int_\Omega \nu_A dP)$$

for each $A = (\mu_A, \nu_A) \in \mathcal{F}$, where $\alpha \in [0,1]$ and $P : \mathcal{S} \to [0,1]$ be a probability (see [14]). We see that the IF-approach cannot be coordinatwisely reduced to the fuzzy approach.

The next important notion is notion of IF-observable.

By **IF-observable on** $\mathcal{F}$ we understand any mapping $x : \mathcal{B}(\mathbf{R}) \to \mathcal{F}$ satisfying the following conditions:

(i) $x(\mathbf{R}) = (1,0)$;

(ii) if $A \cap B = \emptyset$, then $x(A) \odot x(B) = (0,1)$ and $x(A \cup B) = x(A) \oplus x(B)$;

(iii) if $A_n \nearrow A$, then $x(A_n) \nearrow x(A)$.

If we denote $x(A) = \big(x^\flat(A), 1 - x^\sharp(A)\big)$ for each $A \in \mathcal{B}(\mathbf{R})$, then $x^\flat, x^\sharp : \mathcal{B}(\mathbf{R}) \to \mathcal{T}$ are observables, see [11].

By **product operation on** $\mathcal{F}$ we understand any binary operation $\cdot$ satisfying the following conditions:

(i) $(1,0) \cdot (a_1, a_2) = (a_1, a_2)$ for each $(a_1, a_2) \in \mathcal{F}$;

(ii) the operation $\cdot$ is commutative and associative;

(iii) if $(a_1, a_2) \odot (b_1, b_2) = (0,1)$ and $(a_1, a_2), (b_1, b_2) \in \mathcal{F}$, then
$(c_1, c_2) \cdot \big((a_1, a_2) \oplus (b_1, b_2)\big) = \big((c_1, c_2) \cdot (a_1, a_2)\big) \oplus \big((c_1, c_2) \cdot (b_1, b_2)\big)$ and
$\big((c_1, c_2) \cdot (a_1, a_2)\big) \odot \big((c_1, c_2) \cdot (b_1, b_2)\big) = (0,1)$ for each $(c_1, c_2) \in \mathcal{F}$;

(iv) if $(a_{1n}, a_{2n}) \searrow (0,1)$, $(b_{1n}, b_{2n}) \searrow (0,1)$ and $(a_{1n}, a_{2n}), (b_{1n}, b_{2n}) \in \mathcal{F}$, then $(a_{1n}, a_{2n}) \cdot (b_{1n}, b_{2n}) \searrow (0,1)$.

The operation $\cdot$ defined by

$$(x_1, y_1) \cdot (x_2, y_2) = (x_1 \cdot x_2, y_1 + y_2 - y_1 \cdot y_2)$$

for each $(x_1, y_1), (x_2, y_2) \in \mathcal{F}$ is a product operation on $\mathcal{F}$ (see [8]).

3 Lower and upper limits

The aim of this section is a modification of almost everywhere convergence with help of lim sup and lim inf and the construction of translation formulas for these limits. First we define max-min connectives

$$A \vee B = (a_1 \vee b_1, a_2 \wedge b_2),$$

$$A \wedge B = (a_1 \wedge b_1, a_2 \vee b_2)$$

and IF-ordering

$$A \le B \iff a_1 \le b_1 \text{ and } a_2 \ge b_2$$

Definition 3.1 *Given a sequence $x_1, x_2, \ldots$ of IF-observables in a family of IF-events $\mathcal{F}$ with IF-probability $\mathcal{P}$. We write $\overline{x}_{IF} = \limsup_{n\to\infty} x_n$ if $\overline{x}_{IF}$:*

$\mathcal{B}(\mathbf{R}) \to \mathcal{F}$ *is an IF-observable having the following property*

$$\overline{x}_{IF}((-\infty,t)) = \bigvee_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n\Big(\Big(-\infty, t-\frac{1}{p}\Big)\Big)$$

for all $t \in \mathbf{R}$.

Note that if another IF-observable y *satisfies the above condition then* $\mathcal{P}\circ y = \mathcal{P}\circ \overline{x}_{IF}$.

Similarly we write $\underline{x}_{IF} = \liminf_{n\to\infty} x_n$ *if* $\underline{x}_{IF} : \mathcal{B}(\mathbf{R}) \to \mathcal{F}$ *is an IF-observable satisfying the condition*

$$\underline{x}_{IF}((-\infty,t)) = \bigvee_{p=1}^{\infty}\bigwedge_{k=1}^{\infty}\bigvee_{n=k}^{\infty} x_n\Big(\Big(-\infty, t-\frac{1}{p}\Big)\Big)$$

for all $t \in \mathbf{R}$.

Theorem 3.2 *The IF-observables* $\overline{x}_{IF}$, $\underline{x}_{IF}$ *from Definition 3.1 can be expressed in the following form*

$$\overline{x}_{IF}(A) = \Big(\overline{x^{\flat}}(A), 1-\overline{x^{\sharp}}(A)\Big),$$

$$\underline{x}_{IF}(A) = \Big(\underline{x^{\flat}}(A), 1-\underline{x^{\sharp}}(A)\Big),$$

for each $A \in \mathcal{B}(\mathbf{R})$. *Here* $\overline{x^{\flat}}$, $\underline{x^{\flat}}$ *are upper and lower limits of sequence* $(x_n^{\flat})_1^{\infty}$ *of observables in tribe* $\mathcal{T}$ *and* $\overline{x^{\sharp}}$, $\underline{x^{\sharp}}$ *are upper and lower limits of sequence* $(x_n^{\sharp})_1^{\infty}$ *of observables in tribe* $\mathcal{T}$ *(see [18]).*

Proof. Let $(x_n)_1^{\infty}$ be a sequence of IF-observables. Denote

$$x_n(A) = \big(x_n^{\flat}(A), 1-x_n^{\sharp}(A)\big)$$

for each $A \in \mathcal{B}(\mathbf{R})$, then $x_n^{\flat}, x_n^{\sharp} : \mathcal{B}(\mathbf{R}) \to \mathcal{T}$ are observables for every $n \in N$.

Therefore using definition of max-min connectives $\vee, \wedge$ we obtain

$$\overline{x}_{IF}(-\infty,t) = \bigvee_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n\Big(-\infty, t-\tfrac{1}{p}\Big) =$$

$$= \left(\bigvee_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n^{\flat}\Big(-\infty, t-\tfrac{1}{p}\Big), 1-\bigvee_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n^{\sharp}\Big(-\infty, t-\tfrac{1}{p}\Big)\right) = \Big(\overline{x^{\flat}}(-\infty,t), 1-\overline{x^{\sharp}}(-\infty,t)\Big)$$

and

$$\underline{x}_{IF}(-\infty,t) = \bigvee_{p=1}^{\infty}\bigwedge_{k=1}^{\infty}\bigvee_{n=k}^{\infty} x_n\Big(-\infty, t-\tfrac{1}{p}\Big) =$$

$$= \left(\bigvee_{p=1}^{\infty}\bigwedge_{k=1}^{\infty}\bigvee_{n=k}^{\infty} x_n^{\flat}\Big(-\infty, t-\tfrac{1}{p}\Big), 1-\bigvee_{p=1}^{\infty}\bigwedge_{k=1}^{\infty}\bigvee_{n=k}^{\infty} x_n^{\sharp}\Big(-\infty, t-\tfrac{1}{p}\Big)\right) = \Big(\underline{x^{\flat}}(-\infty,t), 1-\underline{x^{\sharp}}(-\infty,t)\Big)$$

for every $t \in \mathbf{R}$. □

Proposition 3.1 *Let* $x_1, x_2, \ldots$ *be a sequence of IF-observables in a family* $\mathcal{F}$ *of IF-events with IF-probability* $\mathcal{P}$. *Suppose that both IF-observables* $\overline{x}_{IF}$ *and* $\underline{x}_{IF}$ *exist. Then for every* $t \in \mathbf{R}$

$$\overline{x}_{IF}((-\infty,t)) \le \underline{x}_{IF}((-\infty,t)).$$

Proof. Let $x_1, x_2, \ldots$ be a sequence of IF-observables. Suppose that both IF-observables $\overline{x}_{IF}$ and $\underline{x}_{IF}$ exist, then by *Theorem 3.2* there exist observables $\overline{x^{\flat}}$, $\underline{x^{\flat}}$, $\overline{x^{\sharp}}$ and $\underline{x^{\sharp}}$ in tribe $\mathcal{T}$.

Since from *Proposition 8.6.4* in [18] the following inequalities hold for every $t \in \mathbf{R}$

$$\overline{x^{\flat}}((-\infty,t)) \le \underline{x^{\flat}}((-\infty,t)), \tag{1}$$

$$\overline{x^{\sharp}}((-\infty,t)) \le \underline{x^{\sharp}}((-\infty,t)),$$

then we obtain

$$1-\overline{x^{\sharp}}((-\infty,t)) \ge 1-\underline{x^{\sharp}}((-\infty,t)) \tag{2}$$

by simply modifications.

Hence by (1), (2) and *Theorem 3.2* we have

$$\overline{x}_{IF}((-\infty,t)) \le \underline{x}_{IF}((-\infty,t)).$$

□

Definition 3.3 *A sequence* $x_1, x_2, \ldots$ *of IF-observables in family of IF-events* $\mathcal{F}$ *with IF-probability* $\mathcal{P}$ *is said to converge* $\mathcal{P}$*-almost everywhere to an IF-observable* x, *if both IF-observables* $\overline{x}_{IF}, \underline{x}_{IF}$ *exist and for each* $t \in \mathbf{R}$

$$\mathcal{P}(\overline{x}_{IF}(-\infty,t)) = \mathcal{P}(x(-\infty,t)) = \mathcal{P}(\underline{x}_{IF}(-\infty,t)).$$

Proposition 3.2 *A sequence* $(x_n)_1^{\infty}$ *of IF-observables in family of IF-event with product* $(\mathcal{F},\cdot)$ *converges* $\mathcal{P}$*-almost everywhere to zero IF-observable* $0_{\mathcal{F}}$ *defined by*

$$0_{\mathcal{F}}(A) = \begin{cases} (1,0), & \text{if } 0 \in A \\ (0,1), & \text{if } 0 \notin A \end{cases}$$

for each $A \in \mathcal{B}(\mathbf{R})$ *if and only if*

$$\mathcal{P}\left(\bigwedge_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n\Big(-\frac{1}{p},\frac{1}{p}\Big)\right) = \langle 1,1\rangle.$$

Proof. "⇒" Let $(x_n)_1^{\infty}$ converges $\mathcal{P}$-almost everywhere to $0_{\mathcal{F}}$. Then by *Definition 3.3* there exist the IF-observables $\overline{x}_{IF}, \underline{x}_{IF}$ such that the following equality holds for each $t \in \mathbf{R}$,

$$\begin{aligned} \mathcal{P}(\overline{x}_{IF}(-\infty,t)) &= \mathcal{P}(0_{\mathcal{F}}(-\infty,t)) = \\ &= \mathcal{P}(\underline{x}_{IF}(-\infty,t)). \end{aligned} \tag{3}$$

Let $t > 0$. Then by (3), *Theorem 3.2* and by definition of IF-observable $0_{\mathcal{F}}$ we have

$$\mathcal{P}\Big(\Big(\overline{x^{\flat}}(-\infty,t) \quad , \quad 1-\overline{x^{\sharp}}(-\infty,t)\Big)\Big) = \mathcal{P}\big((1,0)\big) = \\ = \mathcal{P}\Big(\Big(\underline{x}^{\flat}(-\infty,t), 1-\underline{x}^{\sharp}(-\infty,t)\Big)\Big),$$

$$\Big\langle \mathcal{P}^{\flat}\big(\overline{x^{\flat}}(-\infty,t)\big) \quad , \quad \mathcal{P}^{\sharp}\big(\overline{x^{\sharp}}(-\infty,t)\big)\Big\rangle = \langle 1,1\rangle = \\ = \Big\langle \mathcal{P}^{\flat}\big(\underline{x}^{\flat}(-\infty,t)\big), \mathcal{P}^{\sharp}\big(\underline{x}^{\sharp}(-\infty,t)\big)\Big\rangle.$$

Therefore

$$\mathcal{P}^{\flat}\big(\overline{x^{\flat}}(-\infty,t)\big) = 1 = \mathcal{P}^{\flat}\big(\underline{x}^{\flat}(-\infty,t)\big) \tag{4}$$

$$\mathcal{P}^{\sharp}\big(\overline{x^{\sharp}}(-\infty,t)\big) = 1 = \mathcal{P}^{\sharp}\big(\underline{x}^{\sharp}(-\infty,t)\big). \tag{5}$$

Now let $t \leq 0$. Then by (3), *Theorem 3.2* and by definition of IF-observable $0_{\mathcal{F}}$ we have

$$\mathcal{P}\Big(\Big(\overline{x^{\flat}}(-\infty,t) \quad , \quad 1-\overline{x^{\sharp}}(-\infty,t)\Big)\Big) = \mathcal{P}\big((0,1)\big) = \\ = \mathcal{P}\Big(\Big(\underline{x}^{\flat}(-\infty,t), 1-\underline{x}^{\sharp}(-\infty,t)\Big)\Big),$$

$$\Big\langle \mathcal{P}^{\flat}\big(\overline{x^{\flat}}(-\infty,t)\big) \quad , \quad \mathcal{P}^{\sharp}\big(\overline{x^{\sharp}}(-\infty,t)\big)\Big\rangle = \langle 0,0\rangle = \\ = \Big\langle \mathcal{P}^{\flat}\big(\underline{x}^{\flat}(-\infty,t)\big), \mathcal{P}^{\sharp}\big(\underline{x}^{\sharp}(-\infty,t)\big)\Big\rangle.$$

Therefore

$$\mathcal{P}^{\flat}\big(\overline{x^{\flat}}(-\infty,t)\big) = 0 = \mathcal{P}^{\flat}\big(\underline{x}^{\flat}(-\infty,t)\big) \tag{6}$$

$$\mathcal{P}^{\sharp}\big(\overline{x^{\sharp}}(-\infty,t)\big) = 0 = \mathcal{P}^{\sharp}\big(\underline{x}^{\sharp}(-\infty,t)\big). \tag{7}$$

Finally by (4), (6) and *Proposition 8.6.6* in [18] we obtain that the sequence of observables $(x_n^{\flat})_1^{\infty}$ converges $\mathcal{P}^{\flat}$-almost everywhere to a observable $0_{\mathcal{T}}$ defined by

$$0_{\mathcal{T}}(A) = \begin{cases} 1, & \text{if } 0 \in A \\ 0, & \text{if } 0 \notin A \end{cases}$$

for each $A \in \mathcal{B}(\mathbf{R})$. Similarly by (5), (7) and *Proposition 8.6.6* in [18] we obtain that the sequence of observables $(x_n^{\sharp})_1^{\infty}$ converges $\mathcal{P}^{\sharp}$-almost everywhere to the observable $0_{\mathcal{T}}$, too.

Hence the equalities

$$\mathcal{P}^{\flat}\Big(\bigwedge_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n^{\flat}\Big(-\frac{1}{p},\frac{1}{p}\Big)\Big) = 1$$

and

$$\mathcal{P}^{\sharp}\Big(\bigwedge_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n^{\sharp}\Big(-\frac{1}{p},\frac{1}{p}\Big)\Big) = 1$$

hold from *Theorem 3* in [15] and therefore we have

$$\mathcal{P}\Big(\bigwedge_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n\Big(-\frac{1}{p},\frac{1}{p}\Big)\Big) = \langle 1,1\rangle.$$

The proof of "$\Leftarrow$" is a analogue to the proof of "$\Rightarrow$". □

The *Proposition 3.2* says about equality of two kinds of definition of almost everywhere convergence in family of IF-events $\mathcal{F}$ with product.

Now we recall the notion of **IF-distribution function** defined in [6]. It is a mapping $\mathbf{F} : \mathbf{R} \to \mathcal{J}$ given by formula

$$\mathbf{F}(t) = \mathcal{P} \circ x((-\infty,t)) = \\ = \langle \mathcal{P}^{\flat}((-\infty,t)), \mathcal{P}^{\sharp}((-\infty,t))\rangle = \langle \mathbf{F}^{\flat}(t), \mathbf{F}^{\sharp}(t)\rangle$$

for each $t \in \mathbf{R}$, where $\mathbf{F}^{\flat}, \mathbf{F}^{\sharp} : \mathbf{R} \to \langle 0,1\rangle$ are the distribution functions.

Proposition 3.3 *Let $\mathbf{F} : \mathbf{R} \to \mathcal{J}$ be an IF-distribution function and $(x_n)_1^{\infty}$ be a sequence of IF-observables. Let*

$$\varphi_{IF}(t) = \bigvee_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n\Big(\big(-\infty, t-\frac{1}{p}\big)\Big),$$

$$\psi_{IF}(t) = \bigvee_{p=1}^{\infty}\bigwedge_{k=1}^{\infty}\bigvee_{n=k}^{\infty} x_n\Big(\big(-\infty, t-\frac{1}{p}\big)\Big).$$

If $\mathbf{F}(t) = \mathcal{P} \circ \varphi_{IF}(t) = \mathcal{P} \circ \psi_{IF}(t)$ for each $t \in \mathbf{R}$, then there exist IF-observables $\overline{x}_{IF}, \underline{x}_{IF} : \mathcal{B}(\mathbf{R}) \to \mathcal{F}$ such that

$$\overline{x}_{IF}\big((-\infty,t)\big) = \varphi_{IF}(t),$$

$$\underline{x}_{IF}\big((-\infty,t)\big) = \psi_{IF}(t)$$

for each $t \in \mathbf{R}$.

Proof. Denote

$$\varphi^{\flat}(t) = \bigvee_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n^{\flat}\Big(\big(-\infty, t-\frac{1}{p}\big)\Big),$$

$$\psi^{\flat}(t) = \bigvee_{p=1}^{\infty}\bigwedge_{k=1}^{\infty}\bigvee_{n=k}^{\infty} x_n^{\flat}\Big(\big(-\infty, t-\frac{1}{p}\big)\Big)$$

and analogously

$$\varphi^{\sharp}(t) = \bigvee_{p=1}^{\infty}\bigvee_{k=1}^{\infty}\bigwedge_{n=k}^{\infty} x_n^{\sharp}\Big(\big(-\infty, t-\frac{1}{p}\big)\Big),$$

$$\psi^{\sharp}(t) = \bigvee_{p=1}^{\infty}\bigwedge_{k=1}^{\infty}\bigvee_{n=k}^{\infty} x_n^{\sharp}\Big(\big(-\infty, t-\frac{1}{p}\big)\Big).$$

Then $\varphi_{IF}(t)$, $\psi_{IF}(t)$ can be expressed by formulas

$$\varphi_{IF}(t) = (\varphi^{\flat}(t), 1 - \varphi^{\sharp}(t)),$$

$$\psi_{IF}(t) = (\psi^{\flat}(t), 1 - \psi^{\sharp}(t)).$$

If $\mathbf{F}(t) = \mathcal{P} \circ \varphi_{IF}(t) = \mathcal{P} \circ \psi_{IF}(t)$ for each $t \in \mathbf{R}$, then

$$\begin{aligned} \langle \mathbf{F}^{\flat}(t), \mathbf{F}^{\sharp}(t) \rangle &= \langle \mathcal{P}^{\flat} \circ \varphi^{\flat}(t), \mathcal{P}^{\sharp} \circ \varphi^{\sharp}(t) \rangle = \\ &= \langle \mathcal{P}^{\flat} \circ \psi^{\flat}(t), \mathcal{P}^{\sharp} \circ \psi^{\sharp}(t) \rangle. \end{aligned}$$

Hence

$$\mathbf{F}^{\flat}(t) = \mathcal{P}^{\flat} \circ \varphi^{\flat}(t) = \mathcal{P}^{\flat} \circ \psi^{\flat}(t), \tag{8}$$

$$\mathbf{F}^{\sharp}(t) = \mathcal{P}^{\sharp} \circ \varphi^{\sharp}(t) = \mathcal{P}^{\sharp} \circ \psi^{\sharp}(t), \tag{9}$$

where $\mathbf{F}^{\flat}(t), \mathbf{F}^{\sharp}(t)$ are the distribution functions.

By (8), (9) and from *Proposition* in [15] there exist observables $\overline{x^{\flat}}$, $\underline{x^{\flat}}$, $\overline{x^{\sharp}}$, $\underline{x^{\sharp}} : \mathcal{B}(\mathbf{R}) \to \mathcal{T}$ such that

$$\overline{x^{\flat}}((-\infty, t)) = \varphi^{\flat}(t),\ \underline{x^{\flat}}((-\infty, t)) = \psi^{\flat}(t),$$

$$\overline{x^{\sharp}}((-\infty, t)) = \varphi^{\sharp}(t),\ \underline{x^{\sharp}}((-\infty, t)) = \psi^{\sharp}(t)$$

for each $t \in \mathbf{R}$.

Therefore there exist IF-observables $\overline{x}_{IF}, \underline{x}_{IF}$ given by formulas

$$\overline{x}_{IF}(A) = \big(\overline{x^{\flat}}(A), 1 - \overline{x^{\sharp}}(A)\big),$$

$$\underline{x}_{IF}(A) = \big(\underline{x^{\flat}}(A), 1 - \underline{x^{\sharp}}(A)\big)$$

for each $A \in \mathcal{B}(\mathbf{R})$ such that the following equalities hold

$$\begin{aligned} \overline{x}_{IF}((-\infty, t)) &= \big(\overline{x^{\flat}}((-\infty, t)), 1 - \overline{x^{\sharp}}((-\infty, t))\big) = \\ &= \big(\varphi^{\flat}(t), 1 - \varphi^{\sharp}(t)\big) = \varphi_{IF}(t), \end{aligned}$$

$$\begin{aligned} \underline{x}_{IF}((-\infty, t)) &= \big(\underline{x^{\flat}}((-\infty, t)), 1 - \underline{x^{\sharp}}((-\infty, t))\big) = \\ &= \big(\psi^{\flat}(t), 1 - \psi^{\sharp}(t)\big) = \psi_{IF}(t) \end{aligned}$$

for each $t \in \mathbf{R}$. □

4 Conclusion

The paper is concerned in the probability theory on IF-events with product. We define the notion of upper and lower limits. We compare two kinds of definition almost everywhere convergence for IF-events with product and we show that they are equivalent.

Acknowledgement

This paper was supported by Grant VEGA 1/2002/05.

References

[1] K. Atanassov, Intuitionistic Fuzzy sets: Theory and Applications, Physica Verlag, New York, 1999.

[2] D. Butnariu, E. P. Klement: Triangular norm-based measures and their markov kernel representation, J. Math. Anal. Appl. 162, 1991, 11 - 143.

[3] A. Dvurečenskij, S. Pulmannová, New Trends in Quantum Structures, Kluwer, Dordrecht, 2000.

[4] P. Grzegorzewski, E. Mrówka, Probability of intuistionistic fuzzy events, in: P. Grzegorzewski et al. eds, *Soft Metods in Probability, Statistics and Data Analysis*, Physica Verlag, New York, 2002, pp. 105-115.

[5] K. Lendelová, IF-probability on MV-algebras, Notes on Intuitionistic Fuzzy Sets, Vol. 11, Number 3, (2005) 66-72.

[6] K. Lendelová, Convergence of IF-observables, in: *Issues in the Representation and processing of Uncertain and Imprecise Information - Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized nets and Related Topics*, EXIT, Warszawa, 2005, pp. 232-240.

[7] K. Lendelová, Strong law of large numbers for IF-events, in: *Proceedings of the Eleventh International Conference IPMU 2006, July, 2-7*, Éditions EDK Paris, France, Vol. 3, 2006, pp. 2363-2366.

[8] K. Lendelová, Conditional IF-probability, Advances in Soft Computing: Soft Methods for Integrated Uncertainty Modelling, Springer-Verlag, Berlin Heidelberg, 2006, pp. 275-283.

[9] K. Lendelová, J. Petrovičová, Representation of IF-probability for MV-algebras, Soft Computing - A Fusion of Fundations, Metodologies and Applications, Vol. 10, Number 7, Springer - Verlag, 2006, pp. 564-566.

[10] K. Lendelová, B. Riečan, Weak law of large numbers for IF-events, in: B. De Baets et al. eds., *Current Issues in Data and Knowledge Engineering*, EXIT, Warszawa, 2004, pp. 309-314.

[11] J. Petrovičová, B. Riečan, On the central limit theorem on IFS-events, Mathware Soft Computing 12, 2005, pp. 5-14.

[12] P. Pták, S. Pulmannová, Orthomodular Structures as Quantum Logics, Kluwer Acad. Publ., Dordrecht, 1991.

[13] B. Riečan, A descriptive definition of the probability on intuitionistic fuzzy sets, in: M. Wagenecht, R. Hampet eds., *EUSFLAT '2003*, Zittau-Goerlitz Univ. Appl. Sci., 2003, pp. 263-266.

[14] B. Riečan: On a problem of Radko Mesiar: general from of IF probabilities, Fuzzy Sets and Systems 152, 2006, 1485 - 1490.

[15] B. Riečan, Almost everywhere convergence in probability MV-algebras with product, Soft Computing 5, 2001, pp. 396-399.

[16] B. Riečan, Representation of Probabilities on IFS Events, in: López-Diáz et al. eds., *Soft Methodology and Random Information Systems*, Springer, Berlin Heidelberg New York, 2004, pp. 243-248.

[17] B. Riečan, D. Mundici, Probability in MV-algebras, in: E. Pap ed. *Handbook of Measure Theory*, Elsevier, Amsterdam, 2002, pp. 869-909.

[18] B. Riečan, T. Neubrunn, Integral, Measure, and Ordering, Kluwer Academic Publishers, Dordrecht and Ister Science, Bratislava, 1997.

Outer measure on MV-algebras

Alžbeta Michalíková
Faculty of Natural Sciences
Matej Bel University
Tajovského 40
SK-974 01 Banská Bystrica
michalik@fpv.umb.sk

Abstract

In this paper we study an outer measure on MV-algebras. In *Section 1* the definition of MV-algebra and Mundici theorem are remind. In *Section 2* there is defined an outer measure, measurable elements and there is proved Choquet lema for this structure. In conclusion some properties of measurable elements are resumed.

Keywords: MV-algebra, outer measure, Choquet lema.

1 MV-algebras: basic notions

In this section we recall the basic notions like MV-algebra, ℓ-group, isomorphism between MV-algebra and ℓ-group.

Definition 1.1 *An MV-algebra $M = (M, 0, 1, \neg, \oplus, \odot)$ is a system where $\oplus$ is an associative, commutative operation with neutral element 0, and in addition, $\neg 0 = 1$, $\neg 1 = 0$, $x \oplus 1 = 1$, $x \odot y = \neg(\neg x \oplus \neg y)$ and $y \oplus \neg(y \oplus \neg x) = x \oplus \neg(x \oplus \neg y)$, for each $x, y \in M$.*

The constant 1 and the operation $\odot$ are definable from the remaining operations $0, \neg, \oplus$. In every MV-algebra $M = (M, 0, 1, \neg, \oplus, \odot)$ the binary relation $\leq$ given by $x \leq y \Leftrightarrow x \odot \neg y = 0$ is a partial order.

Definition 1.2 *Structure $(G, \oplus, \vee, \wedge)$ is called ℓ-group if it holds:*

1. *$(G, \oplus)$ is an Abelian group.*
2. *$(G, \vee, \wedge)$ is a lattice.*
3. *If $a \leq b$ then $a \oplus c \leq b \oplus c$ for each $a, b, c \in G$.*

In the book [1] it is presented the following theorem:

Theorem 1.3 *For any ℓ-group G with strong unit u let $\Gamma(G, u)$ be the unit interval $[0, u] = \{h \in G \mid 0 \leq h \leq u\}$, equipped with operations:*

$$\neg g = u - g,\ g \oplus h = u \wedge (g + h),\ g \odot h = 0 \vee (g + h - u).$$

Then the structure $M = ([0, u], 0, u, \neg, \oplus, \odot) = \Gamma(g, u)$ is an MV-algebra. The lattice operations on M agree with those of G. Up to isomorphism every MV-algebra M can be identified with the unit interval of a unique ℓ-group G with strong unit u in symbols

$$M = \Gamma(G, u).$$

We say that G is the ℓ-group corresponding to M.

This theorem is called Mundici theorem.

2 Outer measure on MV-algebras

Because of the Mundici theorem we will use like MV-algebra the structure $M = ([0, u], 0, u, \neg, \oplus, \odot)$.

Definition 2.1 *Let $\mu^* : M \to [0, 1]$ be a function with the following properties:*

1. *$\mu^*(0) = 0,\ \mu^*(u) = 1$.*
2. *If $x, y \in M,\ x \leq y$ then $\mu^*(x) \leq \mu^*(y)$.*
3. *If $x, y \in M$ then $\mu^*(x) + \mu^*(y) \geq \mu^*(x \vee y) + \mu^*(x \wedge y)$.*
4. *If $x, y \in M$ then $\mu^*(x) + \mu^*(y) \geq \mu^*(x \oplus y)$.*

Then μ^* is an outer measure defined on the MV-algebra M.

Proposition 2.1 *Let μ^* be an outer measure on M. Then it holds:*

$$\mu^*(x) \leq \mu^*(x \wedge y) + \mu^*(x - (x \wedge y)),$$

for each $x, y \in M$.

Proof.
Since $x = (x \wedge y) \oplus (x - (x \wedge y))$ then by condition 4 it holds:

$$\mu^*(x) = \mu^*((x \wedge y) \oplus (x - (x \wedge y))) \leq$$

$$\leq \mu^*(x \wedge y) + \mu^*(x - (x \wedge y)).$$

This property is called subadditivity of the measure μ^*.

Definition 2.2 *Let μ^* be an outer measure defined on the set M. An element $x \in M$ is called μ^*-measurable element if it holds:*

$$\mu^*(a) = \mu^*(a \wedge x) + \mu^*(a - (a \wedge x)),$$

for each $a \in M$.

Theorem 2.3 *Denote by $\mathcal{M}$ the set of all μ^*-measurable elements. Then the set $\mathcal{M}$ form a lattice.*

Proof. We need to show, that if $x, y \in \mathcal{M}$, then also $x \wedge y$ and $x \vee y$ are from $\mathcal{M}$.
Let $x, y \in \mathcal{M},\ a \in M$. Then

$$\mu^*(a \wedge (x \wedge y)) = \mu^*((a \wedge x) \wedge (a \wedge y)) \leq$$

$$\leq \mu^*(a \wedge x) + \mu^*(a \wedge y) - \mu^*(a \wedge (x \vee y)).$$

Similarly

$$\mu^*(a - (a \wedge x \wedge y)) = \mu^*((a - a \wedge x) \vee (a - a \wedge y)) \leq$$

$$\leq \mu^*(a - a \wedge x) + \mu^*(a - a \wedge y) - \mu^*(a - a \wedge (x \vee y)).$$

Therefore

$$\mu^*(a \wedge (x \wedge y)) + \mu^*(a - (a \wedge x \wedge y)) \leq$$

$$\mu^*(a \wedge x) + \mu^*(a - a \wedge x) + \mu^*(a \wedge y) + \mu^*(a - a \wedge y) -$$

$$-\mu^*(a \wedge (x \vee y)) - \mu^*(a - a \wedge (x \vee y)).$$

But $x, y \in M$ and therefore for each $a \in M$ holds:

$$a = (a \wedge (x \vee y)) \oplus (a - a \wedge (x \vee y)),$$

hence

$$\mu^*(a \wedge (x \wedge y)) + \mu^*(a - a \wedge (x \wedge y)) \leq$$

$$\leq \mu^*(a) + \mu^*(a) - \mu^*(a) = \mu^*(a).$$

The opposite inequality follows from the subadditivity of the measure μ^*. The proof that $x \vee y$ belongs to $\mathcal{M}$ is analogical.

3 Induced outer measure

Let M_0 be a subalgebra of MV-algebra M, closed to the operations $\neg, \oplus, \odot$ and let $u \in M_0$.

Definition 3.1 *Function $\mu : M_0 \to [0, 1]$ is called a measure on the set M_0 if it holds:*

1. $\mu(0) = 0,\ \mu(u) = 1$.
2. *If $a, b \in M_0, a \leq b$ then $\mu(a) \leq \mu(b)$.*
3. *If $a, b \in M_0$ then $\mu(a) + \mu(b) \geq \mu(a \vee b) + \mu(a \wedge b)$.*
4. *If $a, b \in M_0$ then $\mu(a) + \mu(b) \geq \mu(a \oplus b)$. Moreover, if $a \odot b = 0$ then $\mu(a) + \mu(b) = \mu(a \oplus b)$.*
5. *If $a_n \nearrow a$ (i.e. $a_n \leq a_{n+1},\ a = \bigvee_{n=1}^{\infty} a_n$), $a_n \in M_0$ $(n = 1, 2, \ldots)$, then $a \in M_0$ and $\mu(a) = \lim_{n \to \infty} \mu(a_n)$.*

Definition 3.2 *Let $x \in M$. Define:*

$$\mu^*(x) = \inf \{\mu(a); a \geq x, a \in M_0\}.$$

Remark 3.3 In the follows we shall use the symbol $\bigwedge$ to label the infimum of the set.

Proposition 3.1 *Function μ^* defined in* Definition 3.2 *is an extension of the measure μ. Moreover it satisfied the conditions of the outer measure.*

Proof. Function μ^* is an extension of the measure μ if for each $a \in M_0$ holds $\mu^*(a) = \mu(a)$.
Let $b \in M_0, b \geq a$ then $\mu(b) \geq \mu(a)$, therefore

$$\mu^*(a) = \bigwedge \{\mu(b);\ b \geq a,\ b \in M_0\} \geq \mu(a).$$

On the other hand

$$\mu(a) \in \{\mu(b); b \geq a, b \in M_0\}.$$

Therefore

$$\mu(a) \geq \bigwedge \{\mu(b);\ b \geq a,\ b \in B\} = \mu^*(a).$$

Since $0 \in M_0,\ u \in M_0$ then $\mu^*(0) = \mu(0) = 0$ and $\mu^*(u) = \mu(u) = 1$.
Let $x, y \in M,\ x \leq y$. Then

$$\mu^*(x) = \bigwedge \{\mu(a);\ a \geq x,\ a \in M_0\}$$

and

$$\mu^*(y) = \bigwedge \{\mu(b);\ b \geq y,\ b \in M_0\}.$$

Because $x \leq y$ and μ is non-decreasing function, we can compare the sets and we get:

$$\{\mu(a);\ a \geq x,\ a \in M_0\} \supset \{\mu(b);\ b \geq y,\ b \in M_0\}.$$

We can see that $\mu^*(x)$ is the infimum of the lager set, so this is also lower bound of the smaller set. Therefore

$$\mu^*(x) \leq \mu^*(y).$$

Let $x, y \in M$ then for each $a \in M_0, b \in M_0, a \geq x, b \geq y$ holds:

$$\mu(a) + \mu(b) \geq \mu(a \vee b) + \mu(a \wedge b).$$

Since $a \vee b \in M_0, a \wedge b \in M_0$ and $a \vee b \geq x \vee y, a \wedge b \geq x \wedge y$ then

$$\mu(a \vee b) + \mu(a \wedge b) \geq \mu^*(x \vee y) + \mu^*(x \wedge y).$$

Now fix for a moment b then for each $\mu(a)$ holds:

$$\mu(a) \geq \mu^*(x \vee y) + \mu^*(x \wedge y) - \mu(b).$$

But this inequality holds for each appropriate a so it holds also for the infimum of $\mu(a)$ and $\bigwedge \mu(a) = \mu^*(x)$. Therefore

$$\mu^*(x) \geq \mu^*(x \vee y) + \mu^*(x \wedge y) - \mu(b).$$

Similarly can be proved that inequality holds for $\mu^*(y)$. Therefore

$$\mu^*(x) + \mu^*(y) \geq \mu^*(x \vee y) + \mu^*(x \wedge y).$$

Let $x, y \in M$ then there exist such elements $a, b \in M_0$ that holds $x \leq a$ and $y \leq b$. Then also $x \oplus y \leq a \oplus b$ therefore

$$\mu^*(x \oplus y) \leq \mu(a \oplus b) \leq \mu(a) + \mu(b).$$

Fix for a moment expression $\mu(a)$ then for each $b \geq y$ holds:

$$\mu^*(x \oplus y) - \mu(a) \leq \mu(b).$$

Hence also for $\mu^*(y) = \bigwedge \mu(b)$ holds:

$$\mu^*(x \oplus y) - \mu(a) \leq \mu^*(y).$$

Similarly

$$\mu^*(x \oplus y) - \mu^*(y) \leq \mu^*(x).$$

Therefore for each $x, y \in M$ holds:

$$\mu^*(x \oplus y) \leq \mu^*(x) + \mu^*(y).$$

Definition 3.4 *The function μ^* from* Definition 3.2 *is called an outer measure induced by measure μ.*

Theorem 3.5 *If $x_n \nearrow x$ (i.e. $x_n \leq x_{n+1}$, $x = \bigvee_{n=1}^{\infty} x_n$), $x_n \in M$ $(n = 1, 2, \ldots)$, $x \in M$ then $\mu^*(x_n) \nearrow \mu^*(x)$.*

Proof. Evidently $\lim_{n\to\infty} \mu^*(x_n) \leq \mu^*(x)$.
Let $\varepsilon > 0$. Take $b \geq x$, $b \in M_0$ and $a_n \in M_0$ $(n = 1, 2, \ldots)$ such that $b \geq a_n \geq x_n$ and it holds:

$$\mu^*(x_n) + \frac{\varepsilon}{2^n} > \mu(a_n).$$

Put $b_n = \bigvee_{i=1}^{n} a_i$. Then $x_n \leq b_n \leq b_{n+1} \leq b$ $(n = 1, 2, \ldots)$. By induction can be proved following relation:

$$\mu^*(x_n) + \sum_{i=1}^{n} \frac{\varepsilon}{2^i} > \mu\Big(\bigvee_{i=1}^{n} a_i\Big).$$

The first step is clear.
The second step: We will use the property of measure μ:

$$\mu(a \vee b) + \mu(a \wedge b) \leq \mu(a) + \mu(b).$$

Since $b_n \geq x_n$ and $a_{n+1} \geq x_{n+1}$. Then $(b_n \wedge a_{n+1}) \geq (x_n \wedge x_{n+1}) = x_n$.
Let for some $n \in N$ holds:

$$\mu^*(x_n) + \sum_{i=1}^{n} \frac{\varepsilon}{2^i} > \mu\Big(\bigvee_{i=1}^{n} a_i\Big)$$

then

$$\mu\Big(\bigvee_{i=1}^{n+1} a_i\Big) = \mu(b_{n+1}) = \mu(b_n \vee a_{n+1}) \leq$$

$$\leq \mu(b_n) + \mu(a_{n+1}) - \mu(b_n \wedge a_{n+1}) <$$

$$< \mu^*(x_n) + \sum_{i=1}^{n} \frac{\varepsilon}{2^i} + \mu^*(x_{n+1}) + \frac{\varepsilon}{2^{n+1}} - \mu^*(x_n) =$$

$$= \mu^*(x_{n+1}) + \sum_{i=1}^{n+1} \frac{\varepsilon}{2^i}.$$

Since $\bigvee_{i=1}^{n} a_i \leq b$ $(n = 1, 2, \ldots)$, then there exists $\bigvee_{i=1}^{\infty} a_i$ such that $x \leq \bigvee_{i=1}^{\infty} a_i$ and it holds:

$$\mu\Big(\bigvee_{i=1}^{\infty} a_i\Big) = \lim_{n\to\infty} \mu\Big(\bigvee_{i=1}^{n} a_i\Big).$$

Therefore

$$\mu^*(x) \leq \mu\Big(\bigvee_{i=1}^{\infty} a_i\Big) \leq \lim_{n\to\infty} \mu^*(x_n) + \varepsilon.$$

Since this inequality holds for each $\varepsilon > 0$, then it holds:

$$\mu^*(x) \leq \lim_{n\to\infty} \mu^*(x_n).$$

Therefore $\mu^*(x_n) \nearrow \mu^*(x)$.
This theorem is called Choquet lema.

4 Some notes about outer measure

Theorem 4.1 *Let $\mathcal{M}$ be a set of all μ^*-measurable elements. Then $\mathcal{M}$ is a σ-complete lattice.*

Proof. We have already proved that the set of all μ^*-measurable elements of M form a lattice (see *Theorem 2.3*). Now we will show that if $(x_n)_{n=1}^{\infty}$ is any sequence of μ^*-measurable elements, then also $x = \bigvee_{n=1}^{\infty} x_n$ and $x' = \bigwedge_{n=1}^{\infty} x_n$ are the μ^*-measurable elements. ($\bigwedge_{n=1}^{\infty} x_n$ means that $x_n \geq x_{n+1}$ and $x = \bigwedge_{n=1}^{\infty} x_n$.)

Put $y_n = \bigvee_{i=1}^{n} x_i$ for each $n \in N$. Then $y_n \leq y_{n+1}$ and $\bigvee_{n=1}^{\infty} y_n = \bigvee_{n=1}^{\infty} x_n = x$. Therefore $y_n \nearrow x$. An element y_n is μ^*-measurable element for each $n \in N$. Therefore for any $a \in M$ holds:

$$\mu^*(a) = \mu^*(a \wedge y_n) + \mu^*(a - a \wedge y_n).$$

Since $x \geq y_n$ then $a \wedge x \geq a \wedge y_n$ and $a - a \wedge y_n \geq a - a \wedge x$. Therefore

$$\mu^*(a - a \wedge y_n) \geq \mu^*(a - a \wedge x)$$

and

$$\mu^*(a) \geq \mu^*(a \wedge y_n) + \mu^*(a - a \wedge x).$$

Because $y_n \nearrow x$ then also $a \wedge y_n \nearrow a \wedge x$. When we use Choquet lema we get:

$$\lim_{n \to \infty} \mu^*(a \wedge y_n) = \mu^*(a \wedge x).$$

Therefore

$$\mu^*(a) \geq \mu^*(a \wedge x) + \mu^*(a - a \wedge x).$$

The opposite inequality also holds because outer measure μ^* is subadditive.

We proved that $x = \bigvee_{n=1}^{\infty} x_n$ is μ^*-measurable element. Similarly can be proved that element $x' = \bigwedge_{n=1}^{\infty} x_n$ is μ^*-measurable element. Therefore the set $\mathcal{M}$ is a σ-complete lattice.

Theorem 4.2 *$\mu^*|\mathcal{M}$ is an additive measure.*

Proof. For each $x \in \mathcal{M}$ and for each $a \in M$ holds:

$$\mu^*(a) = \mu^*(a \wedge x) + \mu^*(a - a \wedge x).$$

Take elements $x, y \in \mathcal{M}$ and put $a = x \oplus y$. If $x \odot y = 0$ then $x \oplus y = x + y$. Therefore

$$\mu^*(x+y) = \mu^*((x+y) \wedge x) + \mu^*((x+y) - (x+y) \wedge x) =$$

$$= \mu^*(x) + \mu^*((x+y) - x) = \mu^*(x) + \mu^*(y).$$

Let $x \oplus y \neq 0$. Then $x + y > u$ a $x \oplus y = u$. Therefore

$$\mu^*(x \oplus y) = \mu^*((x \oplus y) \wedge x) + \mu^*((x \oplus y) - (x \oplus y) \wedge x) =$$

$$= \mu^*(u \wedge x) + \mu^*(u - u \wedge x) = \mu^*(x) + \mu^*(u - x).$$

But $u - x < y$ so $\mu^*(u - x) < \mu^*(y)$.
Therefore

$$\mu^*(x \oplus y) \leq \mu^*(x) + \mu^*(y).$$

Acknowledgement

This paper was supported by Grant VEGA 1/2002/05.

References

[1] R. Cignoli, I.M.L. D'Ottaviano, D. Mundici, Foundations of Many - Valued Reasoning, Trends in Logic, Studia Logica Library, 7, Kluwer, Dordrecht, 2000.

[2] D.H. Fremlin, A direct proof of the Matthes - Wright integral extension theorem. J. London Math. Soc. 11, 276 - 284, 1975.

[3] B. Riečan, D. Mundici, Probability on MV-algebras, in: Handbook of Measure Theory, Elsevier, Amsterdam, 869–909, 2002.

[4] B. Riečan, T. Neubrunn, Integral, Measure, and Ordering. *Kluwer Academic Publishers, Dordrecht and Ister Science*, Bratislava, 1997.

A fixed point theorem in probabilistic metric spaces with a convex structure

Tatjana Žikić-Došenović
Faculty of Technology, University of Novi Sad
Bulevar Cara Lazara 1, 21000 Novi Sad, Serbia
tatjanad@tehnol.ns.ac.yu

Abstract

The inequality $F_{fx,fy}(qs) \geq F_{x,y}(s)$ $(s \geq 0)$, where $q \in (0, 1)$, is generalized for multivalued mappings in many directions. Using Hausdorff distance S.B. Nadler in [7] introduced a generalization of Banach contraction principle in metric spaces. In [3] the definition of probabilistic Nadler q-contraction is given. Using some results given in [12] a fixed point theorem on spaces with a convex structure is obtained. Some fixed point theorems in such spaces are proved in [1, 2].

Keywords: multivalued mappings, coincidence point, probabilistic metric space, Menger space, triangular norm, Menger space with a convex structure.

1 Introduction

K. Menger introduced in 1942. the notion of a probabilistic metric space as a natural generalization of the notion of a metric space (M, d) in which the distance $d(p, q)$ $(p, q \in M)$ between p and q is replaced by a distribution function $F_{p,q} \in \Delta^+$. $F_{p,q}(x)$ can be interpreted as the probability that the distance between p and q is less than x. Since then the theory of probabilistic metric spaces has been developed in many directions([9]).

Since Sehgal and Bharucha-Reid proved ([10]) a fixed point theorem in Menger spaces $(S, \mathcal{F}, T_M)$ many fixed point theorem for singlevalued and multivalued mappings on Menger spaces $(S, \mathcal{F}, T)$ are obtained.

Further development of the fixed point theory in a more general Menger space $(S, \mathcal{F}, T)$ was connected with investigations of the structure of the t-norm T. This problem was very soon in some sence completely solved. V. Radu proved that if $f : S \to S$ is a probabilistic q-contraction, where $(S, \mathcal{F}, T)$ is a complete Menger space and T is a continuous t-norm, f has a fixed point if and only if T is of the H-type [8].

S.B. Nadler proved in [7] the generalization of the Banach contraction principle for multivalied mappings $f : X \to CB(X)$, where (X, d) is a metric space, by introducing the condition

$$D(fx, fy) \leq qd(x, y),$$

where D is the Hausdorff metric and $q \in (0, 1)$.

Probabilistic version of Nadler's q-contraction is given in [3].

Definition 1 *Let $(S, \mathcal{F})$ be a probabilistic metric space, M a nonempty subset of S and $f : M \to 2^S$, where 2^S is the family of all nonempty subsets of S. The mapping f is said to be a probabilistic Nadler's q-contraction, where $q \in (0, 1)$ if the following condition is satisfied:*

For every $u, v \in M$, every $x \in fu$ and every $\delta > 0$ there exists $y \in fv$ such that for every $\varepsilon > 0$

$$F_{x,y}(\varepsilon) \geq F_{u,v}(\frac{\varepsilon - \delta}{q}).$$

If f is a singlevalued mapping, then the notion of a probabilistic Nadler's q-contraction coincides with the notion of a probabilistic q-contraction by Sehgal and Bharucha-Reid [10], since the function $F_{u,v}(\cdot)$ is left-continuous.

2 Preliminaries

Let Δ^+ be the set of all distribution functions F such that $F(0) = 0$ (F is a nondecreasing, left continuous

mapping from $\mathbb{R}$ into $[0,1]$ such that $\sup_{x\in\mathbb{R}} F(x) = 1$).

The ordered pair $(S,\mathcal{F})$ is said to be **a probabilistic metric space** if S is a nonempty set and $\mathcal{F}: S\times S \to \Delta^+$ ($\mathcal{F}(p,q)$ written by $F_{p,q}$ for every $(p,q)\in S\times S$) satisfies the following conditions:

1. $F_{u,v}(x) = 1$ for every $x > 0 \Rightarrow u = v$ $(u,v \in S)$.

2. $F_{u,v} = F_{v,u}$ for every $u, v \in S$.

3. $F_{u,v}(x) = 1$ and $F_{v,w}(y) = 1 \Rightarrow F_{u,w}(x+y) = 1$ for $u,v,w \in S$ and $x,y \in \mathbb{R}^+$.

A **Menger space** (see [9]) is a ordered triple $(S,\mathcal{F},T)$, where $(S,\mathcal{F})$ is a probabilistic metric space, T is a triangular norm (abbreviated t -norm) and the following inequality holds

$$F_{u,v}(x+y) \geq T(F_{u,w}(x), F_{w,v}(y))$$

for every $u,v,w \in S$ and every $x > 0, y > 0$.

Recall that a mapping $T : [0,1]\times[0,1] \to [0,1]$ is called a **triangular norm** (a t-norm) if the following conditions are satisfied:
$T(a,1) = a$ for every $a \in [0,1]$; $T(a,b) = T(b,a)$ for every $a,b \in [0,1]$;

$$a \geq b,\ c \geq d \Rightarrow \quad T(a,c) \geq T(b,d) \quad (a,b,c,d \in [0,1]);$$

$$T(a,T(b,c)) = T(T(a,b),c) \quad (a,b,c \in [0,1]).$$

Example 1 The following are the four basic t-norms :

(i) The **minimum** t-norm, T_M, is defined by

$$T_M(x,y) = \min(x,y),$$

(ii) The **product** t-norm, T_P, is defined by

$$T_P(x,y) = x\cdot y,$$

(iii) The **Lukasiewicz** t-norm T_L is defined by

$$T_L(x,y) = \max(x+y-1, 0),$$

(iv) The weakest t-norm, the **drastic product** T_D, is defined by

$$T_D(x,y) = \begin{cases} \min(x,y) & \text{if } \max(x,y) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

As regards the pointwise ordering, we have the inequalities

$$T_D < T_L < T_P < T_M.$$

The $(\epsilon,\lambda)-$ topology in S is introduced by the family of neighbourhoods $\mathcal{U} = \{U_v(\epsilon,\lambda)\}_{(v,\epsilon,\lambda)\in S\times\mathbb{R}_+\times(0,1)}$, where

$$U_v(\epsilon,\lambda) = \{u;\ F_{u,v}(\epsilon) > 1-\lambda\}.$$

If a t-norm T is such that $\sup_{x<1} T(x,x) = 1$, then S is in the (ϵ,λ) topology a metrizable topological space.

Each t-norm T can be extended (see [6]) (by associativity) in a unique way to an n-ary operation taking for $(x_1,\ldots,x_n) \in [0,1]^n$ the values

$$\mathbf{T}^0_{i=1}x_i = 1,\ \mathbf{T}^n_{i=1}x_i = T(\mathbf{T}^{n-1}_{i=1}x_i, x_n).$$

We can extend T to a countable infinitary operation taking for any sequence $(x_n)_{n\in\mathbb{N}}$ from $[0,1]$ the values

$$\mathbf{T}^\infty_{i=1}x_i = \lim_{n\to\infty} \mathbf{T}^n_{i=1}x_i.$$

Limit of right side exists since the sequence $(\mathbf{T}^n_{i=1}x_i)_{n\in\mathbb{N}}$ is nonincreasing and bounded from below.

In the fixed point theory it is of interest to investigate the classes of t-norms T and sequences $(x_n)_{n\in\mathbb{N}}$ from the interval [0, 1] such that $\lim_{n\to\infty} x_n = 1$, and

$$\lim_{n\to\infty} \mathbf{T}^\infty_{i=n}x_i = \mathbf{T}^\infty_{i=1}x_{n+i} = 1.$$

In [5] it is proved that for the Dombi, Aczél-Alsina and Sugeno-Weber family of t-norms exists the sequence $(x_n)_{n\in\mathbb{N}}$ from (0, 1] such that the last condition is satisfied.

Definition 2 : *Let $(S,\mathcal{F},T)$ be a Menger space, $\emptyset \neq M \subset S$, $f: M\to M$ i $A: M \to 2^M$ (the family of nonempty subsets of M). The mapping A is f-strongly demicompact if for every sequence $\{x_n\}_{n\in\mathbb{N}}$ from M, such that $\lim_{n\to\infty} F_{fx_n,y_n}(\epsilon) = 1$, for some sequence $\{y_n\}_{n\in\mathbb{N}}$, $y_n \in Ax_n$, $n\in\mathbb{N}$ and every $\epsilon > 0$, there exists a convergent subsequence $\{fx_{n_k}\}_{k\in\mathbb{N}}$.*

A mapping $A: M \to 2^M$ is weakly commuting with $f: M\to M$ if for every $x \in M$

$$f(Ax) \subset A(fx).$$

In [12] the following theorem is proved.

Theorem 1 : *Let $(S,\mathcal{F},T)$ be a complete Menger space, $\sup_{a<1} T(a,a) = 1$, A is a nonempty and closed subset of S, $f: A\to A$ a continuous mapping, $L, L_1 : A \to 2^{f(A)}_c$ closed, multivalued mappings such*

that the following condition is satisfied:

For every $u, v \in A$, $x \in Lu$ *and* $\delta > 0$, *there exists* $y \in L_1 v$ *such that*

$$F_{x,y}(\varepsilon) \geq F_{fu,fv}(\frac{\varepsilon - \delta}{q}), \text{ for all } \varepsilon > 0, \text{ where } q \in (0, 1).$$

If L *and* L_1 *are weakly commuting with* f *and (i) or (ii) are satisfied, then there exists* $x \in A$ *such that* $fx \in Lx \cap L_1 x$ *where*

(i) L *or* L_1 *is* f*-strongly demicompact.*

(ii) There exists $x_0, x_1 \in A$, $fx_1 \in Lx_0$ *and* $\mu \in (0, 1)$ *such that t-norm* T *satisfies the following condition*

$$\lim_{n\to\infty} \mathbf{T}_{i=n}^{\infty} F_{fx_0,fx_1}(\frac{1}{\mu^i}) = 1.$$

3 Main results

W. Takahashi introduced in [11] the notion of a metric space with a convex structure. This class of metric spaces includes normed linear spaces and metric spaces of the hiperbolic type.

Let us recall that a metric space (S, d) has a convex structure in the sense of Takahashi, if there exists a mapping $W : S \times S \times [0, 1] \to S$ such that for every $(u, x, y, \delta) \in S \times S \times S \times [0, 1]$

$$d(u, W(x, y, \delta)) \leq \delta d(u, x) + (1 - \delta) d(u, y).$$

This definition can be generalized in Menger spaces $(S, \mathcal{F}, T)$. The notion of convex structure in probabilistic metric spaces, as well as Definition 3 and Definition 4 belong to O. Hadžić [4].

A mapping $W : S \times S \times [0, 1] \to S$ is said to be a convex structure on S if for every $(x, y) \in S \times S$

$W(x, y, 0) = y$, $W(x, y, 1) = x$, and for every $\delta \in (0, 1)$, $u \in S$, $\varepsilon > 0$

$$F_{u,W(x,y,\delta)}(2\varepsilon) \geq T(F_{u,x}(\frac{\varepsilon}{\delta}), F_{u,y}(\frac{\varepsilon}{1-\delta})).$$

It is easy to see that every metric space (S, d) with a convex structure W can be considered as a Menger space $(S, \mathcal{F}, T_M)$ with the same function W.

It is well-known that every probabilistic normed space is a probabilistic metric space with a convex structure $W(x, y, \delta) = \delta x + (1 - \delta)y$ $(x, y \in S)$ since for every $\varepsilon > 0$ and $\delta \in [0, 1]$ we have

$$\begin{aligned} F_{u,W(x,y,\delta)} &= F_{u-\delta x-(1-\delta y)}(2\varepsilon) \\ &= F_{\delta u+(1-\delta u)-\delta x-(1-\delta)y}(2\varepsilon) \\ &= F_{\delta(u-x)+(1-\delta)(u-y)}(2\varepsilon) \\ &\geq T(F_{\delta(u-x)(\varepsilon}, F_{(1-\delta)(u-y)}(\varepsilon)) \\ &= T(F_{u-x}(\frac{\varepsilon}{\delta}), F_{u-y}(\frac{\varepsilon}{1-\delta}))). \end{aligned}$$

In this paper we shall suppose that a convex structure W on a Menger space $(S, \mathcal{F}, T)$ satisfies the condition

$$F_{W(x,z,\delta),W(y,z,\delta)}(\varepsilon\delta) \geq F_{x,y}(\varepsilon) \tag{1}$$

for every $(x, y, z) \in S \times S \times S$, every $\varepsilon > 0$ and $\delta \in (0, 1)$.

For the next theorem we shall need the following definitions.

Definition 3 *Let* $(S, \mathcal{F}, T)$ *be Menger space with a convex structure* W. *A nonempty subset* M *of* S *is called* W*-starshaped if there exists* $x_0 \in M$ *such that the set* $\{W(x, x_0, \lambda) : x \in M, \lambda \in (0, 1)\} \subset M$. *The point* x_0 *is said to be the star-centre of* M.

Clearly, every convex set is a starshaped set and the inverse is not true.

Definition 4 *Let* $(S, \mathcal{F}, T)$ *be a Menger space with a convex structure* W *and* M *a nonempty subset of* S. *A mapping* $f : M \to S$ *is said to be* (W, x_0)*-convex if for each* $(x, \lambda) \in M \times [0, 1]$

$$W(fx, x_0, \lambda) = f(W(x, x_0, \lambda)).$$

Lemma 1 *Let* $(S, \mathcal{F}, T)$ *be a Menger space,* M *a nonempty subset of* S *which is* W*-starshaped with the star-centre in* x_0, $f : M \to S$ *is the mapping which is* (W, x_0)*-convex. Then the* $f(M)$ *is the* (W, x_0)*-convex.*

Proof: Let $u \in f(M)$. Then there exists $x \in M$ so that $u = f(x)$. Let us prove that for every $\lambda \in [0, 1]$,

$$z = W(u, x_0, \lambda) \in f(M).$$

As $u = f(x)$ it follows

$$z = W(f(x), x_0, \lambda).$$

Since f is (W, x_0)-convex it follows

$$z = f(W(x, x_0, \lambda)).$$

From the condition we have that the set M is W-starshaped, and it follows that $W(x, x_0, \lambda) \subset M$, i.e. $z \in f(M)$.

Lemma 2 *Let $(S, \mathcal{F}, T)$ be a Menger space, T a t-norm such that $\sup_{x<1} T(x, x) = 1$, M a nonempty subset of S, $f : M \to S$ a continuous mapping, $L : M \to C(S)$ (where $C(S)$ is the family of all nonempty and closed subset of S) and the following inequality is satisfied:*

For every $u, v \in M$, every $x \in Lu$ and every $\delta > 0$, there exists $y \in Lv$ such that

$$F_{x,y}(\varepsilon) \geq F_{fu,fv}(\frac{\varepsilon - \delta}{q}) \tag{2}$$

for every $\varepsilon > 0$. Then the mapping L is closed.

Proof:

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence from M such that $\lim_{n\to\infty} x_n = x$ and let $y_n \in Lx_n$, for every $n \in \mathbb{N}$ such that $\lim_{n\to\infty} y_n = y$. Let us prove that $y \in Lx$.
Since Lx is closed we shall prove that $y \in \overline{L}x$. Let $\varepsilon > 0$ and $\lambda \in (0, 1)$ be given. It remains to be proved that there exists $b \in Lx$ such that $b \in U_y(\varepsilon, \lambda)$, i.e. $F_{b,y}(\varepsilon) > 1 - \lambda$. Using condition 2, where $u = x_n$, $v = x$ and $\delta = \frac{3\varepsilon}{4}$ from $y_n \in Lx_n$ it follows that there exists $b_n \in Lx$ such that

$$F_{y_n,b_n}(\varepsilon) \geq F_{fx_n,fx}(\frac{\varepsilon}{4q}).$$

So,

$$\begin{aligned} F_{y,b_n}(\varepsilon) &\geq T(F_{y,y_n}(\frac{\varepsilon}{2}), F_{y_n,b_n}(\frac{\varepsilon}{2})) \\ &\geq T(F_{y,y_n}(\frac{\varepsilon}{2}), F_{fx_n,fx}(\frac{\varepsilon}{4q})). \end{aligned}$$

From the continuity of the mapping f it follows $\lim_{n\to\infty} fx_n = fx$, i.e.

$\lim_{n\to\infty} F_{fx_n,fx}(\varepsilon) = 1$. We also have $\lim_{n\to\infty} F_{y,y_n}(\varepsilon) = 1$, for every $\varepsilon > 0$. From the condition $\sup_{x<1} T(x, x) = 1$, it follows that for a given $\lambda \in (0, 1)$ there exists $\eta \in (0, 1)$ such that $T(\eta, \eta) > 1 - \lambda$, so there exists $n_0(\eta)$ such that for every $n \geq n_0(\eta)$

$$F_{y,y_n}(\tfrac{\varepsilon}{2}) > \eta \text{ i } F_{fx_n,fx}(\tfrac{\varepsilon}{4q}) > \eta.$$

Hence,

$$\begin{aligned} F_{y,b_n}(\varepsilon) &\geq T(\eta, \eta) \\ &> 1 - \lambda, \end{aligned}$$

i.e. $b_{n_0} \in U_y(\varepsilon, \lambda) \cap Lx$.

Theorem 2 *Let $(S, \mathcal{F}, T)$ be a complete Menger space with a convex structure W and a continuous t-norm T and M is nonempty, closed and W-starshaped subset of S with the star-centre x_0. Let $f : M \to M$ be a continuous , (W, x_0)-convex mapping and $L : M \to 2_c^{f(M)}$ such that $\overline{L(M)}$ is compact set and the following condition is satisfied:*

For every $u, v \in M$, $x \in Lu$ and $\delta > 0$ there exists $y \in Lv$ such that

$$F_{x,y}(\varepsilon) \geq F_{fu,fv}(\varepsilon - \delta), \text{ for every} \varepsilon > 0. \tag{3}$$

If L is weakly comuting with f, then there exists $x \in M$ such that $fx \in Lx$.

Proof: Let $(k_n)_{n\in\mathbb{N}}$ be a sequence from the interval $(0, 1)$ such that $\lim_{n\to\infty} k_n = 1$. For every $n \in \mathbb{N}$, and every $x \in M$ let $L_n x = W(Lx, x_0, k_n)$ i.e.

$$L_n x = \bigcup_{z\in Lx} W(z, x_0, k_n), \quad n \in \mathbb{N}, \quad x \in M.$$

From the Lemma 1 it follows that $f(M)$ is (W, x_0)-convex. We shall prove that $L_n x \subset f(M)$, i.e. that for every $z \in L_n x$ it follows $z \in f(M)$. Since $z \in L_n x = W(Lx, x_0, k_n)$, there exists $u \in Lx$ such that $z = W(u, x_0, k_n)$. Since $Lx \subset f(M)$ it follows that $u \in f(M)$, so $W(u, x_0, k_n) \subset f(M)$ i.e. $z \in f(M)$. It means that $L_n x \subset f(M)$.
From the condition (1) it follows that the mapping W

is continuous in respect to the first variable, and since Lx is closed it follows that Lx is compact (as a subset of $L(M)$) such that $W(Lx, x_0, k_n)$ is closed for every $n \in \mathbb{N}$. It follows that $L_n x$ is closed for every $n \in \mathbb{N}$ and $x \in M$.
Next, we shall prove that for every $u, v \in M$ and every $x \in L_n u$ and $\delta > 0$ there exists $y \in L_n v$ such that

$$F_{x,y}(\varepsilon) \geq F_{fu,fv}(\frac{\varepsilon - \delta}{k_n}), \qquad \varepsilon > 0.$$

Let $u, v \in M$, $\delta > 0$ and $x \in L_n u = W(Lu, x_0, k_n)$. Then, there exists $z \in Lu$ such that $x = W(z, x_0, k_n)$ and from (3) there exists $y' \in Lv$ such that

$$F_{z,y'}(\varepsilon) \geq F_{fu,fv}(\varepsilon - \frac{\delta}{k_n}).$$

Let $y = W(y', x_0, k_n) \in L_n v$. Then

$$\begin{aligned} F_{x,y}(\varepsilon) &= F_{W(z,x_0,k_n),W(y',x_0,k_n)}(\frac{\varepsilon}{k_n}k_n) \\ &\geq F_{z,y'}(\frac{\epsilon}{k_n}) \\ &\geq F_{fu,fv}(\frac{\varepsilon - \delta}{k_n}). \end{aligned}$$

Let us prove that L_m is a f-strongly demicompact. Suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence in M such that for every $\varepsilon > 0$

$$\lim_{n\to\infty} F_{fx_n,y_n}(\varepsilon) = 1$$

for some sequence $(y_n)_{n\in\mathbb{N}}$ $y_n \in L_m x_n$.
Since $L_n(M) = W(L(M), x_0, k_n)$, $n \in \mathbb{N}$ is relatively compact, from $y_n \in L_m x_n$ it follows that $(y_n)_{n\in\mathbb{N}}$ has a convergent subsequence $(y_{n_k})_{k\in\mathbb{N}}$ and let $\lim_{k\to\infty} y_{n_k} = z$. Then we have

$$F_{fx_{n_k},z}(\varepsilon) \geq T(F_{fx_{n_k},y_{n_k}}(\frac{\varepsilon}{2}), F_{y_{n_k},z}(\frac{\varepsilon}{2}))$$

i.e. $\lim_{k\to\infty} fx_{n_k} = z$, which means that the mapping L_m is f-strongly demicompact.
We are going to prove that the mapping L_n is weakly comuting with f, i.e. that for every $x \in M$

$$f(L_n x) \subset L_n(fx) = W(L(fx), x_0, k_n).$$

Let $u \in L_n x = W(Lx, x_0, k_n)$. Then there exists $z \in Lx$ such that $u = W(z, x_0, k_n)$, so

$$\begin{aligned} f(u) &= f(W(z, x_0, k_n)) \\ &= W(fz, x_0, k_n) \\ &\subset W(f(Lx), x_0, k_n). \end{aligned}$$

From the condition that the mapping L is weakly commuting with f it follows $W(f(Lx), x_0, k_n) = W(L(fx), x_0, k_n)$, i.e. $f(u) \in L_n(fx)$.
Hence, all the conditions of the Theorem 1 are satisfied and according to it, for every $n \in \mathbb{N}$, there exists $x_n \in M$ such that

$$fx_n \in L_n x_n.$$

Since $L_n x_n = \bigcup_{z\in Lx_n} W(z, x_0, k_n)$ there exists $z_n \in Lx_n$ such that $fx_n = W(z_n, x_0, k_n)$. Then, for every $n \in \mathbb{N}$

$$\begin{aligned} F_{fx_n,z_n}(\varepsilon) &= F_{z_n,W(z_n,x_0,k_n)}(\varepsilon) \\ &\geq T(F_{z_n,z_n}(\frac{\varepsilon}{2k_n}), F_{z_n,x_0}(\frac{\varepsilon}{2(1-k_n)})) \\ &= T(1, F_{z_n,x_0}(\frac{\varepsilon}{2(1-k_n)})) \\ &= F_{z_n,x_0}(\frac{\varepsilon}{2(1-k_n)}). \end{aligned}$$

Since $\lim_{n\to\infty} \frac{\varepsilon}{2(1-k_n)} = \infty$, it follows that $\lim_{n\to\infty} F_{z_n,x_0}(\frac{\varepsilon}{2(1-k_n)}) = 1$ because $z_n \in L(M)$ which is probabilistically bounded. Then we have $\lim_{n\to\infty} F_{fx_n,z_n}(\varepsilon) = 1$. Since $z_n \in Lx_n$ and $\overline{L(M)}$ is compact it follows that there exists convergent subsequence $(z_{n_k})_{k\in\mathbb{N}}$ of the sequence $(z_n)_{n\in\mathbb{N}}$. From $\lim_{n\to\infty} fx_{n_k} = z$ it follows $\lim_{n\to\infty} z_{n_k} = z$.
It remains to be proved that $fz \in Lz$. From $\lim_{n\to\infty} z_{n_k} = z$ and from the continuity of the mapping f it follows $\lim_{n\to\infty} fz_{n_k} = fz$. Since $z_{n_k} \in Lx_{n_k}$ it follows that

$$fz_{n_k} \in f(Lx_{n_k}) \subset L(fx_{n_k}).$$

From the Lemma 2 it follows that the mapping L is closed i.e.

$$fz \in Lz.$$

Acknowledgement

Research supported by MNTRRS-144012.

References

[1] Shih-sen Chang, Yeol Je Cho, Shin Min Kang, Probabilistic metric spaces and nonlinear operator theory, Sichuan University Press, 1994.

[2] O. Hadžić, A fixed point theorem in Menger spaces, Publ. Inst. Math. Beograd T 20 (1979), 107–112.

[3] O. Hadžić, Fixed point theorems for multivalued mappings in probabilistic metric spaces, Mat. Vesnik 3 (16)(31), (1979), 125–133.

[4] O. Hadžić, Fixed point theory in probabilistic metric spaces, Serbian academy of science and arts branch in Novi Sad, University of Novi Sad, Institute of Mathematics, 1995.

[5] O. Hadžić, E. Pap, Fixed Point Theory in Probabilistic Metric Spaces, Theory in Probabilistic Metric Spaces, Kluwer Academic Publishers, Dordrecht, 2001.

[6] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publishers, Trends in Logic 8, Dordrecht (2000a).

[7] S.B. Nadler, Multivalued contraction mappings, Pacific J. Math. 30 (1969), 475–478.

[8] V. Radu, On the contraction principle in Menger spaces, Seminarul the theoria probabilitatilor siaplicatii, University of Timisoara, Faculty of Natural Science, 1983.

[9] B. Schweizer, A. Sklar, Probabilistic metric spaces, Elsevier North-Holland, New York, 1983.

[10] V.M. Sehgal, A.T. Baharucha-Reid Fixed points of contraction mappings on probabilistic metric spaces, Math. Syst. Theory 6 (1972), 97–102.

[11] W. Takahashi, A convexity in metric spaces and nonexpansive mappings I, Kodai Math. rep. 22 (1970), 142–149.

[12] T. Žikić, Multivalued probabilistic q-contraction, Journal of Electrical Engineering, Vol. 53 No.12/s, (2002), 13–16.

Measuring the similarity of different types of fuzzy sets in FRDB

Aleksandar Takači
Faculty of Technology
University of Novi Sad, Serbia
stakaci@tehnol.ns.ac.yu

Srdjan Škrbić
Faculty of Natural Sciences
University of Novi Sad
skhrba@uns.ns.ac.yu

Abstract

The representation of imprecise, uncertain or inconsistent information is not possible in relational databases, thus they require add-onns to handle these types of information. One possible add-on is to allow the attributes to have values that are fuzzy sets on the attribute domain, which result in fuzzy relational databases (FRDB). From the implementational point of view, values are limited to certain types of fuzzy sets, most often trapezoidal. In this paper we measure the similarity of fuzzy sets when they are values of fuzzy attributes in FRDB. We give a similarity fuzzy relation suitable for this task, which is easy to calculate and implement. The ordering $\preceq_I$ over the set of all fuzzy subsets of a universe ($\mathcal{F}(x)$) which is a generalization of the classical ordering $\leq$ is introduced. This ordering can be used to compare fuzzy sets and it is useful in query implementation. Ordering $\preceq_I$ together with the compatibility fuzzy relation is used to define another fuzzy relation FLQ - fuzzy less or equal.

Keywords: similarity, fuzzy, relation, database.

1 Introduction

Relational databases (RDB) have been developed over the years. The representation of imprecise, uncertain or inconsistent information is not possible in RDB, thus they require add-onns to handle these types of information. One possible add-on is to allow the attributes to have values that are fuzzy sets on the attribute domain, which result in fuzzy relational databases (FRDB). From the implementational point of view, values are limited to certain types of fuzzy sets, most often trapezoidal. In our model that has been developed at the University of Novi Sad (in which the author of this paper has a significant role) we opted for interval values, triangular fuzzy numbers, fuzzy quantities and linguistic labels. Fuzzy values of attributes are not incorporated into existing database management systems, meaning that the database management system should be done from scratch. This is huge task and most often programmers build on existing RDB's which was the case of the system developed by the authors. FRDB usually have their own query language - fuzzy structured query language (FSQL). For more details on FRDB see [3, 2]. In classical database environments the result of a query is a relation over a database. In FRDB the result of a query is a fuzzy relation i.e. each row in the database has its satisfaction degree which is a value form the unit interval. The satisfaction degree is calculated based on the similarity of attribute values in the query and in the database. The focus of the paper is to find a general similarity fuzzy relation between attribute values when they appear in queries and in the database. We will show that a common fuzzy relation FEQ is suitable for this task and can work on any type of fuzzy sets. Moreover, the implementation of this relation is possible.

2 FRDB and FSQL

The relational model uses a collection of tables to represent data and relationships among those data. In our model, data values need not be exact. We can handle imprecise and uncertain information using interval values, fuzzy numbers and quantities. For more details see [5, 6]. Also, each attribute can have linguistic labels as their values. In order for a linguistic label to be used it needs to be predefined by the database administrator. This information is stored in a fuzzy meta knowledge base, a crucial part of FRBD. An example of a table from FRDB is given in Table 1. The value $tri(170, 10, 10)$ represents a linear trian-

gular fuzzy with a center value at 170, and left and right tolerance 10. Values "small", "average", "tall" and "high" represent linguistic labels.

Table 1: An example of a Table in a FRDB

Name	Height	Salary
Steve	tri(170,10,10,lin)	30000
Amy	average	tri(45000,1000,1000,lin)
Jack	[188,190]	small
Jill	tall	high

SQL is the most common commercially marketed database query language. It uses a combination of relational algebra and relational calculus constructs to retrieve desired data from a database. FSQL is SQL that can handle fuzzy attribute values. The main difference between SQL and FSQL is that SQL returns a subset of the database as the query result. On the other hand, FSQL returns a satisfaction degree which is a number from the unit interval. When attributes with fuzzy values appear in the query, it is transformed into a query that can be handled by SQL and finally results obtained from the SQL query are then post processed in order to obtain the desired information. The exact syntax and more information about FSQL used here can be found in [5, 6]. Let us mention that other versions of FSQL exist, the most often used is the one found in [3].

The processing of the FSQL query consists of the following fazes:

1. Checking the query for syntax and semantic errors
2. Query transformation
3. Processing the results of the transformed query

First, the actual query text is checked, because it must follow the FSQL query syntax. Semantically, we must check if fuzzy attribute values are assigned only to attributes that allow them and also if there are any linguistic labels, whether the correct ones are used for particular attribute values. Then, the query is transformed into a query which can be processed by classical SQL. The dataset resulting from the transformed query is post processed in order to obtain a satisfaction degree for each line in the dataset.

The processing of the data set consists of two fazes. First, each attribute value from the database need to be compared with its counterpart in the query. In the following section this will be discussed in detail. The second faze is to calculate the satisfaction degree for each line in the dataset. This is done by using fuzzy logic and adequate generalizations of the conjunction, disjunction and negation operator (see [4, 7]).

3 Query database similarity fuzzy relation

Classical RDB data types are well known and the relations among their elements have been known for years in mathematics, linguistics etc. The implementations of the most common relations $(=, >, <)$ has been done simultaneously with the creation of the first computer. However, in FRDB the relation between fuzzy subsets over a particular domain is more complex which makes its implementation more difficult. Also, allowing various data types for an attribute over a particular domain (crisp, interval, fuzzy subset, linguistic label) creates the necessity to find a similarity fuzzy relation between each pair of data types. In this section we will give some pointers how to create a similarity fuzzy relation between various data types and also suggest a well known fuzzy relation of fuzzy inclusion to be used for as the base for the query db similarity relation. Moreover, we will sketch how to implement this relation in order for it to work on general fuzzy sets.

First, let us review the simplest case of a crisp database attribute value. If the attribute value in the query is also crisp we have the same situation as in classical SQL which makes the comparison trivial. If the attribute value in the query is an interval we need to check whether that value from the dataset belongs to it and return the appropriate value. In case of fuzzy set attribute value the membership function value is returned.

Now, we observe the case of fuzzy set database values and also fuzzy set query values. The object is to find a fuzzy relation $R(db, q)$ where db and q are general fuzzy sets. One of the properties that relation should fulfill is reflexivity i.e. $(\forall x)\ R(x, x) = 1$. The relation that we suggest as the base for the similarity of fuzzy sets is fuzzy inclusion.

Definition 1. For arbitrary $A, B \in \mathcal{F}(x)$ $FINCL(A, B)$ is calculated in the following way.

$FINCL(A, B) = \frac{card(A \cap B)}{card(A)}$.

where $A \cap B$ is the intersection of fuzzy sets A and B, and $card(S)$ is the cardinality of a fuzzy set S.

This relation can be calculated for any fuzzy set. From the implementational point of view all it takes is to implement an algorithm for calculating the intersection of fuzzy sets and the algorithm for calculating the integral of a positive curve. This can lead to computational complexity problems which can easily be solved for particular types of fuzzy sets (triangular, trape-

zoidal, Gaussian).

When we recall the well known relations $\subseteq, =$, the following statement holds for any two sets A, B:

$$\text{if } A \subseteq B \wedge B \subseteq A \text{ then} A = B. \tag{1}$$

If A, B are fuzzy sets, then we use statement (1) to define the fuzzy relation FEQ.

Definition 2. For arbitrary $A, B \in \mathcal{F}(x)$ the similarity relation $FEQ(A, B)$.

$$FEQ(A, B) = T(FINCL(A, B), FINCL(B, A)).$$

where t is a t-norm. If we take $T(x, y) = \min(x, y)$, then:

$$FEQ(A, B) = \frac{card(A \cap B)}{\max(card(A), card(B))}.$$

In Table 2 some results for $FEQ(A, B)$ are given.

Table 2: Similarity between triangular fuzzy sets.

A	**B**	**FEQ(A, B)**
tri(170,10,10,lin)	tri(170,5,5,lin)	0.375
tri(170,5,5,lin)	tri(175,5,5,lin)	0.25
tri(170,5,5,lin)	tri(200,50,50,lin)	0.0635
tri(170,10,10)	tri(150,10,10)	0

Linguistic labels are most often defined as fuzzy sets. We should consider their meaning in order to handle them properly. Our opinion is that when a attribute value in the database is a linguistic label, it should be viewed as possibility distributions. This is not a problem in the case when the support of the set that represents the linguistic label is finite. On the other hand, when there are sets with infinite support the membership function cannot accurately represent the possibility distribution of the actual attribute value, as it is the case for fuzzy quantities. For example, a person that has that has the linguistic value "tall" for the height attribute is not likely to be 235cm tall, but $\mu_{tall}(235) = 1$. This yields that the fuzzy quantity that represents "tall" does not reflect the possibility distribution of the linguistic label. Let $fq(a, b, dec, lin)$ be a linear increasing fuzzy quantity with a membership function is a straight line that connects the points $(-\infty, 1), (a, 1), (b, 0), (0, \infty)$ and similarly let $fq(a, b, inc, lin)$ be a linear decreasing fuzzy quantity. The possibility distribution is found as a simple transformation of parameters:

$$Poss(fq(a, b, inc, lin)) \rightarrow tri(a, b, a + 2 * (b - a), lin),$$

$$Poss(fq(a, b, dec, lin)) \rightarrow tri(a - 2 * (b - a), a, b, lin).$$

where a triangular fuzzy number represents the possibility distribution of a fuzzy quantity. On the other hand, when we have linguistic labels as query data values they should not be transformed since they accurately reflect the essence of the query.

Finally, intervals can be viewed as special possibility distributions-fuzzy sets

$$\pi([a, b]) = \left\{ \begin{array}{ll} 1, & x \in [a, b], \\ 0, & otherwise. \end{array} \right\},$$

which makes calculations with them easy (see Figure 1.).

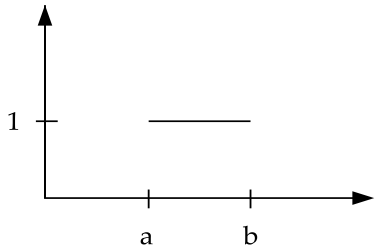


Figure 1. Intervals as possibility distributions.

In Table 3 some examples of compatibility between intervals and fuzzy numbers are given.

Table 3: Similarity between intervals and triangular fuzzy sets.

A	**B**	**FEQ(A, B)**
[10,20]	tri(15,10,10,lin)	0,75
[10,20]	tri(15,15,15,lin)	0,556
tri(15,10,10,lin)	[10,20]	0,75
tri(15,15,15,lin)	[10,20]	0,444

In the previous section we have presented a general similarity relation for any type of fuzzy set, relation FEQ (see Definition 2.). The introduction of a new fuzzy set type is only the question of imagination and patience of the database administrator. We have already implemented triangular fuzzy numbers, fuzzy quantities and intervals. In our future work we hope to implement fuzzy sets that allow inconsistency in data values i.e. multi–trapezoidal fuzzy sets (see Figure 2.).

4 Generalization and fuzzyfication of the operator $\leq$

In order to make possible conditions like $height \leq triangle(180, 5, 5)$, an ordering needs to be introduced

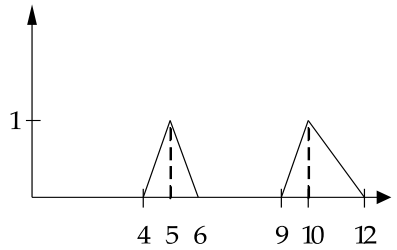


Figure 2. Inconsistent data values

on the set of all fuzzy subset of a universe (attribute domain) denoted $\mathcal{F}(X)$.Orderings on $\mathcal{F}(X)$ have been studied over the years. The most common one is the generalization of the well known ordering $\leq$. In case of intervals the generalization is in the following form. For two intervals [a1,b1] and [a2,b2] we have:

$[a1, b1] \preceq [a2, b2] \;\; iff \; a1 \leq a2 \; and \; b1 \leq b2.$

Now, let us expand the ordering on fuzzy sets.

Definition 3. Let $\preceq$ be an ordering on X and let $A \in \mathcal{F}(x)$. A fuzzy superset of A , denoted by $LTR(A)$ is defined as:

$$\mu_{LTR(A)}(x) = \sup\{\mu_A(y) | y \preceq x\}.$$

$LTR(A)$ is actually the smallest fuzzy superset of A with a non-decreasing membership function. Analogously, $RTL(A)$ is defined as:

$$\mu_{RTL(A)}(x) = \sup\{\mu_A(y) | x \preceq y\}.$$

Likewise, $RTL(A)$ is the smallest fuzzy superset of A with a non-increasing membership function.

Definition 4. If $A, B \in \mathcal{F}(X)$ then we define an ordering $\preceq_I$ **[1]** on $\mathcal{F}(X)$:

$A \preceq_I B$ iff $LTR(A) \supseteq LTR(B) \wedge RTL(A) \subseteq RTL(B)$.

For more details on orderings on the set $\mathcal{F}(X)$ see [1]. It can easily be seen that two fuzzy sets with different heights cannot be compared. Thus, a new ordering $\preceq_I''$ is defined. First, a new set $\lceil A \rceil$ is defined:

$$\mu_{\lceil A \rceil}(x) = \begin{cases} 1 & \mu_A(x) = \text{ height}(A), \\ \mu_A(x) & \text{otherwise.} \end{cases}$$

Definition 5. For arbitrary $A, B \in \mathcal{F}(x)$, a new ordering which can be applied to larger number of fuzzy sets is proposed :

$$A \preceq_I'' B \text{ iff } \lceil A \rceil \preceq_I \lceil B \rceil$$

The ordering $\preceq_I$ ranks fuzzy numbers depending on their horizontal position on the graph. The more the membership function is to the right on the graph the larger the fuzzy number is. This is not a total ordering, incomparable fuzzy sets can be seen on Figure 3. For triangular fuzzy numbers this can be done very easily. For example:

$LTR(tri(180, 10, 10, lin)) = fq(170, 180, inc, lin),$

$RTL(tri(180, 10, 10, lin)) = fq(180, 190, dec, lin).$

Similar relations hold for fuzzy quantities and trapezoidal fuzzy numbers.

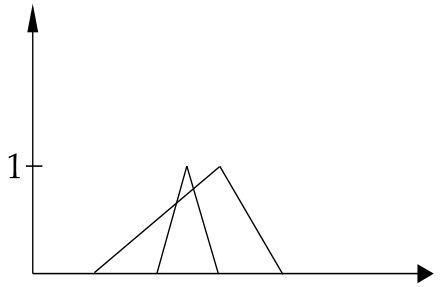


Figure 3. Incomparable fuzzy sets by $\preceq_I$ ordering

Besides using the relation operators many contributions have been written on fuzzy relational operators. It is clear that we can define the relation "fuzzy less or equal" denoted FLQ fuzzyfying the previously defined ordering $\preceq_I$.

Definition 6. Let $A, B \in \mathcal{F}(x)$.The fuzzy relation FLQ is defined in the following way:

$FLQ(A, B) = T(FINCL(LTR^A(B), LTR^B(A)), FINCL(RTL^B(A), RTL^A(B)))$.

where $LTR^M(N) = LTR(M) \setminus LTR(M \cap N)$, $RTL^M(N) = RTL(M) \setminus RTL(M \cap N)$ and T is a t-norm. The introduction of $LTR^M(N)$ and $RTL^M(N)$ was necessary since the fuzzy sets LTR and RTL have infinite support thus it is impossible to calculate their cardinality. Some results of $FLQ(A, B)$ are given in the Table 4.

However, in [3] fuzzy relation on the set of trapezoidal $FLEQ$ is defined, which is the equivalent of FLQ. Let $A = (\alpha_A, \beta_A, \gamma_A, \delta_A)$ and $B = (\alpha_B, \beta_B, \gamma_B, \delta_B)$ be two linear trapezoidal fuzzy numbers. The relation $FLEQ(A, B)$ is defined in the following way:

$$FLEQ(A, B) = \begin{cases} 1 & \text{if} \beta_A \leq \gamma_B, \\ \frac{\delta_B - \alpha_A}{(\beta_A - \alpha_A) - (\gamma_B - \delta_B)} & \text{if} \beta_A > \gamma_B \text{and} \\ & \alpha_A < \delta_B, \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that if $FLQ(A,B) = 1$ then also $FLEQ(A,B) = 1$. Moreover, if

$$FLQ(A(\alpha_A,\beta_A,\gamma_A,\delta_A), B(\alpha_B,\beta_B,\gamma_B,\delta_B) > 0$$

then $\delta_B \geq \delta_A$ which implies $FLEQ(A,B) = 1$. Also we have: If

$$FLEQ(A(\alpha_A,\beta_A,\gamma_A,\delta_A), B(\alpha_B,\beta_B,\gamma_B,\delta_B) < 1$$

then $FLQ(A,B) = 0$. This leads us to conclusion that for each A, B we have $FLEQ(A,B) \geq FLQ(A,B)$ making FLQ a stricter relation than $FLEQ$. In Table 4 a comparison of FLQ and $FLEQ$ is given for trapezoidal fuzzy numbers.

Table 4: FLQ and FLEQ relation.

A	B	FLQ	FLEQ
(10,20,60,80)	(30,40,50,70)	0	1
(10,20,60,80)	(30,40,50,100)	0.5	1
(10,20,60,80)	(11,11.5,12,25)	0	0.65
(10,20,60,80)	(11,15,19,25)	0	0.94

5 Conclusion

In this paper we have presented a similarity relation that can be used in querying of fuzzy relational databases(FRDB). Using the ordering $\preceq_I$ fuzzy relation FLQ is defined. Let us mention that the computer implementation of these relations is very simple for most types of fuzzy sets and the computational complexity is at the same level as the calculation of classical database relations $(=, \leq)$. However, the practical use of these implementations has not been tested on commercial databases. This is one of our aims in the future.

Acknowledgement

The authors would like to acknowledge the support of the Serbian ministry of science, technology and environmental protection project "Mathematical models of non-linearity, uncertainty and decision making", No. 144012 and project "Abstract methods and applications in Computer Science" No 144017A.

References

[1] Bodenhofer U., A Similarity-Based Generalization of Fuzzy Orderings, Phd Theses, Johannes Kepler University, Linz, Austria.(1998)

[2] G. Q. Chen, Fuzzy logic in data modeling: semantics, constraints, and database design, Kluwer Academic, Dordrecht (1998).

[3] J. Galindo, Angelica Urrutia and Mario Piattini, Fuzzy databases : modeling, design and implementation, Idea Group Publishing 2006.

[4] E. Klement, R. Mesiar, E. Pap, Triangular norms, Series: Trends in Logic, Kluwer Academic Publishers, Vol. 8, Dordrecht, 2000.

[5] A. Takači, S. Škrbić, How to Implement FSQL and Priority Queries, Procc. of Sisy 2005 98-104.

[6] A. Takači, Handling Priority Within a Database Scenario, ETF Journal of Electrical Engineering, (2006), 130-134.

[7] Weber S., On general concepts of fuzzy conectives, negations and implications based on t-norms and t-conorms, Fuzzy Sets and Systems 11, (1983) 115-134

A universal integral

Erich Peter Klement
Department of Knowledge-Based Mathematical Systems
Johannes Kepler University
4040 Linz (Austria)
ep.klement@jku.at

Radko Mesiar
Department of Mathematics and Descriptive Geometry
Faculty of Civil Engineering
Slovak University of Technology
81 368 Bratislava (Slovakia)
Institute for Research and Applications of Fuzzy Modeling
University of Ostrava
70103 Ostrava (Czech Republic)
mesiar@math.sk

Endre Pap
Department of Mathematics and Informatics, University of Novi Sad
21000 Novi Sad (Serbia)
pap@im.ns.ac.yu, pape@eunet.yu

Abstract

Based on a minimal set of axioms we introduce a general integral which can be defined on arbitrary measurable spaces. It acts on measures which are only (finite) monotone set functions and on measurable functions whose range is contained in the unit interval. We introduce the notion of integral equivalence of pairs of measures and functions which leads us to a special important general integrals called universal integral. Several special types of such functionals, including extremal ones, are characterized.

Key words. General integral, monotone set function, integral equivalence, universal integral, Choquet integral, semicopula.

1 Introduction

We try to contribute to a classical discussion: "What is an integral?" Based on certain minimal sets of axioms we introduce two concepts of functionals (here called general and universal integrals), which can be defined on arbitrary measurable spaces and which act on measures which are only (finite) monotone set functions and for measurable functions whose range is contained in the unit interval. Several special types of such functionals, including extremal ones, are characterized.

Clearly, the Choquet and the Sugeno integral are well-known examples of universal integrals: for any monotone set function m vanishing at the empty set and for any measurable function f with range $[0,1]$ the Choquet integral (considered earlier in [19]) is given by

$$\mathbf{Ch}(f) = \int_0^1 m(\{x \mid f(x) \geq t\})\, dt.$$

Note that the original proposal in [3] was done for special monotone set functions (capacities) only. The Sugeno integral (originally considered for continuous monotone set functions) is given by

$$\mathbf{Su}(f) = \sup_t \min(t, m(\{x \mid f(x) \geq t\})),$$

and again it can be defined for any m and f.

Usually these integrals are either interpolations (when sets are represented as characteristic functions), e.g., Choquet integral on $[0,1]$, or extrapolations, e.g., the Choquet integral on $[0,\infty]$. We will concentrate on interpolative integrals trying to provide some common framework. Based on a natural (although very weak) axiomatization we shall prove some intrinsic properties.

2 Integral equivalence

Throughout of this paper, let X be a fixed non-empty set, $\mathcal{A}$ a σ-algebra of subsets of X (in the case of a finite set X we usually take $\mathcal{A} = 2^X$), and $\mathcal{F}$ the class of all measurable functions $f\colon X \to [0,1]$. Finally, denote by $\mathcal{M}$ the class of all monotone set functions $m\colon \mathcal{A} \to [0,1]$ (considered, sometimes with additional properties, in [4,6,11,15,18,20]) which satisfy $m(\emptyset) = 0$, $m(X) = 1$ and $m(A) \leq m(B)$ whenever $A \subseteq B$.

Since we will consider integrals depending only on the pair consisting of a measure and a function we introduce a generalization of the equality almost everywhere of two functions.

Definition 2.1 Let $m_1, m_2 \in \mathcal{M}$ and $f_1, f_2 \in \mathcal{F}$. We say that the pairs (m_1, f_1) and (m_2, f_2) are *integral equivalent*, in symbols $(m_1, f_1) \sim (m_2, f_2)$, whenever

$$m_1(\{x \mid f_1(x) \geq t\}) = m_2(\{x \mid f_2(x) \geq t\})$$

for every $t \in [0,1]$.

Remark 2.2

(i) Fixing the measures $m_1 = m_2 = m$ in Definition 2.1 we obtain the m-indistinguishability from

[17] of functions f_1 and f_2, and in this case we will write $f_1 \sim_m f_2$.

(ii) Fixing the functions $f_1 = f_2 = f$ in Definition 2.1 we can introduce the f-indistinguishability of measures m_1 and m_2, and in this case we will write $m_1 \sim_f m_2$.

Example 2.3 (i) Let m be a probability measure. Then the classical equality m-almost everywhere of two functions f and g implies that $f \sim_m g$, but not vice versa, i.e., the relation $\sim_m$ is a generalization of the equality m-almost everywhere.

(ii) Note that for the weakest monotone measure $m_*\colon \mathcal{A} \to [0,1]$ given by

$$m_*(A) = \begin{cases} 1 & \text{if } A = X, \\ 0 & \text{otherwise,} \end{cases}$$

we have $f_1 \sim_{m_*} f_2$ if and only if $\inf f_1 = \inf f_2$.

(iii) Similarly, for the strongest monotone measure $m^*\colon \mathcal{A} \to [0,1]$ given by

$$m^*(A) = \begin{cases} 0 & \text{if } A = \emptyset, \\ 1 & \text{otherwise,} \end{cases}$$

we have $f_1 \sim_{m^*} f_2$ if and only if $\sup f_1 = \sup f_2$ and both function either simultaneously possess the maximal value, or not. Hence, if X is finite, then $f_1 \sim_{m^*} f_2$ if and only if $\max f_1 = \max f_2$.

(iv) We get $m_1 \sim_{\mathbf{c}} m_2$ for all $m_1, m_2 \in \mathcal{M}$ and each constant function $\mathbf{c}\colon X \to [0,1]$ given by $\mathbf{c}(x) = c$. On the other hand, if $f = c \cdot \mathbf{1}_A$ for some $c \in]0,1]$ and $A \in \mathcal{A}$, then we have $m_1 \sim_f m_2$ if and only if $m_1(A) = m_2(A)$.

Proposition 2.4 *Define for $m \in \mathcal{M}$ and for a measurable function $\varphi\colon X \to X$ the measure $m^\varphi \in \mathcal{M}$ by $m^\varphi(A) = m(\varphi^{-1}(A))$. Then for all $f \in \mathcal{F}$ the pairs $(m, f \circ \varphi)$ and (m^φ, f) are integral equivalent.*

We shall compare the notion of m-indistinguishability with the equality almost everywhere based on the notion of null set for the case of monotone set functions given in [9], see [11].

Definition 2.5 A set $N \in \mathcal{A}$ is a *null set* with respect to $m \in \mathcal{M}$ if for all $A \in \mathcal{A}$ we have

$$m(A \cup N) = m(A).$$

Definition 2.6 Let $m \in \mathcal{M}$ and $f, g \in \mathcal{F}$. We say that $f = g$ *almost everywhere with respect to* m if there exists a null set $N \in \mathcal{A}$ with respect to m such that $f(x) = g(x)$ for every $x \in \complement N$.

Theorem 2.7 *Let $m \in \mathcal{M}$ and $f, g \in \mathcal{F}$. If $f = g$ almost everywhere with respect to m, then f and g are m-indistinguishable.*

The converse of Theorem 2.7 is not true in general:

Example 2.8 The only null set with respect to m^* is the empty set. Therefore we have $f_1 = f_2$ almost everywhere with respect to m^* if and only if $f_1(x) = f_2(x)$ for every $x \in X$, but $f_1 \sim_{m^*} f_2$ if and only if $\sup f_1 = \sup f_2$.

3 General integral

We have introduced in the paper [8] the notion of general integral.

Definition 3.1 A mapping $\mathbf{G}\colon \mathcal{M} \times \mathcal{F} \to [0,1]$ is called a *general integral* on $(X, \mathcal{A})$ if it satisfies the following conditions:

(I1) *boundary conditions*, i.e., for each $m \in \mathcal{M}$ we have

$$\mathbf{G}(m, 0) = 0 \quad \text{and} \quad \mathbf{G}(m, 1) = 1,$$

(I2) *monotonicity* in both coordinates, i.e., for $m_1 \le m_2$ and $f_1 \le f_2$ we have

$$\mathbf{G}(m_1, f_1) \le \mathbf{G}(m_2, f_2),$$

(I3) *extension of the measure*, i.e., for each $A \in \mathcal{A}$ and for each $m \in \mathcal{M}$ we have

$$\mathbf{G}(m, \mathbf{1}_A) = m(A).$$

(I4) *idempotency*, i.e., for every $c \in [0,1]$ we have

$$\mathbf{G}(m, c) = c,$$

(I5) *existence of a pseudo-multiplication*, i.e., there exists a binary operation $\otimes\colon [0,1]^2 \to [0,1]$ such that for each $m \in \mathcal{M}$, each $c \in [0,1]$ and each $A \in \mathcal{A}$

$$\mathbf{G}(m, c \cdot \mathbf{1}_A) = c \otimes m(A),$$

(I6) for each measurable function $\varphi\colon X \to X$, and for each $(m, f) \in \mathcal{M} \times \mathcal{F}$ we have

$$\mathbf{G}(m^\varphi, f) = \mathbf{G}(m, f \circ \varphi),$$

where the measure $m^\varphi \in \mathcal{M}$ is given by $m^\varphi(A) = m(\varphi^{-1}(A))$.

Obviously, (I3) as well as (I4) implies (I1); however we prefer to keep axiom (I1) in order to stress the boundary conditions of integral functionals.

Observe that one of the consequences of axiom (I6) is that for a general integral $\mathbf{G}$ and a Dirac measure

$m_{\{x_0\}}$ (all the mass is concentrated in $x_0 \in X$) we have $\mathbf{G}(m_{\{x_0\}}, f) = f(x_0)$ for each $f \in \mathcal{F}$.

Also, the class $\mathcal{G}$ of general integrals is convex, and it is closed under each idempotent aggregation operator.

Observe that the axioms (I1)–(I4) imply that the pseudo-multiplication $\otimes$ required in axiom (I5) is a semicopula (see [1, 5]).

Following the ideas of inner and outer measures in classical measure theory, we obtain the following result:

Theorem 3.2 *Let $\otimes$ be a semicopula. Then the class $\mathcal{G}_\otimes$ of all general integrals related to $\otimes$ is a convex class with smallest element $\mathbf{G}_\otimes$ and greatest element $\mathbf{G}^\otimes$, given by*

$$\mathbf{G}_\otimes(m, f) = \sup\{t \otimes m(\{f \geq t\}) \mid t \in [0, 1]\},$$
$$\mathbf{G}^\otimes(m, f) = (\sup f) \otimes m(\{f > 0\}).$$

Recall that the drastic product $T_\mathbf{D}$ is the weakest and the minimum $T_\mathbf{M}$ is the strongest semicopula. Obviously, for any two semicopulas $\otimes_1$ and $\otimes_2$ with $\otimes_1 \leq \otimes_2$ we have $\mathbf{G}_{\otimes_1} \leq \mathbf{G}_{\otimes_2}$ and $\mathbf{G}^{\otimes_1} \leq \mathbf{G}^{\otimes_2}$.

Corollary 3.3 *The smallest general integral $\mathbf{G}_* = \mathbf{G}_{T_\mathbf{D}}$ and the largest general integral $\mathbf{G}^* = \mathbf{G}_{T_\mathbf{M}}$ are given by*

$$\begin{aligned}\mathbf{G}_*(m, f) &= \sup\{T_\mathbf{D}(t, m(\{f \geq t\})) \mid t \in [0, 1]\} \\ &= \max(\operatorname{essinf}_m f, m(\{f = 1\})),\\ \mathbf{G}^*(m, f) &= \min(\sup f, m(\{f > 0\})).\end{aligned}$$

Here $\operatorname{essinf}_m f = \sup\{t \in [0, 1] \mid m\{f \geq t\} = 1\}$.

Note that the Choquet and the Sugeno integral are examples of general integrals. Moreover, these integrals (and also $\mathbf{G}_*$) give the same result for integral equivalent pairs (m_1, f_1) and (m_2, f_2).

However, a general integral $\mathbf{G}$ does not fulfill $\mathbf{G}(m_1, f_1) = \mathbf{G}(m_2, f_2)$ whenever $(m_1, f_1) \sim (m_2, f_2)$, in general.

Example 3.4 Let $X =]0, 1[$ and $\mathcal{A}$ the σ-algebra of Borel subsets of X. Define $\mathbf{G} : \mathcal{M} \times \mathcal{F} \to [0, 1]$ in the following way

$$\mathbf{G}(m, f) = m(f > 0) \cdot \sup f.$$

Observe that $\mathbf{G} = \mathbf{G}^{T_\mathbf{P}}$ is the strongest general integral based on the product $T_\mathbf{P}$ as the semicopula $\otimes$. Then $(m_*, \mathbf{0}) \sim (m_*, \mathrm{id}_X)$, but $\mathbf{G}(m_*, \mathbf{0}) = 0$ and $\mathbf{G}(m_*, \mathrm{id}_X) = 1$.

4 A universal integral

To overcome the possibility of obtaining different outputs applying a general integral to integral equivalent pairs $(m_1, f_1) \sim (m_2, f_2)$, we introduce universal integrals.

Definition 4.1 A general integral $\mathbf{U} \colon \mathcal{M} \times \mathcal{F} \to [0, 1]$ is called *universal integral* whenever for each integral equivalent pairs (m_1, f_1) and (m_2, f_2) from $\mathcal{M} \times \mathcal{F}$ we have

$$\mathbf{U}(m_1, f_1) = \mathbf{U}(m_2, f_2).$$

Note that there are two other equivalent concepts of defining universal integrals [7, 17].

Again, the Choquet, and the Sugeno integral are examples of universal integrals. The class $\mathcal{U}$ of universal integrals is also a convex set, and it is closed under any idempotent aggregation operator.

A generalization of the construction method of Choquet and Sugeno integrals for a universal integral can be described in the following way (see [17]). Suppose that for a general integral $\mathbf{U} \colon \mathcal{M} \times \mathcal{F} \to [0, 1]$ there exists a monotone function $\mathbf{J} \colon \mathcal{L}([0, 1]) \to [0, 1]$ such that

$$\mathbf{U}(m, f) = \mathbf{J}(h_{m,f}), \tag{1}$$

where $\mathcal{L}([0, 1])$ is the class of all Borel measurable functions from $[0, 1]$ to $[0, 1]$ and $h_{m,f} \colon [0, 1] \to [0, 1]$ is given by $h_{m,f}(t) = m(\{f \geq t\})$. Then evidently $\mathbf{U}$ is a universal integral (observe that $(m_1, f_1) \sim (m_2, f_2)$ if and only if $h_{m_1,f_1} = h_{m_2,f_2}$). Conversely, for a universal integral $\mathbf{U}$ it is enough to put

$$\mathbf{J}(g) = \sup\{\mathbf{U}(m, f) \mid h_{m,f} \leq g\}$$

in order to obtain (1).

Now we recall two properties of the class $\mathcal{U}$ of universal integrals, see also [17]:

(i) For each $\mathbf{U} \in \mathcal{U}$ we have

$$\mathbf{U}_* \leq \mathbf{U} \leq \mathbf{U}^*,$$

where $\mathbf{U}_*$ and $\mathbf{U}^*$ are given by

$$\mathbf{U}_* = \mathbf{G}_*,$$
$$\mathbf{U}^*(m, f) = \min(\operatorname{essup}_m f, m(\{f > 0\})),$$

and $\operatorname{essup}_m f = \sup\{t \in [0, 1] \mid m\{f \geq t\} > 0\}$.

(ii) For each measurable space $(X, 2^X)$, each $\{0, 1\}$-valued measure $m \in \mathcal{M}$ and each $f \in \mathcal{F}$ we have $\mathbf{U}_*(m, f) = \mathbf{U}(m, f) = \mathbf{U}^*(m, f)$ for all $\mathbf{U} \in \mathcal{U}$. Defining the function $L_m \colon \mathcal{F} \to [0, 1]$ by $L_m(f) = \mathbf{U}_*(m, f)$, then L_m is a lattice polynomial on X, and it can be written as

$$L_m(f) = \bigvee_{m(A)=1} \bigwedge_{x \in A} f(x).$$

For universal integrals we have the following counterpart of Theorem 3.2:

Theorem 4.2 *Let $\otimes$ be a semicopula. Then the class $\mathcal{U}_\otimes$ of all universal integrals related to $\otimes$ is a convex class with smallest element $\mathbf{U}_\otimes$ and greatest element $\mathbf{U}^\otimes$, given by*

$$\mathbf{U}_\otimes = \mathbf{G}_\otimes,$$
$$\mathbf{U}^\otimes(m, f) = (\mathrm{essup}_m f) \otimes m(\{f > 0\}).$$

Observe that $\mathbf{U}_{T_\mathbf{M}}$ is the Sugeno integral, while $\mathbf{U}_{T_\mathbf{P}}$ is the Shilkret integral [13] (originally defined for maxitive measures only).

5 Acknowledgment

The second and the third author were supported by a technology transfer project of the Upper Austrian Government. The second author also acknowledges the support of the Österreichische Forschungsgemeinschaft and by the grants VEGA 1/4209/07, APVV-0375-06, and MSM VZ 6198898701, the third author was also supported by Vojvodina PSSTD and by grant MNTRS-144012.

References

[1] B. Bassan and F. Spizzichino, *Relations among univariate aging, bivariate aging and dependence for exchangeable lifetimes*, J. Multivariate Anal. **93** (2005), 313–339.

[2] P. Benvenuti, R. Mesiar, and D. Vivona, *Monotone set functions-based integrals*, in E. Pap [12], pp. 1329–1379.

[3] G. Choquet, *Theory of capacities*, Ann. Inst. Fourier (Grenoble) **5** (1953–1954), 131–292.

[4] D. Denneberg, *Non-additive measure and integral*, Kluwer Academic Publishers, Dordrecht, 1994.

[5] F. Durante and C. Sempi, *Semicopulæ*, Kybernetika (Prague) **41** (2005), 315–328.

[6] G. H. Greco, *Monotone integrals*, Rend. Sem. Mat. Univ. Padova **57** (1977), 149–166.

[7] E. P. Klement, R. Mesiar, and E. Pap, *Measure-based aggregation operators*, Fuzzy Sets and Systems **142** (2004), 3–14.

[8] E. P. Klement, R. Mesiar, and E. Pap, *Integrals which can be defined on arbitrary measurable spaces*, in Abstracts 28th Linz Seminar on Fuzzy Set Theory (Fuzzy Sets, Probability, and Statistics — Gaps and Bridges) 72-77, 2007.

[9] T. Murofushi and M. Sugeno. *Fuzzy t-conorm integrals with respect to fuzzy measures: generalization of Sugeno integral and Choquet integral.* Fuzzy Sets and Systems, 42:57–71, 1991.

[10] R. B. Nelsen, *An introduction to copulas*, Lecture Notes in Statistics 139, Springer, New York, 1999.

[11] E. Pap, *Null-Additive Set Functions*, Kluwer, 1995.

[12] E. Pap (ed.), *Handbook of measure theory*, Elsevier Science, Amsterdam, 2002.

[13] N. Shilkret, *Maxitive measure and integration*, Indag. Math. **33** (1971), 109–116.

[14] M. Sion. *A Theory of Semigroup Valued Measures*, Lecture Notes in Mathematics 355. Springer, Berlin, 1973.

[15] J. Šipoš, *Integral with respect to a pre-measure*, Math. Slovaca **29** (1979), 141–155.

[16] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris **8** (1959), 229–231.

[17] P. Struk, *Extremal fuzzy integrals*, Soft Computing **10** (2006), 502–505.

[18] M. Sugeno, *Theory of fuzzy integrals and its applications*, Ph.D. thesis, Tokyo Institute of Technology, 1974.

[19] G. Vitali. *Sulla definizione di integrale delle funzioni di una variabile*, Ann. Mat. Pura Appl. **2** (1925) 111–121.

[20] Z. Wang and G. J. Klir, *Fuzzy measure theory*, Plenum Press, New York, 1992.

On the k-additive core of capacities

Michel GRABISCH
University of Paris I – Panthéon-Sorbonne,
CERMSEM
106-112 Bd de l'Hôpital, 75013 Paris, France
michel.grabisch@lip6.fr

Pedro MIRANDA
Universidad Complutense de Madrid
Plaza de Ciencias, 3, 28040 Madrid, Spain
pmiranda@mat.ucm.es

Abstract

We investigate in this paper the set of k-additive capacities dominating a given capacity, which we call the k-additive core. We study its structure through achievable families, which play the role of maximal chains in the classical case ($k = 1$), and show that associated capacities are elements (possibly a vertex) of the k-additive core when the capacity is $(k+1)$-monotone. As a particular case, we study the set of k-additive belief functions dominating a belief function. The problem of finding all vertices of the k-additive core is still an open question.

Keywords: k-additive capacity, core, belief function

1 Introduction

The core of a capacity or a game is a fundamental concept, both in decision making theory and in cooperative game theory. In decision making, it is the set of probability measures which are coherent with the information given by the capacity in the representation of uncertainty [15]. In game theory, it is the set of imputations (additive games) that can be given to players so that no subcoalition of the grand coalition has interest to form.

The properties of the core are well known, most of them have been shown by Shapley [13]. In many cases, it happens that the core is empty. A sufficient and necessary condition for nonemptiness is known for capacities and games, which is called balancedness. In particular, convex capacities have a nonempty core.

Since having an empty core is not a favorable situation, either in decision making or in game theory, it may be an alternative solution to look for more general concepts. For example, since the core contains additive capacities or games, we may relax additivity to a weaker notion: k-additivity, proposed by Grabisch [5]. We may call this new notion the *k-additive core*.

Some studies on the k-additive core have already been done by the authors, see, e.g., [6, 11]. It happens that the structure of the k-additive core is much more complex than the one of the classical core. In particular, the set of its vertices is not known. The aim of this paper is to provide new insights in this direction, and to complete results shown in [9].

2 Background

Throughout the paper, we consider a finite universal set X, with $|X| = n$. We use indifferently 2^X or $\mathcal{P}(X)$ to denote the set of subsets of X, and the set of subsets of X containing at most k elements is denoted by $\mathcal{P}^k(X)$, while $\mathcal{P}^k_*(X) := \mathcal{P}^k(X) \setminus \{\emptyset\}$. A *set function* on X is a function $\mu : 2^X \to \mathbb{R}$.

Definition 1 *[3, 14] A* fuzzy measure *or* capacity μ *on X is a nonnegative set function on X such that $\mu(\emptyset) = 0$, and $\mu(A) \leq \mu(B)$ whenever $A \subseteq B$ (*monotonicity*). A capacity is* normalized *if $\mu(X) = 1$.*

We assume in this paper that capacities are normalized. The set of capacities (or fuzzy measures) on X is denoted by $\mathcal{FM}(X)$. For any $A \in 2^X \setminus \{\emptyset\}$, the *unanimity game centered on A* is defined by $u_A(B) = 1$ iff $B \supseteq A$, and 0 otherwise.

Definition 2 *A capacity μ on X is said to be:*

(i) additive *if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$;*

(ii) convex *if $\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B)$, for all $A, B \subseteq X$;*

(iii) k-monotone *for $k \geq 2$ if for any family of k sub-*

sets $A_1, \dots A_k$, it holds

$$\mu(\bigcup_{i=1}^{k} A_i) \geq \sum_{\substack{K \subseteq \{1,\dots,k\} \\ K \neq \emptyset}} (-1)^{|K|+1} \mu(\bigcap_{j \in K} A_j)$$

(iv) totally monotone *if it is k-monotone for all $k \geq 2$.*

Totally monotone capacities are also called *belief functions* [12]. By extension, we define 1-monotone capacities as monotone capacities. We will denote the set of belief functions on X by $\mathcal{BEL}(X)$. Remark that k-monotonicity implies k'-monotonicity for all $2 \leq k' \leq k$. Also, for $n > 3$, $(n-2)$-monotone capacities are totally monotone [1].

Definition 3 *Let μ be a set function on X. The* Möbius transform *of μ is a set function $m : 2^X \to \mathbb{R}$ defined by:*

$$m(A) := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B).$$

The Möbius transform is invertible since one can recover μ from m by:

$$\mu(A) = \sum_{B \subseteq A} m(B).$$

If μ is an additive capacity, then m is non null only for singletons, and $m(\{i\}) = \mu(\{i\})$. The Möbius transform of u_A is given by $m(A) = 1$ and m is 0 otherwise.

Definition 4 *A set function μ vanishing at the empty set is said to be k-*additive *for some integer $k \in \{1, \dots, n\}$ if $m(A) = 0$ whenever $|A| > k$, and there exists some A such that $|A| = k$, and $m(A) \neq 0$.*

Clearly, 1-additive capacities are additive measures, and a k-additive set function needs only $\sum_{i=1}^{k} \binom{n}{i} - 2$ values to be defined. The set of capacities on X being at most k-additive is denoted by $\mathcal{FM}^k(X)$. Similarly, we denote by $\mathcal{BEL}^k(X)$ the set of belief functions being at most k-additive.

We recall the fundamental following result.

Proposition 1 *[2] (i) Let μ be a set function on X such that $\mu(\emptyset) = 0$. Then monotonicity is equivalent to*

$$\sum_{i \in L \subseteq B} m(L) \geq 0, \quad \forall B \subseteq X, \quad \forall i \in B.$$

(ii) Let μ be a capacity. Then for $2 \leq k \leq n$, k-monotonicity is equivalent to

$$\sum_{A \subseteq L \subseteq B} m(L) \geq 0, \quad \forall A \subseteq B \subseteq X, \quad |A| \leq k.$$

Clearly, a totally monotone capacity has a non nonnegative Möbius transform.

3 The core of capacities

Definition 5 *Let μ be a capacity on X. The* core *of μ is defined by:*

$$\mathcal{C}(\mu) := \{\nu \in \mathcal{FM}^1(X) \mid \nu \geq \mu\},$$

where $\nu \geq \mu$ means $\nu(A) \geq \mu(A)$ for all $A \subseteq X$.

A *maximal chain* in 2^X is a sequence of subsets $A_0 := \emptyset, A_1, \dots, A_{n-1}, A_n := X$ such that $A_i \subset A_{i+1}$, $i = 0, \dots, n-1$. The set of maximal chains of 2^X is denoted by $\mathcal{M}(2^X)$.

To each maximal chain $C := \{\emptyset, A_1, \dots, A_n = X\}$ in $\mathcal{M}(2^X)$ corresponds a unique permutation σ on X such that $A_1 = \sigma(1)$, $A_2 \setminus A_1 = \sigma(2)$, $\dots$, $A_n \setminus A_{n-1} = \sigma(n)$. The set of all permutations over X is denoted by $\mathfrak{S}(X)$. Let μ be a capacity. To each permutation σ (or maximal chain C) we assign a *marginal worth vector* p^σ (or p^C) in $\mathbb{R}^n$ defined by:

$$p^\sigma_{\sigma(i)} := \mu(\{\sigma(1), \dots, \sigma(i)\}) - \mu(\{\sigma(1), \dots, \sigma(i-1)\})$$

or

$$p^C_{\sigma(i)} := \mu(A_i) - \mu(A_{i-1})$$

with the above notation. Any marginal worth vector forms a probability distribution over X, and hence defines an additive capacity. The following is immediate.

Proposition 2 *Let μ be a capacity on X, and C a maximal chain of 2^X. Then*

$$p^C(A) = \mu(A), \quad \forall A \in C.$$

Theorem 1 *The following are equivalent.*

(i) *μ is a convex capacity*

(ii) *all marginal worth vectors p^σ, $\sigma \in \mathfrak{S}(X)$ belong to the core of μ*

(iii) $\mathcal{C}(\mu) = \mathrm{co}(\{p^\sigma\}_{\sigma \in \mathfrak{S}(X)})$

(iv) $\mathrm{ext}(\mathcal{C}(\mu)) = \{p^\sigma\}_{\sigma \in \mathfrak{S}(X)}$,

where $\mathrm{co}(K)$ *and* $\mathrm{ext}(K)$ *denote respectively the convex hull and the extreme points of some convex set K.*

(i) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iv) are due to Shapley [13], while (ii) $\Rightarrow$ (i) was proved by Ichiishi [10].

4 The k-additive core of capacities

Definition 6 *[6] Let μ be a capacity on X, and $1 \leq k \leq n-1$. The* k-additive core *of μ is defined by:*

$$\mathcal{C}^k(\mu) := \{\nu \in \mathcal{FM}^k(X) \mid \nu \geq \mu\}.$$

Note that $\mathcal{C}^1(\mu) = \mathcal{C}(\mu)$. Similarly, we introduce $\mathcal{BC}^k(\mu)$ the set of k-additive belief functions dominating μ.

Let us start introducing some notations. For a given k, $1 \leq k \leq n$, we define

$$\Lambda^k_\cap := \{\lambda : \mathcal{P}(X) \times \mathcal{P}^k_*(X) \to \mathbb{R} \,|\, \forall B \in \mathcal{P}(X),$$
$$\sum_{A \cap B \neq \emptyset} \lambda(B, A) = 1, \lambda(B, A) = 0 \text{ if } A \cap B = \emptyset\}.$$

From this set and for a given capacity μ, we define a set of k-additive set functions $\mathcal{M}_{\Lambda^k_\cap}(\mu) := \{\mu_\lambda \mid \lambda \in \Lambda^k_\cap\}$, with μ_λ defined by its Möbius transform m_λ:

$$m_\lambda(A) = \sum_{B \in \mathcal{P}(X)} \lambda(B, A) m(B).$$

Similarly, we can define $\Lambda^k_\subseteq$ and $\Lambda^k_\supseteq$, from which we can derive the corresponding sets $\mathcal{M}_{\Lambda^k_\subseteq}(\mu)$ and $\mathcal{M}_{\Lambda^k_\supseteq}(\mu)$, by replacing $\cap$ by $\subseteq, \supseteq$. When λ is restricted to non-negative values (then it becomes a weight function), we will use the notations $\Lambda^k_{\cap,+}, \Lambda^k_{\subseteq,+}, \Lambda^k_{\supseteq,+}$, and the corresponding $\mathcal{M}_{\Lambda^k_{\cap,+}}(\mu), \mathcal{M}_{\Lambda^k_{\subseteq,+}}(\mu)$ and $\mathcal{M}_{\Lambda^k_{\supseteq,+}}(\mu)$.

Given a capacity μ, the problem of obtaining the set of all probability measures dominating μ has been addressed by several authors [2, 4]. Chateauneuf and Jaffray proved in [2] the following result:

Theorem 2 *[2] Let μ be a capacity on X, m its Möbius transform, and suppose $P \in \mathcal{C}^1(\mu)$. Then, $P \in \mathcal{M}_{\Lambda^1_{\subseteq,+}}(\mu)$.*

Dempster has shown the same result in [4] and also Shapley in [13], but both of them only for belief functions. In general, $\mathcal{C}^1(\mu) \subseteq \mathcal{M}_{\Lambda^1_{\subseteq,+}}(\mu)$, and equality holds when $\mu \in \mathcal{BEL}(X)$.

Grabisch has extended Theorem 2 for the k-additive case in [8].

Theorem 3 *[8] Let μ be a capacity, and suppose that $\mu^* \in \mathcal{C}^k(\mu)$, for some $1 \leq k \leq n-1$. Then, necessarily, μ^* belongs to $\mathcal{M}_{\Lambda^k_\cap}(\mu)$.*

As shown in [11], $\mathcal{C}^k(\mu) \subseteq \mathcal{M}_{\Lambda^k_\cap}(\mu)$, and equality holds only for the unanimity game centered on X. The following can be proved.

Theorem 4 *[11] Let μ be a belief function, and suppose $\mu^* \in \mathcal{BC}^k(\mu)$. Then, necessarily μ^* belongs to $\mathcal{M}_{\Lambda^k_{\cap,+}}(\mu)$.*

Remark that $\mathcal{M}_{\Lambda^k_{\cap,+}}(\mu)$ generalizes $\mathcal{M}_{\Lambda^1_{\subseteq,+}}(\mu)$ for the general k-additive case. However, similarly as Theorem 3, Theorem 4 provides a very large class of functions, as shown by the following result.

Proposition 3 *Let μ be a capacity. Then, $\mathcal{M}_{\Lambda^k_{\cap,+}}(\mu) \subseteq \mathcal{BC}^k(\mu)$ if and only if $\mu(A) = 0$, $\forall A \neq X$.*

Thus, as pointed out in [11], it is not possible to generalize the good properties obtained for probabilities in [2]. For $\Lambda^k_{\subseteq,+}$, the following can be proved.

Proposition 4 *[7] If $\mu \in \mathcal{BEL}(X)$ and $\mu^* \in \mathcal{M}_{\Lambda^k_{\subseteq,+}}(\mu)$, then $\mu^* \in \mathcal{BC}^k(\mu)$.*

On the other hand, it can be seen that this set does not cope in general with the set of all dominating k-additive measures [7] ($k > 1$), i.e., the conditions on λ given in Theorem 4 cannot be strengthened. Then, if μ is a belief function,

$$\mathcal{M}_{\Lambda^k_{\subseteq,+}}(\mu) \subseteq \mathcal{BC}^k(\mu) \subseteq \mathcal{M}_{\Lambda^k_{\cap,+}}(\mu),$$

where inclusion is strict in general. Next result shows that all k-additive belief measures dominating another belief function μ can be generated from $\mathcal{M}_{\Lambda^k_{\subseteq,+}}(\mu)$.

Theorem 5 *[11] Let μ be a belief function, and suppose that $\mu^* \in \mathcal{BC}^k(\mu)$. Then, there exists $\mu' \in \mathcal{BC}^k(\mu)$ such that μ' belongs to $\mathcal{M}_{\Lambda^k_{\subseteq,+}}(\mu)$ and μ^* belongs to $\mathcal{M}_{\Lambda^k_{\supseteq,+}}(\mu')$.*

This result is explained in Figure 1 for $|X| = 3$ and $k = 2$. m, m', and m^* are respectively the Möbius transforms of μ, μ', and μ^*.

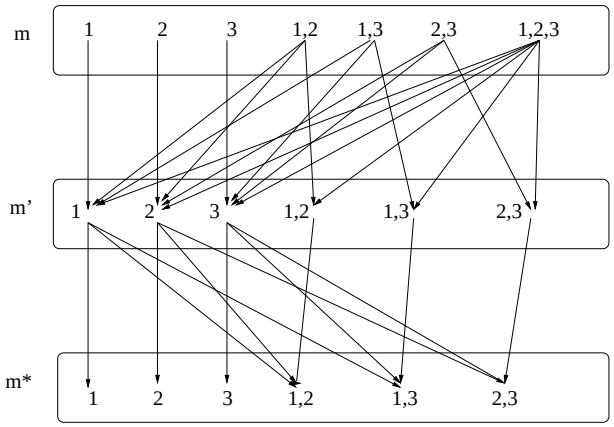


Figure 1: Illustration of Theorem 5.

5 Vertices of the k-additive core

We put $N(k) := \binom{n}{1} + \cdots + \binom{n}{k}$. A first fact is the following.

Proposition 5 *For any capacity μ, $\mathcal{C}^k(\mu)$ and $\mathcal{BC}^k(\mu)$ are closed convex $(N(k)-1)$-dimensional polytopes.*

We denote by $\prec$ a total (strict) order on $\mathcal{P}_*^k(X)$, $\preceq$ denoting the corresponding large order.

Definition 7 *For any $B \in \mathcal{P}_*^k(X)$, we define*

$$\mathcal{A}(B) := \{A \subseteq X \mid A \supseteq B, \forall K \subseteq A, K \in \mathcal{P}_*^k(X), K \preceq B\}$$

the achievable family of B.

For example, taking $n = 3$, $k = 2$ and the order $1 \prec 2 \prec 12 \prec 13 \prec 23 \prec 3$, we get:

$$\mathcal{A}(1) = \{1\}, \quad \mathcal{A}(2) = \{2\}, \quad ,\mathcal{A}(12) = \{12\},$$
$$\mathcal{A}(13) = \emptyset, \quad \mathcal{A}(23) = \emptyset, \quad \mathcal{A}(3) = \{3, 13, 23, 123\}.$$

It is easy to see that $\{\mathcal{A}(B)\}_{B \in \mathcal{P}_*^k(X)}$ is a partition of $\mathcal{P}(X) \setminus \{\emptyset\}$.

A total order $\prec$ on $\mathcal{P}_*^k(X)$ is said to be *compatible* if for all $i, j \in X$, $i \prec j$ implies $S \cup i \prec S \cup j$, for any $S \in \mathcal{P}^{k-1}(X)$, $i, j \notin S$. It is said to be $\subseteq$-*compatible* if $A \subseteq B$ implies $A \prec B$. Lastly, $\prec$ is said to be *strongly compatible* if it is compatible and $\subseteq$-compatible, and *weakly compatible* if only compatible.

Proposition 6 *Assume $\prec$ is compatible. For any $B \in \mathcal{P}_*^k(X)$ such that $\mathcal{A}(B) \neq \emptyset$, $\mathcal{A}(B)$ endowed with inclusion is a Boolean lattice with bottom element B. The top element is denoted by $\check{B}$.*

$\subseteq$-compatibility is a sufficient and necessary condition for the nonemptiness of all achievable families.

Let μ be a capacity on X, m its Möbius transform, and $\prec$ some total order on $\mathcal{P}_*^k(X)$. We define $\mu_\prec$ by its Möbius transform as follows:

$$m_\prec(B) := \begin{cases} \sum_{A \in \mathcal{A}(B)} m(A), & \text{if } \mathcal{A}(B) \neq \emptyset \\ 0, & \text{else} \end{cases} \qquad (1)$$

for all $B \in \mathcal{P}_*^k(X)$, $m_\prec(\emptyset) := 0$. Since achievable families form a partition of 2^X, $m_\prec$ satisfies $\sum_{B \subseteq X} m_\prec(X) = 1$, hence $\mu_\prec(X) = 1$. The following can be shown.

Proposition 7 *If $\prec$ is compatible, then for any nonempty achievable family $\mathcal{A}(B)$, $\mu_\prec(\check{B}) = \mu(\check{B})$.*

(see the analogy with Prop. 2).

Proposition 8 *Let μ be a capacity on X. Then $\mu_\prec$ is a belief function for any compatible order $\prec$ if and only if μ is k-monotone.*

The following propositions are analogous to the Shapley-Ichiishi theorem above.

Proposition 9 *Let μ be a capacity on X. Then $\mu_\prec \in \mathcal{C}^k(\mu)$ for all compatible orders $\prec$ if and only if μ is $(k+1)$-monotone.*

Proposition 10 *Let μ be a $(k+1)$-monotone capacity. Then*

(i) If $\prec$ is strongly compatible, then $\mu_\prec$ is a vertex of $\mathcal{C}^k(\mu)$.

(ii) If $\prec$ is compatible, then $\mu_\prec$ is a vertex of $\mathcal{BC}^k(\mu)$.

However, there are many vertices that are not belief functions. Experiments conducted with the PORTA software finding vertices and facets of polyhedra show that, for example, the set of vertices of $\mathcal{C}^2(\mu)$ of the following 3-monotone capacity μ with $n = 3$

A	1	2	3	12	13	23	123
$m(A)$	0	0.1	0.2	0.1	0	0.2	0.4
$\mu(A)$	0	0.1	0.2	0.2	0.2	0.5	1

has 48 elements, whose only 3 are belief functions.

```
(  1)    0 1/10 1/ 5  1/10     0  3/ 5
(  2)    0 1/10 1/ 5  1/10  2/ 5  1/ 5
(  3)    0 1/10 1/ 5  1/ 2     0  1/ 5
(  4)    0 1/10 1/ 2  1/10  2/ 5 -1/10
(  5)    0 1/10 1/ 2  1/ 2     0 -1/10
(  6)    0 1/10 9/10  1/10     0 -1/10
(  7)    0 1/ 5 1/ 5     0     0  3/ 5
(  8)    0 1/ 5 1/ 5     0  1/ 2  1/10
(  9)    0 1/ 5 1/ 2     0  1/ 2 -1/ 5
( 10)    0 1/ 2 1/ 5     0  1/ 2 -1/ 5
( 11)    0 1/ 2 1/ 5  1/ 2     0 -1/ 5
( 12)    0 1/ 2 1/ 2     0  1/ 2 -1/ 2
( 13)    0 1/ 2 1/ 2  1/ 2     0 -1/ 2
( 14) 1/10 1/10 1/ 5     0 -1/10  7/10
( 15) 1/10 1/10 1/ 2     0  2/ 5 -1/10
( 16) 1/ 5 1/10 1/ 5 -1/10 -1/10  7/10
( 17) 1/ 5 1/10 1/ 5 -1/10  2/ 5  1/ 5
( 18) 1/ 5 1/10 1/ 5     0 -1/ 5  7/10
( 19) 1/ 5 1/10 1/ 5  1/ 2 -1/ 5  1/ 5
( 20) 1/ 5 1/10 2/ 5 -1/10  2/ 5     0
( 21) 1/ 5 1/10 9/10 -1/10 -1/10     0
( 22) 1/ 5 1/ 5 1/ 5 -1/ 5     0  3/ 5
( 23) 1/ 5 1/ 5 1/ 5 -1/ 5  1/ 2  1/10
( 24) 1/ 5 1/ 5 3/10 -1/ 5  1/ 2     0
( 25) 1/ 5 1/ 5 4/ 5 -1/ 5     0     0
( 26) 1/ 5 3/10 1/ 5  1/ 2 -1/ 5     0
( 27) 1/ 5 4/ 5 1/ 5     0 -1/ 5     0
( 28) 3/10 1/10 1/ 5 -1/10 -1/ 5  7/10
( 29) 3/10 3/10 1/ 5 -3/10  1/ 2     0
( 30) 3/10 1/ 2 1/ 5 -3/10  1/ 2 -1/ 5
( 31) 2/ 5 1/10 2/ 5  1/ 2 -2/ 5     0
( 32) 2/ 5 1/10 1/ 2  1/ 2 -2/ 5 -1/10
( 33) 1/ 2 1/ 2 1/ 2     0     0 -1/ 2
( 34) 4/ 5 1/10 1/ 5 -1/10 -1/ 5  1/ 5
( 35) 4/ 5 1/ 2 1/ 5 -3/10     0 -1/ 5
( 36) 4/ 5 4/ 5 1/ 5 -4/ 5     0     0
( 37) 9/10 1/10 1/ 2     0 -2/ 5 -1/10
( 38) 9/10 1/10 9/10     0 -9/10     0
( 39)    0 1/ 5    1     0     0 -1/ 5
```

```
( 40)    0    1 1/ 5     0     0 -1/ 5
( 41) 1/10 1/10    1     0 -1/10 -1/10
( 42) 4/ 5    1 1/ 5 -4/ 5     0 -1/ 5
( 43) 9/10 1/10    1     0 -9/10 -1/10
( 44)    1 1/10 2/ 5 -1/10 -2/ 5     0
( 45)    1 1/10 9/10 -1/10 -9/10     0
( 46)    1 3/10 1/ 5 -3/10 -1/ 5     0
( 47)    1 4/ 5 1/ 5 -4/ 5 -1/ 5     0
( 48)    0    1    1     0     0    -1
```

Let us examine more precisely the vertices induced by strongly compatible orders. In fact, there are much fewer than expected, since many strongly compatible orders lead to the same $\mu_\prec$ (hence the experimental result above). We can show the following.

Proposition 11 *The number of vertices of $\mathcal{C}^k(\mu)$ given by strongly compatible orders is at most $\frac{n!}{k!}$.*

We examine the case of weakly compatible orders.

Proposition 12 *Suppose μ is a $(k+1)$-monotone capacity which satisfies $\mu(\{i\}) > 0$ for all $i \in X$. Then no weakly compatible order can produce a vertex of $\mathcal{C}^k(\mu)$.*

Weakly compatible orders can produce vertices if $\mu(\{i\}) = 0$ for some $i \in X$. It suffices that it exists $B \in \mathcal{P}^k_*(X)$ such that $\mathcal{A}(B) = \emptyset$, and $i \in B$ such that $\mu(\{i\}) = 0$, and all subsets C such that $i \in C \subset B$ satisfy $m_\prec(B) = 0$. The above example with 48 vertices illustrates this. By Prop. 11, we know that 3 vertices are produced by the strongly compatible orders, with corresponding sequences of $\check{B}$'s:

$$1, 2, 3, 12, 13, 123 \quad \text{(this is vertex 1)}$$
$$2, 1, 3, 12, 23, 123 \quad \text{(this is vertex 2)}$$
$$3, 1, 2, 13, 23, 123 \quad \text{(this is vertex 3)}.$$

Take the weakly compatible order $1 \prec 12 \prec 2 \prec 3 \prec 13 \prec 23$. Then achievable families are:

$$\mathcal{A}(1) = \{1\}, \quad \mathcal{A}(12) = \emptyset, \quad \mathcal{A}(2) = \{2, 12\},$$
$$\mathcal{A}(3) = \{3\}, \quad \mathcal{A}(13) = \{13\}, \quad \mathcal{A}(23) = \{23, 123\}.$$

This gives

A	1	2	3	12	13	23
$m_\prec(A)$	0	0.2	0.2	0	0	0.6

which is vertex 7.

6 The case of belief functions

Let us now assume that μ is a belief function. We try to derive results on the vertices $\mathcal{BC}^k(\mu)$. We stress the fact that even if $\mathcal{BC}^k(\mu) \subseteq \mathcal{C}^k(\mu)$, a vertex of $\mathcal{BC}^k(\mu)$ is not necessarily a vertex of $\mathcal{C}^k(\mu)$.

As shown in the previous section, for a compatible order $\prec$, the measure $\mu_\prec$ determines an extreme point of $\mathcal{BC}^k(\mu)$ (Proposition 10). However, not all the vertices can be obtained this way.

Example 1 *Take $|X| = 4$ and consider a belief function μ whose Möbius transform is given by:*

$$m(i) = 0.1,\ m(i, j) = 0.1,\ m(A) = 0 \ \textit{otherwise.}$$

Consider the capacity whose Möbius transform is given by

$$m^*(1) = 0.1, m^*(2) = 0.2, m^*(3) = 0.1, m^*(1, 3) = 0.2,$$

$$m^*(4) = 0.4, m^*(A) = 0 \ \textit{otherwise.}$$

We will show below (Proposition 14) that this is the Möbius transform of an extreme point of $\mathcal{BC}^2(\mu)$. However, it can be checked that this capacity cannot be obtained through a compatible order $\prec$.

Remark that this example does not invalidate Conjecture 1, which is only concerned with vertices of $\mathcal{C}^k(\mu)$.

On the other hand, it can be shown that if $\prec$ is not compatible, the belief function obtained is not necessarily an extreme point:

Example 2 *Consider $|X| = 4$ and the capacity whose Möbius transform is defined by*

$$m(i) = 0.1, m(i, j) = 0.1,\ \forall i, j = 1, 2, 3, 4, i \neq j.$$

Let us consider the 2-additive case and the order $\prec$ given by

$$\{1\} \prec \{1, 2\} \prec \{2\} \prec \{1, 3\} \prec \{2, 3\} \prec$$

$$\{3\} \prec \{3, 4\} \prec \{4\} \prec \{1, 4\} \prec \{2, 4\}.$$

For this order, the corresponding Möbius transform obtained through Equation (1) is:

$$m^*(1) = 0.1, m^*(2) = 0.2, m^*(3) = 0.3, m^*(4) = 0.2,$$

$$m^*(1, 4) = 0.1, m^*(2, 4) = 0.1, m^*(A) = 0, \ \textit{otherwise.}$$

However, this measure is not an extreme point of $\mathcal{BC}^2(\mu)$. It suffices to consider μ_1, μ_2 given by the following Möbius transforms:

$$m_1(4) = 0.1, m_1(1, 4) = 0.2, m_1(A) = m^*(A), \ \textit{otherwise.}$$

$$m_2(4) = 0.3, m_2(1, 4) = 0, m_2(A) = m^*(A), \ \textit{otherwise.}$$

It is clear that $m^*(A) = \dfrac{m_1(A) + m_2(A)}{2}$, $\forall A \subseteq X$. Moreover, they both dominate μ, whence the result.

This leads us to look for an alternative assignment. For a given order $\prec$, let us define recursively the following set function:

$$\underline{m}_{\prec}(B) := \max_{A \in \mathcal{A}(B)} \{\mu(A) - \sum_{K \subseteq A, K \in \mathcal{P}^k_*(X), K \prec B} \underline{m}_{\prec}(K), 0\} \tag{2}$$

if $\mathcal{A}(B) \neq \emptyset$, being 0 otherwise.

Proposition 13 *If $\prec$ is compatible, then the set functions defined in Equations (1) and (2) coincide.*

Let us study the properties of $\underline{\mu}_{\prec}$, the inverse transform of $\underline{m}_{\prec}$. Remark that $\underline{\mu}_{\prec} \geq \mu$ by construction of $\underline{\mu}_{\prec}$. Moreover, as $\underline{m}_{\prec}$ is non-negative, $\underline{\mu}_{\prec}$ is monotone. However, $\underline{\mu}_{\prec}$ is not necessarily a capacity, as it could be the case that $\underline{\mu}_{\prec}(X) > 1$ (see next example).

Example 3 *Let $|X| = 4$ and consider $\mu \in \mathcal{FM}(X)$ whose Möbius transform is given by:*

$$m(i) = 0.1,\ m(i,j) = 0.1,\ m(A) = 0 \text{ otherwise.}$$

Consider the order $\prec$ defined by:

$$1 \prec 1,2 \prec 2 \prec 3 \prec 2,3 \prec 1,3 \prec 2,4 \prec 4 \prec 1,4 \prec 3,4.$$

Then,

$$\underline{m}_{\prec}(1) = 0.1, \underline{m}_{\prec}(1,2) = 0, \underline{m}_{\prec}(2) = 0.2, \underline{m}_{\prec}(3) = 0.1,$$
$$\underline{m}_{\prec}(2,3) = 0, \underline{m}_{\prec}(1,3) = 0.2, \underline{m}_{\prec}(2,4) = 0, \underline{m}_{\prec}(4) = 0.1,$$
$$\underline{m}_{\prec}(1,4) = 0.2, \underline{m}_{\prec}(3,4) = 0.2$$

Therefore, $\underline{\mu}_{\prec}(X) = \sum_{A \subseteq X} \underline{m}_{\prec}(A) = 1.1$.

However, this procedure always leads to a belief function for compatible orders (Proposition 8).

Proposition 14 *Suppose that $\underline{\mu}_{\prec}$, whose Möbius transform $\underline{m}_{\prec}$ is obtained through Eq. (2), is a normalized measure. Then, $\underline{\mu}_{\prec}$ is an extreme point of $\mathcal{BC}^k(\mu)$.*

If the order is compatible, this proposition is just Proposition 10. However, there are vertices of $\mathcal{BC}^k(\mu)$ that cannot be obtained by this procedure, as shown in the following example.

Example 4 *Let $|X| = 4$ and consider $\mu \in \mathcal{BEL}(X)$ whose Möbius transform is given by*

$$m(1,2) = 0.3, m(1,3) = 0.2, m(1,4) = 0.2, m(1,2,3) = 0.1,$$
$$m(1,2,4) = 0.1, m(1,3,4) = 0.1, m(A) = 0 \text{ otherwise.}$$

Consider μ^* whose Möbius transform is given by:

$$m^*(1,2) = 0.35, m^*(1,3) = 0.25, m^*(1,4) = 0.25,$$
$$m^*(X) = 0.15, m(A) = 0 \text{ otherwise.}$$

Thus, $\mu^* \in \mathcal{BEL}(X)$. Moreover, $\mu^* \geq \mu$. Let us now prove that μ^* is a vertex of $\mathcal{BC}^4(\mu)$. For this, let us suppose that there are $\mu_1, \mu_2 \in \mathcal{BC}^4(\mu)$ satisfying

$$\mu^* = \alpha\mu_1 + (1-\alpha)\mu_2,\ \alpha \in (0,1).$$

As μ_1, μ_2 and μ^* are belief functions, it follows that $m_1(A) = m_2(A) = 0$ when $m^*(A) = 0$. Therefore, the only subsets that can attain non-null Möbius inverse for μ_1 and μ_2 are $\{1,2\}, \{1,3\}, \{1,4\}$ and X.

On the other hand, as $\mu^* = \mu$ for $\{1,2,3\}, \{1,3,4\}$ and $\{1,2,4\}$, so are μ_1 and μ_2. But this implies that μ_1 and μ_2 are in the set of capacities whose Möbius transform m' satisfies:

$$\begin{array}{lll} m'(1,2)+ & m'(1,3) = & 0.6 \\ m'(1,2)+ & m'(1,4) = & 0.6 \\ m'(1,3)+ & m'(1,4) = & 0.5 \end{array}.$$

Since the determinant of the system is non-null, there is only one solution for the values of $m'(1,2), m'(1,3)$ and $m'(1,4)$, and this solution is given by the values of m^*.

Finally, as $\sum_{A \subseteq X} m_1(A) = \sum_{A \subseteq X} m_2(A) = \sum_{A \subseteq X} m^*(A) = 1$, we conclude that $m_1(X) = m_2(X) = m^*(X) = 0.15$. Hence, μ^* is an extreme point of $\mathcal{BC}^4(\mu)$.

However, there is no order $\prec$ on $\mathcal{P}^4_*(X)$ such that $\mu^* = \underline{\mu}_{\prec}$. First, remark that if such an order exists, then X is the last subset in the order $\prec$, as $m^*(X) \neq 0$.

The other subsets which need nonempty achievable families are $\{1,2\}, \{1,3\}$ and $\{1,4\}$. On the other hand, if the achievable family is nonempty, then there is a subset A in it such that $\underline{\mu}_{\prec}(A) = \mu(A)$ by construction. For μ^*, the equality $\mu^*(A) = \mu(A)$ is attained at:

$$\begin{array}{ll} \text{Supersets of } \{1,2\} \rightarrow & \{1,2,3\}, \{1,2,4\}, X \\ \text{Supersets of } \{1,3\} \rightarrow & \{1,2,3\}, \{1,3,4\}, X \\ \text{Supersets of } \{1,4\} \rightarrow & \{1,2,4\}, \{1,3,4\}, X. \end{array}$$

Suppose for example that $\{1,2\} \prec \{1,3\} \prec \{1,4\}$. In this case, neither $\{1,2,3\}$ nor $\{1,2,4\}$ belong to $\mathcal{A}(1,2)$, and μ^* cannot be recovered. The same can be done for the other possibilities. Therefore, μ^* cannot be obtained by this procedure.

7 Some results on the derivation of vertices of $\mathcal{BC}^k(\mu)$ from the core

Let us now treat the problem from a different point of view. Based on the results of Section 4, the following can be proved:

Proposition 15 *Let $\mu' \in \bigcup_{\mu^* \in \mathcal{M}_{\Lambda^k_{\subseteq,+}}(\mu)} \mathcal{M}_{\Lambda^k_{\supseteq,+}}(\mu^*)$. Then, there exists $P \in \mathcal{M}_{\Lambda^1_{\subseteq,+}}(\mu)$ such that $m' \in \mathcal{M}_{\Lambda^k_{\supseteq,+}}(P)$.*

Corollary 1 *Let μ' be in the conditions of Proposition 15. Then,*

$$\mu' \in \bigcup_{P \in \mathcal{C}^1(\mu)} \mathcal{M}_{\Lambda^k_{\supseteq,+}}(P).$$

Corollary 2 *Suppose $\mu' \in \mathcal{BC}^k(\mu)$. Then, $\mu' \in \bigcup_{P \in \mathcal{C}^1(\mu)} \mathcal{M}_{\Lambda^k_{\supseteq,+}}(P)$.*

However, it is not true that any extreme point of $\mathcal{BC}^k(\mu)$ can be obtained through $\Lambda^k_{\supseteq,+}$ from a suitable extreme point of the core, as next example shows.

Example 5 *Let us consider $|X| = 4$ and the belief function μ whose Möbius transform is given by:*

$$m(i) = 0.1, m(i,j) = 0.1, \forall i,j \in X,$$

and $m(A) = 0$ otherwise. Consider now μ^* whose Möbius transform is given by

$$m^*(i) = 0.2, \forall i \in X, m^*(X) = 0.2,$$

and $m^*(A) = 0$ otherwise. It is easy to check that $\mu^* \geq \mu$. Note that μ^* cannot be generated by any extreme probability, as if P is a vertex of $\mathcal{BC}^1(\mu)$, we know that P is determined by a permutation σ on X. If i is the first element according to σ, it is $P(i) = m(i) = 0.1$, and then, $m^*(i) = 0.2$ cannot be generated from $m(i)$.

Finally, it can be shown that μ^* is a vertex of $\mathcal{BC}^4(\mu)$. To see this, it suffices to remark that if $\mu^* = \alpha\mu_1 + (1-\alpha)\mu_2$, for some $\mu_1, \mu_2 \in \mathcal{BC}^4(\mu)$, it follows that $m_1(A) = m_2(A)$ when $m^*(A) = 0$. Moreover, $\mu^*(A) = \mu(A)$ if $|A| \geq 3$, and thus, so are μ_1 and μ_2. However, the only measure satisfying these two conditions is μ^*.

Consider $\mu \in \mathcal{BEL}(X)$. We already know by Corollary 2 that any $\mu^* \in \mathcal{BC}^k(\mu)$ can be obtained from the set of dominant probabilities. On the other hand, as remarked in Section 4, $\mathcal{BC}^1(\mu) = \mathcal{M}_{\Lambda^1_{\subseteq,+}}(\mu)$. Moreover, this set is a convex polytope whose vertices are the marginal worth vectors. We can obtain some vertices of $\mathcal{BC}^k(\mu)$ through the following result:

Proposition 16 *Let μ^* be the Möbius transform of an extreme point m^* of $\mathcal{M}_{\Lambda^k_{\supseteq,+}}(P) \cap \mathcal{BC}^k(\mu)$, with P an extreme point of $\mathcal{BC}^1(\mu)$. Then, μ^* is an extreme point of $\mathcal{BC}^k(\mu)$.*

References

[1] J.-P. Barthélemy. Monotone functions on finite lattices: an ordinal approach to capacities, belief and necessity functions. In J. Fodor, B. De Baets, and P. Perny, editors, *Preferences and Decisions under Incomplete Knowledge*, pages 195–208. Physica Verlag, 2000.

[2] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[3] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[4] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339, 1967.

[5] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.

[6] M. Grabisch. On lower and upper approximation of fuzzy measures by k-order additive measures. In *7th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98)*, pages 1577–1584, Paris, France, July 1998.

[7] M. Grabisch. Upper approximation of non-additive measures by k-additive measures — the case of belief functions. In *Proc. of the 1st Int. Symp. on Imprecise Probabilities and Their Applications*, Ghent, Belgium, June 1999.

[8] M. Grabisch. On upper approximation of fuzzy measures by k-order additive measures. In B. Bouchon-Meunier, R. R. Yager, and L. Zadeh, editors, *Information, Uncertainty, Fusion*, pages 105–118. Kluwer Scientific Publ., 2000.

[9] M. Grabisch and P. Miranda. A study of the k-additive core of capacities through achievable families. In *SCIS-ISIS 2006, 3nd Int. Conf. on Soft Computing and Intelligent Systems and 7th Int. Symp. on Advanced Intelligent Systems*, Yokohama, Japan, September 2006.

[10] T. Ichiishi. Super-modularity: applications to convex games and to the greedy algorithm for LP. *J. Econom. Theory*, 25:283–286, 1981.

[11] P. Miranda, M. Grabisch, and P. Gil. Dominance of fuzzy measures by k-additive belief functions. In *Int. Fuzzy Systems Association World Congress (IFSA 2003)*, pages 143–146, Istanbul, Turkey, June 2003.

[12] G. Shafer. *A Mathematical Theory of Evidence*. Princeton Univ. Press, 1976.

[13] L. S. Shapley. Core of convex games. *Int. J. Game Theory*, 1:11–26, 1971.

[14] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

[15] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

Decomposable Signed Fuzzy Measures

Biljana Mihailović
Faculty of Engineering,
University of Novi Sad, Serbia
lica@uns.ns.ac.yu

Endre Pap
Department of Mathematics and Informatics,
University of Novi Sad, Serbia
pape@eunet.yu

Abstract

In this paper some types of decomposable signed fuzzy measures have been presented. We discuss properties of $\bigtriangledown$-and $\oplus$-decomposable signed fuzzy measures. We consider their relationship with bi-capacities.

Keywords: signed fuzzy measure, bi-capacity, pseudo-addition

1 Introduction

There are many generalizations of the concept of a classical (σ-additive) measure studied by various authors, [7, 15, 14, 2, 8]. A generalization of measures by allowing that m can take negative values, leads to the notion of signed measures, see [7]. In [8], authors presented two concepts of generalized measures and integrals, one of them is the concept of S-measures, i.e. S-decomposable fuzzy measures, and (S,T)-integrals, based on an appropriate t-conorm S and a t-norm T. On the other hand, in [9, 4] generalizations of a fuzzy measure, i.e. of a non-negative, monotone set function, vanishing at the empty set, have been introduced. A generalized fuzzy measure, *a signed fuzzy measure*, has been introduced by Liu in [9]. Murofushi et al. in [10] used the term *a non-monotonic fuzzy measure* to denote a general real-valued set function, vanishing at the empty set. As an another generalization of fuzzy measures, the concepts of bi-capacities and bi-cooperative games have been introduced in [4].
This papers discusses decomposable signed fuzzy measures. In the next section the short overview of basic notions and definitions is given. Section 3 introduces $\oplus$- and $\bigtriangledown$-decomposable set functions. Several properties of such set functions are investigated and concrete examples are given. Section 4 discusses the relationship of decomposable signed fuzzy measures with special classes of bi-capacities.

2 Preliminaries

Let X be a universal set. Let $\mathcal{P}(X)$ be the class of subsets of the universal set X.

Definition 1 *[9, 11] A real-valued set function $m : \mathcal{P}(X) \to \mathbb{R}$, is a signed fuzzy measure if it satisfies*
(i) $m(\emptyset) = 0$
(ii) (RM) *If $E, F \in \mathcal{P}(X)$, $E \cap F = \emptyset$, then*
a) $m(E) \geq 0$, $m(F) \geq 0$, $m(E) \vee m(F) > 0 \Rightarrow$

$$m(E \cup F) \geq m(E) \vee m(F);$$

b) $m(E) \leq 0$, $m(F) \leq 0$, $m(E) \wedge m(F) < 0 \Rightarrow$

$$m(E \cup F) \leq m(E) \wedge m(F);$$

c) $m(E) > 0$, $m(F) < 0 \Rightarrow$

$$m(F) \leq m(E \cup F) \leq m(E).$$

The property (RM) of m is called the revised monotonicity, see [11, 13].

Definition 2 *The dual set function of a real-valued set function m, $m : \mathcal{P}(X) \to \mathbb{R}$ is defined by $\bar{m}(E) = m(X) - m(\bar{E})$, where $\bar{E}$ denotes the complement set of E, $\bar{E} = X \setminus E$.*

Obviously, if m is a fuzzy measure, $\bar{m}$ is a fuzzy measure, too. However, if m is a signed fuzzy measure, its dual set function $\bar{m}$ need not be a signed fuzzy measure and this fact will be discussed in the next section.

The *symmetric maximum* $\bigtriangledown : [-1,1]^2 \to [-1,1]$, originally introduced in [3], can be represented by:

$$a \bigtriangledown b = (|a| \vee |b|) \operatorname{sign}(a+b),$$

The *pseudo-addition* $\oplus : [-1,1]^2 \to [-1,1]$, associated to a continuous t-conorm S, introduced in [6], is defined by:

$$a \oplus b = \begin{cases} S(a,b) & (a,b) \in [0,1]^2 \\ -S(-a,-b) & (a,b) \in [-1,0]^2 \\ d & (a,b) \in [0,1] \times (-1,0], a \geq -b \\ e & (a,b) \in [0,1] \times (-1,0], a \leq -b \\ f & a = 1, b = -1 \end{cases}$$

where $f = 1$ or $f = -1$, $d = inf\{\, c \,|\, S(-b, c) \geq a\}$ and $e = -inf\{\, c \,|\, S(a, c) \geq -b\}$, and the remaining cases being determined by the commutativity of $\oplus$.
The binary operations, $\bigtriangledown$ and $\oplus$ are commutative, isotonic, with neutral element 0. The second one is associative and the first one is not. For more details we recommend [4, 6].

In the next example, we consider a set function which satisfies the conditions given in Definition 1, i.e. a signed fuzzy measure.

Example 1 *Let X be a set of k elements. Let $A, B \subset X$ such that $X = A \cup B$, $A \cap B = \emptyset$, A, $B \neq \emptyset$ and $card(A) = n$, $card(B) = k - n$. We define a set function $m : \mathcal{P}(X) \to [-1, 1]$ by:*

$$m(E) = \begin{cases} 0, & E = X \\ \left(\frac{card(A\cap E)}{n}\right) \oplus_{S_P} \left(-\frac{card(B\cap E)}{k-n}\right), & else. \end{cases}$$

$S_P : [0, 1]^2 \to [0, 1]$ is the probabilistic sum, defined by $S_P(x, y) = x + y - xy$. The set function m is a signed fuzzy measure.

As an example, we consider $A = \{a_1, a_2, a_3\}$ and $B = \{b_1, b_2, b_3, b_4\}$. We have, e.g. $m(A) = 1$, $m(B) = -1$, $m(X) = 0$, $m(\{a_1, b_1\}) = \frac{1}{3} \oplus_{S_P} -\frac{1}{4} = \frac{1}{9}$, $m(\{a_1, a_2, b_1\}) = \frac{2}{3} \oplus_{S_P} -\frac{1}{4} = \frac{5}{9}$, $m(\{a_1, a_2\}) = \frac{2}{3} \oplus_{S_P} 0 = \frac{2}{3}$, etc. Let us indicate the sets A and B, the sets of two types of drugs influential in duration of some medical procedures(e.g. the duration of lungs harmful fluidity throwing out). It is known that the first one, A, decreases, and the second one, B, increases the duration of the considered procedure. It is also known that simultaneously taking drugs A and B has no effects in the duration of the procedure, i.e. the duration is unchanged. For the reasons of patients health, sometimes the procedure must be slower and sometimes, faster. The set function m in the above example is a mathematical model for the described situation.

3 Decomposable signed fuzzy measures

Definition 3 *A set function $m : \mathcal{P}(X) \to [-1, 1]$, is*
i) a $\bigtriangledown$-decomposable set function if it satisfies

$$m(E \cup F) = m(E) \bigtriangledown m(F)$$

for all E, $F \in \mathcal{P}(X)$, $E \cap F = \emptyset$.
ii) a $\oplus$-decomposable set function if it satisfies

$$m(E \cup F) = m(E) \oplus m(F)$$

for all E, $F \in \mathcal{P}(X)$, $E \cap F = \emptyset$.

Example 2 *Let $X = \{x_1, x_2, \ldots, x_n\}$. Let m be a set function $m : \mathcal{P}(X) \longrightarrow [-1, 1]$ with $m(\emptyset) = 0$, defined by:*

$$m(E) = \begin{cases} \frac{1}{\min\limits_{x_i \in E} i} & if \ \min\limits_{x_i \in E} i = 2k \\ -\frac{1}{\min\limits_{x_i \in E} i} & if \ \min\limits_{x_i \in E} i = 2k + 1 \end{cases}$$

m is a $\bigtriangledown$-decomposable set function.

Example 3 *Let m be a set function defined on $\mathcal{B}([-1, 1])$, the class of Borel subsets of $[-1, 1]$, and λ the usual Lebesgue measure. We denote $\lambda_1(E) = \lambda(E \cap [0, 1])$ and $\lambda_2(E) = \lambda(E \cap [-1, 0))$. We define $m : \mathcal{B}([-1, 1]) \longrightarrow [-1, 1]$ by*

$$m(E) = sign(\lambda_1(E) - \lambda_2(E)) \left(1 - e^{-|\lambda_1(E) - \lambda_2(E)|}\right).$$

m is a $\oplus_{S_P}$-decomposable signed fuzzy measure.

In the sequel of this section we will consider a signed fuzzy measure m with $m(X) = 0$. We will examine when its dual set function $\bar{m}$ is a signed fuzzy measure, too. Note that for a non-negative (non-positive) signed fuzzy measure m, the condition $m(X) = 0$ implies $m(E) = 0$ for all $E \in \mathcal{P}(X)$. We suppose that $m : \mathcal{P}(X) \to \mathbb{R}$ is a signed fuzzy measure of a non-constant sign. We easily obtain the next lemma by the definition of a signed fuzzy measure and the condition $m(X) = 0$.

Lemma 1 *Let m be a signed fuzzy measure, $m(X) = 0$. $m(E)$ and $m(\bar{E})$ are of opposite sign values, i.e. $(\forall E \in \mathcal{P}(X))\ (m(E) > 0 \ \Leftrightarrow \ m(\bar{E}) < 0)$.*

Definition 4 *We say that a real-valued set function m, $m(\emptyset) = 0$ satisfies the intersection property if for all $E, F \in \mathcal{P}(X)$, $E \cap F \neq \emptyset$ and $E \cup F = X$ we have*

$$\begin{aligned} &a)\ m(E) \geq 0,\ m(F) \geq 0,\ m(E) \vee m(F) > 0 \quad \Rightarrow \\ &\qquad m(E \cap F) \geq m(E) \vee m(F); \\ &b)\ m(E) \leq 0,\ m(F) \leq 0,\ m(E) \wedge m(F) < 0 \quad \Rightarrow \\ &\qquad m(E \cap F) \leq m(E) \wedge m(F); \\ &c)\ m(E) > 0,\ m(F) < 0 \quad \Rightarrow \\ &\qquad m(F) \leq m(E \cap F) \leq m(E). \end{aligned}$$

We have the next theorem.

Proposition 1 *Let m be a signed fuzzy measure, $m(X) = 0$. m satisfies the intersection property if and only if the dual set function $\bar{m}$ of m is a signed fuzzy measure.*

`Proof.` Let m be a signed fuzzy measure such that $m(X) = 0$.

($\Longrightarrow$) First, we suppose that m satisfies the intersection property. We will prove that $\bar{m}$ is a signed fuzzy measure.
(i) Directly by the definition of $\bar{m}$ we have $\bar{m}(\emptyset) = 0$.
(ii) In order to prove condition (RM) *a)* let $E, F \in \mathcal{P}(X)$ such that $E \cap F = \emptyset$ and $\bar{m}(E) \geq 0$, $\bar{m}(F) \geq 0$, $\bar{m}(E) \vee \bar{m}(F) > 0$. We have $\bar{E} \cup \bar{F} = X$ and $m(\bar{E}) \leq 0$, $m(\bar{F}) \leq 0$ and $m(\bar{E}) \wedge m(\bar{F}) < 0$. $(*)$ If we suppose that $\bar{E} \cap \bar{F} = \emptyset$ then we have $F = \bar{E}$. By Lemma 1. we obtain that the values $m(F)$ and $m(\bar{F})$ are of opposite sign values and it is a contradiction with $(*)$. Therefore, $\bar{E} \cap \bar{F} \neq \emptyset$. By the intersection property of m we have:

$$\begin{aligned} & m(\bar{E} \cap \bar{F}) \leq m(\bar{E}) \wedge m(\bar{F}) \Longleftrightarrow \\ \Longleftrightarrow \quad & m(\overline{E \cup F}) \leq m(\bar{E}) \wedge m(\bar{F}) \\ \Longleftrightarrow \quad & -\bar{m}(E \cup F) \leq (-\bar{m}(E)) \wedge (-\bar{m}(F)) \\ \Longleftrightarrow \quad & \bar{m}(E \cup F) \geq \bar{m}(E) \vee \bar{m}(F). \end{aligned}$$

Hence, we have that $\bar{m}$ satisfies condition (RM) a). Similarly we obtain that $\bar{m}$ satisfies conditions (RM) b) and c), hence, $\bar{m}$ is a signed fuzzy measure.
($\Longleftarrow$) Let $\bar{m}$ be a signed fuzzy measure, i.e. $\bar{m}$ is a revised monotone set function and $\bar{m}(\emptyset) = 0$. We obtain the claim directly by the definition of the intersection property and the above consideration. □

Directly by the definitions we have that $\oplus$- and $\bigotimes$-decomposable set functions are signed fuzzy measures. The condition of the intersection property we will replace by the condition $m(E \cap F) = m(E) \oplus m(F)$ for all $E, F \in \mathcal{P}(X)$, $E \cup F = X$ for a $\oplus$-decomposable set function and respectively with $m(E \cap F) = m(E) \bigotimes m(F)$ for all $E, F \in \mathcal{P}(X)$, $E \cup F = X$ for a $\bigotimes$-decomposable set function. Hence, we have the next corollaries of Proposition 1.

Corollary 1 *The dual set function $\bar{m}$ of a $\bigotimes$-decomposable set function m, $m(\emptyset) = m(X) = 0$ is a $\bigotimes$-decomposable iff $m(E \cap F) = m(E) \bigotimes m(F)$ for all $E, F \in \mathcal{P}(X)$, $E \cup F = X$.*

Corollary 2 *Let m be a $\bigotimes$-decomposable set function, $m(\emptyset) = m(X) = 0$, such that $m(E \cap F) = m(E) \bigotimes m(F)$ for all $E, F \in \mathcal{P}(X)$, $E \cup F = X$. Then m is a self-dual set function, i.e. $m = \bar{m}$.*

Analogously, we have the next two corollaries related to $\oplus$- decomposable set functions.

Corollary 3 *The dual set function $\bar{m}$ of a $\oplus$-decomposable set function m, $m(\emptyset) = m(X) = 0$, $1 \notin Ran(m)$, is a $\oplus$-decomposable iff $m(E \cap F) = m(E) \oplus m(F)$ for all $E, F \in \mathcal{P}(X)$, $E \cup F = X$.*

Corollary 4 *Let m be a $\oplus$-decomposable set function, $m(\emptyset) = m(X) = 0$, and $1 \notin Ran(m)$, such that $m(E \cap F) = m(E) \oplus m(F)$ for all $E, F \in \mathcal{P}(X)$, $E \cup F = X$. Then m is a self-dual set function, i.e. $m = \bar{m}$.*

As an illustration, in the next example a ternary signed fuzzy measure m which is $\oplus$-decomposable, where $\oplus$ is the pseudo-addition related to an arbitrary continuous t-conorm S, is given. The values of a decomposable set function m are determined by its values on singletons. The considered signed fuzzy measure m is a self-dual set function.

Example 4 *Let $X = \{x_1, x_2, x_3, x_4\}$ and let m be a $\oplus$-decomposable set function with $m(\emptyset) = 0$, $m : \mathcal{P}(X) \to [-1, 1]$, defined by:*

$$\begin{aligned} m(\{x_1\}) = a, \qquad & m(\{x_2\}) = 0, \\ m(\{x_3\}) = -a, \qquad & m(\{x_4\}) = 0. \end{aligned}$$

where $0 < a < 1$. Obviously, m is a self-dual signed fuzzy measure.

4 Decomposable signed fuzzy measures and bi-capacities

In this section we consider the relationship of a decomposable signed fuzzy measure m and a bi-capacity $\mathbf{m}$. Let X be a finite set. Bi-capacities are a real-valued functions defined on the set $\mathcal{Q}(X) = \{(E, F) \in \mathcal{P}(X) \times \mathcal{P}(X) | \, E \cap F = \emptyset\}$, which are non-decreasing in the first variable and non-increasing in the second one.

Definition 5 *[4] A bi-capacity is a real-valued function $\mathbf{m} : \mathcal{Q}(X) \longrightarrow \mathbb{R}$ with the following properties:*
(BC1) $\mathbf{m}(\emptyset, \emptyset) = 0$,
(BC2) $E \subset F \Longrightarrow \mathbf{m}(E, \cdot) \leq \mathbf{m}(F, \cdot)$ *and* $\mathbf{m}(\cdot, E) \geq \mathbf{m}(\cdot, F)$ *for all* $(E, F) \in \mathcal{Q}(X)$

The structure $(\mathcal{Q}(X), \leq)$ is the lattice of the type 3^n, for more details see [4]. For $n = 2$ and $X = \{1, 2\}$ the Hasse diagram of $(\mathcal{Q}(X), \leq)$ is

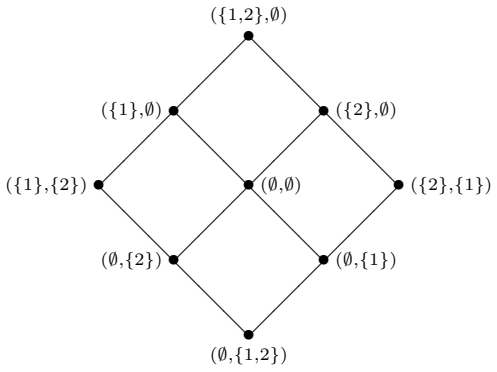


Figure 1. $(\mathcal{Q}(X), \leq)$ for $n = 2$.

A bi-capacity $\mathbf{m}$ is $\vee$-CPT type if there exist two fuzzy measures m_1 and m_2 such that $\mathbf{m}(E,F) = m_1(E) \bigcirc\!\!\!\!\vee (-m_2(F))$ for all $(E,F) \in \mathcal{Q}(X)$ (see [5]).

Proposition 2 *Let $m : \mathcal{P}(X) \to [-1,1]$ be a $\bigcirc\!\!\!\!\vee$-decomposable signed fuzzy measure. Then the set function $\mathbf{m} : \mathcal{Q}(X) \to [-1,1]$ defined by*

$$\mathbf{m}(E,F) = \max_{A\subset E, m(A)\geq 0} m(A) \bigcirc\!\!\!\!\vee \min_{B\subset F, m(B)\leq 0} m(B) \quad (1)$$

is a bi-capacity.

Proof. We will show that $\mathbf{m}$ defined as above satisfies conditions (BC1) and (BC2).
(BC1) $\mathbf{m}(\emptyset,\emptyset) = 0$,
(BC2) Let $E \subset F$. For all $G \subset X$ such that (E,G), $(F,G) \in \mathcal{Q}(X)$ we have:

$$\begin{aligned} \mathbf{m}(E,G) &= \max_{A\subset E, m(A)\geq 0} m(A) \bigcirc\!\!\!\!\vee \min_{B\subset G, m(B)\leq 0} m(B) \\ &\leq \max_{A\subset F, m(A)\geq 0} m(A) \bigcirc\!\!\!\!\vee \min_{B\subset G, m(B)\leq 0} m(B) \\ &= \mathbf{m}(F,G) \end{aligned}$$

For all $H \subset X$, such that (H, E),$(H, F) \in \mathcal{Q}(X)$

$$\begin{aligned} \mathbf{m}(H,E) &= \max_{A\subset H, m(A)\geq 0} m(A) \bigcirc\!\!\!\!\vee \min_{B\subset E, m(B)\leq 0} m(B) \\ &\geq \max_{A\subset H, m(A)\geq 0} m(A) \bigcirc\!\!\!\!\vee \min_{B\subset F, m(B)\leq 0} m(B) \\ &= \mathbf{m}(H,F) \end{aligned}$$

□

We shall give the next simple example.

Example 5 *Let $X = \{1,2\}$ and let $m : \mathcal{P}(X) \to [-1,1]$ be a $\bigcirc\!\!\!\!\vee$-decomposable set function, defined by $m(\emptyset) = 0$, $m(\{1\}) = 0.2$, $m(\{2\}) = -0.3$ and $m(\{1,2\}) = -0.3$. By Proposition 2, the set function $\mathbf{m}$ determined by Eq.(1) is a bi-capacity. The bi-capacity $\mathbf{m}$ associated to m, is defined on $\mathcal{Q}(X)$, $card(\mathcal{Q}(X)) = 3^2 = 9$, illustrated by Figure 1.*
$\mathbf{m}(\{1,2\},\emptyset) = \mathbf{m}(\{1\},\emptyset) = 0.2$,
$\mathbf{m}(\emptyset,\{2\}) = \mathbf{m}(\emptyset,\{1,2\}) = \mathbf{m}(\{1\},\{2\}) = -0.3$, *and* $\mathbf{m}(E,F) = 0$ *in the remaining cases.*

Proposition 3 *Let $\mathbf{m} : \mathcal{Q}(X) \to [-1,1]$ be a bi-capacity such that for any nonempty index set $\mathcal{I} \subset \{1,2,\ldots,n\}$, we have*

$$\max_{i\in\mathcal{I}} \mathbf{m}(\{x_i\},\emptyset) \neq -\min_{i\in\mathcal{I}} \mathbf{m}(\emptyset,\{x_i\}). \quad (2)$$

Then the set function $m : \mathcal{P}(X) \to [-1,1]$ defined by

$$m(E) = \max_{x_i\in E} \mathbf{m}(\{x_i\},\emptyset) \bigcirc\!\!\!\!\vee \min_{x_i\in E} \mathbf{m}(\emptyset,\{x_i\})$$

is a $\bigcirc\!\!\!\!\vee$-decomposable signed fuzzy measure.

Proof. Let us suppose that $\mathbf{m} : \mathcal{Q}(X) \to [-1,1]$ is such that inequality (2) holds for any nonempty set $\mathcal{I} \subset \{1,2,\ldots,n\}$. Let E, $F \in \mathcal{P}(X)$, $E \cap F = \emptyset$. Then we have

$$\begin{aligned} &m(E\cup F) = \\ &= \max_{x_i\in E\cup F} \mathbf{m}(\{x_i\},\emptyset) \bigcirc\!\!\!\!\vee \min_{x_i\in E\cup F} \mathbf{m}(\emptyset,\{x_i\}) = \\ &= \left(\max_{x_i\in E} \mathbf{m}(\{x_i\},\emptyset) \bigcirc\!\!\!\!\vee \max_{x_i\in F} \mathbf{m}(\{x_i\},\emptyset)\right) \bigcirc\!\!\!\!\vee \\ &\qquad \bigcirc\!\!\!\!\vee \left(\min_{x_i\in E} \mathbf{m}(\emptyset,\{x_i\}) \bigcirc\!\!\!\!\vee \min_{x_i\in F} \mathbf{m}(\emptyset,\{x_i\})\right) = \\ &= \left(\max_{x_i\in E} \mathbf{m}(\{x_i\},\emptyset) \bigcirc\!\!\!\!\vee \min_{x_i\in E} \mathbf{m}(\emptyset,\{x_i\})\right) \bigcirc\!\!\!\!\vee \\ &\qquad \bigcirc\!\!\!\!\vee \left(\max_{x_i\in F} \mathbf{m}(\{x_i\},\emptyset) \bigcirc\!\!\!\!\vee \min_{x_i\in F} \mathbf{m}(\emptyset,\{x_i\})\right) = \\ &\qquad = m(E) \bigcirc\!\!\!\!\vee m(F) \end{aligned}$$

□

Corollary 5 *Let $\mathbf{m} : \mathcal{Q}(X) \to [-1,1]$ be a bi-capacity of $\vee$-CPT type, such that m_1 and m_2 are two $\vee$-decomposable fuzzy measures and $m_1(E) \neq m_2(E)$ for all $E \in \mathcal{P}(X)$. Then the set function $m : \mathcal{P}(X) \to [-1,1]$ defined by*

$$m(E) = \mathbf{m}(E,\emptyset) \bigcirc\!\!\!\!\vee \mathbf{m}(\emptyset,E)$$

is a $\bigcirc\!\!\!\!\vee$-decomposable signed fuzzy measure.

Proof. Obviously, two $\vee$-decomposable fuzzy measures m_k, $k = 1,2$ satisfy for all $E \in \mathcal{P}(X)$:

$$m_k(E) = \bigvee_{x_i\in E} m_k(\{x_i\}).$$

It is clear that $\forall\, i$, $m_1(\{x_i\}) = \mathbf{m}(\{x_i\},\emptyset)$ and $m_2(\{x_i\}) = -\mathbf{m}(\emptyset,\{x_i\})$. Hence, by Proposition 3, we have the claim. □

We present two Propositions for a $\oplus$-decomposable signed fuzzy measure, where S is an arbitrary continuous t-conorm. The proofs are similarly to the proofs of Proposition 2 and Proposition 3, we remark that $\oplus$ is associative on $[-1,1]^2$, so in this case, a condition similar to condition (2) is omitted here.

Proposition 4 *Let $m : \mathcal{P}(X) \to [-1,1]$ be a $\oplus$-decomposable signed fuzzy measure. Then the set function $\mathbf{m} : \mathcal{Q}(X) \to [-1,1]$ defined by*

$$\mathbf{m}(E,F) = \left(\bigoplus_{A\subset E, m(A)\geq 0} m(A)\right) \oplus \left(\bigoplus_{B\subset F, m(B)\leq 0} m(B)\right)$$

is a bi-capacity.

Proposition 5 *Let $\mathbf{m} : \mathcal{Q}(X) \to [-1,1]$ be a bi-capacity, then the set function $m : \mathcal{P}(X) \to [-1,1]$ defined by*

$$m(E) = \left(\bigoplus_{x_i\in E} \mathbf{m}(\{x_i\},\emptyset)\right) \oplus \left(\bigoplus_{x_i\in E} \mathbf{m}(\emptyset,\{x_i\})\right)$$

is a $\oplus$-decomposable signed fuzzy measure.

Corollary 6 *Let $\mathbf{m} : \mathcal{Q}(X) \to [-1, 1]$ be a bi-capacity of $\oplus$-CPT type, i.e. $\mathbf{m}(E, F) = m_1(E) \oplus (-m_2(F))$, for all $(E, F) \in \mathcal{Q}(X)$, such that m_1 and m_2 are two S-decomposable fuzzy measures, then the set function $m : \mathcal{P}(X) \to [-1, 1]$ defined by*

$$m(E) = \mathbf{m}(E, \emptyset) \oplus \mathbf{m}(\emptyset, E)$$

is a $\oplus$-decomposable signed fuzzy measure.

Acknowledgement

The work has been supported by the project MNTRS 144012 and the project "Mathematical Models for Decision Making under Uncertain Conditions and Their Applications" supported by Vojvodina Provincial Secretariat for Science and Technological Development.

References

[1] D. Denneberg: *Non-additive Measure and Integral.* Kluwer Academic Publishers, Dordrecht, 1994.

[2] D. Dubois, H. Prade: *A class of fuzzy measures based on triangular norms- a general framework for the combination of uncertain information*, Internat. J. Gen. Systems **8**, 1982, 43-61.

[3] M. Grabisch: *The Symmetric Sugeno Integral.* Fuzzy Sets and Systems **139**, 2003, 473-490.

[4] M. Grabisch, Ch. Labreuche: *Bi-capacities I:-definition, Möbius transform and interaction*, Fuzzy Sets and Systems **151**, 2005, 211-236.

[5] M. Grabisch: *Aggregation on bipolar scales*, in: H. de Swart and E. Orlowska and M. Roubens and G. Schmidt (Eds.): *Theory and Applications of Relational Structures as Knowledge Instruments II*, Springer, 2006.

[6] M. Grabisch, J. L. Marichal, R. Mesiar, E. Pap: *Aggregation Functions for Decision Making* (book under preparation)

[7] P. R. Halmos: *Measure Theory.* Springer-Verlag New York-Heidelberg-Berlin, 1950.

[8] E. P. Klement, R. Mesiar, E. Pap: *Triangular Norms.* Kluwer Academic Publishers, Dordrecht, 2000.

[9] X. Liu: *Hahn decomposition theorem for infinite signed fuzzy measure,* Fuzzy Sets and Systems **57**, 1993, 189-212.

[10] T. Murofushi, M. Sugeno and M. Machida: *Non-monotonic fuzzy measures and the Choquet integral*, Fuzzy Sets and Systems **64**, 1994,73-86.

[11] E. Pap: *Null-Additive Set Functions.* Kluwer Academic Publishers, Dordrecht, 1995.

[12] E. Pap, B. Mihailović: *A representation of a comonotone-$\bigtriangledown$-additive and monotone functional by two Sugeno integrals*, Fuzzy Sets and Systems **155**, 2005, 77-88.

[13] E. Pap, B. Mihailović: *Representation of the utility functional by two fuzzy integrals,* Proc. IPMU'06, Paris, France, 2006, 1710-1717.

[14] S. Weber: *$\perp$-decomposable measure and integrals of Archimedian t-conorms $\perp$*, J. Math. Anal. Appl. **101**, 1984, 114-138.

[15] L. Zadeh: *Fuzzy sets as a basis for a theory of possibility*, Fuzzy Sets ans Systems **1**, 1978, 3-28.

Session 5

Fuzzy Measures and Integrals: Applications in Decision Making and Game Theory – M. Grabisch

Extending the Choquet integral

Giovanni Rossi
Department of Computer Science, University of Bologna
Mura Anteo Zamboni 7, 40126 Bologna, Italy, giorossi@cs.unibo.it

Abstract

In decision under uncertainty, the Choquet integral yields the expectation of a random variable with respect to a fuzzy measure (or non-additive probability or capacity). In general, for the discrete setting, this technique allows to integrate functions taking values on a finite n-set with respect to a (fuzzy) measure taking values on subsets of such a set. Yet, the integrand may well be treated as an additive function taking values on subsets itself: the value associated with each subset is simply the sum of the values associated with all the atoms (or 1-cardinal subsets) in that subset. The Choquet technique is here extended to the case where the integrand, just like the measure, is a non-additive function taking values on subsets itself. The resulting aggregation operator is an extension of the Choquet integral: the former coincides with the latter whenever the integrand is additive. Four such extensions are provided, two of which are obtained by means of the Möbius inversion of the integrand and the (fuzzy) measure with respect to which integration is performed. In all cases, the resulting integral is an extension of the measure: it coincides with this latter on the vertices of the n-dimensional unit hypercube. Yet, one of these extensions also inherits another main feature of the (traditional) Choquet integral: if the fuzzy measure is convex, then this extended Choquet integral equals its minimum over all probabilities in the core of the measure. The general technique applies to both monotone and antitone integrands, and when the integrand is real-valued (i.e., taking both positive and negative values) it allows for both a symmetric and an asymmetric form. Two conceivable applications are provided. One furnishes an expectation of diversity in a random sample of a known population. Here the integrand is a *diversity function*, which is monotone by construction. The other application furnishes the certainty equivalent for a problem of decision making under uncertainty where the decision maker has some belief about what information (in the form of an event containing the 'true' state) will be available before taking action. Here the integrand is antitone.

MSC2000 classification numbers: 49N30, 62C10.

Keywords: fuzzy measure, Choquet integral, Möbius inversion, uncertainty, information, certainty equivalent, diversity functions.

1 Introduction

The (discrete) Choquet integral is an aggregation operator very useful in DUU (decision under uncertainty) and MCDM (multicriteria decision making). In the former case, a DM (decision maker) has to take action in response to an unknown state of Nature. Preferences take the form of a utility function associating a real number to each pair consisting of a state and an action. The set of utility values attained on all pairs consisting of a state and a fixed action is a random variable. This leads to rank actions according to the associated expectation, requiring, in turn, some beliefs about what state will occur. If beliefs take the form of a probability or additive measure, then actions get ranked according to the EU (expected utility) model. On the other hand, if beliefs take the form of a capacity or fuzzy measure, then actions get ranked according to the CEU (Choquet expected utility) model (see [3] and [10]). In both cases, for each action, the DM performs an aggregation, based on beliefs, of all the utility values such an action may yield. An important fact inspiring the present paper is that the CEU model constitutes, in fact, an extension of the EU one. In particular, fuzzy measures comprehend additive ones as special cases, and whenever the measure is additive the Choquet expectation coincides with the traditional one.

In MCDM, the DM has to choose within a given (finite) set of alternatives according to their score on different evaluation criteria. In this case, for each alternative, the DM has to aggregate the scores that such an alternative attains on each criterion. In particular, if criteria display no interaction, then aggregation can be performed most naturally through a weighted average. More precisely, for the DM each criterion has its own importance or (positive) weight. Such weights may well be normalized so to add up to unity. Accordingly, each alternative can be evaluated according to the weighted average of the scores it attains on criteria. On the other hand, the situation where criteria display interaction is formalized by specifying a weight for each subset of criteria. In this case, in order to keep into account such an interaction, aggregation can no longer be performed through a weighted average. Conversely, the Choquet integral

becomes the needed aggregation operator, as the collection of subsets' weights constitutes a fuzzy measure. Here again, this latter aggregation technique appears as an extension: as soon as the fuzzy measure is additive, in which case criteria display no interaction, the Choquet integral reduces to a weighted average (see [6]). A great deal of attention has been paid to the evaluation of negative scores, in which case the Möbius inversion of the measure yields a useful representation. In fact, if the integrand is real-valued, then the Choquet integral allows for both a symmetric and an asymmetric form (see [7]).

The aim of this paper is to present a novel aggregation technique which constitutes, in fact, an extension of the Choquet integral. More precisely, given some n-set N, this technique allows to aggregate the 2^n values taken by a (monotone or antitone) function defined on subsets $A \subseteq N$ w.r.t. (with respect to) a fuzzy measure, taking values on such subsets as well. Yet, if the former function (i.e., the integrand) is additive, then this technique yields the same result as the Choquet integral.

Such an extension is shown to inherit two main features from the Choquet integral. Firstly, for given integrand, the extended Choquet integral (as defined below) w.r.t. a convex measure equals the minimum of this integral over all probabilities in the core of the measure. Secondly, this novel aggregation operator is an extension of the measure: it coincides with this latter on the vertices of the n-dimensional unit hypercube. In addition, it may be represented in terms of the Möbius inversions of both the integrand and the measure, and when the former is real-valued it allows for both a symmetric and an asymmetric form.

2 Preliminaries

Let $N = \{1, \ldots, n\}$ be a finite set (of states or criteria). Any function $f : N \to \mathbb{R}$ is, in fact, a collection $f(1), \ldots, f(n)$ of n numbers. A weighted average or expectation (if f is a random variable) of these numbers takes form $E_p(f) = \sum_{i \in N} p(i) \cdot f(i)$, where $p(i) \geq 0$ for all $i \in N$ and $\sum_{i \in N} p(i) = 1$, that is, p is a probability. Perhaps, this is the most natural way of aggregating such n numbers $f(1), \ldots, f(n)$. Also, for $g, f : N \to \mathbb{R}$ and given p, if $E_p(g) \geq E_p(f)$, then g is ranked no lower than f.

The Choquet integral is a more sophisticated aggregation technique. It uses fuzzy measures (or capacities, see [3], or non-additive probabilities, see [8]). These are *monotone* $\gamma : 2^N \to [0, 1]$ (i.e., $A \subseteq B \in 2^N$ implies $\gamma(A) \leq \gamma(B)$, where $2^N = \{A : A \subseteq N\}$) such that $\gamma(\emptyset) = 1 - \gamma(N) = 0$. In DUU subsets $A \in 2^N$ are *events*, and fuzzy measures are interpreted as follows: $\gamma(A)$ quantifies the belief that the 'true' state i (that will occur) satisfies $i \in A$. On the other hand, in MCDM $\gamma(A)$ quantifies the importance (or worth) of all criteria $i \in A$ considered together.

Given that for functions taking both positive and negative values the Choquet integral may be computed either symmetrically or else asymmetrically (see [7] and below), firstly consider the case of integrands $f : N \to \mathbb{R}_+$ taking only positive values. Formally, the Choquet integral of f w.r.t. γ is

$$\int_N^C f d\gamma = \sum_{1 \leq i \leq n} \big[f((i)) - f((i-1))\big] \cdot \gamma(\{(i), \ldots, (n)\})$$

$$= \sum_{1 \leq i \leq n} f((i)) \cdot \big[\gamma(\{(i), \ldots, (n)\}) - \gamma(\{(i+1), \ldots, (n)\})\big]$$

where $(\cdot) : N \to N$ is any permutation such that $f((n)) \geq \cdots \geq f((i)) \geq \cdots \geq f((1)) \geq f((0)) := 0$. Also, $\gamma(\{(i+1), \ldots, (n)\})$ clearly vanishes for $i = n$. Measure γ is additive when for all $A, B \in 2^N$

$$\gamma(A \cap B) + \gamma(A \cup B) = \gamma(A) + \gamma(B),$$

i.e., when $\gamma(A) = \sum_{i \in A} \gamma(\{i\})$ for all $A \in 2^N$. If γ is additive, then the Choquet integral reduces to

$$\int_N^C f d\gamma = \sum_{1 \leq i \leq n} \gamma(\{i\}) \cdot f(i).$$

Out of the above two expressions for the Choquet integral, the latter links this aggregation operator and the *core* of *convex* (or *supermodular*) measures (see [5], theorem 2.2). The core $\mathcal{C}(\gamma)$ of a measure γ is the set of probabilities p such that $\sum_{i \in A} p(i) \geq \gamma(A)$ for all $A \in 2^N$ (and thus $\sum_{i \in N} p(i) = \gamma(N) = 1$). In general, the core may well be empty. Yet, if the measure is convex, that is, if

$$\gamma(A \cup B) + \gamma(A \cap B) \geq \gamma(A) + \gamma(B)$$

for all $A, B \in 2^N$, then $\mathcal{C}(\gamma) \neq \emptyset$ and

$$\int_N^C f d\gamma = \bigwedge_{p \in \mathcal{C}(\gamma)} E_p(f)$$

for any integrand f, where $\wedge$ (and $\vee$) denote the min or inf (and max or sup) operator.

The characteristic function $\chi_A : N \to \{0, 1\}$ associated with each subset $A \in 2^N$ is defined by $\chi_A(i) = 1$ if $i \in A$ and $\chi_A(i) = 0$ if $i \notin A$. Hence, there is a bijection between 2^N and the set $\{0, 1\}^n$ of vertices of the n-dimensional unit hypercube $[0, 1]^n$. This yields that the Choquet integral is an extension of

the measure w.r.t. which integration is performed, as the former coincides with the latter on vertices $\chi_A \in \{0,1\}^n, A \in 2^N$. That is, for $A = \{i_1, \ldots, i_{|A|}\}$, $\int_N^C \chi_A d\gamma =$

$$\sum_{1 \leq h \leq |A|} \left[\gamma(\{i_h, \ldots, i_{|A|}\}) - \gamma(\{i_{h+1}, \ldots, i_{|A|}\})\right] = \gamma(A)$$

(with $\gamma(\{i_{h+1}, \ldots, i_{|A|}\})$ vanishing for $h = |A|$). In words, as the integrand varies over all $\chi_A, A \in 2^N$, the Choquet integral of such functions w.r.t. to any measure γ yields precisely $\gamma(A), A \in 2^N$. Also, there are $|A|!(n-|A|)!$ distinct permutations $(\cdot)$ that may be interchangeably used for computing the Choquet integral of χ_A. This is the number of distinct ways of putting all $j \in A^c = N \backslash A$ first and all $i \in A$ last.

Permutations bijectively correspond to *maximal chains* $K = \{A_0, A_1, \ldots, A_n\} \subset 2^N$, where

$$N = A_n \supset^* \cdots \supset^* A_k \supset^* \cdots \supset^* A_0 = \emptyset$$

and $\supset^*$ is the *covering relation* (see [1]), that is to say, $A \supset^* B \Leftrightarrow A \supset B, |A| = |B| + 1$ and reads "A covers B", with $A \supset B \Leftrightarrow A \supseteq B, A \neq B$.

Let $\mathcal{V}_m \subset \mathbb{R}_+^{2^n}$ denote the vector space of monotone $v : 2^N \to \mathbb{R}_+$, with $\gamma \in \mathcal{V}_m$ for all measures γ. The *Möbius inversion* $\mu^v : 2^N \to \mathbb{R}$ of $v \in \mathcal{V}_m$ is given by $\mu^v(A) = \sum_{B \subseteq A} (-1)^{|A \backslash B|} \cdot v(B)$ for all $A \in 2^N$. A basis of $\mathcal{V}_m$ is $\{u_A : A \in 2^N\}$, where $u_A(B) = 1$ if $A \subseteq B$ and 0 otherwise. In fact, $v = \sum_{A \in 2^N} \mu^v(A) \cdot u_A$. That is, $v(A) = \sum_{B \in 2^N} \mu^v(B) \cdot u_B(A) = \sum_{B \subseteq A} \mu^v(B)$ for all $A, B \in 2^N$ (see [1]). The Choquet integral of f w.r.t. γ may be expressed in terms of μ^γ as follows

$$\int_N^C f d\gamma = \sum_{A \in 2^N} \mu^\gamma(A) \bigwedge_{i \in A} f(i),$$

where $\mu^\gamma(\emptyset) = \gamma(\emptyset) = 0$ (see [2] and [7]).

3 The extension

The aggregation technique proposed below enables to integrate monotone set functions $v \in \mathcal{V}_m$ w.r.t. fuzzy measures γ. Firstly note that, given monotonicity, $0 \leq v(\emptyset) \leq v(A)$ for all $A \in 2^N$. In fact, setting $v(\emptyset) = 0$ yields no loss of generality.

For any $v \in \mathcal{V}_m$, let $\mathcal{K}^v$ denote the set of *admissible maximal chains* $K^v = \{A_0^v, A_1^v, \ldots, A_n^v\}$, defined to be those satisfying

$$v(A_i^v) = \bigwedge_{A \supset^* A_{i-1}^v} v(A) \text{ for } 1 \leq i \leq n,$$

where $\supset^*$ is the covering relation defined above.

Remark 1 *By construction, $\mathcal{K}^v \neq \emptyset$. In particular, $1 \leq |\mathcal{K}^v| \leq n!$. To see this, let $N = \{1,2,3\}$ and consider the two extreme examples $v, w : 2^N \to \mathbb{R}_+$ where $v(A) = \left(\sum_{i \in A} i\right)^2$ and $w(A) = \left(|A|\right)^2$ for all $A \in 2^N$. In the former case, $|\mathcal{K}^v| = 1$, and the unique maximal chain $K^v \in \mathcal{K}^v$ is $K^v = \{\emptyset, \{1\}, \{1,2\}, N\}$. Conversely, $\mathcal{K}^w$ contains all the 3! available maximal chains of lattice 2^N.*

For the i-th increment $\Delta(v(A_i^v)) = v(A_i^v) - v(A_{i-1}^v)$ of v along any $\{A_0^v, A_1^v, \ldots, A_n^v\} = K^v \in \mathcal{K}^v$, define

$$\Delta^2(v(A_i^v)) := \Delta(v(A_i^v)) - \Delta(v(A_{i-1}^v)) =$$

$$= v(A_i^v) - 2v(A_{i-1}^v) + v(A_{i-2}^v).$$ This is the i-th difference, $1 \leq i \leq n$, between consecutive increments.

Now, the values taken by set function v may be aggregated w.r.t. fuzzy measure γ by means of the *extended Choquet integral* defined as follows:

$$\int_{2^N}^{EC} v d\gamma := \bigwedge_{K^v \in \mathcal{K}^v} \sum_{1 \leq i \leq n} \Delta^2(v(A_i^v)) \cdot \gamma(N \backslash A_{i-1}^v),$$

with $K^v = \{A_0^v, A_1^v, \ldots, A_n^v\}$ and $v(A_{-1}^v) := 0$.

Notice immediately that $\wedge$ is over the set $\mathcal{K}^v$ of *admissible* maximal chains only. In particular, $\int_{2^N}^{EC} v d\gamma$ works as follows. Given monotonicity, the integrand v takes non-decreasing values along any maximal chain. Firstly, the focus is placed on the set $\mathcal{K}^v$ of maximal chains along which increments $\Delta(v(A_i^v)), 1 \leq i \leq n$ are minimal. Secondly, precisely in order to minimize over such chains $\{A_0^v, A_1^v, \ldots, A_n^v\} = K^v \in \mathcal{K}^v$, the focus turns on those satisfying $\gamma(N \backslash A_i^v) \geq \gamma(N \backslash B_i^v)$ for every $\{B_0^v, B_1^v, \ldots, B_n^v\} \in \mathcal{K}^v$ and $1 \leq i \leq n$. That is to say, increments

$$\Delta(\gamma(N \backslash A_i^v)) = \gamma(N \backslash A_i^v) - \gamma(N \backslash A_{i+1}^v)$$

must be maximal for (increasing) $0 \leq i \leq n-1$. Once a maximal chain $\{A_0^v, A_1^v, \ldots, A_n^v\}$ satisfying this sort of $\min - \max$ criterion is found (there surely exists at least one), $\int_{2^N}^{EC}$ is simply the sum of all differences $\Delta^2(v(A_i^v))$ between consecutive increments of the integrand, multiplied each by the value $\gamma(N \backslash A_{i-1}^v)$ taken by the measure on A_{i-1}^v's complement.

Also note that, although both the integrand and the measure take 2^n values, only $n+1$ values (for each) are used for aggregation. This is because both the integrand and the measure are not defined on a generic set. Conversely, they are defined on a subset lattice. In fact, this technique may be adapted to any lattice (or even any poset) satisfying the Jordan-Dedekind condition (that is, any two maximal chains must have equal length, see [1]).

Remark 2 *Note that there may be two or more maximal chains satisfying the above min−max criterion. In fact, if both the integrand and the measure are symmetric (that is, if $v(A) = \eta_v(|A|), \gamma(A) = \eta_\gamma(|A|)$ for some $\eta_v, \eta_\gamma : \{0, 1, \ldots, n\} \to \mathbb{R}_+$ and all $A \in 2^N$), then all $n!$ maximal chains satisfy such a criterion (see remark 1 above). Still, given v and γ, for any such a maximal chain $\int_{2^N}^{EC} v d\gamma$ yields the same result.*

Any $f : N \to \mathbb{R}_+$ extends to the whole power set 2^N through additivity, i.e., as $v_f : 2^N \to \mathbb{R}_+$ defined by $v_f(A) = \sum_{i \in A} f(i), A \in 2^N$. Also, such an extension v_f is monotone (for all f with positive range). Let $\mathcal{V}_m^a \subset \mathcal{V}_m$ denote the space of additive set functions, and for every $v \in \mathcal{V}_m^a$ let f_v denote its restriction to singletons, that is to say, $f_v(i) = v(\{i\}), 1 \leq i \leq n$.

Claim 3 *If $v \in \mathcal{V}_m^a$, then $\int_{2^N}^{EC} v d\gamma = \int_N^C f_v d\gamma$.*

Proof: If $v \in \mathcal{V}_m^a$, then any permutation $(\cdot) : N \to N$ satisfying $v(\{(1)\}) \leq \cdots \leq v(\{(n)\})$ yields that maximal chain $K^v = \{A_0^v, \ldots, A_n^v\}$ obtained by setting $A_i^v = \{(1), \ldots, (i)\}$ for $1 \leq i \leq n$ satisfies $K^v \in \mathcal{K}^v$. Also, $\Delta(v(A_i^v)) = v(\{(i)\})$ for $1 \leq i \leq n$. Substitution completes the proof. ■

Hence, $\int^{EC}$ extends $\int^C$, regarded as an aggregation operator, from N to 2^N. For this reason, in the sequel the former shall be termed extended Choquet (EC) integral.

Claim 4 *If γ is convex, then for all $v \in \mathcal{V}_m$*

$$\int_{2^N}^{EC} v d\gamma = \bigwedge_{p \in \mathcal{C}(\gamma)} \int_{2^N}^{EC} v dp,$$

where $p(A) = \sum_{i \in A} p(i), A \in 2^N$.

Proof: Any convex γ has a non-empty core whose extreme points $p_K \in ex(\mathcal{C}(\gamma))$ get defined each through a maximal chain $K = \{A_0, \ldots, A_n\}$ by

$$p_K(i) = \gamma(A_k) - \gamma(A_{k-1}) : A_k \backslash A_{k-1} = i, 1 \leq i \leq n$$

(see [11]). Given any $v \in \mathcal{V}_m$ with the associated set $\mathcal{K}^v \ni K^v = \{A_0^v, \ldots, A_n^v\}$ of maximal chains as above (i.e., $v(A_k^v) = \bigwedge_{A \supset^* A_{k-1}^v} v(A), 1 \leq k \leq n$), consider those corresponding $\overline{K^v} = \{N \backslash A_n^v, \ldots, N \backslash A_0^v\}$ obtained by substituting each element with its complement. Clearly, $\gamma(N \backslash A_k^v) = p_{\overline{K^v}}(N \backslash A_k^v), 1 \leq k \leq n$. It only remains to observe that the EC integral obtains by minimizing over $\mathcal{K}^v$. ■

The EC integral inherits another main property from the Choquet integral: it constitutes an extension of the measure, as it coincides with this latter on the vertices of the hypercube. Yet, as integrands must be set functions, each $\chi_A : N \to \{0, 1\}, A \in 2^N$ must be turned into some v_{χ_A} taking values on 2^N. This is most naturally achieved through additivity:

$$v_{\chi_A}(B) = \sum_{i \in B} \chi_A(i) \text{ for all } A, B \in 2^N.$$

In this way, $v_{\chi_A} : 2^N \to \{0, 1, \ldots, |A|\}, A \in 2^N$.

Claim 5 *For any measure γ,*

$$\int_{2^N}^{EC} v_{\chi_A} d\gamma = \gamma(A) \text{ for all } A \in 2^N.$$

Proof: For any integrand $v_{\chi_A}, A \in 2^N$, admissible maximal chains $K^{v_{\chi_A}} \in \mathcal{K}^{v_{\chi_A}}$ bijectively correspond to those $(n - |A|)!|A|!$ permutations $(\cdot) : N \to N$ where all $j \notin A$ come first and all $i \in A$ come last, that is to say, $\{(1), \ldots, (n - |A|)\} = N \backslash A$ as well as $\{(n - |A| + 1), \ldots, (n)\} = A$. Also, along any such a maximal chain $\{B_0^v, \ldots, B_n^v\} \in \mathcal{K}^{v_{\chi_A}}$,

$$\sum_{1 \leq k \leq n} \Delta^2(v_{\chi_A}(B_k^v)) \cdot \gamma(N \backslash B_{k-1}^v) =$$

$$\sum_{n - |A| + 1 \leq h \leq n} \left[\gamma(\{(h), \ldots, (n)\}) - \gamma(\{(h+1), \ldots, (n)\})\right]$$

$= \gamma(A)$ (with $\gamma(\{(n + 1), \ldots, (n)\})$ vanishing). ■

3.1 An extension through averaging

Let $N^{(k)} = \{A \in 2^N : |A| = k\}$ denote the k-th level of subset lattice $(2^N, \cap, \cup)$, containing all k-cardinal subsets $A \subseteq N$, with $|N^{(k)}| = \binom{n}{k}$ for $0 \leq k \leq n$ (see [1]). An integrand $v \in \mathcal{V}_m$ may well attain the same value on several elements of each level, for each level. That is, for each k, it may be $v(A_1) = \cdots = v(A_h)$ with $A_1, \ldots, A_h \in N^{(k)}$ and $1 << h \leq \binom{n}{k}$. If this is the case, then the set $\mathcal{K}^v$ of admissible maximal chains is rather likely to contain many elements (at most, $|\mathcal{K}^v| = n!$ for symmetric v; see remarks 1 and 2 above). Hence, for given γ, define $g_\gamma : \mathcal{K}^v \to \mathbb{R}_+$ by

$$g_\gamma(K^v) = \sum_{1 \leq k \leq n} \Delta^2(v(B_k^v)) \cdot \gamma(N \backslash B_{k-1}^v)$$

for every $\{B_0^v, \ldots, B_n^v\} = K^v \in \mathcal{K}^v$. In general, $g_\gamma(K^v)$ yields different values for different admissible chains. Clearly, in order to define an integration technique, these values must be aggregated. Above, this is achieved through the min operator $\wedge$, as the EC integral is conceived for dealing with DUU, where $\wedge$ seems the most natural way of formalizing risk-aversion. An

alternative aggregation technique is the following: if g_γ takes $q \leq |\mathcal{K}^v|$ different values $g^1_\gamma, \ldots, g^q_\gamma$, then set

$$\widetilde{\int_{2^N}^{EC}} v d\gamma = \sum_{1 \leq h \leq q} \frac{|\mathcal{K}^v_h|}{|\mathcal{K}^v|} \cdot g^h_\gamma,$$

where $\mathcal{K}^v_h = \{K^v \in \mathcal{K}^v : g_\gamma(K^v) = g^h_\gamma\}$. This is the weighted average where the weight of each value $g^h_\gamma, 1 \leq h \leq q$ is the ratio of the number of admissible chains yielding that value to the total number of admissible chains. A somehow extreme case is that where v is symmetric but γ is not. Then, $|\mathcal{K}^v| = n!$, and it might well be $q = n!$. Accordingly, the EC integral as defined above through $\wedge$ considers only the minimum over all these $n!$ values, while this (modified) $\widetilde{\text{EC}}$ integral considers the average of all these values, associating weight $(n!)^{-1}$ to each of them.

Inspection reveals that while claims 2 and 4 remain valid for this $\widetilde{\text{EC}}$ integral, claim 3 does not. Formally, for given γ, if v is additive (through f_v), then

$$\widetilde{\int_{2^N}^{EC}} v d\gamma = \int_N^C f_v d\gamma.$$

$$\text{Also, } \widetilde{\int_{2^N}^{EC}} v_{\chi_A} d\gamma = \gamma(A) \text{ for all } A \in 2^N.$$

$$\text{Yet, } \widetilde{\int_{2^N}^{EC}} v d\gamma \neq \bigwedge_{p \in \mathcal{C}(\gamma)} \widetilde{\int_{2^N}^{EC}} v dp$$

for generic convex γ.

3.2 Antitone integrands

The EC technique may also be applied to antitone integrands $v : 2^N \backslash \{\emptyset\} \to \mathbb{R}_+$, where $A \subseteq B \in 2^N$ implies $v(A) \geq v(B)$ for all $A, B \in 2^N \backslash \{\emptyset\}$, with $v(\emptyset) = 0$. (Note that $v(\emptyset) = 0$ implies that v is antitone on $2^N \backslash \{\emptyset\}$ but not on the whole power set 2^N).

In order to achieve this, the general procedure above is somehow reversed. Firstly, for taking into account both the maximum (i.e., $\bigvee_{i \in N} v(\{i\})$) and the minimum (i.e., $v(N)$) values taken by v on $2^N \backslash \{\emptyset\}$, admissible chains $\{A^v_0, \ldots, A^v_n\} = K^v \in \mathcal{K}^v$ satisfy

$$v(A^v_i) = \bigvee_{A \supset^* A^v_{i-1}} v(A) \text{ for } \textit{increasing } 1 \leq i \leq n.$$

Next, for $0 < i \leq n$, define

$$\Delta^2(v(A^v_i)) = v(A^v_i) - 2v(A^v_{i+1}) + v(A^v_{i+2}),$$

with $v(A^v_{n+1}) = v(A^v_{n+2}) = 0$. Then, given any capacity γ, the EC integral of v w.r.t. γ is

$$\int_{2^N}^{EC} v d\gamma = \bigwedge_{K^v \in \mathcal{K}^v} \sum_{0 \leq i < n} \Delta^2(v(A^v_{n-i})) \cdot \gamma(A^v_{n-i}),$$

with $K^v = \{A^v_0, \ldots, A^v_n\}$.

This EC integral, with integrand v antitone on $2^N \backslash \{\emptyset\}$, may be used for quantifying the certainty equivalent of a seemingly sophisticated DUU problem (see subsection 4.2 below).

3.3 Real-valued integrands

There are two alternative techniques for integrating real-valued functions $f : N \to \mathbb{R}$ w.r.t. fuzzy measures γ, namely, the Šipoš or symmetric Choquet integral, and the asymmetric Choquet integral. In order to formalize these operators, define $f^+(i) = 0 \vee f(i)$ and $f^-(i) = 0 \vee -f(i)$ for $1 \leq i \leq n$. Also, for any measure γ, consider its conjugate $\bar{\gamma}$ defined, for every $A \in 2^N$, by $\bar{\gamma}(A) = \gamma(N) - \gamma(N \backslash A)$. Then, the symmetric Choquet integral is

$$\int_N^{C_{sy}} f d\gamma = \int_N^C f^+ d\gamma - \int_N^C f^- d\gamma,$$

while the asymmetric Choquet integral is

$$\int_N^{C_{asy}} f d\gamma = \int_N^C f^+ d\gamma - \int_N^C f^- d\bar{\gamma}$$

(see [7]).

Paralleling this in terms of the EC integral seems rather straightforward: given a real-valued, monotone $v : 2^N \to \mathbb{R}$, define $v^+(A) = 0 \vee v(A)$ as well as $v^-(A) = 0 \vee -v(A)$ for every $A \in 2^N$. Then, the symmetric EC integral is

$$\int_{2^N}^{EC_{sy}} v d\gamma = \int_{2^N}^{EC} v^+ d\gamma - \int_{2^N}^{EC} v^- d\gamma,$$

while the asymmetric EC integral is

$$\int_{2^N}^{EC_{asy}} v d\gamma = \int_{2^N}^{EC} v^+ d\gamma - \int_{2^N}^{EC} v^- d\bar{\gamma}.$$

The same argument may be applied, *mutatis mutandis*, to antitone integrands.

3.4 Möbius inversion

Like the Choquet integral, the EC integral may be expressed in terms of Möbius inversion. Yet, while in the former case only the measure has a non-trivial such an inversion (as the integrand is additive), in the

latter case the integrand displays a non-trivial such an inversion as well.

Given $v, \gamma \in \mathcal{V}_m$, the EC integral of v w.r.t. γ may be written as follows $\int_{2^N}^{EC} v d\gamma =$

$$\bigwedge_{K^v \in \mathcal{K}^v} \sum_{1 \leq i \leq n} \Delta(v(A_i^v)) \cdot \big[\gamma(N \backslash A_{i-1}^v) - \gamma(N \backslash A_i^v)\big],$$

with $K^v = \{A_0^v, \ldots, A_n^v\}$. Also, for all $\emptyset \neq A \in 2^N$,

$$v(A) - v(A \backslash i) = \sum_{B \subseteq A : B \ni i} \mu^v(B) = \sum_{B \in 2^A \backslash 2^{A \backslash i}} \mu^v(B)$$

for all $i \in A$. Accordingly, $\int_{2^N}^{EC} v d\gamma =$

$$\bigwedge_{K^v \in \mathcal{K}^v} \sum_{1 \leq i \leq n} \Big[\sum_{\substack{B \subseteq A_i^v \\ B \not\subseteq A_{i-1}^v}} \mu^v(B) \Big] \cdot \Big[\sum_{\substack{B' \subseteq N \backslash A_{i-1}^v \\ B' \not\subseteq N \backslash A_i^v}} \mu^\gamma(B') \Big],$$

with $K^v = \{A_0^v, \ldots, A_n^v\}$.

Möbius inversion yields two further extensions $\widehat{\int_{2^N}^{EC}}$ and $\overline{\int_{2^N}^{EC}}$ of the Choquet integral. The $\widehat{\text{EC}}$ integral of v w.r.t. γ is

$$\widehat{\int_{2^N}^{EC}} v d\gamma = \sum_{A \in 2^N} \mu^\gamma(A) \bigwedge_{i \in A} \big(v(A) - v(A \backslash i)\big)$$

$$= \sum_{A \in 2^N} \mu^\gamma(A) \bigwedge_{i \in A} \Big(\sum_{B \in 2^A \backslash 2^{A \backslash i}} \mu^v(B) \Big).$$

To see that this actually constitutes an extension (i.e., when the integrand is additive it reduces to the Choquet integral), note that if v is additive through f_v (see above), then $v(A) - v(A \backslash i) = f_v(i)$ for all $A \in 2^N$ and all $i \in A$, so that the former expression above reduces to a well known representation of the Choquet integral in terms of the Möbius inversion of the measure (see [2], [7] and section 2 above). Concerning the latter expression, the Möbius inversion of additive set functions v satisfying $v(\emptyset) = 0$ lives only on 1-cardinal subsets (or atoms), i.e., $|A| \neq 1 \Rightarrow \mu^v(A) = 0$ (see [1] on valuations of distributive lattices, pp. 189-191). In particular, $v_{\chi_A}, A \in 2^N$ is an additive set function (see above), and it may be easily checked that

$$\widehat{\int_{2^N}^{EC}} v_{\chi_A} d\gamma = \sum_{B \in 2^A} \mu^\gamma(B) = \gamma(A).$$

On the other hand, the $\overline{\text{EC}}$ integral of v w.r.t. γ is

$$\overline{\int_{2^N}^{EC}} v d\gamma = \sum_{A \in 2^N} \mu^\gamma(A) \bigwedge_{i \in A} \Big(v\big((N \backslash A) \cup i\big) - v(N \backslash A)\Big).$$

Here again, $v\big((N \backslash A) \cup i\big) - v(N \backslash A) = f_v(i)$ for all $A \in 2^N$ and all $i \in A$ whenever v is additive (through f_v), in which case this latter expression reduces to the same well known representation of the Choquet integral (in terms of the Möbius inversion of the capacity) as above. Also note that

$$\bigwedge_{i \in B} \Big(v_{\chi_A}\big((N \backslash B) \cup i\big) - v_{\chi_A}(N \backslash B)\Big) = 1$$

if $B \subseteq A$ and 0 otherwise, with $A, B \in 2^N$. Therefore, for any integrand $v_{\chi_A}, A \in 2^N$, only the values $\mu^\gamma(B)$ such that $B \subseteq A$ are summed. Hence,

$$\overline{\int_{2^N}^{EC}} v_{\chi_A} d\gamma = \sum_{B \subseteq A} \mu^\gamma(B) = \gamma(A).$$

3.5 An example

In order to appreciate the difference between these four extensions of the Choquet integral (i.e., each coincides with this latter whenever the integrand is additive), consider the simple case where $N = \{1, 2, 3\}$ and the integrand v is symmetric; in particular, $v(A) = \big(\frac{|A|}{3}\big)^2$ for all $A \subseteq N$. (Note that this is a capacity itself.) Let γ denote the (fuzzy) measure w.r.t. which integration is performed, with

$$\gamma(\{1\}) = 0.1 \ ; \ \gamma(\{2\}) = 0.4 \ ; \ \gamma(\{3\}) = 0.6$$

$$\gamma(\{1,2\}) = 0.6 \ ; \ \gamma(\{1,3\}) = 0.8 \ ; \ \gamma(\{2,3\}) = 0.9.$$

Accordingly, $\mu^\gamma(A) = \gamma(A)$ if $|A| = 1$, and

$$\mu^\gamma(\{1,2\}) = 0.6 - 0.1 - 0.4 = 0.1$$

$$\mu^\gamma(\{1,3\}) = 0.8 - 0.1 - 0.6 = 0.1$$

$$\mu^\gamma(\{2,3\}) = 0.9 - 0.4 - 0.6 = -0.1$$

$$\mu^\gamma(\{1,2,3\}) = v(\{1,2,3\}) - \sum_{A \subset \{1,2,3\}} \mu^\gamma(A) =$$

$$= 1 - (0.1 + 0.4 + 0.6 + 0.1 + 0.1 - 0.1) = -0.2.$$

Simple computations yield

$$\int_{2^N}^{EC} v d\gamma = \frac{24}{90},$$

where the (admissible) maximal chain along which integration is performed is $\{\emptyset, \{3\}, \{2,3\}, N\}$. Also,

$$\widehat{\int_{2^N}^{EC}} v d\gamma = \frac{1}{3!} \cdot \frac{40 + 36 + 38 + 28 + 30 + 24}{90} = \frac{98}{3 \cdot 90},$$

$$\text{while } \widehat{\int_{2^N}^{EC}} v d\gamma = \frac{4}{90} \text{ and } \overline{\int_{2^N}^{EC}} v d\gamma = \frac{56}{90}.$$

4 Applications

This final section is devoted to possible applications of the EC integral (or any variation $\widehat{\text{EC}}$, $\widetilde{\text{EC}}$ or $\overline{\text{EC}}$). More generally, the focus turns on conceivable situations where integrating (i.e., aggregating) set functions w.r.t. capacities might be useful. In particular, two such applications are provided below. In both cases, the aggregation operator yields an expectation. In the former case the integrand is monotone, while in the latter it is antitone.

4.1 An expectation of diversity

The idea of measuring diversity within any subset $A \subseteq N$ of a finite population $N = \{1, \ldots, n\}$ has been recently formalized through *diversity functions* $v : 2^N \to \mathbb{R}_+$ which, by construction, are monotone (see [9]). This approach relies upon the multi-attribute model of diversity: any attribute characterizing any member of the population is identified with the set of members characterized by that attribute and, conversely, any subset A of the population identifies a conceivable attribute, i.e., belonging to A. Accordingly, the set of conceivable attributes is 2^N. For each attribute $A \in 2^N$ there is a weight $\lambda_A \geq 0$. In fact, $v : 2^N \to \mathbb{R}_+$ is defined to be a diversity function if there is a $\lambda : 2^N \to \mathbb{R}_+$ such that

$$v(B) = \sum_{\substack{A \in 2^N \\ B \cap A \neq \emptyset}} \lambda_A \text{ for all } B \in 2^N,$$

with $v(\emptyset) = \lambda_\emptyset = 0$. Furthermore, if this is the case, then λ is the *conjugate Möbius inverse* of v, given by

$$\lambda_A = \sum_{B \subseteq A} (-1)^{|A \setminus B|+1} \cdot v(N \setminus B) \text{ for all } A \in 2^N.$$

In turn, v has a positive conjugate Möbius inversion if and only if it is monotone and *totally submodular*, that is, for all collections $\{A_1, \ldots, A_k\} \subseteq 2^N$

$$v\Big(\bigcap_{1 \leq k' \leq k} A_{k'}\Big) \leq \sum_{\emptyset \subset L \subseteq \{1,\ldots,k\}} (-1)^{|L|+1} \cdot v\Big(\bigcup_{l \in L} A_l\Big),$$

(see [9] and [2]).

Hence, a diversity function v is a perfect candidate as a monotone integrand of the EC integral above. In particular, consider the case where one is interested in the expectation of diversity within a random sample of population N. For the sake of concreteness, assume, for example, that the population of a national park has to be created (or increased) by moving some random sample of an existing, known population from another park. Similarly, assume one has a given amount of time and resources to devote to survey a known population of animals moving freely within a wide region, and is interested in forming an expectation of the diversity that will be observed. In this case, the capacity γ w.r.t. which integration is performed is to be interpreted as follows: $\gamma(A), A \in 2^N$ quantifies the belief that the random sample will be precisely A. Clearly enough, there is no reason why such a capacity should be additive. Then, $\int_{2^N}^{EC} v d\gamma$ (or $\widetilde{\int_{2^N}^{EC}} v d\gamma$ or $\widehat{\int_{2^N}^{EC}} v d\gamma$ or $\overline{\int_{2^N}^{EC}} v d\gamma$) furnishes an expectation of the diversity in the random sample.

4.2 Certainty equivalent in DUU

Consider again the DUU situation described above, where a DM has to take action in response to an unknown state of Nature $i \in N$. In particular, let $\mathcal{A} = \{a_1, \ldots, a_n\}$ denote the set of available actions, and formalize preferences by $u : N \times \mathcal{A} \to \mathbb{R}$, with $u(i, a) = u_a(i)$ for all $(i, a) \in N \times \mathcal{A}$. For the sake of concreteness, for $1 \leq i \leq n$, assume $u_a(i) < u_{a_i}(i)$ for all $a \in \mathcal{A} \setminus a_i$. In words, every state $i \in N$ has its own, distinct optimal action $a_i \in \mathcal{A}$. Finally, beliefs are quantified by capacity γ.

The DM chooses some action a^* maximizing her CEU, that is,

$$\int_N^C u_{a^*} d\gamma = \bigvee_{a \in \mathcal{A}} \int_N^C u_a d\gamma,$$

where $\int_N^C u_a d\gamma$ is the Choquet integral of u_a w.r.t. γ (see section 2 above). In fact, this is the certainty equivalent of the decision making problem. More precisely, assume that u is in money terms. Accordingly, if $\int_N^C u_{a^*} d\gamma > 0$, then the DM is willing to pay, at most, such a quantity in order to face this problem. Conversely, if $\int_N^C u_{a^*} d\gamma < 0$, then the DM is willing to pay, at most, such a quantity in order to avoid this problem. Finally, if $\int_N^C u_{a^*} d\gamma = 0$, then the DM is indifferent between facing this problem or not. In particular, as u is real-valued, the Choquet integral may be computed either symmetrically, or else asymmetrically (see subsection 3.3 above and [7]).

This framework can be enriched by introducing the idea of information, which may be formalized in alternative ways and raises the issue of updating non-additive beliefs (see [4]). Information may be modeled as a *field* of events and, in particular, as a field of events generated by some *partition* of N (see [8]). Yet, the focus here is placed on the situation where the DM knows, before taking action, that the 'true' state of Nature belongs to some subset $\emptyset \neq B \in 2^N$. Then, what action will be taken? One possibility is

to choose an action that maximizes the conditional expectation defined by the Choquet integral (over B) of the restricted utility function $u : B \times \mathcal{A} \to \mathbb{R}$ w.r.t. the normalized restricted capacity $\gamma : 2^B \to [0,1]$. That is, the DM may choose an action a_B^* satisfying

$$\int_B^C u_{a_B^*} d\gamma_B = \bigvee_{a \in \mathcal{A}} \int_B^C u_a d\gamma_B,$$

where $\int_B^C u_a d\gamma_B =$

$$\sum_{1 \leq i \leq |B|} \left[u_a((i)) - u_a((i-1)) \right] \cdot \frac{\gamma(\{(i), \ldots, (|B|)\})}{\gamma(B)}$$

and $(\cdot) : B \to B$ is any permutation such that $u_a((|B|)) \geq \cdots \geq u_a((i)) \geq \cdots \geq u_a((1)) \geq 0$, with $u_a((0)) := 0$ and where the integrand is assumed to take only positive values for the sake of simplicity (and reasons of space). With a real-valued integrand, this conditioned Choquet integral may be computed either symmetrically or else asymmetrically (again, see subsection 3.3 above and [7]). Now, for $\emptyset \neq B \in 2^N$, define $v : 2^N \backslash \{\emptyset\} \to \mathbb{R}_+$ by

$$v(B) := \bigvee_{a \in \mathcal{A}} \int_B^C u_a d\gamma_B.$$

It seems rather evident that if $\emptyset \subset A \subseteq B \in 2^N$, then $v(A) \geq v(B)$. That is, v is antitone on $2^N \backslash \{\emptyset\}$. In fact, roughly speaking, the smaller the (non-void) subset, the better the DM can choose a corresponding optimal action. In particular, $v(\{i\}) = u_{a_i}(i), i \in N$.

Finally, consider the issue of quantifying some belief about what information (in the form of an event) will be available before taking action. In other terms, focus on how to define some capacity $\eta : 2^N \to [0,1]$ such that $\eta(A)$ quantifies the belief that before taking action the DM *will know* that the 'true' state belongs to A, with $\emptyset \neq A \in 2^N$. In fact, $\gamma(A)$ already quantifies the belief that the 'true' state belongs to A. Nevertheless, $\eta(A)$ must quantify the belief that, *in addition*, this will be known before taking action. Accordingly, it must be $\eta(A) \leq \gamma(A)$. For example, one may set $\eta(A) = (\gamma(A))^2$.

Then, the EC integral of v w.r.t. η as defined in subsection 3.2, i.e., $\int_{2^N}^{EC} v d\eta$, furnishes an expectation of the utility that the DM may receive by facing this DUU problem with (non-additive) beliefs about what information (in the form of an event) will be available before taking action. Accordingly, $\bigvee_{a \in \mathcal{A}} \int_N^C u_a d\gamma$ may be interpreted as the certainty equivalent of this DUU problem when the DM believes that no information will be available before taking action. That is, the certainty equivalent with trivial $\eta_{\mathcal{T}}$ such that $\eta_{\mathcal{T}}(A) = 0$ for all $A \subset N$ and $\eta_{\mathcal{T}}(N) = \gamma(N) = 1$, as

$$\int_{2^N}^{EC} v d\eta_{\mathcal{T}} = v(N) = \bigvee_{a \in \mathcal{A}} \int_N^C u_a d\gamma.$$

Conversely, $\int_{2^N}^{EC} v d\eta$ seems a suitable candidate as the certainty equivalent of this DUU problem with non-trivial beliefs $\eta \neq \eta_{\mathcal{T}}$ about what information (i.e., event) will be available before taking action.

References

[1] M. Aigner. *Combinatorial Theory*. Springer-Verlag, 1979.

[2] A. Chateauneuf and J. Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[3] G. Choquet. Theory of capacity. In *Ann. l'Institute Fourier*, volume 5, pages 131–295, 1953-1954.

[4] I. Gilboa and D. Schmeidler. Updating ambiguous beliefs. *Journal of Economic Theory*, 59:33–49, 1993.

[5] I. Gilboa and D. Schmeidler. Canonical representation of set functions. *Mathematics of Operations Research*, 20:197–212, 1995.

[6] M. Grabisch. The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89(33):445–456, 1996.

[7] M. Grabisch, C. Labreuche, and J.-C. Vansnick. On the extension of pseudo-Boolean functions for the aggregation of interacting criteria. *European Journal of Operational Research*, 148:28–47, 2003.

[8] E. Lehrer. Updating non-additive probabilities - a geometric approach. *Games and Economic Behavior*, 50:42–57, 2004.

[9] K. Nehring and C. Puppe. A theory of diversity. *Econometrica*, 70(3):1155–1198, 2002.

[10] D. Schmeidler. Subjective probabilities without additivity. *Econometrica*, 57:571–587, 1989.

[11] L. S. Shapley. Cores of convex games. *International Journal of Game Theory*, 1(1):11–26, 1971.

Shortfall-dependant Risk Measures (and Previsions)

Pietro Baroni
Dip. Elettronica per l'Automazione
Univ. of Brescia, Italy
baroni@ing.unibs.it

Renato Pelessoni
Dip. Mat. Appl. "B. de Finetti"
Univ. of Trieste, Italy
renato.pelessoni@econ.units.it

Paolo Vicig
Dip. Mat. Appl. "B. de Finetti"
Univ. of Trieste, Italy
paolo.vicig@econ.units.it

Abstract

Because of their simplicity, risk measures are often employed in financial risk evaluations and related decisions. In fact, the risk measure $\rho(X)$ of a random variable X is a real number customarily determining the amount of money needed to face the potential losses X might cause. At a sort of second-order level, the adequacy of $\rho(X)$ may be investigated considering the part of the losses it does not cover (its shortfall). This may suggest employing a further, more prudential risk measure, taking the shortfall of $\rho(X)$ into account. In this paper a family of shortfall-dependant risk measures is proposed, investigating its consistency properties and its utilization in insurance pricing. These results are obtained and subsequently extended within the framework of imprecise previsions, of which risk measures are an instance. This also leads us to investigate properties of a rather weak consistency notion for imprecise previsions, termed 1–convexity.

Keywords: Risk Measures, Imprecise Previsions, Shortfall.

1 Introduction

Risk measures are an important and widely studied tool in the field of financial evaluations and decisions. In spite of their significant relationships with other uncertainty theories, in particular with precise and imprecise previsions [2, 14], steps towards bridging the gap and enabling cross-fertilization between these research areas have been taken only relatively recently (see for instance [8, 9, 13]). In this paper we contribute to this effort with focus on *shortfall-dependant* risk measures, and, in particular, on the so-called *Dutch risk measures*.

To give a quick idea of the whole context, we first recall that, among its interpretations, the risk measure $\rho(X)$ of a random variable X supplies a real number determining how much money should be reserved to face potential losses arising from X. Several kinds of *consistency* requirements have been considered in the literature in order to ensure that $\rho(X)$ is "sound" (in some sense). Defining (and satisfying) these requirements is particularly important (and not so obvious) when the domain of a risk measure is a *set* of random quantities. Consistency requirements, in general, allow more alternative risk measures. Selecting a specific measure is, in a sense, arbitrary and may reflect the risk attitude or other subjective features of the agent evaluating the risk. For instance, a coherent but extremely cautious choice for $\rho(X)$ is $\rho(X) = -\inf X$. Clearly, this is questionable from another point of view, since it is likely to reserve an excessive amount of money. For this reason, it is rather common (and reasonable) that the chosen risk measure does not entirely cover all potential losses.

The amount of losses not covered by a risk measure is called its *shortfall*. Given a random variable X and a risk measure ρ, the shortfall is a non-negative random variable which is a function of X and $\rho(X)$ and quantifies the possible losses exceeding $\rho(X)$. Given a risk measure, a derived risk measure can then be defined which depends on the original one by taking account of its shortfall in some way. Examples of motivations for defining this kind of "second-order" risk measure (which can be regarded as an adjustment of the first one) will be better discussed in Section 3.3. A significant class of shortfall-dependant risk measures is represented by Dutch risk measures, which can be given a practically important interpretation in the domain of insurance pricing.

This paper provides some new definitions and results concerning shortfall-dependant risk measures and previsions and their consistency properties.

After recalling basic concepts and results about pre-

visions and risk measures in Section 2, we define in Section 3 a generalized family of Dutch risk measures, relate it to a family of lower previsions and investigate its consistency properties. In particular, we show that this family of measures preserves the consistency property of *coherence* (alternatively *convexity*) provided that the original measure and the measure evaluating its shortfall satisfy it. In insurance pricing, generalized Dutch risk measures allow (unlike previous proposals) certain premium policies (double loading), while preserving the above mentioned properties; this is discussed in Section 3.3. In Section 4 we provide a result concerning the ability of "Dutch-like" imprecise previsions to preserve the weaker consistency property of *1-convexity*. In this context, we also explore some properties of 1-convexity, including its close relationships with the notions of capacity and niveloid.

2 Preliminaries

2.1 Precise and imprecise previsions

Let D be an arbitrary (non-empty) set of *bounded* random variables (unbounded variables will not be considered in this paper). We shall use the term *prevision* to denote a mapping $P : D \to \mathbb{R}$ which is understood to be, unless otherwise stated, a *coherent* (precise) prevision in the sense of de Finetti [2]. As well known, this means that P satisfies a certain no-arbitrage condition in an idealized betting scheme. For each $X \in D$, $P(X)$ "summarizes" X, also meaning that whenever an expectation $E(X)$ is given, $P(X) = E(X)$.

In many practical situations it may be more appropriate to assess an imprecise evaluation on each $X \in D$: a *lower* ($\underline{P}$) or an *upper* ($\overline{P}$) prevision. A precise prevision, coherent or not, corresponds to the special case $\overline{P}(X) = \underline{P}(X) = P(X)$, $\forall X \in D$. The upper (lower) prevision $\overline{P}(X)$ ($\underline{P}(X)$) for X has been given in [14] the meaning of infimum selling price (supremum buying price) for X; this interpretation is relevant also in relating imprecise previsions and risk measures [8]. Consistency requirements for imprecise previsions were proposed in [9, 14] by modifying de Finetti's betting scheme. In all instances, it is sufficient to refer to either lower or upper previsions only (on D or on $D^- = \{X : -X \in D\}$, respectively) because of the conjugacy equality

$$\overline{P}(X) = -\underline{P}(-X). \tag{1}$$

In particular, *coherent* lower previsions may be defined as follows [14]:

Definition 1 *$\underline{P}$ is a* coherent lower prevision *on D iff* $\forall n \in \mathbb{N}^+$, $\forall X_0, \ldots, X_n \in D$, $\forall s_0, \ldots, s_n \geq 0$, *defining* $\underline{G} = \sum_{i=1}^n s_i(X_i - \underline{P}(X_i)) - s_0(X_0 - \underline{P}(X_0))$, *it holds that* $\sup \underline{G} \geq 0$.

The weaker notion of lower prevision that avoids sure loss [14] may be regarded as a minimal consistency requirement. The concept of *centered convex* (or *C-convex*) lower prevision was introduced in [9], and is somewhat intermediate between those of lower prevision that is coherent and that avoids sure loss. The definition of the still weaker notion of convexity is obtained from Definition 1 by adding there the constraint $\sum_{i=1}^n s_i = s_0$.

Definition 2 *$\underline{P}$ is a* convex lower prevision *on D iff* $\forall n \in \mathbb{N}^+$, $\forall X_0, \ldots, X_n \in D$, $\forall s_0, \ldots, s_n \geq 0$ *such that* $\sum_{i=1}^n s_i = s_0$, *defining* $\underline{G} = \sum_{i=1}^n s_i(X_i - \underline{P}(X_i)) - s_0(X_0 - \underline{P}(X_0))$, *it holds that* $\sup \underline{G} \geq 0$.

A convex lower prevision such that ($0 \in D$ and) $\underline{P}(0) = 0$ is C-convex.

Recalling that, given an event A, its lower probability $\underline{P}(A)$ is interpreted as the lower prevision of the indicator function of A, the extra assumption $\underline{P}(0) = 0$ is quite natural: $\underline{P}(0)$ represents just the lower uncertainty evaluation we would assign to the impossible event $\emptyset$.

Referring to [2, 9, 14] for a detailed study of the notions presented in this subsection, we recall now some results for later use:

Proposition 1 *Let $\mu : D \to \mathbb{R}$.*

a) Let μ be a prevision (alternatively, a coherent, convex or C-convex lower prevision). Whatever is $D' \supset D$, there exists an extension of μ on D' which is a prevision (alternatively, which is a coherent, convex or C-convex lower prevision, respectively).

b) When D is a linear space, μ is a coherent lower prevision on D iff:

- *b1) $\mu(X) \geq \inf X$, $\forall X \in D$*
- *b2) $\mu(\lambda X) = \lambda\mu(X)$, $\forall X \in D$, $\lambda > 0$*
- *b3) $\mu(X + Y) \geq \mu(X) + \mu(Y)$, $\forall X, Y \in D$*

c) When D is a linear space containing real constants, μ is a convex lower prevision on D iff:

- *c1) $\mu(X+k) = \mu(X)+k$, $\forall X \in D$, $\forall k \in \mathbb{R}$ (translation invariance)*
- *c2) $\forall X, Y \in D$, if $X \leq Y$ then $\mu(X) \leq \mu(Y)$ (monotonicity)*
- *c3) $\mu(\lambda X + (1-\lambda)Y) \geq \lambda\mu(X) + (1-\lambda)\mu(Y)$, $\forall X, Y \in D$, $\forall \lambda \in [0, 1]$ (concavity)*

d) If μ_1 and μ_2 are both coherent (alternatively, convex or C-convex) lower previsions then $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$,

$\lambda \in [0,1]$ *is a coherent (alternatively, convex or C-convex) lower prevision.*

It ensues from Proposition 1 that the conditions in c), in particular c1) and c2), are necessary for convexity of a lower prevision $\underline{P}$ even when D is not a linear space but the relevant quantities are well-defined. In fact, if $\underline{P}$ is convex it allows for a convex extension on a linear space $\mathcal{L} \supset D$ where c1) and c2) must hold. Since coherence implies convexity, c1) and c2) hold for coherent lower previsions, and for previsions too. It can be shown that b1) (but generally not b2), nor b3)) holds for any C-convex $\underline{P}$. Precise previsions are linear: $P(X+Y) = P(X)+P(Y)$ whenever $X, Y, X+Y \in D$. Using (1), the results of Proposition 1 can be easily restated in their specular version for upper previsions.

Coherent imprecise previsions are characterized as envelopes of (precise) previsions, by means of envelope theorems [14]. We recall the result for coherent upper previsions:

Proposition 2 *Let* $\overline{P} : D \to \mathbb{R}$, $\mathcal{M} = \{P : P(X) \leq \overline{P}(X),\ \forall X \in D,\ P$ *is a prevision on* $D\}$. *Then* $\overline{P}$ *is a coherent upper prevision on* D *iff* $\overline{P}(X) = \max_{P\in\mathcal{M}}\{P(X)\},\ \forall X \in D$.

Generalizations of Proposition 2 characterize similarly convex and C-convex previsions [9].

2.2 Risk measures

The risk measure $\rho(X)$ of a random variable X is a real number measuring how "risky" X is. Usually X is the future value, in some currency, of a financial asset and $\rho(X)$ corresponds to the amount of money to be reserved to cover losses potentially arising from (negative values of) X. When $\rho(X) < 0$, this means that we could add $\rho(X)$ to X, i.e. subtract $|\rho(X)|$ from X, keeping the resulting $X - |\rho(X)|$ desirable (no reserving is believed to be necessary to cover risks from $X - |\rho(X)|$). A risk measure ρ on D is thus a real map $\rho : D \to \mathbb{R}$. A risk measure is a relatively simple tool to take basic financial decisions, and this explains the popularity of such instruments in risk theory and practice. Hence risk measures have been extensively studied, but their relationship with other uncertainty theories was mostly overlooked. It was shown in [8] that the risk measure $\rho(X)$ can be interpreted as an upper prevision for $-X$, being the infimum price one would ask to shoulder X, or to sell $-X$. Because of this and (1), the following fundamental equality holds

$$\rho(X) = \overline{P}(-X) = -\underline{P}(X) \tag{2}$$

This equality lets us transfer results from risk measures to upper or lower previsions or vice versa. In Section 3 most statements are given for risk measures, but proven in their corresponding version for lower previsions, following the prevailing custom in the relevant literature. Given equation (2), the consistency notions of coherence, convexity and C-convexity are easily reworded for risk measures [8, 9]. Thus, for instance:

Definition 3 $\rho : D \to \mathbb{R}$ *is a* coherent risk measure *on* D *iff* $\forall n \in \mathbb{N}^+,\ \forall X_0, \ldots, X_n \in D,\ \forall s_0, \ldots, s_n \geq 0$, *defining* $G_\rho = \sum_{i=1}^n s_i(X_i + \rho(X_i)) - s_0(X_0 + \rho(X_0))$, *it holds that* $\sup G_\rho \geq 0$.

Definition 3 includes as a special case the notion of coherent risk measure defined in [1] through a set of axioms and assuming that the domain is a linear space. Analogously, the concept of convex risk measure [9], obtained adding condition $\sum_{i=1}^n s_i = s_0$ in Definition 3, generalizes to arbitrary domains a notion developed in [5] for linear spaces only. For an overview of the many interactions between imprecise previsions and risk measures, see [13].

3 Shortfall-based risk measures

Whatever the risk measure ρ is, it might be inadequate to fully cover losses. Suppose for instance we assess a priori $\rho(X) = 5$ while a posteriori X assumes the value -8: ρ covers only partly the loss arising from X, since there remains a *residual loss* or *shortfall* of 3 in *absolute value*, after employing the reserve money of 5. If instead, a posteriori, $X = -2$, the protection ensured by $\rho(X) = 5$ is full and the shortfall assumes the value 0.

Formally, given a random variable X and a risk measure $\rho(X)$, the *shortfall* of $\rho(X)$ is the random variable $\max(-\rho(X) - X, 0)$.

In the following we will use the shortened notation $(Y)_+ \triangleq \max(Y, 0)$ where Y is a random variable. Accordingly the shortfall will be denoted as $(-\rho(X) - X)_+$. We will also use the dual notation $(Y)_- \triangleq \min(Y, 0)$.

In this paper we focus on risk measures which take account of the shortfall arising from a previously assessed risk measure. More specifically, we shall generalize the family of Dutch risk measures.

3.1 Dutch risk measures

Suppose a (precise) prevision P_0 is assessed on D. We call *Dutch risk measure* the measure

$$\rho_D(X) = P_0(-X) + cP_1[(P_0(X) - X)_+], c \in [0,1] \tag{3}$$

where P_1 is a prevision on a set D_1 such that (3) is well-defined (in particular the set D_1 must include the

random variables $(P_0(X) - X)_+$, $\forall X \in D$). Since $P_0(X) - X = -P_0(-X) - X$, $(P_0(X) - X)_+$ is the shortfall arising from using the prevision P_0 as a risk measure for X ($\rho(X) = P_0(-X)$). It is intuitively clear that this choice for ρ is inadequate since a risk measure should be typically asymmetric, giving higher weight to lower values of X. However, P_0 can be taken as a basis for building a more appropriate risk measure. The new risk measure takes account of the former one through prevision P_1, which evaluates the size of ρ's shortfall. Thus P_1 should typically be assessed *independently* of P_0, at a later stage and on a possibly different domain D_1. The measure $\rho_D(X)$ is coherent: a direct proof may be found in [13]. An earlier version of (3), to be discussed in Section 3.3, appeared in [6] and later in [3, 7]. The measures discussed in these papers may be written as

$$\rho'_D(X) = E(-X) + cE[(dE(X) - X)_+], c \in]0, 1], d > 0 \quad (4)$$

In (4), P_0 and P_1 are replaced by an expectation. It was shown in [3] that if $c = d = 1$, ρ'_D is defined on a linear space, and the expectations are computed with respect to a *common* probability measure, then ρ'_D is coherent.

A distinguishing feature of (3) with respect to (4) is its emphasizing that the uncertainty evaluations P_0 and P_1 could be assessed independently, while this is not possible in (4) if the same underlying probability is used to compute all expectations. To highlight the relevance of this distinction, let us consider the following extreme example.

Example. Let $D = \{X\}$, $X \leq 0$ and assign $\rho(X) = P_0(-X) = P_0(X) = 0$. This is a coherent but highly unbalanced choice: no reserve money is required in a case where no gain is possible, whatever value X will have. Here $(P_0(X) - X)_+ = -X$, hence using (3) we may correct the evaluation if $P_1[(P_0(X) - X)_+] = P_1(-X) > 0$. However no correction is possible if we require that $P_0 = P_1$, $D = D_1$, since then $P_1(-X) = P_0(-X) = 0$.

3.2 Generalized Dutch risk measures

We introduce now a new family of risk measures, which generalizes the risk measures in (3) in a twofold way. First, a natural idea is to replace $P_0(-X)$ with a risk measure $\rho(X)$. Further, we might be unable to precisely evaluate the shortfall $(-\rho(X) - X)_+$, therefore P_1 could be substituted by an imprecise evaluation: given that the new risk measure, say $\rho_c(X)$, should be a prudential correction of $\rho(X)$, an upper prevision $\overline{P}$ seems more appropriate than a lower one. We therefore propose

$$\rho_c(X) = \rho(X) + c\overline{P}[(-\rho(X) - X)_+], c \in [0, 1] \quad (5)$$

What are the consistency properties of $\rho_c(X)$? We shall now prove the following proposition.

Proposition 3 *Let ρ be a coherent risk measure on D, $\overline{P}$ a coherent upper prevision on a set D_U such that (5) is well-defined . Then $\rho_c(X)$ as defined by (5) is a coherent risk measure on D.*

We shall prove Proposition 3 in its corresponding version for lower previsions which, using (2) and elementary properties of max and min, is stated as follows:

Proposition 4 *Let $\underline{P}_1$, $\underline{P}_2$ be two coherent lower previsions on D_1, $D_2 \supset \{Y : Y = \min(X + h, k), X \in D_1, h, k \in \mathbb{R}\}$ respectively. Then*

$$\underline{P}_c(X) = \underline{P}_1(X) + c\underline{P}_2[(X - \underline{P}_1(X))_-], c \in [0, 1] \quad (6)$$

is a coherent lower prevision on D_1.

The proof relies on the following Lemma.

Lemma 1 *Given $\underline{P}_1$, $\underline{P}_2$ as in Proposition 4,*

$$\underline{P}^*(X) = \underline{P}_2[\min(X, \underline{P}_1(X))] \quad (7)$$

is a coherent lower prevision on D_1.

Proof. By Proposition 1,a), there exist coherent lower previsions extending, respectively, $\underline{P}_1$ and $\underline{P}_2$ on some linear space $\mathcal{L} \supset D_1 \cup D_2$. Using such extensions and (7), $\underline{P}^*$ may be extended on $\mathcal{L}$ too. Consider one such extension (also named $\underline{P}^*$): if it is coherent on $\mathcal{L}$, its restriction on D_1 (our starting $\underline{P}^*$) is coherent too.

Coherence of $\underline{P}^*$ on $\mathcal{L}$ may be proved by checking the axioms in Proposition 1,b). Recall for this that (the extensions of) $\underline{P}_1$ and $\underline{P}_2$, being coherent on $\mathcal{L}$, satisfy all axioms listed in Proposition 1, b) and c).

To check b1) for $\underline{P}^*$, we apply b1) to $\underline{P}_1$, c2) to $\underline{P}_2$ and property $\underline{P}_2(k) = k, \forall k \in \mathbb{R}$ ([14], sec. 2.6.1,(b)). Then $\underline{P}^*(X) = \underline{P}_2[\min(X, \underline{P}_1(X))] \geq \underline{P}_2[\min(X, \inf X)] = \underline{P}_2(\inf X) = \inf X$.

To check b2) for $\underline{P}^*$, apply b2) to $\underline{P}_1$, $\underline{P}_2$: $\underline{P}^*(\lambda X) = \underline{P}_2[\min(\lambda X, \underline{P}_1(\lambda X))] = \underline{P}_2[\lambda \min(X, \underline{P}_1(X))] = \lambda \underline{P}^*(X)$, $\forall \lambda > 0$.

Finally we check b3) for $\underline{P}^*$, using b3), c2) and property

$$\min(a + b, c + d) \geq \min(a, c) + \min(b, d). \quad (8)$$

Then, $\underline{P}^*(X + Y) = \underline{P}_2[\min(X + Y, \underline{P}_1(X + Y))] \geq \underline{P}_2[\min(X + Y, \underline{P}_1(X) + \underline{P}_1(Y))] \geq \underline{P}_2[\min(X, \underline{P}_1(X))] + \underline{P}_2[\min(Y, \underline{P}_1(Y))] = \underline{P}^*(X) + \underline{P}^*(Y)$. □

Proof of Proposition 4. Using, at the second equality, c1) (with $k = \underline{P}_1(X)$) and property $\min(f, 0) + k =$

$\min(f+k,k)$, we can write (6) as follows: $\underline{P}_c(X) = (1-c)\underline{P}_1(X) + c\underline{P}_1(X) + c\underline{P}_2[(X - \underline{P}_1(X))_-] = (1-c)\underline{P}_1(X) + c\underline{P}_2[(X-\underline{P}_1(X))_- + \underline{P}_1(X)] = (1-c)\underline{P}_1(X) + c\underline{P}_2[\min(X,\underline{P}_1(X))] = (1-c)\underline{P}_1(X) + c\underline{P}^*(X)$. Hence $\underline{P}_c$ is coherent, by Proposition 1,d). □

The result in Proposition 3 can be further generalized to the case of convex or C-convex ρ and $\overline{P}$.

Proposition 5 *Let ρ be a convex risk measure on D, $\overline{P}$ a convex upper prevision on D_1. Then $\rho_c(X)$ defined by (5) is a convex risk measure. If ρ and $\overline{P}$ are C-convex, $\rho_c(X)$ is C-convex too.*

Proof. The proof resembles that of Proposition 3: we prove that if $\underline{P}_1$ and $\underline{P}_2$ are convex $\underline{P}_c(X)$ in (6) is convex too, by preliminarily proving that $\underline{P}^*(X)$ in (7) is convex. Much like the proof of Lemma 1, we can check convexity of an extension of $\underline{P}^*$ on a linear space containing real constants $\mathcal{L} \supset D_1 \cup D_2$. This is tantamount to verifying the axioms in Proposition 1,c) for the extended $\underline{P}^*$.

As for c1), we get $\underline{P}^*(X+k) = \underline{P}_2[\min(X+k, \underline{P}_1(X+k))] = \underline{P}_2[\min(X,\underline{P}_1(X))+k] = \underline{P}^*(X)+k$.

To prove c2), let $X \le Y$. Then $\underline{P}_1(X) \le \underline{P}_1(Y)$, $\min(X,\underline{P}_1(X)) \le \min(Y,\underline{P}_1(Y))$ and c2) follows from monotonicity of $\underline{P}_2$, which implies $\underline{P}_2[\min(X,\underline{P}_1(X))] \le \underline{P}_2[\min(Y,\underline{P}_1(Y))]$.

To prove c3), apply: c3) itself, properties of min (including (8)) and c2), getting: $\underline{P}^*(\lambda X + (1-\lambda)Y) = \underline{P}_2[\min(\lambda X + (1-\lambda)Y, \underline{P}_1(\lambda X + (1-\lambda)Y))] \ge \underline{P}_2[\min(\lambda X + (1-\lambda)Y, \lambda\underline{P}_1(X) + (1-\lambda)\underline{P}_1(Y))] \ge \underline{P}_2[\lambda\min(X,\underline{P}_1(X)) + (1-\lambda)\min(Y,\underline{P}_1(Y))] \ge \lambda\underline{P}^*(X) + (1-\lambda)\underline{P}^*(Y)$.

Having thus established that $\underline{P}^*$ is convex, we write, as in the proof of Proposition 4,

$$\underline{P}_c(X) = (1-c)\underline{P}_1(X) + c\underline{P}^*(X). \tag{9}$$

We can do this because the only property of imprecise previsions exploited in the derivation of (9) is c1), which holds for convex lower previsions too. From (9), convexity of $\underline{P}_c$ is immediate using Proposition 1,d). Finally, it is trivial to see that if $\underline{P}_1$ and $\underline{P}_2$ are C-convex then $\underline{P}_c(0) = 0$. □

3.3 Implications for insurance pricing

From the preceding subsection we know that (5) can be employed to get a sort of "second-order" risk measure $\rho_c(X)$ from a previously assessed $\rho(X)$, taking account of the potential inadequacy of $\rho(X)$ to cover residual losses. The measure $\rho_c(X)$ is coherent, alternatively convex, if $\rho(X)$ and $\overline{P}$ are so. There may be many reasons for applying (5): for instance, the use of $\rho(X)$ may be imposed by some regulatory authority but an agent may wish to consider a different, even more prudential measure for certain purposes. Or, conversely, it is the regulatory authority that computes $\rho_c(X)$ on the basis of its own evaluation $\overline{P}$ of the shortfall of the measure ρ adopted by the firm management. This situation is not uncommon, since the management may tend to reserve little money, favouring more profitable (and risky) investments.

To explore yet another interpretation of (5), recall that $\rho(X)$ has the meaning of the infimum price an agent would ask to shoulder X [8], and suppose now $X \le 0$. This is not unusual in insurance, where the insurer asks for a premium to run the risk of paying $-X \ge 0$. Here $\rho(X)$ represents the premium and a rule for determining it is named *premium principle*. A common procedure to obtain a premium principle starts from a fair value for $-X$ (i.e. an expectation or prevision $P(-X)$) and introduces a *loading*, often in a multiplicative form, getting in this case a final price $\overline{P}(-X) = (1+k)P(-X)$, $k > 0$. In [6], the term Dutch premium principle identifies a "double loading" rule, which in our setting can be written as:

$$\rho_{DL} = (1+k)E[\min(-X, d_X)] + (1+c)E[(-X-d_X)_+] \tag{10}$$

with $k, c \ge 0$.

The idea in (10) is that the risk ensuing from X is split between an insurer, which is liable until the threshold d_X, and a reinsurer liable for the residual risk, and both ask for their own loading to be payed by the insured. It is shown in [6] that requiring some reasonable properties reduces ρ_{DL} to ρ'_D with $d = 1$ in (4) and $d_X = E(-X)$, $k = 0$ in (10).

The last constraint, $k = 0$, was interpreted in [6] as impossibility of double loading without violating a condition (*no rip-off*) corresponding to b1) in Proposition 1.

What does Equation (5) tell us about this problem? *If* $\rho(X)$ is greater than the fair value $P(-X)$ it incorporates a loading on X. Then double loading is feasible while obtaining a final measure $\rho_c(X)$ (a premium) which is either coherent or convex, under the assumptions of Propositions 3 or 5, respectively. That is, under these assumptions $\rho_c(X)$ is guaranteed to keep adequate consistency properties and is a generalization of the Dutch risk measure. It is intuitively plausible that the condition for double loading, $\rho(X) > P(-X)$, should hold. The argument may be made more precise when ρ is coherent. In fact, $\rho(X)$ is an upper prevision for $-X$, $\rho(X) = \overline{P}(-X)$. From Proposition 2 we know that $\rho(X) \ge P(-X)$, $\forall P \in \mathcal{M}$, where $\mathcal{M}$ is naturally interpreted as a set containing the "true" (although possibly unknown) prevision P_0 for $-X$. Thus typi-

cally $\rho(X) > P_0(-X)$.

Let us now consider the risk measure $\rho'_D(X)$ in (4), with $d \neq 1$, which was also employed in some papers, including [7]. It is known that $\rho'_D(X)$ satisfies the translation invariance property c1) in Proposition 1 if and only if $d = 1$ [6]. Therefore this measure does not meet the consistency requirement of convexity if $d \neq 1$. Further, its not following translation invariance prevents $\rho'_D(X)$ (and potential analogous generalizations of (5) with $-d\rho(X)$ replacing $-\rho(X)$) from meeting even the much weaker (and in a sense minimal, since it generalizes properties of capacities) consistency notion of centered 1-convexity considered in the next section. Therefore, this kind of generalization does not seem adequate for risk measurement.

Finally, we note that putting $d = 1$ in (4) reduces the second expectation to $E[(E(X) - X)_+]$, which is also a (mild generalization of a) *deviation* measure, following [12]. The correspondence is not necessarily true for the more general term $\overline{P}[(-\rho(X) - X)_+]$ in (5); for instance it does not hold when $\overline{P}$ is C-convex.

4 1-convex and shortfall-dependant imprecise previsions

In the realm of imprecise previsions, the shortfall-dependant measures ρ_c obtained by equation (5) correspond to the lower previsions $\underline{P}_c$ defined by equation (6). Thus equation (6) displays a method for getting a more prudential lower prevision $\underline{P}_c$ from and by means of a previously assessed $\underline{P}_1$.

We might employ for instance (6) when $\underline{P}_1$ is someone else's prevision, considered not fully reliable by us.

The question we are concerned with in this section is: can Proposition 4 be generalized, meaning that asking $\underline{P}_1$, $\underline{P}_2$ to obey weaker consistency requirements than coherence (or convexity), $\underline{P}_c$ satisfies the same consistency conditions? We shall see that the answer is affirmative for a rather mild consistency notion, namely (centered) 1-convexity.

Definition 4 *A map $\underline{P} : D \to \mathbb{R}$ is a 1-*convex *lower prevision on D iff, $\forall X, Y \in D$*

$$\sup[(X - \underline{P}(X)) - (Y - \underline{P}(Y))] \geq 0 \qquad (11)$$

A 1-convex lower prevision is centered *if ($0 \in D$ and) $\underline{P}(0) = 0$.*

We chose the wording "1-convex", exhibiting some alikeness with "1-coherence" in [14](Appendix B), because Definition 4 is actually obtained by imposing $n = 1$ in Definition 2 (putting $s_0 = s_1 = 1$ instead of $s_0 = s_1 = k > 0$ is immaterial for the condition $\sup \underline{G} \geq 0$), i.e. it corresponds to checking convexity only when $n = 1$ (similarly, 1-coherence requires $n = 1$ in Definition 1, plus an extra condition not involved here).

There are other ways of expressing 1-convexity:

Lemma 2 *Given $\underline{P} : D \to \mathbb{R}$,*
a) Condition (11) is equivalent to:

$$X \geq Y + c \Rightarrow \underline{P}(X) \geq \underline{P}(Y) + c, \forall X, Y \in D, \forall c \in \mathbb{R} \qquad (12)$$

b) if D is a linear space, condition (11) is equivalent to translation invariance plus monotonicity, i.e. axioms c1) and c2) in Proposition 1.

Proof. We prove a) (a proof of b) was given in [4]). Suppose (12) holds. Applying it to $X - \sup(X - Y) \leq Y$ we get $\underline{P}(X) - \sup(X - Y) \leq \underline{P}(Y)$, from which (11) follows.

Conversely, let $X \geq Y + c$, hence $-c - \underline{P}(Y) + \underline{P}(X) \geq Y - \underline{P}(Y) - (X - \underline{P}(X))$. Using also (11), $-c - \underline{P}(Y) + \underline{P}(X) \geq \sup[(Y - \underline{P}(Y)) - (X - \underline{P}(X))] \geq 0$, which implies (12). □

Condition (12) is helpful in making a direct comparison between 1-convexity and coherence when D is a convex cone. In fact, in this instance coherence of $\underline{P}$ is equivalent to its jointly satisfying conditions b2), b3) (cf. Proposition 1, b)) and (12) (see [14], p.76).

A remarkable consequence of Lemma 2 is that:

Proposition 6 *Let $\underline{P}$ be a centered 1-convex lower prevision defined on the powerset 2^Ω of a finite partition Ω of events. Then $\underline{P}$ is a capacity.*

Proof. Since $\underline{P}$ is centered, $\underline{P}(0) = 0$. By putting $X = 1$ and $Y = 0$ (0, 1 are the indicators of $\emptyset$, Ω) in Definition 4, we get easily $\underline{P}(1) \leq 1$, whilst the reverse inequality is established by interchanging X and Y. Hence $\underline{P}(1) = 1$. Monotonicity is implied by Lemma 2,a), with $c = 0$. □

Since the proof above is independent of the domain on which $\underline{P}$ is defined, any centered 1-convex $\underline{P}$ is normalized and monotone. It is however not necessarily lower or upper semicontinuous, thus being generally not a fuzzy measure when Ω is infinite.

Functionals defined from a linear space into the compact real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and for which translation invariance and monotonicity hold were termed *niveloids* in [4]. By Lemma 2,b) finite-valued niveloids are 1-convex previsions.

In order to generalize Proposition 4, we preliminarily check whether the results needed for its proof hold for 1-convex previsions too.

It is easy to realize that:

Lemma 3 *If $\underline{P}_1$, $\underline{P}_2$ are 1-convex (or centered 1-convex) on D, then so is $\underline{P} = \lambda\underline{P}_1 + (1-\lambda)\underline{P}_2$, $\lambda \in [0,1]$ (cf. Proposition 1,d)).*

Proof. Condition (11) in Definition 4 is equivalent to $\underline{P}(X) - \underline{P}(Y) \leq \sup(X-Y)$ and holds for $\underline{P}_1$ and $\underline{P}_2$. Then, since $\underline{P}(X) - \underline{P}(Y) = \lambda(\underline{P}_1(X) - \underline{P}_1(Y)) + (1-\lambda)(\underline{P}_2(X) - \underline{P}_2(Y)) \leq \lambda \sup(X-Y) + (1-\lambda)\sup(X-Y) = \sup(X-Y)$, $\forall X, Y \in D$, condition (11) holds also for $\underline{P}$. □

The corresponding generalization of Proposition 1,a) is less immediate. We cannot simply apply an extension theorem for niveloids given in [4] to 1-convex lower previsions, because it does not guarantee that the extension is *finite*, as we need it to be. We proceed then by proving that there always exists a special extension, the *1-convex natural extension* (the notion resembles that of natural extension in [14] or of convex natural extension in [9]):

Definition 5 *Given $\underline{P} : D \to \mathbb{R}$ and a linear space $\mathcal{L} \supset D$, define, for each $Z \in \mathcal{L}$,*
$L(Z) = \{\alpha \in \mathbb{R} \mid Z - \alpha \geq X - \underline{P}(X)$, for some $X \in D\}$.
$\underline{E}(Z) = \sup_\alpha L(Z)$ is the 1-convex natural extension *of $\underline{P}$ on Z ($\underline{E}(Z) = -\infty$ if $L(Z) = \emptyset$).*

When $\underline{P}$ is 1-convex, it is easy to prove that the notion of 1-convex natural extension coincides with that of upper projection of a niveloid given in [4] and, therefore, $\underline{E}$ is a niveloid such that $\underline{E}(X) = \underline{P}(X), \forall X \in D$. We prove that $\underline{E}(Z) \in \mathbb{R}$ $\forall Z \in \mathcal{L}$, hence $\underline{E}$ is a 1-convex extension of $\underline{P}$ to $\mathcal{L}$.

Proposition 7 *Given a 1-convex lower prevision $\underline{P} : D \to \mathbb{R}$, and a linear space $\mathcal{L} \supset D$, $\underline{E}(Z) \in \mathbb{R}$ $\forall Z \in \mathcal{L}$ and $\underline{E}$ is a 1-convex extension of $\underline{P}$.*

Proof. Let $Z \in \mathcal{L}$, $X \in D$ and $\overline{\alpha} = \inf Z - \sup X + \underline{P}(X)$. Hence, $Z - \overline{\alpha} = Z - \inf Z + \sup X - \underline{P}(X) \geq X - \underline{P}(X)$, which implies $\overline{\alpha} \in L(Z)$ (hence $L(Z)$ is non-empty) and $\underline{E}(Z) \geq \inf Z - \sup X + \underline{P}(X) > -\infty$. We show now that $\underline{E}(Z) < +\infty$. If $\alpha \in L(Z)$, there exists $X \in D$ such that $Z - \alpha \geq X - \underline{P}(X)$. Therefore, for any $Y \in D$, $\sup Z - \alpha \geq \sup X - \underline{P}(X) = \sup(X - \inf Y) - \underline{P}(X) + \inf Y \geq \sup(X-Y) - \underline{P}(X) + \inf Y \geq \underline{P}(X) - \underline{P}(Y) - \underline{P}(X) + \inf Y = \inf Y - \underline{P}(Y)$, where (11) is employed in the last inequality. It follows $\underline{E}(Z) \leq \sup Z + \underline{P}(Y) - \inf Y < +\infty$. Since $\underline{E}$ is a niveloid coinciding with $\underline{P}$ on D [4] and $\underline{E}(Z) \in \mathbb{R}$ $\forall Z \in \mathcal{L}$, $\underline{E}$ is a 1-convex extension of $\underline{P}$ to $\mathcal{L}$. □

We can now prove the final result of the section.

Proposition 8 *Let $\underline{P}_1$, $\underline{P}_2$ be two 1-convex lower previsions on D_1, $D_2 \supset \{Y : Y = \min(X+h, k), X \in D_1, h, k \in \mathbb{R}\}$ respectively.*
Then $\underline{P}_c(X) = \underline{P}_1(X) + c\underline{P}_2[(X - \underline{P}_1(X))_-]$, $c \in [0,1]$ is a 1-convex lower prevision on D_1. If $\underline{P}_1$ and $\underline{P}_2$ are centered, then so is $\underline{P}_c$.

Proof. We can follow the guidelines of the proof of Proposition 5. Observe first that we can always suppose that the relevant 1-convex lower previsions are defined and 1-convex on a linear space $\mathcal{L}$ by extending them to $\mathcal{L}$, if necessary. By Proposition 7, there exists a 1-convex extension to $\mathcal{L}$ of a 1-convex lower prevision. Thus $\underline{P}^*(X)$ in equation (7) may be defined on $\mathcal{L}$ with $\underline{P}_1$, $\underline{P}_2$ 1-convex on $\mathcal{L}$. Then $\underline{P}^*(X)$ is 1-convex by Lemma 2,b), because it satisfies c1) and c2) (a proof for this fact is already contained in the proof of Proposition 5). We note then that $\underline{P}_c(X)$ can be decomposed as in (9) when $\underline{P}_1$ and $\underline{P}^*$ are 1-convex. Hence $\underline{P}_c$ is 1-convex by Lemma 3. Finally, if $\underline{P}_1(0) = \underline{P}_2(0) = 0$ then also $\underline{P}^*(0) = \underline{P}_c(0) = 0$. □

5 Conclusions

Resorting to the theory of imprecise previsions, we have generalized a method, originally devised in an insurance pricing framework, for obtaining a second–choice uncertainty measure on the basis of the potential inadequacy of a formerly defined measure.

In particular, as shown in Section 3.1, it is possible (and convenient) to assess independently the initial measure and the measure of its shortfall, which jointly determine the final measure through equation (5).

As a further advantage, the resulting measure conforms to the well established consistency notions of coherence or (centered) convexity, provided that the other measures in the procedure comply with the same (or more stringent) requirements.

As a practical application in the framework of insurance pricing, this approach provides a formal support to the policy of double loading, while ensuring the desirable properties guaranteed by the various consistency notions.

The choice of the consistency criterion to be employed may depend on many factors. Undoubtedly coherence seems preferable [1, 14], but arguments in favour of convexity or centered convexity were also brought forth [5, 9]. 1–convexity is probably too weak for risk measurement, but could be useful for other kinds of applications in the realm of imprecise previsions. It is anyway interesting to notice that the method summarized by (5) can be applied as far as to consider 1-convexity. The generalization of (5) in Proposition 8 seems the largest operationally relevant: 1-convexity is a really minimal consistency requirement, as can be seen from the displayed comparisons with the concepts

of capacity and niveloid.

Further extensions of this work should therefore address different questions. An appealing and largely unexplored area is that of investigating shortfall-based conditional risk measures (and previsions). Here the set D should be made of conditional random variables like $X|B$, where B is a non-impossible event and the conditioning events for the variables in D are generally different. Notions of coherence and C-convexity with related fundamental properties are available in such a framework [10, 11, 14], and equation (2) easily generalises to $\rho(X|B) = \overline{P}(-X|B) = -\underline{P}(X|B)$. The point is how should equation (5) be generalised to guarantee some properties that are similar to those of Propositions 4 and 5. An immediate difficulty is that the proof of these propositions relies on Proposition 1 d), which is known to admit no analogue in the conditional environment.

Acknowledgement

We wish to thank the anonymous referees for their helpful comments.

References

[1] P. Artzner, F. Delbaen, S. Eber, D. Heath, Coherent measures of risk, Mathematical Finance 9 (1999), 203–228.

[2] B. de Finetti, Theory of Probability, Wiley, 1974.

[3] J. Dhaene, S. Vanduffel, Q. Tang, M.J. Goovaerts, R. Kaas, D. Vyncke, Solvency capital, risk measures and comonotonicity: a review, Research Report 0416, Department of Applied Economics, Catholic University of Leuven (www.kuleuven.ac.be/insurance), 2004.

[4] S. Dolecki, G.H. Greco, Niveloids, Topological Methods in Nonlinear Analysis 5 (1995), 1–22.

[5] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance Stochast. 6 (2002), 429–447.

[6] A.E. van Heerwarden, R. Kaas, The Dutch premium principle, Insurance: Mathematics and Economics 11 (1992), 129–133.

[7] W. Hürlimann, A note on experience rating, reinsurance and premium principle, Insurance: Mathematics and Economics 14 (1994), 197–204.

[8] R. Pelessoni, P. Vicig, Imprecise Previsions for Risk Measurement, International Journal of Uncertainty, Fuzziness and Knowledge-based Systems 11 (2003), 393–412.

[9] R. Pelessoni, P. Vicig, Convex imprecise previsions, Reliable Computing 9 (2003), 465–485.

[10] R. Pelessoni, P. Vicig, Uncertainty modelling and conditioning with convex imprecise previsions, International Journal of Approximate Reasoning, 39 (2005), 297-319.

[11] R. Pelessoni, P. Vicig, Envelope Theorems and Dilation with Convex Conditional Previsions, in: F. G. Cozman, R. Nau, T. Seidenfeld (Eds.) *Proc. of 4th International Symposium on Imprecise Probabilities and Their Applications (ISIPTA 05)*. SIPTA, Pittsburgh, Pennsylvania, 2005, pp. 266-275.

[12] R.T. Rockafellar, S. Uryasev, M. Zabarankin, Generalized deviations in risk analysis, Finance Stochast. 10 (2006), 51–74.

[13] P. Vicig, Financial risk measurement with imprecise probabilities, submitted to International Journal of Approximate Reasoning.

[14] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, 1991.

Representation of two bipolar decision strategies with generalized Choquet integrals

Christophe Labreuche
THALES Research and Technology
RD 128, 91767 Palaiseau Cedex, France
christophe.labreuche@thalesgroup.com

Michel Grabisch
LIP 6, Université Paris I La Sorbonne
8, rue du Capitaine Scott 75015 Paris
michel.grabisch@lip6.fr

Abstract

We study the notion of conditional relative importance in a quantitative framework. This is an important notion in the context of the Choquet integral since this latter is usually motivated by this decisional behavior. A systematic investigation on generalizations of the Choquet integral is performed. The determination of the utility function is central in this analysis.

Keywords: Multiple criteria analysis, Choquet integral, bi-capacity, reference point.

1 Introduction

In Multiple Criteria Decision Aiding (MCDA), it is very seldom that an option dominates the other ones over all or most of the decision criteria. The selection of the *best* option is often performed by prioritizing the criteria. However, the semantics of the prioritization depend on the framework that is considered. We are interested here in options that are a good compromise between the criteria. More precisely, we look for *compensation* between criteria, that is that well-satisfied criteria can compensate ill-satisfied ones. It is well-known that quantitative Multi-Attribute Utility Theory (MAUT) [11] models such as weighted sums are the best suited for representing compensation. The prioritization of the criteria is then performed, based on the notion of "relative importance" of one or several criteria compared to other ones.

Yet, simple models such as the weighted sum are not always sufficient. The decision model has indeed to be rich enough to model the decisional behaviors of the decision maker. This view tends to imply the use of very elaborate models. Two trends have been recently observed. Linear aggregation functions are often not satisfactory, in particular due to the existence of interaction between criteria. Veto or favor are examples of extreme interaction. This has lead to the introduction of the Choquet integral as an aggregation function in MCDA. Another complexity is the possible existence of a special element on the scale, called neutral element, such that the decision behaviors are quite different for values above or below this element. Bipolar aggregation functions must then be used in such situations.

Bi-capacities have been introduced in MCDA for taking into account both interaction between criteria and bipolarity. More generally, bi-capacities are supposed to represent bipolar decision strategies, i.e. decision strategies which are not the same when the value of an attribute is above or below the neutral element.

We investigate in this paper two particular decision behaviors: bipolar conditional relative importance, and bipolar veto. Such strategies are very interesting for two reasons. The first one is that these behaviors are usually easier to obtain from a decision maker than quantitative preferential information. In some cases, the decision maker can naturally express statements of this kind. The second one is that we want to make an in-depth analysis of bipolar conditional relative importance and investigate its close link with interaction between criteria.

The two decision strategies that are considered here are described in Section 2 and motivated by examples. The qualitative and quantitative notion of independence and conditional relative importance are given in Section 3. The definition of bipolarity and of the Choquet integral is recalled in Section 4. The representation of the first bipolar behavior, which is conditional relative importance, is investigated in Section 5. A general analysis of this behavior under piecewise linear aggregation is performed. We present a general framework for the construction of utility functions. It is shown that the bipolar conditional relative importance cannot be strictly satisfied by any piecewise linear aggregation function. However, by relaxing this decision behavior, the non-representation result disap-

pears. It turns out that the Choquet integral w.r.t. a bi-capacity can represent that relaxed decision behavior, whereas the Choquet integral w.r.t. a usual capacity cannot. Finally, Section 6 presents an analysis of the bipolar veto decision behavior.

2 Two examples of decision strategies

2.1 Bipolar conditional relative importance

Consider the situation of financial analysts in a bank that would like to decide on customers credit granting. Three factors have been retained: the debt ratio (denoted by D) given in %, an indicator P reflecting the behavior of the customer in the past (for instance the number of late payments) and the capital contribution ratio (denoted by C) given in %. The main factor (i.e. the most important one) is the debt ratio D. However, the preference of the analysts over P and C is not so simple. For a customer that has an attractive debt ratio, the contribution rate is not so important so that P is more important than C. On the other hand, for a customer that has a bad debt ratio, there is relative substitutability between D and P so that the analysts hopes at least that the contribution ratio is good. Hence, C becomes more important than P.

Let us give another example. Consider the director of a university that decides on students who are applying for graduate studies in Economics on the basis of an assessment of their skills in Mathematics (M), Statistics (S) and Languages(L). The director feels that Mathematics is the most important criteria. However, his preference over S and L is not that simple. There is a relative substitutability between M and S. Hence, for an applicant good in Mathematics, the director prefers if he is furthermore good in L than if he is also good in S, so that L is more important than S in this case. On the contrary, for candidates bad in Mathematics, the director hopes they are at least good in S since the director basically looks for applicants with strong scientific background. Hence, S becomes more important than L.

The previous two examples exhibit the same decisional pattern, in which the relative importance of one attribute k^+ compared to another one k^- may depend on the value of a third attribute k being good or bad. This type of behavior is called "*conditional relative importance*". One can express the statement of this behavior as follows.

> **(S1)**: If the value w.r.t. criterion k is very well-satisfied, then criterion k^+ is more important than criterion k^-. If the value w.r.t. criterion k is very ill-satisfied, then criterion k^+ is less important than criterion k^-.

Such statements are often intuitive for actors. The Choquet integral has been shown to represent this type of statement [7]. The relative importance of criteria k^+ and k^- is specified in statement **(S1)** only for the two extreme levels of performance (very good and very bad respectively) on criterion k. One wonders what happens for intermediate values on criterion k. One should whenever possible specify the preferences between criteria k^+ and k^- for all values of criterion k. A generalization of statement **(S1)** is the following one :

> **(S2)**: If value w.r.t. criterion k is "well-satisfied", then criterion k^+ is more important than criterion k^-. If value w.r.t. criterion k is "ill-satisfied", then criterion k^+ is less important than criterion k^-.

In the previous statement, the values judged as well-satisfied and ill-satisfied are supposed to form a partition of the scale. There thus exists on attribute k a particular element called "neutral element" such that better elements are considered as well-satisfied and worse elements are considered as ill-satisfied for the actor. Hence, the scale underlying the criterion k is of *bipolar nature*. The classical Choquet integral fails to represent **(S2)** [8, 14].

2.2 Bipolar veto

The engineering of complex systems is a difficult task since all components of the system interact together in a hard-to-predict way. An analysis of each component separately is not enough. There are many consequences that have to be analyzed when considering a system as a whole. Some of these aspects concern the measure of the performance on the system. This often requires large simulations run on several scenarios. The indicators on which the analysis of the results of the simulations is performed are called *metrics*. They correspond to the so-called *functional* criteria. On the other hand, there are also *non-functional* criteria, i.e. criteria that can be assessed without the use of these simulators. One can mention for instance, acquisition and possession costs, and the technical readiness levels of the components of the system.

The functional criteria often correspond to fuzzy requirements given by the customer. Hence, if the functional criteria are ill-satisfied, then the customer will be ill-satisfied whatever the value on the non-functional attributes. This means that the functional criteria behaves like a veto. Now, when the functional criteria are well-satisfied, the customer now seeks for

systems that also have good figures in non-functional parts. Hence there is compensation between the functional and the non-functional criteria. Assuming that the functional criteria have been gathered in one attribute denoted by x_F, and that all non-functional criteria have been gathered in one attribute x_{NF}, we obtain the following rule.

> **(R1)**: If the value w.r.t. criterion k_F is "ill-satisfied", then criterion k_F is a veto. If the value w.r.t. criterion k_F is "well-satisfied", then criterion k_{NF} can compensate k_F.

3 Representation of the preferences

3.1 Construction of the preferences on each attribute

Consider a problem of selecting one option among several, where each option is described by several attributes. $N = \{1, \ldots, n\}$ is the set of attributes and the set of possible values of attribute $i \in N$ is denoted by X_i. Options are thus elements of the product set $X := X_1 \times \cdots \times X_n$. The preferences of the DM over the options can be described by a preference relation $\succeq$ over X. For $x, y \in X$, $x \succeq y$ means that x is at least as good as y according to the DM.

Considering two acts $x, y \in X$ and $S \subseteq N$, we use the notation (x_S, y_{-S}) to denote the compound act $w \in X$ such that $w_i = x_i$ if $i \in S$ and y_i otherwise. Likewise, options $(x_S, y_T, z_{-S\cup T})$ denotes the compound act $w \in X$ such that $w_i = x_i$ if $i \in S$, $w_i = y_i$ if $i \in T$, and $w_i = z_i$ otherwise.

Representing $\succeq$ by a numerical or graphical model demands to address two issues: the preferences over each attribute and the aggregation of these preferences. Speaking of a preference relation focusing only on one attribute implies that the other attributes could have been removed. The existence of a preference relation on each attribute classically relies on *weak separability* [12].

Definition 1 *A relation $\succeq$ is said to be* weakly separable *if for every $i \in N$, and for all $x_i, x'_i \in X_i$, $y_{N\setminus\{i\}}, y'_{N\setminus\{i\}} \in X_{N\setminus\{i\}}$,*

$$(x_i, y_{-i}) \succeq (x'_i, y_{-i}) \iff (x_i, y'_{-i}) \succeq (x'_i, y'_{-i}) \ .$$

Then, for $i \in N$, the marginal preference relation $\succeq_i$ on attribute i is defined on X_i as follows

$$x_i \succeq_i y_i \iff \forall z_{N\setminus\{i\}} \in X_{N\setminus\{i\}} \ , \ (x_i, z_{-i}) \succeq (y_i, z_{-i}) \ .$$

This property can be interpreted as a weak independence between attributes. This assumption is essential for quantitative models based on an overall utility function. From [12], $\succeq$ can be represented by functions $u_i : X_i \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}$:

$$x \succeq y \iff U(x) \geq U(y) \tag{1}$$

where

$$U(x) = F(u_1(x_1), \ldots, u_n(x_n)) \tag{2}$$

iff $\succeq$ is a weak order, is weakly separable and satisfies a technical topological assumption.

There are representations in which the weak-separability assumption is relaxed. Weak separability implies that the elements of an attribute can be ranked independently of the values w.r.t. the other attributes. One can indeed say that $x_i \in X_i$ is preferred to $x'_i \in X_i$ "*everything else being equal*" (or "*Ceteris Paribus*"), that is if $(x_i, y_{-i}) \succeq (x'_i, y_{-i})$ for any $y_{N\setminus\{i\}} \in X_{N\setminus\{i\}}$. However, it may happen that the partial preferences over an attribute $i \in N$ are conditional on the value of some other attributes. This leads to generalizing Definition 1, yielding the notion of partial preferential independence [11].

Definition 2 *Let $S \subseteq N$. Attribute $i \subseteq N$ is said to be* conditionally preferentially independent *of $N \setminus (S \cup \{i\})$ for S if for all $x_i, x'_i \in X_i$, $y_{N\setminus(S\cup\{i\})}, y'_{N\setminus(S\cup\{i\})} \in X_{N\setminus(S\cup\{i\})}$ and $z_S \in X_S$,*

$$(x_i, z_S, y_{-(S\cup\{i\})}) \succeq (x'_i, z_S, y_{-(S\cup\{i\})}) \iff (x_i, z_S, y'_{-(S\cup\{i\})}) \succeq (x'_i, z_S, y'_{-(S\cup\{i\})}) \ .$$

Given an attribute i, we denoted by $\mathrm{Pa}(i) \subseteq N \setminus \{i\}$ the attributes under which the partial preferences over an attribute $i \in N$ depend. We note $\succeq_i^{z_{\mathrm{Pa}(i)}}$ on X_i as follows

$$x_i \succeq_i^{z_{\mathrm{Pa}(i)}} y_i \iff \forall t_{N\setminus(\mathrm{Pa}(i)\cup\{i\})} \in X_{N\setminus(\mathrm{Pa}(i)\cup\{i\})} \ (x_i, z_{\mathrm{Pa}(i)}, t_{-(\mathrm{Pa}(i)\cup\{i\})}) \succeq (y_i, z_{\mathrm{Pa}(i)}, t_{-(\mathrm{Pa}(i)\cup\{i\})})$$

A *Conditional Preference network* (CP-net in short) is a network of conditional preferences on the attributes. It is defined as follows [1].

Definition 3 *A* CP-net *over attributes N is a directed graph G over N whose nodes are annotated with conditional preference tables. For a node $i \in N$,* $\mathrm{Pa}(i)$ *is the set of all antecedents of i in the graph. The conditional preference table for i is composed of the conditional preferences $\succeq_i^{z_{\mathrm{Pa}(i)}}$ over X_i for all $z_{Pa(i)} \in X_{\mathrm{Pa}(i)}$.*

This defines a qualitative representation of $\succeq$ that is compact.

3.2 Aggregation of the preferences on each attribute

From a knowledge of only the preferences with respect to each attribute separately, one is interested in all the preference relation compatible with this a priori information. If $\succeq$ is assumed to satisfy weak separability, the *least specific* preference relation compatible with $\succeq_i$ is the Pareto ordering relation

$$x \succeq_{\text{Pareto}} y \iff \forall i \in N \ , \ x_i \succeq_i y_i$$

Most of the interesting couples of options of X are incomparable according to the ordering. In CP-nets, one looks for the preference relations that are compatible with a given CP-net. Depending on the graph structure, usually there are also many couples of options that are incomparable.

In order to reduce the number of incomparabilities in the preference relation, information on how to combine several attributes together must be added. Looking at Definition 2, one sees that it is not sure that an actor can compare two options that are not equal on all attributes except one. This is quite restrictive. This is why CP-nets have been extended to allow more general preference representations than conditionally preferentially independence. The extension is called *Tradeoff CP-nets* (*TCP-nets* in short) [3]. The notion of qualitative importance between two attributes is introduced in this formalism. Attribute $i \in N$ is said to be *more important* than attribute $j \in N$ (we note $i \rhd j$) iff

$$\begin{array}{l} \forall w_{N\setminus\{i,j\}} \in X_{N\setminus\{i,j\}} \ , \ \forall x_i, y_i \in X_i \ , \ \forall z_j, t_j \in X_j \\ \text{with } x_i \succeq_i^{w_{N\setminus\{i,j\}}} y_i \ , \\ \qquad (x_i, z_j, w_{N\setminus\{i,j\}}) \succeq (y_i, t_j, w_{N\setminus\{i,j\}}) \end{array}$$

In other words, when i is more important than j, this implies that, for two options differing only on attributes i and j, one prefers the options that has the preferred value on attribute i whatever the value on attribute j.

An attribute can be more important than an another one, depending from the value w.r.t. some other attributes. This leads to the notion of *conditional relative importance* between attributes. Attribute $i \in N$ is said to be *more important* than attribute $j \in N$ given $v_S \in X_S$ (we note $i \rhd^{v_S} j$), with $S \cap \{i,j\} = \emptyset$, iff

$$\begin{array}{l} \forall w_{N\setminus(S\cup\{i,j\})} \in X_{N\setminus(S\cup\{i,j\})} \ , \ \forall x_i, y_i \in X_i \ , \\ \forall z_j, t_j \in X_j \text{ with } x_i \succeq_i^{(v_S, w_{N\setminus(S\cup\{i,j\})})} y_i \ , \\ \qquad (x_i, z_j, v_S, w_{-(S\cup\{i,j\})}) \succeq (y_i, t_j, v_S, w_{-(S\cup\{i,j\})}) \end{array}$$

One can have for instance $i \rhd^{v_S} j$ for some $v_S \in X_S$ and $j \rhd^{v'_S} i$ for some other $v'_S \in X_S$.

There is no compensation in the previous qualitative notion of relative importance. For instance, if $1 \rhd 2 \rhd \cdots \rhd n$, then $\succeq$ is the lexicographic ordering. In a quantitative setting, the relative importance of a criterion in an aggregation function F is clearly defined when F is a weighted sum $F(s_1, \ldots, s_n) = \sum_{k=1}^n w_k s_k$, where all weights w_k are non-negative and they sum up to one. The importance of criterion k is then its associated weight w_k. More generally, if F is continuous piecewise linear, then there exists a partition of $\mathbb{R}^n$ such that F is a weighted sum on each domain of the partition. The importance of a criterion is thus defined in each domain as its relative weight. Since any smooth function can be approximated by continuous piecewise linear functions, this leads to defining the importance of criterion k at point $s \in \mathbb{R}^n$ as the partial derivative $\frac{\partial F}{\partial s_k}(s)$ (since the weight of criterion k in the continuous piecewise linear approximation converges to $\frac{\partial F}{\partial s_k}(s)$ when the size of the mesh describing the approximation tends to zero). Yet, this interpretation is not so simple for an actor.

So, we will stick to continuous piecewise linear aggregation functions so that the notion of relative importance clearly makes sense to the DM. He can indeed elicit a statement such as **(S2)** only if the notion of relative importance is clear to him. From a measurement standpoint [12], piecewise linearity is derived from the property of invariance to linear changes of the scales [14]. The weight of criterion l at point $s \in \Omega$ is denoted by $w_l(s)$:

$$F(s) = \sum_{l \in N} w_l(s)\, s_l \ ,$$

where $w_l(s)$ is piecewise constant. These weights are said to be *normalized* if, for any s, they are non-negative and sum up to one.

A piecewise linear function F is characterized by a partition $\Phi(F)$ of $\mathbb{R}^n$ on which F is a weighted sum on each element of the partition. Hence

$$\Phi(F) = \left\{ \Phi_F^1, \ldots, \Phi_F^{\phi(F)} \right\}$$

where for all $l \in \{1, \ldots, \phi(F)\}$, there exists normalized weights $w_1^{\Phi_F^l}, \ldots, w_n^{\Phi_F^l}$ such that one has, for all $s \in \Phi_F^l$, $F(s) = \sum_{k \in N} w_k^{\Phi_F^l} s_k$. The partition is unique if each subset Φ_F^l is connected and

$$\overline{\Phi_F^l} \cap \overline{\Phi_F^{l'}} \neq \emptyset \implies \exists k \in N \ , \ w_k^{\Phi_F^l} \neq w_k^{\Phi_F^{l'}} \ . \quad (3)$$

Criterion i is said to be more important than j relatively to point $v \in \mathbb{R}^n$ if

$$w_i^{\Phi_F^l} > w_j^{\Phi_F^l}$$

where $v \in \Phi_F^l$.

4 Bipolarity and Choquet integral

In this paper, we are interested in the quantitative model (1). The notion of relative importance in quantitative models implies that the attributes must be made comparable. Let us consider indeed the following relation

$$(x_i, y'_j, z_{-\{i,j\}}) \sim (x'_i, y_j, z_{-\{i,j\}})$$

where $x_i \neq x'_i$ and $y_j \neq y'_j$. The previous relation expresses that the change from x_i to x'_i on attribute i is similar to the change from y_j to y'_j on attribute j. In order to be able to say that the criteria i and j have the same importance, the change from x_i to x'_i on attribute i shall represent the same difference of satisfaction as the change from y_j to y'_j on attribute j. In other word, one can define the notion of importance of criteria only once the attributes have been make commensurable.

Since the aggregation functions we use are based on sums, the utility functions must be constructed as interval scales [12].

4.1 Bipolar and unipolar scales

It may exists in X_k a particular element or level $\mathbf{0}_k$, called *neutral level*, such that if $x_k \succ_k \mathbf{0}_k$, then x_k is considered as "good", while if $x_k \prec_k \mathbf{0}_k$, then x_k is considered as "bad" for the actor.

Such a neutral level exists whenever relation $\succeq_k$ corresponds to two opposite notions of common language. For example, this is the case when $\succeq_k$ means "more attractive than", "better than", etc., whose pairs of opposite notions are respectively "attractiveness/repulsiveness", and "good/bad". By contrast, relations "more satisfactory than", "more allowed than", "belongs more to category C than" do not clearly exhibit a neutral level.

A scale is said to be *bipolar* if X_k contains such a neutral level. A *unipolar scale* has no neutral level. As an example, preference statement **(S2)** implies that criterion i is of bipolar nature.

4.2 Aggregation of unipolar scales

A *capacity* [4], also called *fuzzy measure*, is a set function $\nu : 2^N \to \mathbb{R}$ satisfying $\nu(\emptyset) = 0$, $\nu(N) = 1$, and $A \subset B$ implies $\nu(A) \leq \nu(B)$. In MCDA, $\nu(A)$ is interpreted as the overall assessment of the binary act $(1_A, 0_{-A})$.

The Choquet integral [4] defined w.r.t. a capacity ν has the following expression :

$$\mathcal{C}_\nu(s_1, \ldots, s_n) = s_{\pi(1)}\nu(N) + \sum_{i=2}^{n} \left(s_{\pi(i)} - s_{\pi(i-1)}\right) \nu\left(A_{\pi(i)}\right) \tag{4}$$

where $s_{\pi(1)} \leq s_{\pi(2)} \leq \cdots \leq s_{\pi(n)}$, $A_{\pi(i)} = \{\pi(i), \cdots, \pi(n)\}$ and $s_1, \ldots, s_n \in \mathbb{R}_+$. We also have

$$\mathcal{C}_\nu(s) = \sum_{i=1}^{n} s_{\pi(i)} \left[\nu(A_{\pi(i)}) - \nu(A_{\pi(i+1)})\right] \ .$$

Clearly, the Choquet integral w.r.t. a capacity is continuous piecewise linear. Moreover, one has

$$\Phi\left(\mathcal{C}_\nu\right) = \{\Omega_\pi \ , \ \pi \text{ permutation on } N\}$$

where $\Omega_\pi = \left\{s \in \mathbb{R}^n \ , \ s_{\pi(1)} \leq s_{\pi(2)} \leq \cdots \leq s_{\pi(n)}\right\}$.

4.3 Aggregation on bipolar scales

The Choquet integral has a natural extension to bipolar scales. The limitation of the Choquet integral w.r.t. capacities is that the overall evaluation at any point is computed from information coming only from the attractive part (i.e. the parameters of the capacity correspond to the overall assessment of the positive binary acts $(1_A, 0_{-A})$). Hence, the notion of capacity is not suited to deal with real bipolar scales. The idea is thus to generalize the notion of capacity. Let

$$\mathcal{Q}(N) = \{(A, B) \in \mathcal{P}(N) \times \mathcal{P}(N) \mid A \cap B = \emptyset\} \ .$$

A *bi-capacity* [8, 14] is a function $\mu : \mathcal{Q}(N) \to \mathbb{R}$ satisfying $\mu(\emptyset, \emptyset) = 0$, $\mu(N, \emptyset) = 1$, $\mu(\emptyset, N) = -1$, $A \subseteq A'$ implies $\mu(A, B) \leq \mu(A', B)$, and $B \subseteq B'$ implies $\mu(A, B) \geq \mu(A, B')$. The last two properties depict increasing monotonicity. In MCDA, $\mu(A, B)$ is interpreted as the overall assessment of the ternary act $(1_A, -1_B, 0_{-A\cup B})$.

The Choquet integral w.r.t. a bi-capacity μ proposed in [8] is now given. For any $A \subseteq N$, $s \in \Sigma_A$,

$$\mathcal{BC}_\mu(s) := \mathcal{C}_\nu\left(s_A, -s_{-A}\right) = \mathcal{C}_\nu\left(|s|\right)$$

where $\nu(C) := \mu\left(C \cap A, C \setminus A\right)$ and $\Sigma_A := \{s \in \mathbb{R}^n \ , s_A \geq 0 \ , s_{-A} < 0\}$. Let τ be a permutation such that $|s_{\tau(1)}| \leq \ldots \leq |s_{\tau(n)}|$, and

$$\begin{aligned} A^+_{\tau(i)} &= \{\tau(i), \cdots, \tau(n)\} \cap A \\ &= \left\{\tau(j) \ , \ j \geq i \text{ and } s_{\tau(j)} \geq 0\right\} \ , \end{aligned}$$

$$\begin{aligned} A^-_{\tau(i)} &= \{\tau(i), \cdots, \tau(n)\} \cap (N \setminus A) \\ &= \left\{\tau(j) \ , \ j \geq i \text{ and } s_{\tau(j)} < 0\right\} \ . \end{aligned}$$

Then one can write

$$\mathcal{BC}_\mu(s) = \sum_{i=1}^{n} \left|s_{\tau(i)}\right| \left[\mu(A^+_{\tau(i)}, A^-_{\tau(i)}) - \mu(A^+_{\tau(i+1)}, A^-_{\tau(i+1)})\right] \quad (5)$$

Clearly, the Choquet integral w.r.t. a bi-capacity is continuous piecewise linear. Moreover, one has

$$\Phi\left(\mathcal{BC}_\mu\right) = \{\Omega_{A,\tau}\ ,\ A \subseteq N \text{ and } \tau \text{ permutation on } N\}$$

where $\Omega_{A,\tau} = \left\{s \in \Sigma_A\ ,\ |s_{\tau(1)}| \leq \ldots \leq |s_{\tau(n)}|\right\}$.

5 Representation of Statement (S2) by an aggregation function

5.1 Construction of the utility functions

As we have seen in Section 4.1, statement **(S2)** clearly exhibits a bipolar behavior on criterion k. Due to commensurateness between the criteria, all criteria are considered as bipolar. Statement **(S2)** shows that there are interaction between criteria in preference relation over options. The construction of the utility functions u_i is thus more complex than for the case where all criteria are independent (see for instance the *utility independence* assumption [11]). Utility functions u_i have a priori no intrinsic meaning, and only make sense through the overall utility U thanks to (2). When all criteria are independent, $U(x_i, z_{-i}) = F(u_i(x_i), u_{-i}(z_{-i}))$ and $u_i(x_i)$ are two equivalent interval scales for any $z_{-i} \in X_{-i}$ fixed. This relation gives a sense to u_i. In this relation, the presence of z_{-i} is not essential so that u_i can be considered as a utility representation of a preference relation over attribute X_i *all else being equal* (i.e. the value w.r.t. the other attributes being fixed to any value). The utility functions can thus be considered and constructed separately.

When there are some interactions among criteria, the "all else being equal" assumption does not hold anymore. The choice of the reference z_{-i} from which the utility function is constructed becomes essential. For a given z_{-i}, if the following assumption

$$\exists l \in \{1, \ldots, \phi(F)\}\ ,\ \ \forall x_i \in X_i \quad (u_i(x_i), u_{-i}(z_{-i})) \in \Phi^l_F \quad (6)$$

holds then the utility function u_i can be constructed as for the weighted sum since the options used in the construction of u_i remain in the same domain Φ^l_F of linearity of F. Hence F does not alter the perception of u_i through $U(x_i, z_{-i})$. For instance, if F is a Choquet integral w.r.t. a capacity, z_{-i} is considered at either the lowest or the highest satisfaction level on each attribute $l \neq i$ [13]. These two extreme values correspond to two reference levels. If F is a Choquet integral w.r.t. a bi-capacity, the attractive and repulsive parts of u_i must necessarily be constructed separately in order to have (6) (see the end of Section 4.3). The neutral level becomes an essential point in the construction of utility functions. Apart from the neutral level, one reference level is required on the attractive part and the repulsive one in order to normalize the scale. Hence three reference levels are necessary [14].

The actor is first asked to identify on each attribute X_i a neutral element $\mathbf{0}_i$ that is considered as neither good nor bad [16]. Since statement **(S2)** clearly relies on the existence of such level on attribute k, one can assume that the actor who provides **(S2)** can identify the value of $\mathbf{0}_k$. We assume here that this neutral element can also be identified on the other attributes. It is assumed furthermore that there exists an element denoted by $\mathbf{1}_i$ that is considered as good and completely satisfactory, even if more attractive elements could exist on this point of view [13]. The existence of such reference level comes from the theory of satisficing bounded rationality [17]. We assume finally that there exists an element denoted by $-\mathbf{1}_i$ that is considered as bad and unsatisfactory. Element $-\mathbf{1}_i$ is somehow symmetric to $\mathbf{1}_i$. More precisely, $-\mathbf{1}_i$ corresponds to the same level of appreciation in the repulsive scale than $\mathbf{1}_i$ in the attractive scale. All levels have the same absolute meaning throughout the criteria, so we impose:

$$\begin{aligned} &u_1(-\mathbf{1}_1) = \cdots = u_n(-\mathbf{1}_n) = -1\ , \\ &u_1(\mathbf{0}_1) = \cdots = u_n(\mathbf{0}_n) = 0\ , \\ &u_1(\mathbf{1}_1) = \cdots = u_n(\mathbf{1}_n) = 1\ . \end{aligned}$$

Since the attractive and repulsive values refer to different affect stimuli [18], it may be more appropriate to construct separately the positive and the negative parts of the partial utility functions in order to make the actor compare attractive values with repulsive ones. If F satisfies for all $i \in N$

$$\begin{aligned} &\exists l \in \{1, \ldots, \phi(F)\}\ ,\ \ \forall s_i \geq 0\ \ (s_i, 0_{-i}) \in \Phi^l_F \\ &\exists l' \in \{1, \ldots, \phi(F)\}\ ,\ \ \forall s_i \leq 0\ \ (s_i, 0_{-i}) \in \Phi^{l'}_F \end{aligned} \quad (7)$$

then one can construct the utility function u_i from $U(x_i, \mathbf{0}_{-i})$:

$$\begin{aligned} &\forall x_i \succeq_i \mathbf{0}_i\ ,\ \ u_i(x_i) = \frac{U(x_i, \mathbf{0}_{-i}) - U(\mathbf{0}_N)}{U(\mathbf{1}_i, \mathbf{0}_{-i}) - U(\mathbf{0}_N)} \\ &\forall x_i \preceq_i \mathbf{0}_i\ ,\ \ u_i(x_i) = \frac{U(x_i, \mathbf{0}_{-i}) - U(\mathbf{0}_N)}{U(\mathbf{0}_N) - U(-\mathbf{1}_i, \mathbf{0}_{-i})} \end{aligned}$$

Relation (7) is satisfied by the Choquet integral w.r.t. a bi-capacity.

5.2 General analysis of rule (S2)

Rule **(S2)** can be stated in the more precise form

> **(S2')**: If the value w.r.t. criterion k is attractive (i.e. $x_k \succ_k \mathbf{0}_k$) then criterion k^+ is more important than criterion k^-. If the value w.r.t. criterion k is repulsive (i.e. $x_k \prec_k \mathbf{0}_k$) then criterion k^+ is less important than criterion k^-.

Aggregation functions are henceforth assumed to be continuous since a slight modification in the argument shall also result in a slight change in the overall utility [6].

Theorem 1 shows that statement **(S2')** cannot be thoroughly modeled in all situations. Some restrictions will thus be made. One can show the following result.

Theorem 1 *There does not exist any continuous and piecewise linear aggregation function F, for which statement* **(S2')** *is satisfied.*

Let $\mathbb{R}^n_{+i} = \{s \in \mathbb{R}^n\ ,\ s_i \geq 0\}$, $\mathbb{R}^n_{-i} = \{s \in \mathbb{R}^n\ ,\ s_i \leq 0\}$ and $\mathbb{R}^n_{0\,i} = \{s \in \mathbb{R}^n\ ,\ s_i = 0\}$. More precisely, let $\Phi^+, \Phi^- \in \Phi(F)$ such that $\Phi^+ \subseteq \mathbb{R}^n_{+i}$ and $\Phi^- \subseteq \mathbb{R}^n_{-i}$. Assume that $\Gamma := \overline{\Phi^+} \cap \overline{\Phi^-} \neq \emptyset$. If $\Gamma \subseteq \mathbb{R}^n_{0\,i}$ is parallel to the axis of criteria k^+ and k^-, then statement **(S2')** *cannot be satisfied in both Φ^+ and Φ^-.*

Under the condition of the previous theorem, the variables s_{k^+} and s_{k^-} are free on the boundary Γ, even though criterion k vanishes. The following Corollary provides a special case of this relation between criteria k, k^+, k^-.

Corollary 1 *Assume that the domains of $\Phi(F)$ correspond to that of a bi-capacity, i.e. $\Phi(\mathcal{BC})$. If criterion k is the one closest to the neutral level among criteria k, k^+, k^-, then the weights of criteria k^- and k^+ are not conditional on the fact that criterion i is attractive or repulsive (i.e. statement* **(S2')** *cannot be satisfied).*

The result of Theorem 1 is not true when F is not continuous piecewise linear. Consider indeed the following nonlinear aggregation function

$$F(s) = \frac{1 + s_k}{2} \times s_{k^+} + \frac{1}{2} \times s_{k^-}\ .$$

Then the weight of criterion j^+ is $\frac{\partial F}{\partial s_{k^+}}(s) = \frac{1+s_k}{2}$ and that of criterion k^- is $\frac{\partial F}{\partial s_{k^-}}(s) = \frac{1}{2}$. Hence, statement **(S2')** is perfectly satisfied by F.

Theorem 1 can be interpreted in the following way. This result states that when x_k is close to the neutral level $\mathbf{0}_k$ (relatively to criteria k^+ and k^-), the relative preference of the actor over criteria k^+ and k^- is not so clear. This is a hesitation area.

Let us show as an example that bi-capacities satisfy to the restriction imposed by the previous corollary. By (5), when the scores w.r.t. criteria are all different in absolute value, the weight of criterion l for an act $s \in \Omega$ for a bi-capacity μ is given by

$$w_l(s) = \begin{cases} \mu\left(\{l\} \cup C_l^+(s), C_l^-(s)\right) - \mu\left(C_l^+(s), C_l^-(s)\right) & \\ \qquad\qquad \text{if } s_l \geq 0 & \\ \mu\left(C_l^+(s), C_l^-(s)\right) - \mu\left(C_l^+(s), \{l\} \cup C_l^-(s)\right) & \\ \qquad\qquad \text{if } s_l < 0 & \end{cases}$$

where

$$C_l^+(s) = \{m \neq l\ ,\ s_m \geq 0 \text{ and } |s_m| \geq |s_l|\}$$

and

$$C_l^-(s) = \{m \neq l\ ,\ s_m < 0 \text{ and } |s_m| \geq |s_l|\}\ .$$

If $|s_i| < |s_l|$ then $i \notin C_l^+(s) \cup C_l^-(s)$. Hence, if criterion k is the one closest to the neutral level among criteria k, k^+, k^-, then $k \notin C_{k^+}^+(s) \cup C_{k^+}^-(s)$ and $k \notin C_{k^-}^+(s) \cup C_{k^-}^-(s)$. This means that $w_{k^+}(s)$ and $w_{k^-}(s)$ do not depend on whether $s_k \geq 0$. Henceforth, **(S2')** cannot be modeled in this case. Now when $|s_i| \geq |s_l|$, then $i \in C_l^+(s)$ if $s_i \geq 0$ and $i \in C_l^-(s)$ if $s_i < 0$. Hence, the weight $w_l(s)$ can change between the two cases $s_i \geq 0$ and $s_i < 0$. As a consequence, one cannot model **(S2')** with a bi-capacity whenever criterion k is the one closest to the neutral level among criteria k, k^+, k^-.

We restrict **(S2')** according to Theorem 1:

> **(S3)**: If the value w.r.t. criterion k is attractive (> 0), and k is not the one closest to the neutral level among criteria k, k^+, k^-, then criterion k^+ is more important than criterion k^-. If the value w.r.t. criterion k is repulsive (< 0), and k is not the one closest to the neutral level among criteria k, k^+, k^-, then criterion k^+ is less important than criterion k^-.

Let us give the requirements on a bi-capacity imposed by the previous statement **(S3)**. One has the following lemma.

Lemma 1 *There exists a bi-capacity such that the corresponding Choquet integral satisfies to* **(S3)**.

We have shown in this section that the general statement **(S2)** cannot be satisfied by a continuous piecewise linear aggregation function. We introduce then

a restriction of **(S2)** - namely **(S3)**. Finally we have seen that **(S3)** can be thoroughly fulfilled at least by the Choquet integral w.r.t. some bi-capacity. It turns out that there is no capacity such that its associated Choquet integral satisfies statement **(S2)**.

6 Representation of Statement (R1) by an aggregation function

Consider Rule **(R1)**. Rule **(R1)** becomes

$$w_{NF}(u_F, u_{NF}) = 0 \text{ if } u_F < 0 \quad (8)$$

$$w_{NF}(u_F, u_{NF}) > 0 \text{ if } u_F > 0 \quad (9)$$

The previous two relations cannot be satisfied when F is the Choquet integral w.r.t. a capacity. Indeed, if (8) is satisfied with a Choquet integral, then (8) is satisfied for all $u_F \in \mathbb{R}$ such that $u_F \leq u_{NF}$. A similar result is obtained with (9).

Theorem 2 *There does not exist any continuous and piecewise linear aggregation function F, for which statement* **(R1)** *is satisfied.*

Moreover, one can show that the Choquet integral w.r.t. a bi-capacity cannot do better than the Choquet integral w.r.t. a capacity. Indeed, if (8) is satisfied with a Choquet integral w.r.t. a bi-capacity, then (8) is satisfied for all $u_F \in \mathbb{R}$ such that $u_F \leq u_{NF}$.

References

[1] C. Boutilier, R.I. Brafman, C. Domshlak, H.H. Hoos and D. Poole. CP-nets: a tool for representing and reasoning with conditional *Ceteris Paribus* preference statements. J. of Artificial Intelligence Research 21 (2004) 135-191.

[2] D. Bouyssou, M. Pirlot. Preferences for multi-attributed alternatives: traces, dominance, and numerical representations. J. of Mathematical Psychology 48 (2004) 167-185.

[3] R. Brafman, C. Domshlak, S. Shimony. Introducing variable importance tradeoffs into CP-nets. AIPS'02 Workshop on Planning and Scheduling with Multiple Criteria.

[4] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1953) 131-295.

[5] B. De Finetti. Theory of Probability. Wiley, London, 1974.

[6] J.C. Fodor and M. Roubens. Fuzzy preferences modeling and multicriteria decision aid. Kluwer Academic Publisher, 1994.

[7] M. Grabisch, T. Murofushi and M. Sugeno, Fuzzy measures and integrals, Physica-Verlag, Heidelberg, New York, 2000.

[8] M. Grabisch, Ch. Labreuche. Bi-Capacities for decision making on bipolar scales. In: EUROFUSE Workshop on Information Systems, Varenna, Italy, September 2002.

[9] M. Grabisch, Ch. Labreuche. Fuzzy Measures and Integrals in MCDA. In: J. Figueira, S. Greco, M. Ehrgott. Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, New York, 2005. pp. 563-608.

[10] D. Kahneman, D. Tversky. Propect theory: an analysis of decision under risk. Econometrica 47 (1979) 263-291.

[11] R.L. Keeney, H. Raiffa, Decision with Multiple Objectives, Wiley, New York, 1976.

[12] D.H. Krantz, R.D. Luce, P. Suppes, A. Tversky, Foundations of measurement, vol 1: Additive and Polynomial Representations, Academic Press, San Diego, 1971.

[13] Ch. Labreuche, M. Grabisch. The Choquet integral for the aggregation of interval scales in multicriteria decision making, Fuzzy Sets and Systems 137 (2003) 11-26.

[14] C. Labreuche, M. Grabisch. Generalized Choquet-like aggregation functions for handling bipolar scales. European J. of Operational Research 172 (2006) 931-955.

[15] J. Pennings, A. Smidts. Assessing the construct validity of risk attitude. Management Science 46 (2000) 1337-1348 .

[16] Nicolas Rescher, An introduction to value theory. Prentice-Hall edition, New York, 1969.

[17] H. Simon. Rational choice and the structure of the environment. Psychological Review 63(2) (1956) 129-138.

[18] P. Slovic, M. Finucane, E. Peters, D.G. MacGregor, The affect heuristic. in: T. Gilovitch, D. Griffin, D. Kahneman (Eds.), Heuristics and biases: the psychology of intuitive judgment, Cambridge University Press, 2002, pp. 397–420.

Various representations and algebraic structure of linear imprecision indices

Alexander E. Lepskiy
Taganrog State University
RUSSIA
lepskiy@mail.ru

Andrey G. Bronevich
Taganrog State University
RUSSIA
brone@mail.ru

Abstract

The paper is devoted to the investigation of imprecision indices, introduced in [7]. They are used for evaluation of uncertainty (or more exactly imprecision), which is contained in information given by fuzzy (non-additive) measures, in particular, by lower or upper probabilities. We argue that there exist various types of uncertainty, for example, randomness, investigated in probability theory, imprecision, described by interval calculi, inconsistency, incompleteness, fuzziness and so on. In general these types of uncertainty have very complex behavior, caused by their interaction. Therefore, the choice of uncertainty measures is not unique, and is defined by the problems addressed. The classical uncertainty measures are Shannon's entropy and Hartley's measure. In the paper imprecision indices and their linear representatives are introduced axiomatically. The system of axioms enables to define various imprecision indices. So we investigate the algebraic structure of all imprecision indices and describe their families with best properties.

Keywords: Imprecision indices lower and upper probabilities, uncertainty based information.

1 Introduction

Measuring of uncertainty plays a major role in uncertainty theories, in particular, probability theory, information theory, fuzzy sets theory and so on. There are some ways of defining such measures in the theory of evidence, in the theory of fuzzy (non-additive) measures and in the theory of imprecise probabilities. However, one can see that in such general theories an uncertainty measure with the best properties has no yet been found. This situation is explained by very complex interaction among various types of uncertainty, including randomness, inconsistence, imprecision, incompleteness of analyzed information. We recall classical uncertainty measures, used in information theory and probability theory. Let X be a finite set of alternatives. Assigning to each alternative $x \in X$ some probability $P(\{x\})$, we have information, which is described by probability measure P, and in this case Shannon's entropy $S(P) = -\sum_{x \in X} P(\{x\}) \log_2 P(\{x\})$ can be used. Let we know only that the "true" alternative is in a non-empty set $B \subseteq X$. This situation can be described by the set function $\eta_{\langle B \rangle}(A) = \begin{cases} 1, B \subseteq A \\ 0, B \not\subseteq A \end{cases}$, $A \subseteq X$, which gives lower probability of an event A, and Hartley's measure $H(\eta_{\langle B \rangle}) = \log_2 |B|$ can be justified. It is easily seen that in the first case uncertainty has a type that one call randomness, and the second case is more connected with imprecision of the information. The generalization of these two cases consists in the following. Consider a pair $(\underline{g}, \overline{g})$ of set functions $\underline{g} : 2^X \to [0,1]$, $\overline{g} : 2^X \to [0,1]$ defined on the power-set 2^X. We suggest that $\underline{g}(A) \le \overline{g}(A)$ for all $A \in 2^X$, $\underline{g}(\varnothing) = \overline{g}(\varnothing) = 0$, and there is a "true" probability measure P on 2^X with $\underline{g}(A) \le P(A) \le \overline{g}(A)$ for all $A \in 2^X$. In other words, set functions $\underline{g}, \overline{g}$ give us upper and lower bounds of probabilities, and for any event $A \in 2^X$ we have only the interval $[\underline{g}(A), \overline{g}(A)]$ of possible values of a "true" probability $P(A)$. In practical issues it is sufficient to define

the lower probability $\underline{g}$, the upper probability can be calculated by $\overline{g}(A) = 1 - \underline{g}(\overline{A})$, where $A \in 2^X$ and $\overline{A}$ is the complement of A. Due to works of Klir, Higashi, Harmanec and others (see [4,5,6]), there are two important uncertainty measures, which show the best properties in a sense of obeying axioms, which are similar to the axioms of Shannon's entropy. They are generalized Hartley's measure, and aggregate measure of uncertainty. Let $\underline{g}$ be a belief function, i.e. it can be represented by

$$\underline{g} = \sum_{B \in 2^X} m(B)\eta_{\langle B \rangle},$$

where $m(\varnothing) = 0$, $m(B) \geq 0$ for all $B \in 2^X$, and $\sum_{B \in 2^X} m(B) = 1$. Then generalized Hartley's measure is defined by

$$GH\left(\underline{g}\right) = \sum_{B \in 2^X \setminus \{\varnothing\}} m(B) \log_2 |B|.$$

The aggregate measure of uncertainty is calculated by

$$Au\left(\underline{g}\right) = \sup_{P \geq \underline{g}} S(P),$$

where sup is taken over all probability measures on 2^X, which are consistent with $\underline{g}$, i.e. $P(A) \geq \underline{g}(A)$ for all $A \in 2^X$. It is worth to mention that generalized Hartley's measure can be used for measuring imprecision and aggregate measure of uncertainty for total uncertainty. It is easy to check that aggregate measure of uncertainty coincides with Shannon's entropy for probability measures and with Hartley's measure for $\underline{g} = \eta_{\langle B \rangle}$, $B \neq \varnothing$. These uncertainty measures can be also used in the case, where $\underline{g}$ is a 2-monotone set function [2]. There is a possibility to extend our consideration to the case, where the set of probability measures, which are consistent with $\underline{g}$, is empty. Then we say that the information in our disposal is inconsistent and we should analyze three types of uncertainty: randomness, imprecision, and inconsistence.

The paper has the following structure. First we remind some definitions and results from the theory of non-additive measures and axiomatics of imprecision indices, formulated in [7]. Then we analyze so called linear imprecision indices on the set of upper and lower probabilities, giving their detailed description, and introducing their important families with symmetrical properties.

2 Basic definitions and problem statement

Let X be a finite set. In the sequel we will use the following notations:

1. $M(X)$ is the set of all real-valued set functions on the powerset 2^X;

2. $M_0(X) = \{g \in M(X) \mid g(\varnothing) = 0\}$;

3. We write $g_1 \leq g_2$ for $g_1, g_2 \in M(X)$ if $g_1(A) \leq g_2(A)$ for all $A \in 2^X$.

4. $M_{mon}(X) \subset M_0(X)$ is the set of all normalized monotone set functions on 2^X. It means that $g \in M_0(X)$ implies $g(\varnothing) = 0$, $g(X) = 1$, and $g(A) \leq g(B)$ if $A \subseteq B$.

5. $M_{\Pr}(X)$ is the set of all probability measures on 2^X;

6. $M_{low}(X) = \{g \in M_0 \mid \exists P \in M_{\Pr} : g \leq P\}$ is the set of all lower probabilities on 2^X.

6. $M_{up}(X) = \{g \in M_0 \mid \exists P \in M_{\Pr} : g \geq P\}$ is the set of all upper probabilities on 2^X.

7. Let $g \in M(X)$ then the dual of g is denoted by $\overline{g}$ and by definition $\overline{g}(A) = g(X) - g(\overline{A})$, $A \in 2^X$.

8. $M_{bel}(X)$ is the set of all belief functions on 2^X. Any $g \in M_{bel}(X)$ has the following unique representation: $g = \sum_{B \in 2^X} m(B)\eta_{\langle B \rangle}$, where $m(B) \geq 0$ for all $B \in 2^X$, $m(\varnothing) = 0$, and $\sum_{B \in 2^X} m(B) = 1$.

9. $M_{pl}(X)$ is the set of all plausibility functions on 2^X. Any $g \in M_{pl}(X)$ is uniquely represented by $g = \sum_{B \in 2^X} m(B)\overline{\eta}_{\langle B \rangle}$, where $m(B) \geq 0$ for all $B \in 2^X$, $m(\varnothing) = 0$, and $\sum_{B \in 2^X} m(B) = 1$.

We can consider the set $M(X)$ (or $M_0(X)$) as a linear space w.r.t. to usual sum of set functions and usual product of set functions and real numbers. In the non-additive measure theory, the basis consisting of functions $\eta_{\langle B \rangle}$, $B \in 2^X$, is of interest. Let $g \in M(X)$ and $g = \sum_{B \in 2^X} m_g(B)\eta_{\langle B \rangle}$ then the set function m_g is called the Möbius transform of g.

The function m_g is expressed by $m_g(B)=\sum_{A:A\subseteq B}(-1)^{|B\setminus A|}g(A)$. We will also use so-called dual Möbius transform of g. This transform is connected with the basis consisting of set functions $\eta^{\langle B\rangle}$, $B\in 2^X$, defined by $\eta^{\langle B\rangle}(A)=\eta_{\langle \overline{B}\rangle}(\overline{A})$. Let $g=\sum_{B\in 2^X} m^g(B)\eta^{\langle B\rangle}$ then the set function m^g is called the dual Möbius transform of g. It is calculated by $m^g(B)=\sum_{A:B\subseteq A}(-1)^{|A\setminus B|}g(A)$.

We remind now some definitions, introduced in [7].

Definition 1. *A functional* $f: M_{low}(X)\to[0,1]$ *is called imprecision index if the following conditions are fulfilled: 1)* $g\in M_{\Pr}(X)$ *implies* $f(g)=0$*; 2)* $f(g_1)\geq f(g_2)$ *for all* $g_1,g_2\in M_{low}(X)$ *such that* $g_1\leq g_2$*; 3)* $f\left(\eta_{\langle X\rangle}\right)=1$.Some important examples of imprecision indices are:

1) $v_B(g)=\overline{g}(B)-g(B)$ for a fixed $B\in 2^X/\{\varnothing,X\}$ (We call it a primitive imprecision index in the sequel.);

2) $v_p(g)=\left(2^{|X|}-2\right)^{-1/p}\left(\sum_{B\in 2^X}\left|\overline{g}(B)-g(B)\right|^p\right)^{1/p}$, $p\geq 1$;

3) $v_\infty(g)=\sup\left\{\left|\overline{g}(B)-g(B)\right|\,|\,B\in 2^X\right\}$.

It is clear that there are many ways for defining imprecision indices. One class of them consisting of linear imprecision indices is described in the following definition.

Definition 2. An imprecision index f on $M_{low}(X)$ is called *linear* if for any linear combination $\sum_{j=1}^k\alpha_j g_j\in M_{low}(X)$, $\alpha_j\in\mathbb{R}$, $g_j\in M_{low}(X)$, $j=1,\ldots,k$, we have $f\left(\sum_{j=1}^k\alpha_j g_j\right)=\sum_{j=1}^k\alpha_j f\left(g_j\right)$.

3. An investigation of linear imprecision indices

We notice first that any linear functional f on $M(X)$ is defined uniquely by its values on a chosen basis of $M(X)$. It enables to define f by a set function $\mu_f: 2^X\to\mathbb{R}$ with the following property $\mu_f(B)=f\left(\eta_{\langle B\rangle}\right)$, $B\in 2^X$. Since any $g\in M_{low}$ is represented as a linear combination of $\left\{\eta_{\langle B\rangle}\right\}_{B\in 2^X\setminus\{\varnothing\}}$, we take by definition that $\mu_f(\varnothing)=0$ (or $f\left(\eta_{\langle\varnothing\rangle}\right)=0$) for any linear imprecision index f.

Proposition 1. *Let* f *be a linear imprecision index on* M_{low} *then* $\mu_f\in M_{mon}(X)$ *with* $\mu_f\left(\{x\}\right)=0$ *for any* $x\in X$.

Proof. By definition $\mu_f(\varnothing)=0$ and $\mu_f(X)=1$; $\mu_f\left(\{x\}\right)=0$ for any $x\in X$ because $\eta_{\langle\{x\}\rangle}\in M_{\Pr}(X)$. Further we see that $\eta_{\langle B\rangle}\in M_{low}$ for $B\neq\varnothing$ and $\eta_{\langle B\rangle}\geq\eta_{\langle C\rangle}$ if $B\subseteq C$. It implies that $\mu_f(B)\leq\mu_f(C)$ for $B\subseteq C$, i.e. μ_f is monotone. This finishes the proof of the proposition.

The following proposition gives the expression of any linear functional through the values of the transformed set function.

Proposition 2. *Let* f *be a linear functional on* $M(X)$ *then* $f(g)=\sum_{B\in 2^X}m^{\mu_f}(B)g(B)$ *for any* $g\in M(X)$.

Proof. By definition $\mu_f=\sum_{B\in 2^X}m^{\mu_f}(B)\eta^{\langle B\rangle}$ and $g=\sum_{C\in 2^X}m_g(C)\eta_{\langle C\rangle}$, therefore,

$$
\begin{aligned}
f(g)&=\sum_{C\in 2^X}m_g(C)\mu_f(C)\\
&=\sum_{C\in 2^X}m_g(C)\sum_{B\in 2^X}m^{\mu_f}(B)\eta^{\langle B\rangle}(C)\\
&=\sum_{C\in 2^X}m_g(C)\sum_{B\in 2^X}m^{\mu_f}(B)\eta_{\langle C\rangle}(B)\\
&=\sum_{B\in 2^X}m^{\mu_f}(B)\sum_{C\in 2^X}m_g(C)\eta_{\langle C\rangle}(B)\\
&=\sum_{B\in 2^X}m^{\mu_f}(B)g(B)
\end{aligned}
$$

The following theorem gives necessary and sufficient conditions on a linear functional to be an imprecision index through the dual Möbius transform of μ_f.

Theorem 1. *Let* f *be a linear functional on* $M(X)$ *then it is an imprecision index on* $M_{low}(X)$ *iff*

a) $m^{\mu_f}(X)=1$; $\sum_{D\in 2^X}m^{\mu_f}(D)=0$;

b) $\sum_{D:\,x\in D}m^{\mu_f}(D)=0$ *for all* $x\in X$;

c) $m^{\mu_f}(D)\leq 0$ *for all* $D\in 2^X\setminus\{\varnothing,X\}$.

Proof. It is clear that the condition a) guarantees that $f\left(\eta_{\langle X\rangle}\right)=1$ and $f\left(\eta_{\langle\varnothing\rangle}\right)=0$. It is easy to show that b) is the necessary and sufficient condition that $f(g)=0$ for any $g\in M_{\Pr}(X)$. Indeed, since

$\eta_{\langle\{x\}\rangle} \in M_{\Pr}(X)$ then $\mu_f(\{x\}) = f\left(\eta_{\langle\{x\}\rangle}\right) = 0$ and $\mu_f(\{x\}) = \sum_{D:\, x\in D} m^{\mu_f}(D) = 0$. On the other hand, any $g \in M_{\Pr}(X)$ can be represented as a convex combination of $\eta_{\langle\{x\}\rangle}$, i.e. $g = \sum_{x\in X} m_g(\{x\})\eta_{\langle\{x\}\rangle}$, therefore,

$$f(g) = \sum_{x\in X} m_g(\{x\}) f\left(\eta_{\langle\{x\}\rangle}\right)$$
$$= \sum_{x\in X} m_g(\{x\})\mu_f(\{x\}) = 0.$$

So b) is proved. c) is the sufficient and necessary condition of antimonotonicity of f on $M_{low}(X)$. Let c) be fulfilled and $g_1 \le g_2$ for $g_1, g_2 \in M_{low}(X)$ then by Proposition 2

$$f(g_1) - f(g_2) = \sum_{B\in 2^X} m^{\mu_f}(B)\left(g_1(B) - g_2(B)\right)$$
$$= \sum_{B\in 2^X\setminus\{\varnothing, X\}} m^{\mu_f}(B)\left(g_1(B) - g_2(B)\right).$$

Since $g_1(B) - g_2(B) \le 0$ for any $B \in 2^X$ and $m^{\mu_f}(B) \le 0$ for any $B \in 2^X \setminus \{\varnothing, X\}$, we get $f(g_1) \ge f(g_2)$, i.e. c) implies antimonotonicity of f. Vice versa, let f be antimonotone on $M_{low}(X)$ then for any $D \in 2^X \setminus \{\varnothing, X\}$ we can always find such $g_1, g_2 \in M_{low}(X)$ with $g_1(B) = g_2(B)$ for all $B \ne D$, and $g_1(D) < g_2(D)$. According to Proposition 2 $0 \le f(g_1) - f(g_2) = m^{\mu_f}(D)(g_1(D) - g_2(D))$, i.e. $m^{\mu_f}(D) \le 0$. The theorem is proved.

Conditions of Theorem 1 can be transformed to the expression, which is very close to the condition "avoiding sure loss" from the theory of imprecise probabilities [10]. It enables to get the implicit expression for an arbitrary linear imprecision index. We will further use the functions 1_B, $B \subseteq X$, on X defined by $1_B(x) = 1$ if $x \in B$, and $1_B(x) = 0$ otherwise.

Theorem 2. *Any linear imprecision index f on M_{low} can be uniquely represented by*

$$f(g) = 1 - \sum_{B\in 2^X} q(B) g(B),$$

where the set function q obeys the following conditions:

1) $q(\varnothing) = 0$, $q(X) = 0$, $q(B) \ge 0$ for all $B \in 2^X$;

2) $\sum_{B\in 2^X} q(B) 1_B = 1_X$.

Proof. By Proposition 2 $f(g) = \sum_{B\in 2^X} m^{\mu_f}(B) g(B)$, since $m^{\mu_f}(X) = 1$, $g(X) = 1$, $g(\varnothing) = 0$, we get the required representation, choosing $q(B) = -m^{\mu_f}(B)$ for $B \in 2^X \setminus \{\varnothing, X\}$. The condition b) from Theorem 1 is reformulated for q as $\sum_{B:\, x\in B} q(B) = 1$ for all $x \in X$. We show that this condition is equivalent to 2). Actually, $\sum_{B\in 2^X} q(B) 1_B = 1_X$ iff $\sum_{B\in 2^X} q(B) 1_B(x) = 1$ for any $x \in X$, on the other hand, $\sum_{B\in 2^X} q(B) 1_B(x) = \sum_{B:\, x\in B} q(B)$. The theorem is proved.

Remark 1. The condition of "avoiding sure loss" from the theory of imprecise probabilities can be formulated with the help of the set function q from the Theorem 2 as follows: let $g \in M_0(X)$ then $g \in M_{low}(X)$ iff for any set function q obeying 1), 2) from Theorem 2, we have $\sum_{B\in 2^X} q(B) g(B) \le 1$.

Theorem 3. *Let f be a linear functional on $M(X)$ then it is an imprecision index on $M_{low}(X)$ iff $\mu_f = a\mu - b\overline{\eta}_{\langle X\rangle}$, where $b > 0$, $a = 1 + b$, and $\mu \in M_{pl}(X)$ with $\mu(\{x\}) = b/a$ for all $x \in X$.*

Proof. *Necessity.* Let f be a linear imprecision index on $M_{low}(X)$ then

$$\mu_f(B) = \sum_{A\in 2^X} m^{\mu_f}(A)\eta_{\langle\overline{A}\rangle}(\overline{B})$$
$$= \sum_{A\in 2^X\setminus\{X,\varnothing\}} m^{\mu_f}(A)\eta_{\langle\overline{A}\rangle}(\overline{B})$$
$$+ m^{\mu_f}(X)\eta_{\langle\varnothing\rangle}(\overline{B}) + m^{\mu_f}(\varnothing)\eta_{\langle X\rangle}(\overline{B}),$$

where $m^{\mu_f}(A) \le 0$ for any $A \in 2^X \setminus \{X, \varnothing\}$ and $\eta_{\langle\varnothing\rangle} \equiv 1$, $m^{\mu_f}(X) = 1$. Let $a = -\sum_{A\in 2^X\setminus\{X,\varnothing\}} m^{\mu_f}(A)$ then taking $q(A) = -\frac{1}{a} m^{\mu_f}(\overline{A})$ for $A \in 2^X \setminus \{X, \varnothing\}$ and $q(A) = 0$ for $A \in \{X, \varnothing\}$, we get

$$\mu_f(B) = -a\sum_{A\in 2^X} q(\overline{A})\eta_{\langle\overline{A}\rangle}(\overline{B}) + 1 + m^{\mu_f}(\varnothing)\eta_{\langle X\rangle}(\overline{B})$$
$$= a\sum_{A\in 2^X} q(A)\left(1 - \eta_{\langle A\rangle}(\overline{B})\right)$$
$$- m^{\mu_f}(\varnothing)\left(1 - \eta_{\langle X\rangle}(\overline{B})\right) + m^{\mu_f}(\varnothing) + 1 - a.$$

It is clear $m^{\mu_f}(\varnothing) + 1 - a = \sum_{A\in 2^X} m^{\mu_f}(A) = \mu_f(\varnothing) = 0$, hence, we get the representation required

$$\mu_f(B) = a\sum_{A\in 2^X} q(A)\overline{\eta}_{\langle A\rangle}(B) - b\overline{\eta}_{\langle X\rangle}(B),$$

where $\mu = \sum_{A\in 2^X} q(A)\overline{\eta}_{\langle A\rangle}$, $b = m^{\mu_f}(\varnothing)$, $a = 1 + b$.

It is easy to show that $\mu(\{x\}) = b/a$, $x \in X$, and $b > 0$. Actually, by Proposition 1 $\mu_f(\{x\}) = 0$ for all $x \in X$, i.e. $\mu(\{x\}) = b/a$ for all $x \in X$. On the other hand,

$$\mu_f(\{x\}) = a\sum_{A:x\in A} q(A) - b = 0,$$

i.e. $b \geq 0$ and if $b = 0$ then $q \equiv 0$ and this contradicts to the definition of imprecision index.

Sufficiency. Assume that we have the representation of μ_f from the theorem. We prove sufficiency if we check all conditions from Theorem 1. We see that $\mu_f(\varnothing) = 0$, $\mu_f(X) = 1$, and $\mu_f(\{x\}) = 0$ for all $x \in X$, i.e. conditions a), b) are true. We will further prove that $m^{\mu_f}(A) \leq 0$ for all $A \in 2^X \setminus \{\varnothing, X\}$.

Since μ is a plausibility function, it is represented by $\mu = \sum_{A\in 2^X} m(A)\overline{\eta}_{\langle A\rangle}$, where $m(A) \geq 0$ for all $A \in 2^X$, $m(\varnothing) = 0$, and $\sum_{A\in 2^X} m(A) = 1$. We can write

$$\mu_f(B) = a\sum_{A\in 2^X} m(A)\overline{\eta}_{\langle A\rangle}(B) - b\overline{\eta}_{\langle X\rangle}(B)$$

$$= a\sum_{A\in 2^X} m(A)\left(1 - \eta_{\langle A\rangle}(\overline{B})\right) - b\left(1 - \eta_{\langle X\rangle}(\overline{B})\right)$$

$$= a\sum_{A\in 2^X} m(\overline{A})\left(1 - \eta_{\langle \overline{A}\rangle}(\overline{B})\right) - b\left(1 - \eta_{\langle X\rangle}(\overline{B})\right).$$

The last expression implies $m^{\mu_f}(A) = -am(\overline{A}) \leq 0$ for all $A \in 2^X \setminus \{\varnothing, X\}$, i.e. c) is also true. The theorem is proved.

From the proof of Theorem 3, we see that we can use the basis $\left\{\overline{\eta}_{\langle B\rangle}\right\}_{B\in 2^X\setminus\{\varnothing\}}$ of $M_0(X)$ for defining other sufficient and necessary conditions on linear imprecision index. We formulate them in

Corollary 1. *Let f be a linear functional on $M(X)$ and $\mu_f = \sum_{A\in 2^X\setminus\{\varnothing\}} m(A)\overline{\eta}_{\langle A\rangle}$ then f is an imprecision index iff 1) $\mu_f \in M_0(X)$; 2) $\mu_f(\{x\}) = 0$ for all $x \in X$; 3) $m(A) \geq 0$ for all $A \in 2^X \setminus \{\varnothing, X\}$.*

Theorem 4. *Let f be a linear functional on $M(X)$ then it is an imprecision index on $M_{low}(X)$ iff 1) $\mu_f \in M_0(X)$; 2) $\mu_f(\{x\}) = 0$ for all $x \in X$; 3) the set function $\mu_f^{\{x\}}$, defined by $\mu_f^{\{x\}}(B) = \mu_f(B \cup \{x\})$, $B \in 2^X$, is in $M_{pl}(X)$ for any $x \in X$.*

Proof. *Necessity.* Let f be a linear imprecision index on $M_0(X)$. It is sufficient to show that 3) is true. By Theorem 3 we can use the following representation

$$\mu_f = a\sum_{A\in 2^X} m(A)\overline{\eta}_{\langle A\rangle}(B) - b\overline{\eta}_{\langle X\rangle},$$

where $b > 0$, $a = 1 + b$, $m(A) \geq 0$ for all $A \in 2^X$, $m(\varnothing) = 0$, $m(X) = 0$, and $\sum_{A\in 2^X} m(A) = 1$. Then

$$\mu_f(B \cup \{x\}) = a\sum_{A:x\in A} m(A) - b$$

$$+ a\sum_{A:x\notin A} m(A)\overline{\eta}_{\langle A\rangle}(B \cup \{x\}).$$

Since $a\sum_{A:x\in A} m(A) - b = \mu_f(\{x\}) = 0$, $\overline{\eta}_{\langle A\cup\{x\}\rangle}(B) = \overline{\eta}_{\langle A\rangle}(B \cup \{x\})$, we get $\mu_f^{\{x\}} = a\sum_{A:x\notin A} m(A)\overline{\eta}_{\langle A\cup\{x\}\rangle}$, i.e. $\mu_f^{\{x\}}$ is a plausibility function for any $x \in X$.

Sufficiency. Assume that the all conditions of the theorem are fulfilled, however, f is not linear imprecision index on $M_{low}(X)$. In this case at least one of inequalities $m(A) \geq 0$, $A \in 2^X \setminus \{\varnothing, X\}$, from Corollary 1 is not true, and there is a $B \in 2^X \setminus \{\varnothing, X\}$ such that $m(B) < 0$. Let $x \in X \setminus B$. In the similar way as in the proof of Theorem 3 we get

$$\mu_f^{\{x\}} = \sum_{A:x\notin A} m(A)\overline{\eta}_{\langle A\cup\{x\}\rangle},$$

Since in the last sum $m(A) < 0$ for $A = B$, we conclude that $\mu_f^{\{x\}}$ is not a plausibility function. This implies that our assumption is wrong, and f is a linear imprecision index. The theorem is proved in the whole.

It seems to be logical in some problems that the quantity of imprecision in the situation, where we only know that the true alternative belongs to the set B, depends on $|B|$ and does not depend on other factors. In this case we assume that $f(\eta_{\langle B\rangle}) = f(\eta_{\langle C\rangle})$ or $\mu_f(B) = \mu_f(C)$ if $|B| = |C|$, and we call such linear imprecision indices symmetrical. In the sequel we will use the fact that such symmetrical monotone set functions can be viewed as distorted probabilities [9]. Let P be a probability measure on $X = \{x_1, ..., x_N\}$; let $\lambda : [0,1] \to [0,1]$ be non-decreasing function with $\lambda(0) = 0$, $\lambda(1) = 1$, then the set function $g = \lambda \circ P$ ($g(A) = \lambda(P(A))$, $A \in 2^X$) is called distorted probability. We are interested in the case, where $P(\{x_i\}) = 1/N$, $i = 1, ..., N$. Further we will use the following sufficient condition of total monotonicity [1]: let $g = \lambda \circ P$ then it is a belief function if λ is infinitely differentiable on $[0,1)$ and $d^n\lambda(t)/dt^n \geq 0$, $n = 1, 2, ...$, for any $t \in [0,1)$.

Theorem 5. *Let f be a linear functional on $M(X)$ and $\mu_f = \lambda \circ P$, i.e. μ_f is a distorted probability,*

mentioned above, and $P(\{x_i\}) = 1/N$, $i = 1,...,N$. *Then* f *is an imprecision index if*

1) $\lambda(1/N) = 0$;

2) λ *is infinitely differentiable on* $\left[\frac{1}{N},1\right)$ *and* $(-1)^{n-1} d^n\lambda(t)/dt^n \geq 0$, $n = 1,2,...$, *for any* $t \in \left[\frac{1}{N},1\right)$.

Proof. We will check that the all conditions from Theorem 4 are true. It is clear that $\mu_f \in M_0(X)$ and $\mu_f(\{x\}) = 0$ for all $x \in X$. Now we prove that 3) is also true. In this case $\mu_f^{\{x\}}(B) = \lambda(P(B \cup \{x\}))$, $B \in 2^{X\setminus\{x\}}$, $\mu_f^{\{x\}}$ can be considered as a distorted probability on $2^{X\setminus\{x\}}$, and $\mu_f^{\{x\}} = \lambda_1 \circ P_1$, where $\lambda_1(t) = \lambda\left(\frac{1+t(N-1)}{N}\right)$, $t \in [0,1]$, and $P_1(\{y\}) = 1/(N-1)$, $y \in X \setminus \{x\}$. We find that $\overline{\mu_f^{\{x\}}}(A) = 1 - \lambda_1(P_1(\overline{A})) = 1 - \lambda_1(1 - P_1(A))$, i.e. $\overline{\mu_f^{\{x\}}} = \lambda_2 \circ P_1$ is a distorted probability and $\lambda_2(t) = 1 - \lambda_1(1-t) = 1 - \lambda\left(1 - \frac{t(N-1)}{N}\right)$. It is clear $\mu_f^{\{x\}} \in M_{pl}(X)$ iff $\overline{\mu_f^{\{x\}}} \in M_{bel}(X)$. Then we can argue that $\mu_f^{\{x\}}$ is a plausibility function if $d^n\lambda_2(t)/dt^n \geq 0$, $n = 1,2,...$, for any $t \in [0,1)$, or $(-1)^{n-1} d^n\lambda(t)/dt^n \geq 0$, $n = 1,2,...$, for any $t \in \left[\frac{1}{N},1\right)$.

In some cases it is suitable to define symmetrical μ_f by a non-decreasing function $\varphi : [1,+\infty) \to [0,+\infty)$ with $\varphi(1) = 0$ assuming that $\mu_f(A) = \varphi(|A|)/\varphi(|X|)$ for $A \neq \varnothing$. Then $\lambda(t) = \varphi(tN)/\varphi(N)$ for $t \in \left[\frac{1}{N},1\right]$, where $N = |X|$. It is easy to see that according to Theorem 5, μ_f determines a linear imprecision index if φ is is infinitely differentiable on $[1,N)$ and $(-1)^{n-1} d^n\varphi(t)/dt^n \geq 0$, $n = 1,2,...$, for any $t \in [1,N)$.

Example 1. Let $\varphi(t) = \ln(t)$ then $\mu_f(A) = \ln(|A|)/\ln(|X|)$. In this case the corresponding linear imprecision index can be considered as the analog of generalized Hartley's measure. We see that $(-1)^{n-1} d^n \ln(t)/dt^n = t^{-n} \geq 0$ for $t \geq 1$, i.e. μ_f determines a linear imprecision index on $M_{low}(X)$.

4. The algebraic structure of the set of all linear imprecision indices

Let f_1, f_2 be linear functionals on $M(X)$ then their linear combination $f = af_1 + bf_2$, $a,b \in \mathbb{R}$ is also a linear functional. If we take into consideration set functions $\mu_{f_1}, \mu_{f_2}, \mu_f$, we see that $\mu_f = a\mu_{f_1} + b\mu_{f_2}$, i.e. the set of all linear functionals on $M(X)$ is a linear space and this space is isomorphic to the linear space $M(X)$ of all set functions on 2^X. It is easy to show that if f_1, f_2 are linear imprecision indices then their convex sum $f = af_1 + bf_2$, where $a,b \geq 0$, $a + b = 1$, is also linear imprecision index, i.e. the set of all linear imprecision indices is a convex set. We denote by $M_I(X)$ the set of all set functions μ_f, which correspond to linear imprecision indices on $M_{low}(X)$. One can say that we understand the algebraic structure of a convex set if we have description of its extreme points. The following theorem gives the necessary and sufficient condition on an arbitrary $\mu \in M_I$ to be an extreme point.

Theorem 6. *Let* $\mu \in M_I(X)$, $\mu = \sum_{A\in\mathfrak{B}} m(A)\overline{\eta}_{\langle A\rangle} - b\overline{\eta}_{\langle X\rangle}$, *where* $\mathfrak{B} \subseteq 2^X \setminus \{\varnothing, X\}$, $m(A) > 0$ *for all* $A \in \mathfrak{B}$, $b > 0$, *then* μ *is an extreme point of* $M_I(X)$ *iff functions* $\{1_A\}_{A\in\mathfrak{B}}$ *are linearly independent.*

Proof. Notice first that any $\mu \in M_I(X)$ has the representation $\mu = \sum_{A\in\mathfrak{B}} m(A)\overline{\eta}_{\langle A\rangle} - b\overline{\eta}_{\langle X\rangle}$ by Corollary 1, $b > 0$, and $\mathfrak{B}$ is not empty. Secondly, $\mu(\{x\}) = 0$ for all $x \in X$, i.e.

$$\sum_{A\in\mathfrak{B}} m(A)1_A = b1_X .$$

We will show that μ is not an extreme point of $M_I(X)$ iff functions $\{1_A\}_{A\in\mathfrak{B}}$ are linearly dependent. This implies evidently the theorem statement. Assume that functions $\{1_A\}_{A\in\mathfrak{B}}$ are linearly dependent. Then there exist two different solutions of $\sum_{A\in\mathfrak{B}} \alpha_A 1_A = 1_X$ w.r.t. α_A, $A \in \mathfrak{B}$. We choose one of them as $\alpha_A^{(1)} = m(A)/b$, $A \in \mathfrak{B}$. Since $\alpha_A^{(1)} > 0$ for all $A \in \mathfrak{B}$, we can choose another solution $\alpha_A^{(2)}$ with $\alpha_A^{(2)} \geq 0$, $A \in \mathfrak{B}$. Let $b_2 = 1\Big/\left(\left(\sum_{A\in\mathfrak{B}} \alpha_A^{(2)}\right) - 1\right)$, then it is easy to see that $b_2 > 0$ and the set function μ_2, defined by

$$\mu_2 = \sum_{A\in\mathfrak{B}} b_2\alpha_A^{(2)}\overline{\eta}_{\langle A\rangle} - b_2\overline{\eta}_{\langle X\rangle},$$

is in M_I. Defining

$$c = \sup\left\{r \in \mathbb{R} \mid rb_2\alpha_A^{(2)} \le m(A), A \in \mathfrak{B}, rb_2 \le b\right\},$$

we confirm that $c \in (0,1)$, $\mu \ge c\mu_2$. Then

$$\mu_2 = \frac{1}{1-c}\left(\mu - c\mu_1\right) = \sum_{A\in\mathfrak{B}} m_1(A)\overline{\eta}_{\langle A\rangle} - b_1\overline{\eta}_{\langle X\rangle}.$$

where $m_1(A) = \frac{1}{1-c}\left(m(A) - cb_2\alpha_A^{(2)}\right)$, $b_1 = \frac{1}{1-c}\left(b - cb_2\right)$, is in $M_I(X)$. We see that $\mu \ge (1-c)\mu_1 + c\mu_2$, i.e. we have proved that μ is not an extreme point of $M_I(X)$.

Vice versa assume that μ is not an extreme point of $M_I(X)$. Then there exist set functions $\mu_1, \mu_2 \in M_I(X)$ such that $\mu = a\mu_1 + b\mu_2$, where $a, b > 0$ and $a + b = 1$. Since $\mu_1, \mu_2 \in M_I(X)$ we have $\sum_{A\in\mathfrak{B}} m_i(A)1_A = b_i 1_X$, where $b_i > 0$, $i = 1,2$. Therefore, the equation $\sum_{A\in\mathfrak{B}} \alpha_A 1_A = 1_X$ has more than one solution w.r.t. $\alpha_A \in \mathbb{R}$, $A \in \mathfrak{B}$, hence, functions $\{1_A\}_{A\in\mathfrak{B}}$ are linearly dependent if μ is not an extreme point of M_I. The theorem is proved.

Theorem 6 implies that the set $M_I(X)$ has the finite number of extreme points. According to the Theorem by Krein-Milman [8], any $\mu \in M_I(X)$ can be represented as a convex sum of extreme points. However, it is a very hard problem to describe such extreme points explicitly. Further we consider one convex subset of $M_I(X)$, for which this problem can be solved.

Definition 3. Let f be a linear imprecision index on $M_{low}(X)$, then we call it *complementarily symmetrical* if $m^{\mu_f}(A) = m^{\mu_f}(\overline{A})$ for all $A \in 2^X \setminus \{\varnothing, X\}$.

Important examples of complementarily symmetrical linear imprecision indices are primitive imprecision indices. We see that

$$\nu_B(g) = g(X) - g(B) - g(\overline{B}) + g(\varnothing),$$

$$\begin{aligned}\mu_{\nu_B}(A) &= \eta_{\langle A\rangle}(X) - \eta_{\langle A\rangle}(B) - \eta_{\langle A\rangle}(\overline{B}) + \eta_{\langle A\rangle}(\varnothing)\\ &= \eta^{\langle\varnothing\rangle}(A) - \eta^{\langle\overline{B}\rangle}(A) - \eta^{\langle B\rangle}(A) + \eta^{\langle X\rangle}(A).\end{aligned}$$

Therefore, $m^{\mu_{\nu_B}}(A) = 1$ if $A \in \{\varnothing, X\}$, $m^{\mu_{\nu_B}}(A) = -1$ if $A \in \{B, \overline{B}\}$, and $m^{\mu_{\nu_B}}(A) = 0$ otherwise. We can also express μ_{ν_B} through plausibility functions. In this case

$$\begin{aligned}\mu_{\nu_B}(A) &= 1 - \eta_{\langle\overline{B}\rangle}(\overline{A}) - \eta_{\langle B\rangle}(\overline{A}) + \eta_{\langle X\rangle}(\overline{A})\\ &= \overline{\eta}_{\langle B\rangle}(A) + \overline{\eta}_{\langle\overline{B}\rangle}(A) - \overline{\eta}_{\langle X\rangle}(A).\end{aligned}$$

By Theorem 6 it is easy to show that primitive indices ν_B, $B \in 2^X \setminus \{\varnothing, X\}$, are extreme points of $M_I(X)$. Actually, it follows from the fact that functions $\{1_B, 1_{\overline{B}}\}$ are linearly independent.

The role of primitive indices for describing the set of all complementarily symmetrical linear indices shows the following theorem.

Theorem 7. *The set of all complementarily symmetrical linear indices is convex. Any complementarily symmetrical linear index can be uniquely represented by a convex sum of primitive indices.*

Proof. Let f_1, f_2 be complementarily symmetrical imprecision indices, then

$$f_i(g) = \sum_{B\in 2^X} m^{f_i}(B)g(B),\ i = 1,2,\ g \in M_{low}(X),$$

and by Definition 3 $m^{\mu_{f_i}}(B) = m^{\mu_{f_i}}(\overline{B})$, $i = 1,2$, for all $B \in 2^X \setminus \{\varnothing, X\}$. Let $f = af_1 + cf_2$, where $a, c \ge 0$, and $a + c = 1$. Then it is easy to see that $m^{\mu_f} = am^{\mu_{f_1}} + cm^{\mu_{f_2}}$. This implies $m^{\mu_f}(B) = m^{\mu_f}(\overline{B})$ for all $B \in 2^X \setminus \{\varnothing, X\}$, i.e. the set of all complementarily symmetrical linear indices is convex.

Now we will prove that any complementarily symmetrical linear index can be represented by a convex sum of primitive indices. Let f be a complementarily symmetrical linear index and $g \in M_{low}(X)$ then

$$f(g) = \sum_{B\in 2^X} m^{\mu_f}(B)g(B),$$

where $m^{\mu_f}(B) = m^{\mu_f}(\overline{B})$ for all $B \in 2^X \setminus \{\varnothing, X\}$. Let $\mathfrak{D} = \left\{B \in 2^X \setminus \{X\} \mid x \in B\right\}$, $\overline{\mathfrak{D}} = \left\{B \in 2^X \mid \overline{B} \in \mathfrak{D}\right\}$ for some $x \in X$ then $\mathfrak{D} \cup \overline{\mathfrak{D}} = 2^X \setminus \{\varnothing, X\}$, $\mathfrak{D} \cap \overline{\mathfrak{D}} = \varnothing$.

$$\begin{aligned}f(g) &= m^{\mu_f}(X)g(X) + m^{\mu_f}(\varnothing)g(\varnothing)\\ &\quad + \sum_{B\in\mathfrak{D}} m^{\mu_f}(B)\left(g(B) + g(\overline{B})\right)\\ &= -\sum_{B\in\mathfrak{D}} m^{\mu_f}(B)\left(g(X) - g(B) - g(\overline{B}) + g(\varnothing)\right)\\ &\quad + \sum_{B\in\mathfrak{D}} m^{\mu_f}(B)\left(g(X) + g(\varnothing)\right)\\ &\quad + m^{\mu_f}(X)g(X) + m^{\mu_f}(\varnothing)g(\varnothing).\end{aligned}$$

We see $\sum_{B\in\mathfrak{D}} m^{\mu_f}(B) = \sum_{B:x\in B} m^{\mu_f}(B) - m^{\mu_f}(X) = -m^{\mu_f}(X) = -1$. The equality $\sum_{B\in 2^X} m^{\mu_f}(B) = 0$ implies that $m^{\mu_f}(\varnothing) = -\sum_{B\in\mathfrak{D}}\left(m^{\mu_f}(B) + m^{\mu_f}(\overline{B})\right) -$

$m^{\mu_f}(X) = 1$. Hence,

$$f(g) = \sum_{B \in \mathfrak{D}} (-1) m^{\mu_f}(B) \nu_B ,$$

where $(-1)m^{\mu_f}(B) \geq 0$ for all $B \in \mathfrak{D}$, and $\sum_{B \in \mathfrak{D}} (-1) m^{\mu_f}(B) = 1$, i.e. f can be represented by a convex sum of primitive indices.

We prove that the found representation is unique if we show that system $\{\nu_B\}_{B \in \mathfrak{D}}$ of all primitive indices is linearly independent in $M(X)$, or we show the same property for set functions $\{\mu_{\nu_B}\}_{B \in \mathfrak{D}}$. It is easy to see that set functions $\mu_{\nu_B} = \overline{\eta}_{\langle B \rangle} + \overline{\eta}_{\langle \overline{B} \rangle} - \overline{\eta}_{\langle X \rangle}$, $B \in \mathfrak{D}$, are linearly independent, this follows immediately from the fact that set functions $\{\overline{\eta}_{\langle B \rangle}\}_{B \in 2^X \setminus \{\varnothing\}}$ are also linear independent in $M(X)$. The theorem is proved in the whole.

Example 2. Let $\xi : X \to \mathbb{R}$, $\max_{x \in X} \xi(x) - \min_{x \in X} \xi(x) = 1$. Then we can define the linear imprecision index with the help of Choquet integral [3] $f(g) = \int_X \xi d\overline{g} - \int_X \xi dg$, where $g \in M_{low}$. Then $\mu_f(B) = \max_{x \in B} \xi(x) - \min_{x \in B} \xi(x)$ for $B \neq \varnothing$. It is easy to show that such defined index f is complementarily symmetrical. It is worth to mention that in the theory of imprecise probabilities $\int_X \xi d\overline{g}$ can be viewed as an upper estimation of the expectation $E[\xi]$, and $\int_X \xi dg$ as a lower estimation of the expectation $E[\xi]$.

Conclusion

Although, measuring uncertainty plays a central role in various uncertainty theories, there is no possibility to find one true uncertainty measure. This can be explained by the fact that there are many various types of uncertainty, they have different interpretations; it is very difficult to understand their mutual interaction. One way for overcoming this problem is to find families of suitable uncertainty measures, satisfying some justified properties. The choice of the best uncertainty measure depends considerably on the problem solved. In this paper we have proposed how imprecision can be measured if uncertain information is described by lower probabilities. We have treated the case, where uncertainty consists of some randomness and imprecision. The introduced axiomatics enables us to give detailed description of linear imprecision indices, and investigate some of them with symmetrical properties.

Acknowledgments

The authors express their sincere thanks to the anonymous reviewers, whose excellent work allows to increase the paper's quality. Dr. Bronevich on his behalf expresses his gratitude to the Fulbright Program, Binghamton University, and Prof. George Klir for research opportunity provided.

References

[1] A. G. Bronevich, On the closure of families of fuzzy measures under eventwise aggregations, Fuzzy sets and systems 153 (2005) 45-70.

[2] A. Chateauneuf, J. Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Mathematical Social Sciences 17 (1989) 263-283.

[3] D. Denneberg, Non-additive measure and integral, basic concepts and their role for applications, in M. Grabisch, T. Murofushi, M. Sugeno (Eds.), *Fuzzy Measures and Integrals - Theory and Applications.* Physical-Verlag, Heidelberg (Studies in Fuzzyness and Soft Computing, Vol. 40), 2000.

[4] D. Harmanec, G. J. Klir, Measuring total uncertainty in Dempster-Shafer theory: A novel approach, International Journal of General Systems 22(4) (1994) 405-419.

[5] M. Higashi, G. J. Klir, Measures of uncertainty and information based on possibility distributions, International Journal of General Systems 9(1) (1983) 43-58.

[6] G. J. Klir, Uncertainty and information: foundations of generalized information theory, John Wiley & Sons, Inc., 2006.

[7] A. E. Lepskiy, The linear imprecision indices on the lower probabilities, Proc. of the 11th IPMU Conference (Paris, July 2-7, 2006) 1724-1731.

[8] R. R. Phelps, Lectures on Choquet's theorem, Springer-Verlag, Heidelberg, 2000.

[9] D. Schmeidler, Subjective probability and expected utility without additivity. Econometrica 57 (1989).

[10] P. Walley, Statistical reasoning with imprecise probabilities, Chapman & Hall, London, 1991.

On the moments and the distribution of the Choquet integral

Ivan Kojadinovic
LINA CNRS FRE 2729
Site école polytechnique de l'univ. de Nantes
Rue Christian Pauc
44306 Nantes, France
ivan.kojadinovic@univ-nantes.fr

Jean-Luc Marichal
Applied Mathematics Unit
Univ. of Luxembourg
162A, avenue de la Faïencerie
L-1511 Luxembourg, G.D. Luxembourg
jean-luc.marichal@uni.lu

Abstract

We investigate the distribution functions and the moments of the so-called *Choquet integral*, also known as the *Lovász extension*, when regarded as a real function of a random sample drawn from a continuous population. Since the Choquet integral includes weighted arithmetic means, ordered weighted averaging operators, and lattice polynomials as particular cases, our results encompass the corresponding results for these aggregation operators. After recalling the results obtained by the authors in the uniform case, we present approaches that can be used in the non-uniform case to obtain moment approximations.

Keywords: Discrete Choquet integral; Lovász extension; Order statistic; Distribution function; Moment; B-Spline; Divided difference; Moment approximation; Asymptotic distribution.

1 Introduction

Aggregation operators are of central importance in many fields such as statistics or decision theory. Among such commonly used operators, the most frequently employed is probably the weighted arithmetic mean because of its simplicity and its very intuitive interpretation.

Although very attractive in many fields, the weighted arithmetic mean is not suited for situations where the values to be aggregated display some interaction. Let us choose the framework of multi-criteria decision aid to elaborate this in more detail. We consider a set $N := \{1, \ldots, n\}$ of *criteria* and a set $\mathcal{A}$ of *alternatives* evaluated according to these criteria. As classically done, we assume that with each alternative $a \in \mathcal{A}$ a vector $(a_1, \ldots, a_n) \in \mathbb{R}^n$ is associated, where, for any $i \in N$, a_i represents the *partial score* of a related to criterion i. The partial scores are further assumed to be defined on the same interval scale.

From the vector of scores of any alternative, one can compute an overall evaluation by means of an aggregation operator. Once the overall evaluations are computed, they can be used to rank the alternatives. In such a context, it is very frequent in applications to have criteria that are *substitutive* or *complementary*. Substitutivity between two criteria arises when an alternative can be assigned a high overall score when only one of the two criteria has a high partial evaluation. Complementarity means that it is necessary that the two criteria have simultaneously a high evaluation for the alternative to receive an overall high score. A natural extension of the weighted arithmetic mean that is able to deal with such situations (and many others) is the so-called *Choquet integral* w.r.t. a *capacity* [1, 2, 3].

Also called *Lovász extension* [4] in the context of the extension of pseudo-Boolean functions, the Choquet integral includes weighted arithmetic means, ordered weighted averaging operators [5], and lattice polynomials as particular cases [6, 7].

In this paper, we investigate the distribution and the moments of the *Choquet integral* when considered as a real function of a random sample drawn from a continuous population. In the uniform case, we recall the results obtained by the authors in [7] and we provide algorithms for computing the probability density function (p.d.f.) and the cumulative distribution function (c.d.f.) of the Choquet integral. In the non-uniform case, we present approaches that can be used to obtain approximations of the moments of this functional.

In order to avoid a cumbersome notation, cardinality of subsets $S, T, \ldots$ will be denoted whenever possible by the corresponding lower case letters $s, t, \ldots$, otherwise by the standard notation $|S|, |T|, \ldots$. Moreover,

we will often omit braces for singletons, e.g., writing $\nu(i)$, $N \setminus i$ instead of $\nu(\{i\})$, $N \setminus \{i\}$. Finally, the set of permutations on N will be denoted by $\mathfrak{S}_n$.

2 The Choquet integral and its particular cases

A set function $\nu : 2^N \to [0,1]$ is a *capacity* [1] on $N := \{1,\dots,n\}$ if it is monotone with respect to (w.r.t.) inclusion and satisfies $\nu(\emptyset) = 0$ and $\nu(N) = 1$. In the context of aggregation by the Choquet integral, for any $T \subseteq N$, the coefficient $\nu(T)$ is to be interpreted as the weight of importance of the combination T of criteria, or better, its importance or power to make the decision alone (without the remaining criteria).

Definition 1. The Choquet integral of $x \in \mathbb{R}^n$ w.r.t. a capacity ν on N is defined by

$$C_\nu(x) := \sum_{i=1}^{n} p_i^{\nu,\sigma}\, x_{\sigma(i)}, \tag{1}$$

where σ is a permutation on N such that $x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}$, where

$$p_i^{\nu,\sigma} := \nu_i^\sigma - \nu_{i-1}^\sigma, \qquad \forall\, i \in N, \tag{2}$$

and where $\nu_i^\sigma := \nu\big(\{\sigma(1),\dots,\sigma(i)\}\big)$ for any $i = 0,\dots,n$. In particular, $\nu_0^\sigma := 0$.

The Choquet integral can therefore be regarded as a piecewise linear function that coincides with a weighted arithmetic mean on each n-dimensional region

$$R_\sigma := \{x \in \mathbb{R}^n \mid x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}\} \qquad (\sigma \in \mathfrak{S}_n), \tag{3}$$

whose union covers $\mathbb{R}^n$.

The Choquet integral satisfies very appealing properties for aggregation. For instance, it is continuous, non decreasing, comprised between min and max, stable under the same transformations of interval scales in the sense of the theory of measurement, and coincides with the weighted arithmetic mean whenever the capacity is additive. An axiomatic characterization is provided in [3].

We now present some subclasses of Choquet integrals. Any vector $\omega \in [0,1]^n$ such that $\sum_i \omega_i = 1$ will be called a *weight vector* as we continue.

2.1 The weighted arithmetic mean

Definition 2. For any weight vector $\omega \in [0,1]^n$, the weighted arithmetic mean operator WAM_ω associated to ω is defined by

$$\mathrm{WAM}_\omega(x) := \sum_{i=1}^{n} \omega_i\, x_i, \qquad \forall\, x \in \mathbb{R}^n.$$

We can easily see that WAM_ω is a Choquet integral C_ν with respect to the additive capacity defined by $\nu(T) := \sum_{i\in T} \omega_i$ for all $T \subseteq N$. Conversely, the weights associated to WAM_ω are defined by $\omega_i := \nu(i)$ for all $i \in N$.

The class of weighted arithmetic means WAM_ω includes two important special cases, namely:

- the arithmetic mean $\mathrm{AM}(x) := \frac{1}{n}\sum_{i=1}^n x_i$, when $\omega_i = 1/n$ for all $i \in N$. In this case, we have $\nu(T) := t/n$ for all $T \subseteq N$.
- the k-th projection $\mathrm{P}_k(x) := x_k$, when $\omega_k = 1$ for some $k \in N$. In this case, we have $\nu(T) := 1$ if $T \ni k$ and 0 otherwise.

2.2 The ordered weighted averaging operator

The concept of ordered weighted averaging operator was proposed in aggregation theory by Yager [5] and corresponds, in statistics, to that of linear combination of order statistics.

Definition 3. For any weight vector $\omega \in [0,1]^n$, the ordered weighted averaging operator OWA_ω associated to ω is defined by

$$\mathrm{OWA}_\omega(x) := \sum_{i=1}^{n} \omega_i\, x_{\sigma(i)}, \qquad \forall\, x \in \mathbb{R}^n,$$

where σ is a permutation on N such that $x_{\sigma(1)} \geqslant \cdots \geqslant x_{\sigma(n)}$.

It is easy to verify that an OWA operator is a Choquet integral w.r.t. a capacity that depends only on the cardinality of subsets. The capacity ν associated to OWA_ω is defined by

$$\nu(T) := \sum_{i=1}^{t} \omega_i, \qquad T \subseteq N,\ T \neq \emptyset.$$

Conversely, the weights associated to OWA_ω are defined by $\omega_t := \nu(T) - \nu(T \setminus i)$ for all $T \subseteq N$ and all $i \in T$.

The class of ordered weighted averaging operators OWA_ω includes some important special cases, namely:

- the arithmetic mean when $\omega_i = 1/n$ for all $i \in N$.
- the k-th order statistic when $\omega_{n-k+1} = 1$ for some $k \in N$. In this case, we have

$$\nu(T) := \begin{cases} 1 & \text{if } t \geqslant n-k+1, \\ 0 & \text{otherwise,} \end{cases}$$

- the min operator

$$\min(x) = \min_{i\in N} x_i,$$

when $\omega_n = 1$. In this case, we have $\nu(T) := 1$ if $T = N$ and 0 otherwise.

- the max operator

$$\max(x) = \max_{i \in N} x_i,$$

when $\omega_1 = 1$. In this case, we have $\nu(T) := 1$ for all $T \neq \emptyset$.

2.3 Partial minimum and maximum

Definition 4. For any non-empty subset $A \subseteq N$, the partial minimum operator $\min_A$ and the partial maximum operator $\max_A$, associated to A, are respectively defined by

$$\min_A(x) = \min_{i \in A} x_i,$$
$$\max_A(x) = \max_{i \in A} x_i.$$

For the operator $\min_A$ (resp. $\max_A$), for any $T \subseteq N$, we have

$$\nu(T) := \begin{cases} 1 & \text{if } T \supseteq A, \\ 0 & \text{otherwise.} \end{cases}$$
$$\left(\text{resp. } \nu(T) := \begin{cases} 1 & \text{if } T \cap A \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases} \right)$$

These operators are particular cases of *lattice polynomials* that also correspond to special classes of Choquet integrals; see [6, 7] for more details.

3 Distributional relationships with linear combination of order statistics

From Definition 1, it is clear that the Choquet integral is a linear combination of order statistics whose coefficients depend on the order of the arguments. We state hereafter immediate relationships between the moments (resp. the c.d.f.) of the Choquet integral and the moments (resp. the c.d.f.) of linear combination of order statistics.

Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with p.d.f. f and let $X_{1:n} \leqslant \cdots \leqslant X_{n:n}$ denote the corresponding order statistics. Further, let $Y_\nu := C_\nu(X_1, \ldots, X_n)$ and let h be any function. By definition of the expectation, we have

$$\begin{aligned} \mathbf{E}[h(Y_\nu)] &= \int_{\mathbb{R}^n} h(C_\nu(x_1, \ldots, x_n)) \prod_{i=1}^n f(x_i) \mathrm{d}x_i \\ &= \sum_{\sigma \in \mathfrak{S}_n} \int_{R_\sigma} h\left(\sum_{i=1}^n p_i^{\nu,\sigma} x_{\sigma(i)} \right) \prod_{i=1}^n f(x_i) \mathrm{d}x_i \end{aligned}$$

Using the well-known fact that the joint p.d.f. of $X_{n:n} \geqslant \cdots \geqslant X_{1:n}$ is

$$n! \prod_{i=1}^n f(x_i), \qquad x_n \geqslant \cdots \geqslant x_1,$$

we obtain

$$\begin{aligned} \mathbf{E}[h(Y_\nu)] &= \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbf{E}\left[h\left(\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n} \right) \right] \\ &= \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbf{E}[h(Y_\nu^\sigma)]. \end{aligned} \tag{4}$$

where $Y_\nu^\sigma := \sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n}$ are linear combinations of order statistics. Clearly, the special cases

$$h(x) = x^r, \ [x - \mathbf{E}(Y_\nu)]^r, \text{ and } e^{tx}$$

provide similar relationships, respectively, for raw moments, central moments, and moment-generating functions.

Now, consider the *minus* (resp. *plus*) *truncated power function* x_-^n (resp. x_+^n) defined to be x^n if $x < 0$ (resp. $x > 0$) and zero otherwise. Given $y \in \mathbb{R}$, taking $h(x) = (x - y)_-^0$ in (4) provides a relationship between the c.d.f. F_ν of Y_ν and those of the random variables $\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n}$. Indeed, we clearly have

$$F_\nu(y) := \mathbf{Pr}[Y_\nu \leqslant y] = \mathbf{E}[(Y_\nu - y)_-^0],$$

and, denoting by F_ν^σ the c.d.f. of Y_ν^σ,

$$\begin{aligned} F_\nu^\sigma(y) :=& \mathbf{Pr}\left[\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n} \leqslant y \right] \\ =& \mathbf{E}\left[\left(\sum_{i=1}^n p_i^{\nu,\sigma} X_{n-i+1:n} - y \right)_-^0 \right], \end{aligned}$$

which immediately gives

$$F_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} F_\nu^\sigma(y) \tag{5}$$

As one could have expected from the definition of the Choquet integral, the determination of the moments and the distribution functions of the Choquet integral is closely related to the determination of the moments and the distribution functions of linear combinations of order statistics.

4 The uniform case

In this section, we are interested in the moments and distribution functions of Y_ν when the random sample

$X_1, \ldots, X_n$ is drawn from the standard uniform distribution. In order to emphasize this last point, as classically done, we shall denote the random sample as $U_1, \ldots, U_n$ and the corresponding order statistics by $U_{1:n} \leqslant \cdots \leqslant U_{n:n}$.

Before yielding the main results obtained in [7], let us recall some basic material related to divided differences. See for instance [8, 9, 10] for further details.

4.1 Divided differences

Let $\mathcal{A}^{(n)}$ be the set of $n-1$ times differentiable one-place functions g such that $g^{(n-1)}$ is absolutely continuous. The nth *divided difference* of a function $g \in \mathcal{A}^{(n)}$ is the symmetric function of $n+1$ arguments defined inductively by $\Delta[g : a_0] := g(a_0)$ and

$$\Delta[g : a_0, \ldots, a_n] := \begin{cases} \dfrac{\Delta[g : a_1, \ldots, a_n] - \Delta[g : a_0, \ldots, a_{n-1}]}{a_n - a_0}, & \text{if } a_0 \neq a_n, \\ \dfrac{\partial}{\partial a_0} \Delta[g : a_0, \ldots, a_{n-1}], & \text{if } a_0 = a_n. \end{cases}$$

The *Peano representation* of the divided differences, which can be obtained by a Taylor expansion of g, is given by

$$\Delta[g : a_0, \ldots, a_n] = \frac{1}{n!} \int_{\mathbb{R}} g^{(n)}(t)\, M(t \mid a_0, \ldots, a_n)\, \mathrm{d}t, \tag{6}$$

where $M(t \mid a_0, \ldots, a_n)$ is the *B-spline* of order n, with knots $\{a_0, \ldots, a_n\}$, defined as

$$M(t \mid a_0, \ldots, a_n) := n\, \Delta[(\cdot - t)_+^{n-1} : a_0, \ldots, a_n]. \tag{7}$$

We also recall the *Hermite-Genocchi formula*: For any function $g \in \mathcal{A}^{(n)}$, we have

$$\Delta[g : a_0, \ldots, a_n] = \int_{R_{id} \cap [0,1]^n} g^{(n)} \Big[a_0 + \sum_{i=1}^{n} (a_i - a_{i-1}) x_i \Big]\, \mathrm{d}x, \tag{8}$$

where R_{id} is the region defined in (3) when σ is the identity permutation.

For distinct arguments $a_0, \ldots, a_n$, we also have the following formula, which can be verified by induction,

$$\Delta[g : a_0, \ldots, a_n] = \sum_{i=0}^{n} \frac{g(a_i)}{\prod_{j \neq i} (a_i - a_j)}. \tag{9}$$

4.2 Moments and distribution

Let $g \in \mathcal{A}^{(n)}$. From (8), we immediately have that

$$\mathbf{E}\left[g^{(n)} \left(\sum_{i=1}^{n} p_i^{\nu,\sigma}\, U_{n-i+1:n} \right) \right] = n! \Delta[g : \nu_0^\sigma, \ldots, \nu_n^\sigma]$$

since the joint p.d.f. of $U_{n:n} \geqslant \cdots \geqslant U_{1:n}$ is equal to $\frac{1}{n!}$ on $R_{id} \cap [0,1]^n$ and is zero elsewhere. Combining the previous expression with (4), we obtain

$$\mathbf{E}[g^{(n)}(Y_\nu)] = \sum_{\sigma \in \mathfrak{S}_n} \Delta[g : \nu_0^\sigma, \ldots, \nu_n^\sigma]. \tag{10}$$

Eq. (10) provides the expectation $\mathbf{E}[g^{(n)}(Y_\nu)]$ in terms of the divided differences of g with arguments $\nu_0^\sigma, \ldots, \nu_n^\sigma$ ($\sigma \in \mathfrak{S}_n$). An explicit formula can be obtained by (9) whenever the arguments are distinct for every $\sigma \in \mathfrak{S}_n$.

Clearly, the special cases

$$g(x) = \frac{r!}{(n+r)!}\, x^{n+r},\ \frac{r!}{(n+r)!}\, [x - \mathbf{E}(Y_\nu)]^{n+r},\ \text{and}\ \frac{e^{tx}}{t^n} \tag{11}$$

give, respectively, the raw moments, the central moments, and the moment-generating function of Y_ν. As far as the raw moments are concerned, the following result was obtained in [7].

Proposition 1. *For any integer $r \geqslant 1$, setting $A_0 := N$, we have,*

$$\mathbf{E}[Y_\nu^r] = \frac{1}{\binom{n+r}{r}} \sum_{\substack{A_1 \subseteq N \\ A_2 \subseteq A_1 \\ \cdots \\ A_r \subseteq A_{r-1}}} \prod_{i=1}^{r} \frac{1}{\binom{|A_{i-1}|}{|A_i|}}\, \nu(A_i).$$

Proposition 1 provides an explicit expression for the rth raw moment of Y_ν as a sum of $(r+1)^n$ terms. For instance, the first two moments are

$$\begin{aligned} \mathbf{E}[Y_\nu] &= \frac{1}{n+1} \sum_{A \subseteq N} \frac{1}{\binom{n}{|A|}}\, \nu(A), \\ \mathbf{E}[Y_\nu^2] &= \frac{2}{(n+1)(n+2)} \times \\ &\qquad \sum_{A_1 \subseteq N} \frac{1}{\binom{n}{|A_1|}}\, \nu(A_1) \sum_{A_2 \subseteq A_1} \frac{1}{\binom{|A_1|}{|A_2|}}\, \nu(A_2). \end{aligned}$$

As far as the distribution function $F_\nu(y) := \mathbf{Pr}[Y_\nu \leqslant y]$ of Y_ν is concerned, using (10) with $g(x) = \frac{1}{n!}(x-y)_-^n$, the following result was obtained in [7].

Theorem 1. *There holds*

$$\begin{aligned} F_\nu(y) &= \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \Delta[(\cdot - y)_-^n : \nu_0^\sigma, \ldots, \nu_n^\sigma] \\ &= 1 - \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \Delta[(\cdot - y)_+^n : \nu_0^\sigma, \ldots, \nu_n^\sigma]. \end{aligned} \tag{12}$$

It follows from (12) that the distribution function of Y_ν is absolutely continuous and hence its probability density function is simply given by

$$f_\nu(y) = \frac{1}{(n-1)!} \sum_{\sigma \in \mathfrak{S}_n} \Delta[(\cdot - y)_+^{n-1} : \nu_0^\sigma, \ldots, \nu_n^\sigma], \tag{13}$$

or, using the B-spline notation (7),

$$f_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} M(y \mid \nu_0^\sigma, \ldots, \nu_n^\sigma).$$

Remark 1. (i) When the arguments $\nu_0^\sigma, \ldots, \nu_n^\sigma$ are distinct for every $\sigma \in \mathfrak{S}_n$, then combining (9) with (12) immediately yields the following explicit expression

$$F_\nu(y) = 1 - \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=0}^{n} \frac{(\nu_i^\sigma - y)_+^n}{\prod_{j \neq i}(\nu_i^\sigma - \nu_j^\sigma)},$$

or, using the minus truncated power function,

$$F_\nu(y) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=0}^{n} \frac{(\nu_i^\sigma - y)_-^n}{\prod_{j \neq i}(\nu_i^\sigma - \nu_j^\sigma)}.$$

(ii) The case of linear combinations of order statistics, called *ordered weighted averaging operators* in aggregation theory (see § 2.2), is of particular interest. In this case, each ν_i^σ is independent of σ, so that we can write $\nu_i := \nu_i^\sigma$. The main formulas then reduce to (see for instance [11] and [12])

$$\begin{aligned} \mathbf{E}[g^{(n)}(Y_\nu)] &= n!\, \Delta[g : \nu_0, \ldots, \nu_n], \\ F_\nu(y) &= \Delta[(\cdot - y)_-^n : \nu_0, \ldots, \nu_n], \\ f_\nu(y) &= M(y \mid \nu_0, \ldots, \nu_n). \end{aligned}$$

We also note that the Hermite-Genocchi formula (8) provides nice geometric interpretations of $F_\nu(y)$ and $f_\nu(y)$ in terms of volumes of slices and sections of canonical simplices (see also [13] and [14]).

4.3 Algorithms for computing divided differences

Both functions F_ν and f_ν require the computation of divided differences of truncated power functions. On this issue, we recall a recurrence equation, due to de Boor [15] and rediscovered independently by Varsi [16] (see also [13]), which allows to compute $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n]$ in $O(n^2)$ operations.

Rename as $b_1, \ldots, b_r$ the elements a_i such that $a_i < y$ and as $c_1, \ldots, c_s$ the elements a_i such that $a_i \geqslant y$ so that $r + s = n + 1$. Then, the unique solution of the recurrence equation

$$\alpha_{k,l} = \frac{(c_l - y)\alpha_{k-1,l} + (y - b_k)\alpha_{k,l-1}}{c_l - b_k} \qquad (k \leqslant r,\ l \leqslant s),$$

with initial values $\alpha_{1,1} = (c_1 - b_1)^{-1}$ and $\alpha_{0,l} = \alpha_{k,0} = 0$ for all $l, k \geqslant 2$, is given by

$$\alpha_{k,l} := \Delta[(\cdot - y)_+^{k+l-2} : b_1, \ldots, b_k, c_1, \ldots, c_l], \ (k{+}l \geqslant 2).$$

Algorithm 1 Algorithm for the computation of $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n]$.

Require: n, $a_0, \ldots, a_n$, y
  $S \leftarrow 0$, $R \leftarrow 0$
  for $i = 0, 1, \ldots, n$ **do**
    if $x_i - y \geqslant 0$ **then**
      $S \leftarrow S + 1$
      $C_S \leftarrow x_i - y$
    else
      $R \leftarrow R + 1$
      $B_R \leftarrow x_i - y$
    end if
  end for
  $A_0 \leftarrow 0$, $A_1 \leftarrow 1/(C_1 - B_1)$ {Initialization of the unidimensional temporary array of size $S+1$ necessary for the computation of the divided difference}
  for $j = 2, \ldots, S$ **do**
    $A_j \leftarrow -B_1 A_{j-1}/(C_j - B_1)$
  end for
  for $i = 2, \ldots, R$ **do**
    for $j = 1, \ldots, S$ **do**
      $A_j \leftarrow (C_j A_j - B_i A_{j-1})/(C_j - B_i)$
    end for
  end for
  return A_R {Contains the value of $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n]$.}

In order to compute $\Delta[(\cdot - y)_+^{n-1} : a_0, \ldots, a_n] = \alpha_{r,s}$, it suffices therefore to compute the sequence $\alpha_{k,l}$ for $k + l \geqslant 2$, $k \leqslant r$, $l \leqslant s$, by means of 2 nested loops, one on k, the other on l. We detail this computation in Algorithm 1; see also [13, 16].

We can compute $\Delta[(\cdot - y)_-^n : a_0, \ldots, a_n]$ similarly. Indeed, the same recurrence equation applied to the initial values $\alpha_{0,l} = 0$ for all $l \geqslant 1$ and $\alpha_{k,0} = 1$ for all $k \geqslant 1$, produces the solution

$$\alpha_{k,l} := \Delta[(\cdot - y)_-^{k+l-1} : b_1, \ldots, b_k, c_1, \ldots, c_l] \qquad (k{+}l \geqslant 1).$$

Example. As we have already mentioned, the Choquet integral is widely used in non-additive expected utility theory, cooperative game theory, complexity analysis, measure theory, etc. (see [17] for an overview.) For instance, when a discrete Choquet integral is used as an aggregation tool in a given decision making problem, it is then very informative for the decision maker to know its distribution. In that context, the most natural *a priori* density on $[0, 1]^n$ is the uniform one, which makes the results presented in this section of particular interest.

Let ν be the capacity on $\{1, 2, 3\}$ defined by $\nu(\{1\}) = 0.1$, $\nu(\{2\}) = 0.6$, $\nu(\{3\}) = \nu(\{1, 2\}) = \nu(\{1, 3\}) = \nu(\{2, 3\}) = 0.9$, and $\nu(\{1, 2, 3\}) = 1$. The density of the Choquet integral w.r.t. ν, which can be computed through (13) and by means of Algorithm 1, is

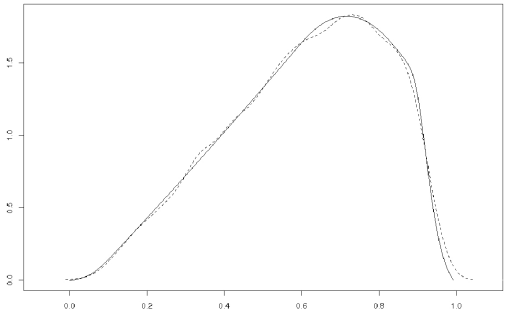


Figure 1: Density of a discrete Choquet integral (solid line).

represented in Figure 1 by the solid line. The dotted line represents the density estimated by the kernel method from 10 000 randomly generated realizations. The typical value and standard deviation can also be calculated through the raw moments: we have

$$\mathbf{E}[Y_\nu] \approx 0.608 \quad \text{and} \quad \sqrt{\mathbf{E}[Y_\nu^2] - \mathbf{E}[Y_\nu]^2} \approx 0.204.$$

From a practical perspective, routines for computing the p.d.f. and the c.d.f. of the Choquet integral in the uniform case have been implemented in the Kappalab package for GNU R [18].

5 The non-uniform case

We now turn to the non-uniform case. Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with c.d.f. F. Unlike in the uniform case, in this section we will only be able to present results allowing to compute approximations of moments of $Y_\nu := C_\nu(X_1, \ldots, X_n)$.

5.1 Expectation and variance of the Choquet integral

We first focus on the expectation and the variance of Y_ν. Starting from (4) with $h(x) = x$, we immediately obtain

$$\mathbf{E}[Y_\nu] = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^{n} p_i^{\nu,\sigma}\, \mathbf{E}[X_{n-i+1:n}]. \tag{14}$$

Similarly, for $h(x) = x^2$, we get

$$\mathbf{E}[Y_\nu^2] = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i^{\nu,\sigma} p_j^{\nu,\sigma}\, \mathbf{E}[X_{n-i+1:n} X_{n-j+1:n}]. \tag{15}$$

It immediatelty follows that the expectation and the variance of the Choquet integral can be computed in the non-uniform case only if the first product moments of order statistics from the same underlying distribution can be computed. As we shall see in the next subsection, it is possible to obtain approximations of these product moments provided the inverse of F and the derivatives of the inverse can be computed.

5.2 Moments of order statistics and their approximation

Let $U_1, \ldots, U_n$ be a random sample from the standard uniform distribution. The product moments of the corresponding order statistics are then given by the following formula (see e.g. [19, Chap. 3] and the references therein):

$$\mathbf{E}\left(\prod_{j=1}^{l} U_{i_j:n}^{m_j}\right) = \frac{n!}{\left(n + \sum_{j=1}^{l} m_j\right)!} \times \prod_{j=1}^{l} \frac{(i_j + m_1 + \cdots + m_j - 1)!}{(i_j + m_1 + \cdots + m_{j-1} - 1)!}. \tag{16}$$

Now, it is well-known that the c.d.f. of $X_{i:n}$ is given by

$$\mathbf{Pr}(X_{i:n} \leqslant x) = \sum_{j=i}^{n} \binom{n}{j} F^j(x)[1 - F(x)]^{n-j}.$$

It immediately follows that

$$\mathbf{Pr}(F^{-1}(U_{i:n}) \leqslant x) = \mathbf{Pr}(U_{i:n} \leqslant F(x)) = \mathbf{Pr}(X_{i:n} \leqslant x),$$

i.e. that $F^{-1}(U_{i:n})$ and $X_{i:n}$ are equal in distribution.

Starting from this distributional equality, David and Johnson [20] expanded $F^{-1}(U_{i:n})$ in a Taylor series around the point $\mathbf{E}(U_{i:n}) = i/(n+1)$ in order to obtain approximations of product moments of non-uniform order statistcs; see also [19, §4.6]. Setting $p_i := i/(n+1)$ and $G := F^{-1}$, we have

$$X_{i:n} = G(p_i) + (U_{i:n} - p_i)G'(p_i) + \frac{1}{2}(U_{i:n} - p_i)^2 G''(p_i) + \frac{1}{6}(U_{i:n} - p_i)^3 G'''(p_i) + \ldots \tag{17}$$

Taking the expectation of the previous expression and using (16), the following approximation for the expectation of $X_{i:n}$ can be obtained to order $(n+2)^{-2}$ [19, §4.6]:

$$\mathbf{E}[X_{i:n}] \approx G_i + \frac{p_i q_i}{2(n+2)} G_i'' + \frac{p_i q_i}{(n+2)^2}\left[\frac{1}{3}(q_i - p_i)G_i''' + \frac{1}{8} p_i q_i G_i''''\right], \tag{18}$$

where $q_i := 1 - p_i$ and $G_i := G(p_i)$, $G_i' := G'(p_i)$, etc. Similarly, for the first product moment, we have, to

order $(n+2)^{-2}$,

$$\begin{aligned}\mathbf{E}[X_{i:n}X_{j:n}] \approx\ & G_iG_j + \frac{p_iq_j}{n+2}G_i'G_j' + \frac{p_iq_i}{2(n+2)}G_jG_i'' \\ & + \frac{p_jq_j}{2(n+2)}G_iG_j'' + \frac{p_iq_j}{(n+2)^2}\Big[(q_i-p_i)G_i''G_j' \\ & + (q_j-p_j)G_i'G_j'' + \frac{1}{2}p_iq_iG_i'''G_j' + \frac{1}{2}p_jq_jG_i'G_j''' \\ & + \frac{1}{2}p_iq_jG_i''G_j''\Big] + \frac{p_ip_jq_iq_j}{4(n+2)^2}G_i''G_j'' \\ & + \frac{p_iq_iG_j}{(n+2)^2}\Big[\frac{1}{8}p_iq_iG_i'''' + \frac{1}{3}(q_i-p_i)G_i''\Big] \\ & + \frac{p_jq_jG_i}{(n+2)^2}\Big[\frac{1}{8}p_jq_jG_j'''' + \frac{1}{3}(q_j-p_j)G_j''\Big]. \qquad (19)\end{aligned}$$

The accuracy of the above approximations is discussed in [19, §4.6]. Note that Childs and Balakrishnan [21] have recently proposed MAPLE routines facilitating the computations and permitting the inclusion of higher order terms.

From a practical perspective, the previous expressions are useful only if $G := F^{-1}$ and its derivates can be easily computed. This is the case for instance when F is the c.d.f. of the standard normal distribution. Indeed, there exists algorithms that enable an accurate computation of F^{-1} (see [22] and the references therein) and it can be verified (see [19, p 85]) that $G' = (f \circ G)^{-1}$,

$$G'' = \frac{G}{f^2 \circ G}, G''' = \frac{1+2G^2}{f^3 \circ G} \text{ and } G'''' = \frac{G(7+6G^2)}{f^4 \circ G},$$

where $f = F'$.

5.3 Back to the two first moments of the Choquet integral

Combining (18) and (19) with (14) and (15), it is therefore possible to obtain approximate values for the expectation and the variance of the Choquet integral provided F^{-1} and its derivates can be easily computed.

6 Future work

Using the expresssions given in Section 3 and distributional results on linear combinations of order statistics [19], it is possible to obtain the exact distribution of the Choquet integral for certain non-uniform distributions and also conditions under which the Choquet integral is asymptotically normal. These aspects will be studied in a forthcoming paper.

References

[1] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

[2] M. Grabisch. Fuzzy integral in multicriteria decision making. *Fuzzy Sets and Systems*, 69:279–298, 1995.

[3] J.-L. Marichal. An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems*, 8(6):800–807, 2000.

[4] L. Lovász. Submodular function and convexity. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical programming: The state of the art*, pages 235–257. Springer-Verlag, 1983.

[5] R. D. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. on Systems, Man and Cybernetics*, 18:183–190, 1988.

[6] J.-L. Marichal. Cumulative distribution functions and moments of lattice polynomials. *Statistics and Probability Letters*, page in press, 2005.

[7] J.-L. Marichal and I. Kojadinovic. The distribution of linear combinations of lattice polynomials from the uniform distribution. submitted, 2006.

[8] Philip J. Davis. *Interpolation and approximation. 2nd ed.* Dover Books on Advanced Mathematics. New York: Dover Publications, 1975.

[9] R.A. DeVore and G.G. Lorentz. *Constructive approximation*. Berlin: Springer-Verlag, 1993.

[10] M.J.D. Powell. *Approximation theory and methods*. Cambridge University Press, 1981.

[11] J. A. Adell and C. Sangüesa. Error bounds in divided difference expansions. A probabilistic perspective. *J. Math. Anal. Appl.*, 318(1):352–364, 2006.

[12] Girdhar G. Agarwal, Rohan J. Dalpatadu, and Ashok K. Singh. Linear functions of uniform order statistics and B-splines. *Commun. Stat., Theory Methods*, 31(2):181–192, 2002.

[13] M. M. Ali. Content of the frustum of a simplex. *Pacific Journal of Mathematics*, 48(2):313–322, 1973.

[14] L. Gerber. The volume cut off a simplex by a half space. *Pacific Journal of Mathematics*, 94(2):311–314, 1981.

[15] C. de Boor. On calculating with B-splines. *Journal of Approximation Theory*, 6:50–62, 1972.

[16] G. Varsi. The multidimensional content of the frustrum of the simplex. *Pacific Journal of Mathematics*, 46(1):303–314, 1973.

[17] M. Grabisch, T. Murofushi, and M. Sugeno, editors. *Fuzzy measures and integrals*, volume 40 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2000. Theory and applications.

[18] M. Grabisch, I. Kojadinovic, and P. Meyer. `kappalab`*: Non additive measure and integral manipulation functions*, 2006. R package version 0.3.

[19] H.A. David and H.N. Nagaraja. *Order statistics. 3rd ed.* Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons., 2003.

[20] F.N. David and N.L. Johnson. Statistical treatment of censored data, I. Fundamental formulae. *Biometrika*, 41:228–240, 1954.

[21] A. Childs and N. Balakrishnan. Series approximation for moments of order statistics using maple. *Computational Statistics and Data Analysis*, 38:331–347, 2002.

[22] P. J. Acklam. An algorithm for computing the inverse normal cumulative distribution function. `http://home.online.no/~pjacklam/notes/invnorm/`.

The Core of Games on Distributive Lattices

Lijue XIE
University of Paris I – Panthéon-Sorbonne
Centre d'Économie de la Sorbonne
Xielijue@yahoo.fr

Michel GRABISCH
University of Paris I – Panthéon-Sorbonne
Centre d'Économie de la Sorbonne
michel.grabisch@lip6.fr

Abstract

Cooperative games under precedence constraints have been introduced by Faigle and Kern [1] , [4] as a generalization of classical cooperative games. An important notion in cooperative game theory is the core of the game, as it contains the rational imputations for players. We propose two definitions for the core of a distributive game, the first one is called the precore and is a direct generalization of the classical definition. It contains the set of imputations and may be unbounded, which makes its application questionable. A second definition is proposed, imposing normalization at each stage, causing the core to be a convex bounded set. We study its properties, introducing balancedness and marginal worth vectors, and defining the Weber set and the pre-Weber set. We show that the classical properties of inclusion of the (pre)core into the (pre)-Weber set as well as their coincidence in the convex case remain valid.

Keywords: core, distributive lattice, stage, cooperative game, Weber set.

1 Introduction

In cooperative game theory a major problem is to find a rational way to share the total worth $v(N)$ among all players. To avoid the formation of subcoalitions, we impose the condition $\phi(S) \leq v(S)$ for all $S \subseteq N$. It leads to the classical notion of the core. Technically, the $n!$ marginal vectors form the set of all vertices of the core whenever the game is convex [5]. Necessary and sufficient conditions for the nonemptiness of the core lead to the notion of balancedness.

These classical results of game theory apply as well for capacities and fuzzy measures, especially in the context of decision making under uncertainty. The core is then the set of dominating probability measures, and the same results apply.

In this paper, we focus on the game theoretical interpretation, although our results remain mathematically valid for capacities and fuzzy measures, which are monotonic normalized games. We are interested in situation where all coalitions cannot form, that is, a game is a function defined on some family of subsets of N, not necessarily the whole power set. We make the assumption that this family forms a distributive lattice. As we will show, this case arises whenever a partial order (like a hierarchy) among players is defined, under the rule that a coalition can form if and only if for any player i in the coalition, all players below i belong to the coalition too.

We will propose a definition of the core suited to this situation, so that most of classical results about the core still hold. In particular, we will exhibit all vertices of the core when the game is convex. The extension of the classical case is called the precore here, it coincides with the definition of Faigle and Kern.

2 Distributive Lattices

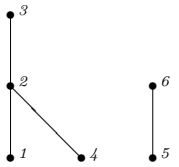
A set P with a binary relation $\leq$ is *a poset* (or *a partially ordered set* [3], [2]) if the binary relation $\leq$ satisfies reflexivity, antisymmetry and transitivity. For any two elements $x, y \in P$, $x < y$ means $x \leq y$ and $x \neq y$. Let x be any element of P. *The principal ideal of* x is defined by $\downarrow x := \{y \in P \mid y \leq x\}$ and *the principal filter of* x is $\uparrow x := \{y \in P \mid y \geq x\}$. If there exists no $y \in P$ such that $y < x$, we call x *a minimal element* of this poset; if there exists no $y \in P$ such that $y > x$, we call x *a maximal element* of this poset.

Let $x, y \in P$ and $x < y$. If there is no $z \in P$, such that $x < z < y$, we say that y *covers* x, denoted by $x \prec y$. If there exists $C := \{z_0 := x, z_1, z_2, \ldots, z_{k-1}, z_k := y\} \subseteq P$, such that $z_0 \prec z_1 \prec z_2 \prec \ldots \prec z_k$, we say that C is a *maximal chain from* x *to* y, and $k = |C| - 1$ is called

the length of this chain. All chains from a minimal element to a maximal element of P are called *maximal chains of* P. Denote the set of maximal chains of P by $\mathcal{C}(P)$.

We partition P into some subsets $Q_1, \cdots, Q_q$ of P: $x \in Q_{i+1}$ if and only if i is the largest length of chains of P from any minimal element to x. Evidently, Q_1 is the set of all minimal elements of P and Q_q is a subset of maximal elements of P. The set Q_i is called *the i-th stage*, every element of Q_i is said to *be in the i-th stage*.

Example 1 *Let us consider the following poset.*



$P = \{1, 2, 3, 4, 5, 6\}$

This poset has 3 stages: $Q_1 = \{1, 4, 5\}, Q_2 = \{2, 6\}$ *and* $Q_3 = \{3\} \subseteq \{3, 6\}$.

Let P be a poset with the collection of stages $\mathcal{Q} = \{Q_1, \cdots, Q_q\}$. We can easily show that for any $x \in Q_i$, $y \in Q_j$, if $x < y$ then $i < j$. But its converse is not always true. Even if $x \in Q_i$, $y \in Q_j$ and $i < j$, x and y may be noncomparable.

For any two element $x, y \in P$, we call $x \vee y$ the *supremum of x and y* if it is the least element of all those greater than x and y, and $x \wedge y$ the *infimum of x and y* if it is the greatest element of all those less than x and y. The *top element* $\top$ of P is the greatest element of P, and the *bottom element* $\bot$ of P is the least element of P. For any $x, y \in P$, $x \vee y$, $x \wedge y$, $\top$ and $\bot$ are unique whenever they exist. If for any $x, y \in P$, both $x \vee y$ and $x \wedge y$ exist, the poset P is called a *lattice*. Clearly, in a finite lattice, $\top, \bot$ always exist. In addition, P is *distributive* if $\vee, \wedge$ satisfy the distributive laws: for all $x, y, z \in P$,

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$$

or equivalently

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z).$$

Let L be a finite lattice and $x \in L$. If $x \neq \bot$ can not be written as a supremum, i.e., $x = y \vee z$ implies $y = x$ or $z = x$, then x is said to be *join-irreducible*. Denote the set of all join-irreducible elements of L by $\mathcal{J}(L)$, and the set of join-irreducible elements less than or equal to an element $x \in L$ by $\eta(x)$.

A *downset* of a poset P is a subset S of P such that, if $x \in S$, then $y \leq x$ implies $y \in S$. A downset of P can be also represented by $\cup_{x \in T} \downarrow x$ for some $T \subseteq P$. We denote the set of all downsets of P by $\mathcal{O}(P)$. G. Birkhoff proved that, for any distributive lattice L, $\mathcal{J}(L)$ is isomorphic to L by the isomorphism η. Put otherwise, any poset P generates a distributive lattice $\mathcal{O}(P)$, whose set of join-irreducible elements is isomorphic to P.

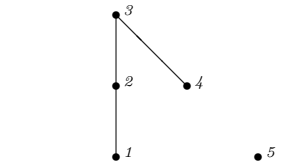
Let P be a poset. Then every stage $Q_i, i = 1, \cdots, q$ of P corresponds to a subset S_i of the distributive lattice $\mathcal{O}(P)$, defined by

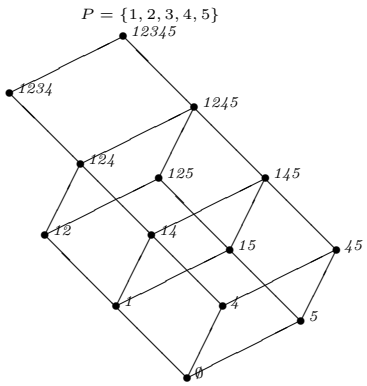
$$S_i := \{x \in \mathcal{O}(P) \mid x \subseteq \cup_{j=1}^{i} Q_j\} \setminus \cup_{j=1}^{i-1} S_j \qquad \forall i = 1, \cdots, q.$$

Proposition 1 *Let P be a poset and $\mathcal{Q} = \{Q_1, \cdots, Q_q\}$ the collection of stages of P. Then $\top_i := \cup_{j=1}^{i} Q_j$ is the greatest element of S_i for all $i = 1, \ldots, q$.*

The collection of all $\top_i$'s is denoted by $\top_P$. Hence $\top$ of $L := \mathcal{O}(P)$ belongs to $\top_P$. If there are some maximal chains passing through all $\top_i$'s in L, we call these maximal chains *restricted maximal chains*. Let us denote the set of restricted maximal chains by $\mathcal{C}_r(L)$.

Example 2 *The poset P and the corresponding distributive lattice $\mathcal{O}(P)$ are given as follows. (For ease of notation,$\{i, j\}$ is denoted by ij and so on.)*





$\mathcal{O}(P)$

Then

$$Q_1 = \{1, 4, 5\}, Q_2 = \{2\}, Q_3 = \{3\}.$$

$$S_1 = \{1, 4, 5, 14, 15, 45, 145\},$$
$$S_2 = \{12, 124, 125, 1245\},$$
$$S_3 = \{1234, 12345\}.$$
$$\top_1 = 145, \top_2 = 1245, \top_3 = 12345.$$

3 Distributive Games

Faigle and Kern introduced cooperative games under precedence constraints [4]. We use the same method to introduce distributive games as follows.

Let $N = \{1, \cdots, n\}$ be the set of players endowed with a partial order $\leq$. The relation $i \leq j$, with $i, j \in N$, indicates the player i is below the player j, we say that, if j participates to the game, all players i below j must also participate to it. All these players compose a subset of N: $\{k \in N : k \leq j\}$, called a feasible coalition. More generally, any downset of $(N, \leq)$ is called a feasible coalition, in other words, $S \subseteq N$ is a feasible coalition if $\forall j \in S, i \leq j$ implies $i \in S$.

Definition 1 *Let $L := \mathcal{O}(N)$ be the collection of all feasible coalitions (all downsets of N). A* game on the distributive lattice *L is a real-valued function $v : L \to \mathbb{R}$ such that $v(\emptyset) = 0$. More simply, we call it a* distributive game.

We introduce the following properties.

Definition 2 *Let v be a distributive game on $\mathcal{O}(N)$.*

(i) v is additive *if $v(S \cup T) = v(S) + v(T)$, for all disjoint $S, T \in \mathcal{O}(N)$.*

(ii) v is convex *if $v(S \cup T) + v(S \cap T) \geq v(S) + v(T)$, for all $S, T \in \mathcal{O}(N)$.*

(iii) v is monotone *if $v(S) \leq v(T)$ whenever $S \subseteq T, S, T \in \mathcal{O}(N)$.*

A monotone distributive game is also called a distributive capacity or a distributive fuzzy measure. We denote the set of all additive distributive games by $\mathcal{A}(\mathcal{O}(N))$.

3.1 Precore and Core

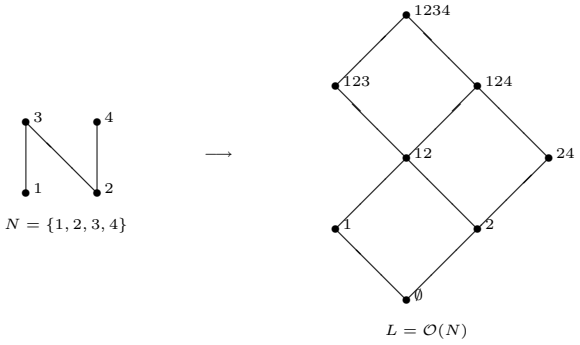
To share the total outcome among all of players so that all peoples satisfy their gains, we define first the precore like in the classical case.

Definition 3 The precore of a distributive game v on $\mathcal{O}(N)$ *is defined by the following set.*

$$\mathcal{PC}(v) := \{\phi \in \mathcal{A}(\mathcal{O}(N)) \mid \phi(N) = v(N) \text{ and } \phi(S) \geq v(S)\, \forall S \in \mathcal{O}(N)\}.$$

It is equivalent to definitions of cores defined by Faigle *et al.* [1] and by Tijs *et al.* [6].

Let us consider another example: A poset N is given in the left. It corresponds to the distributive lattice in the right.



Let v be a distributive game on $\mathcal{O}(N)$. By the definition of the precore, any element ϕ of the precore must satisfy:

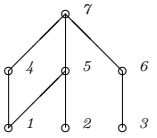
$$\begin{aligned}
\phi(1) + \phi(2) + \phi(3) + \phi(4) = v(\top) = v(1234) \\
\phi(1) \geq v(1) \\
\phi(2) \geq v(2) \\
\phi(1) + \phi(2) \geq v(12) \\
\phi(2) + \phi(4) \geq v(24) \\
\phi(1) + \phi(2) + \phi(3) \geq v(123) \\
\phi(1) + \phi(2) + \phi(4) \geq v(124)
\end{aligned}$$

Whenever $\phi(1), \phi(2)$ are large enough, we can always find out some $\phi(3), \phi(4)$ to satisfy all conditions, i.e., $\phi(1), \phi(2)$ can be arbitrarily large. Hence the precore of this game have four infinite directions: two positive infinite directions of $\phi(1), \phi(2)$ and two negative infinite directions of $\phi(3), \phi(4)$.

Hence the precore of a given game is generally a polyhedron. Denote the set of vertices of some convex set by $\mathrm{Ext}(\cdot)$, and the convex hull of some set by $\mathrm{co}(\cdot)$, the finite part of the precore by $\mathcal{PC}^F(v)$. We have, $\mathcal{PC}^F(v) = \mathrm{co}(\mathrm{Ext}(\mathcal{PC}(v)))$. Hence the finite part of the precore has no more infinite directions. It is a polytope.

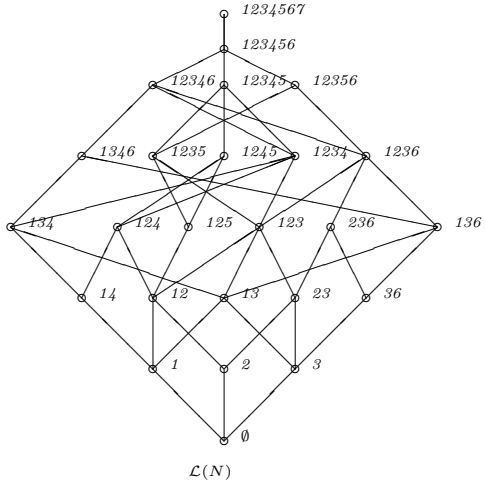
Now we examine the following example.

Example 3 *Suppose that there are 7 employees $N = \{1, 2, 3, 4, 5, 6, 7\}$ in a company, we denote the order of their ranks by the symbole $\leq$. We allow that one worker has more than one rank. We show all of orders among all of workers by the following poset:*



$N = \{1, 2, 3, 4, 5, 6, 7 \mid 1 < 4 < 7, 2 < 5 < 7, 3 < 6 < 7\}$

A downset of $(N, \leq)$ is called a feasible team of the company, i.e., for any work, if the worker p attends it, then every worker with a rank below that of p must attend also this work. All downsets of N compose the set $\mathcal{L}(N)$.



$\mathcal{L}(N)$

Let $\mathcal{Q}_k$ denote the collection of all of workers in the k-th stage, $\top_k$ denote the k-th principal team, i.e., the union of workers up to k-th stage, we have

$$\mathcal{Q}_1 = \{1, 2, 3\}, \mathcal{Q}_2 = \{4, 5, 6\}, \mathcal{Q}_3 = \{7\},$$

$$\top_1 = 123, \top_2 = 123456, \top_3 = 1234567 = N.$$

In the end of each year, we decide how to distribute the total benefit of this year. We denote $v(S)$ the benefit brought by the feasible team S, $\phi(S)$ the factual benefit of S. By the importance of ranks, the benefit is shared going from the highest (top) principal team to lowest (bottom) principal team. That is,

$$v(N) - v(123456) = v(\top_3) - v(\top_2)$$
is given to the group $\mathcal{Q}_3 = \{7\}$,

$$v(123456) - v(123) = v(\top_2) - v(\top_1)$$
is given to the group $\mathcal{Q}_2 = \{4, 5, 6\}$,

$v(123) = v(\top_1)$ *is given to the group* $\mathcal{Q}_1 = \{1, 2, 3\}$

i.e., $\phi(\mathcal{Q}_k) = v(\top_k) - v(\top_{k-1})$, for all k. ϕ must also satisfy the following condition:

$$\phi(T) \geq v(T) \text{ for all feasible teams } T.$$

Otherwise, if for some T, $\phi(T) < v(T)$, then the team T may split from N and build a new independent company, because the money brought by T is more than that T gains.

From the example 3, we can find one another method to share the total outcome and in order to avoid infinite directions, we give the following definition.

Definition 4 The core of a distributive game on $\mathcal{O}(N)$ *is defined by*

$$\mathcal{C}(v) := \{\phi \in \mathcal{PC}(v) \mid \phi(\top_i) = v(\top_i), \forall \top_i \in \top_N\}.$$

Evidently, the precore and the core all are convex.

Theorem 1 *The core and the precore of a game on a distributive lattice are both convex sets. Moreover, the core is bounded, hence it is a polytope.*

3.2 Balancedness

To find out necessary and sufficient conditions for the nonemptiness of the precore, we introduce the notion of pre-balancedness.

Definition 5 *(i) A collection $\mathcal{B}$ of elements of $\mathcal{O}(N) \setminus \{\emptyset\}$ is* pre-balanced *if it exists positive coefficients $\mu(S)$, $S \in \mathcal{B}$, such that $\sum_{S:S\ni i} \mu(S) = 1$, for all $i \in N$. (ii) A distributive game v is* pre-balanced *if for every pre-balanced collection $\mathcal{B}$ of elements of $L \setminus \{\emptyset\}$ with coefficients $\mu(S)$, $S \in \mathcal{B}$, it holds*

$$\sum_{S \in \mathcal{B}} \mu(S) v(S) \leq v(\top).$$

Proposition 2 *A distributive game has a nonempty precore if and if only it is pre-balanced.*

Let $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ be the collection of stages of N and $\top_N = \{\top_1, \ldots, \top_q\}$ be the collection of top elements of every stage of $\mathcal{O}(N)$. Similarly, we introduce the notion of balancedness as follows.

Definition 6 *(i) A collection $\mathcal{B}$ of elements of $\mathcal{O}(N) \setminus \{\emptyset\}$ is* balanced *if it exists positive coefficients $\mu(S)$, $S \in \mathcal{B}$, such that $\sum_{S:S\ni i} \mu(S) = q - k + 1$, for all $i \in Q_k, k = 1, \ldots, q$. (ii) A distributive game v is* balanced *if for every balanced collection $\mathcal{B}$ of elements of $\mathcal{O}(N) \setminus \{\emptyset\}$ with coefficients $\mu(S)$, $S \in \mathcal{B}$, it holds*

$$\sum_{S \in \mathcal{B}} \mu(S) v(S) \leq v(\top).$$

Proposition 3 *A distributive game has a nonempty core if and if only it is balanced.*

3.3 Marginal Worth Vectors

Let $C = \{S_0 := \emptyset \prec S_1 \prec \cdots \prec S_n := N\}$ be a maximal chain in a distributive lattice $L := \mathcal{O}(N)$. Hence, to each maximal chain we associate a permutation on N, $\pi : N \to N$, such that the additional element between any two consecutive coalitions S_{i-1}, S_i of C is $\pi(i)$. So we have $S_i = \{\pi(1), \pi(2), \ldots, \pi(i)\}$.

Definition 7 *The* marginal worth vector $\psi^C \in \mathbb{R}^n$ *associated to* C *and* v *is defined by*

$$\psi^C(j) := v(S_i) - v(S_{i-1}), \quad \forall i \in N,$$

with $j = S_i \setminus S_{i-1}$.

The set of all marginal worth vectors ψ^C for all maximal chains is denoted by $\mathcal{M}(v)$. We can easily get

$$\begin{aligned}\psi^C(S_i) &:= \sum_{k=1}^{i} \psi^C(\pi(k)) \\ &= \sum_{k=1}^{i} (v(S_k) - v(S_{k-1})) = v(S_i), \forall S_i \in C.\end{aligned}$$

Definition 8 *The* Weber set $\mathcal{W}(v)$ *of* v *is defined as the convex hull of all vectors in* $\mathcal{M}(v)$*:*

$$\mathcal{W}(v) := \text{co}(\mathcal{M}(v)).$$

Theorem 2 *For any distributive game* v*, the polytope of the precore is included in the Weber set, i.e,* $\mathcal{PC}^F(v) \subseteq \mathcal{W}(v)$.

We call all marginal worth vector ψ^{C_r} associating just to the restricted chains C_r the *real marginal worth vectors*. The set of all real marginal worth vectors is denoted by $\mathcal{M}^r(v)$.

Definition 9 *The* real Weber set *is defined as the convex hull of all real marginal worth vectors* ψ^{C_r}*:*

$$\mathcal{W}^r(v) := \text{co}(\mathcal{M}^r(v)).$$

Theorem 3 *For any distributive game* v*, the core is included in the real Weber set, i.e.,* $\mathcal{C}(v) \subseteq \mathcal{W}^r(v)$.

3.4 Core of Convex Distributive Games

We give our main results as follows.

Theorem 4 *For any distributive game* v*, it is convex if and only if* $\mathcal{PC}^F(v) = \mathcal{W}(v)$*, i.e.,* $\text{Ext}(\mathcal{PC}(v)) = \mathcal{M}(v)$.

To prove this theorem, we must show some lemmas.

Lemma 1 *If a distributive game* v *is convex, then the pre-Weber set belongs to the precore:*

$$\mathcal{W}(v) \subseteq \mathcal{PC}(v).$$

Lemma 2 *If a distributive game* v *is convex, then any marginal worth vector in* $\mathcal{M}(v)$ *is a vertex of the precore:*

$$\mathcal{M}(v) = \text{Ext}(\mathcal{W}(v)) \subseteq \text{Ext}(\mathcal{PC}(v)).$$

Lemma 3 *If a distributive game* v *is convex, then* $\mathcal{PC}^F(v) = \mathcal{W}(v)$*, or equivalently* $\text{Ext}(\mathcal{PC}(v)) = \mathcal{M}(v)$.

For the core, we have a similar result.

Theorem 5 *If a distributive game* v *is convex, then any payoff vector in* $\mathcal{M}^r(v)$ *is a vertex of the core:*

$$\mathcal{M}^r(v) = \text{Ext}(\mathcal{W}^r(v)) \subseteq \text{Ext}(\mathcal{C}(v)).$$

Corollary 1 *If a distributive game* v *is convex, then* $\mathcal{C}(v) = \mathcal{W}^r(v)$*, or equivalently* $\text{Ext}(\mathcal{C}(v)) = \mathcal{M}^r(v)$

But $\mathcal{C}(v) = \mathcal{W}^r(v)$ can not imply that v is convex, i.e., $\mathcal{C}(v) = \mathcal{W}^r(v)$ is not equivalent that $\mathcal{PC}^F(v) = \mathcal{W}(v)$.

References

[1] U. Faigle. Cores of games with restricted cooperation. *ZOR-Methods and Models of Operations Research*, 33:405–422, 1989.

[2] B.A. Davey and H.A. Priestley. Introduction to Lattices and Orders. *Cambridge University Press*, 1990.

[3] Birkhoff-1967. Lattice Theory. *Americain Mathematical Society*, 3rd Edition, 1967.

[4] U. Faigle and W. Kern. The Shapley value for cooperative games under precedence constraints. *International Journal of Game Theory*, 21:249–266, 1992.

[5] L.S. Shapley. Core of convex games. *International Journal of Game Theory*, 1:11–26, 1971.

[6] A. van den Nouweland, S. Tijs, J. Potters, and J. Zarzuelo. Cores and related solution concepts for multi-choice games. *ZOR - Mathematical Methods of Operations Research*, 41:289–311, 1995.

Session 6

Fuzzy Methods in Learning and Data Mining – E. Hüllermeier and F. Klawonn

Towards Robust Rank Correlation Measures for Numerical Observations on the Basis of Fuzzy Orderings

Ulrich Bodenhofer
Institute of Bioinformatics
Johannes Kepler University Linz
4040 Linz, Austria
bodenhofer@bioinf.jku.at

Frank Klawonn
University of Applied Sciences
Braunschweig/Wolfenbüttel
38302 Wolfenbüttel, Germany
f.klawonn@fh-wolfenbuettel.de

Abstract

This paper aims to demonstrate that established rank correlation measures are not ideally suited for measuring rank correlation for numerical data that are perturbed by noise. We propose a robust rank correlation measure on the basis of fuzzy orderings. The superiority of the new measure is demonstrated by means of illustrative examples.

Keywords: Fuzzy Orderings, Rank Correlation, Robust Statistics.

1 Introduction

Correlation measures are among the most basic tools in statistical data analysis and machine learning. They are applied to pairs of observations ($n \geq 2$)

$$(x_i, y_i)_{i=1}^n \tag{1}$$

to measure to which extent the two observations comply with a certain model. The most prominent representative is surely *Pearson's product moment coefficient* [1, 14], often nonchalantly called *correlation coefficient* for short. Pearson's product moment coefficient is applicable to numerical data and assumes a linear relationship as the underlying model; therefore, it can be used to detect linear relationships, but no non-linear ones.

Rank correlation measures [9, 11, 13] are intended to measure to which extent a monotonic function is able to model the inherent relationship between the two observables. They neither assume a specific parametric model nor specific distributions of the observables. They can be applied to ordinal data and, if some ordering relation is given, to numerical data too. Therefore, rank correlation measures are ideally suited for detecting monotonic relationships, in particular, if more specific information about the data is not available. The two most common approaches are *Spearman's rank correlation coefficient* (short *Spearman's rho*) [16, 17] and *Kendall's tau (rank correlation coefficient)* [2, 10, 11].

This paper argues why the well-known rank correlation measures are not ideally suited for measuring rank correlation for numerical data that are perturbed by noise. Consequently, we propose a robust rank correlation measure on the basis of fuzzy orderings. The superiority of the new measure is demonstrated by means of illustrative examples.

2 An Overview of Rank Correlation Measures

Assume that we are given a family of pairs as in (1), where all x_i and y_i are from linearly ordered domains X and Y, respectively. *Spearman's rho* is computed as

$$\rho = 1 - 6\frac{\sum_{i=1}^n (r(x_i) - r(y_i))^2}{n(n^2-1)},$$

where $r(x_i)$ is the rank of value x_i if we sort the list $(x_1, \ldots, x_n)$; $r(y_i)$ is defined analogously. So, Spearman's rho measures the sum of quadratic distances of ranks and scales this measure to the interval $[-1, 1]$. It can be checked easily that a value of 1 is obtained if the two rankings coincide and that a value of -1 is obtained if one ranking is the reverse of the respective other. Note that the above definition of $r(x_i)$ and $r(y_i)$ was simplified, because it did not take coinciding values, so-called *ties*, into account. In such a case, the values $r(x_i)$ are usually defined as the mean value of all ranks of consecutive coinciding values in the sorted list.

With the same assumptions as above, *Kendall's tau* is computed as the quotient

$$\tau_a = \frac{C - D}{\frac{1}{2}n(n-1)},$$

where C and D denote the numbers of *concordant* and *discordant pairs*, respectively:

$$C = |\{(i,j) \mid x_i < x_j \text{ and } y_i < y_j\}|$$
$$D = |\{(i,j) \mid x_i < x_j \text{ and } y_i > y_j\}|$$

As above, if we have no ties and the two rankings coincide, we have $\frac{1}{2}n(n-1)$ concordant and no discordant pairs, so $\tau_a = 1$; if we have no ties and one ranking is the reverse of the respective other, we have no concordant and $\frac{1}{2}n(n-1)$ discordant pairs, so a value of $\tau_a = -1$ is obtained.

In the above definition of τ_a, ties, no matter whether in the first or in the second list, are not counted. So ties lower the absolute value of τ_a. Therefore, τ_a is best suited for detecting strictly monotonic relationships, but not ideally suited in the presence of ties. A well-established second variant [11],

$$\tau_b = \frac{C - D}{\sqrt{\frac{1}{2}n(n-1) - T}\sqrt{\frac{1}{2}n(n-1) - U}},$$

where

$$T = |\{(i,j) \mid x_i = x_j\}|, \quad U = |\{(i,j) \mid y_i = y_j\}|,$$

takes ties into account, but is still not fully robust to ties. A simple and tie-robust rank correlation measure is the *gamma rank correlation measure* according to Goodman and Kruskal [9] that is defined as

$$\gamma = \frac{C - D}{C + D}.$$

3 Motivation

All rank correlation measures highlighted above have been introduced with the aim to measure rank correlation of ordinal data (e.g. natural numbers, marks, quality classes, ranks). The measurement of rank correlation for *real-valued data*, however, is equally important in statistics and machine learning, but raises completely new issues. Depending on the source, numerical data are almost always subject to random perturbations—noise. The concepts introduced above do not take this into account. Pairs are counted as concordant or discordant only on the basis of ordering relations, but without taking into account that only minimal differences may decide whether a pair is concordant or discordant. If one observable depends on the other in a clearly monotonic way and if the level of noise is low, then the rank correlation measures introduced above will still reveal this strictly monotonic relationship and will not be compromised by minor local effects of noise. In the presence of a larger percentage of ties, however, already the slightest perturbations may lead to situations in which the above rank

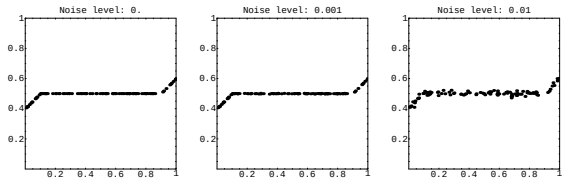


Figure 1: Scatter plots of a simple monotonic relationship with different noise levels.

correlation coefficients cannot yield meaningful results anymore. Consider the data sets in Figure 1. We see a monotonic, yet not strictly monotonic, relationship. The left plot shows data without noise, i.e. $y_i = f(x_i)$ for a non-decreasing function f. For these data, we obtain $\rho = 0.737$, $\tau_b = 0.639$ and $\gamma = 1$ (which confirms that γ is most robust to ties). The middle plot shows the same data, but with additive normally distributed noise with zero mean and $\sigma = 0.001$. Although it is hard to see the noise at all, we obtain $\rho = 0.519$ and $\tau_b = \gamma = 0.387$. These results indicate that none of the three measures can adequately handle a large proportion of ties in the presence of noise. For $\sigma = 0.01$ (right plot), the values are slightly lower, but not significantly: $\rho = 0.456$ and $\tau_b = \gamma = 0.331$. So we can conclude that it is rather the presence of noise in general than the magnitude of noise that distracts the three rank correlation measures.

The obvious reason for the weakness described above is the fact that all measures only take ordering relationships into account, but neglect similarities of data points. To illustrate that, consider the two pairs (a,c) and (b,c), where $b > a$. Obviously, this is a tie in the second component. If we add some noise to the second component of the second pair, i.e., if we replace (b,c) by $(b, c+\varepsilon)$, then ε decides whether $((a,c),(b,c+\varepsilon))$ is a tie (for $\varepsilon = 0$), concordant ($\varepsilon > 0$), or discordant $\varepsilon < 0$), where the magnitude of ε plays no role at all. So we observe a discontinuous behavior. This toy example thereby serves as a proof that all measures introduced above depend on the data in a discontinuous way.

The question arises how we can define a robust rank correlation measure that depends continuously on the data by taking similarities into account, but still serves as a meaningful measure of rank correlation. Obviously, the measure should be designed such that close-to-tie pairs receive less attention than pairs that are clearly concordant or discordant. A reasonable idea would be to base such a concept on the probabilities to which concordant/discordant pairs are observed as such compared to the probabilities that they are falsely observed as something else. That may be a reasonable approach. Note, however, that such probabilities can

only be computed if we know the joint distribution of x and y values or at least if we make distribution assumptions. In practice, such information is most often unavailable and, surely, we do not want to sacrifice the unique feature of rank correlation measures that they are *distribution-free*.

In our opinion, *fuzzy orderings* provide a meaningful way to overcome the difficulties explained above.

4 Fuzzy Orderings

Before we can introduce a fuzzy ordering-based rank correlation coefficient, we need to provide some basics of fuzzy orderings. We restrict to an absolutely necessary minimum and refer to literature for details. We assume that the reader is aware of the most basic concepts of triangular norms and fuzzy relations.

A fuzzy relation $L : X^2 \to [0,1]$ is called *fuzzy ordering* with respect to a t-norm T and a T-equivalence E, for brevity *T-E-ordering*, if and only if it is T-transitive and fulfills the following two axioms for all $x, y \in X$:

(i) E-Reflexivity: $E(x,y) \leq L(x,y)$

(ii) T-E-antisymmetry: $T\big(L(x,y), L(y,x)\big) \leq E(x,y)$

Moreover, we call a T-E-ordering L *strongly complete* if $\max\big(L(x,y), L(y,x)\big) = 1$ for all $x, y \in X$ [4].

Several correspondences between distances and fuzzy equivalence relations are available [6, 7, 12, 18]. From these results, we can easily infer that (assume $r > 0$ in the following)

$$E_r(x,y) = \max(0, 1 - \tfrac{1}{r}|x-y|)$$

is a $T_{\mathbf{L}}$-equivalence on $\mathbb{R}$, where $T_{\mathbf{L}}(x,y) = \max(0, x+y-1)$ denotes the Łukasiewicz t-norm. Analogously,

$$E'_r(x,y) = \exp(-\tfrac{1}{r}|x-y|)$$

is a $T_{\mathbf{P}}$-equivalence on $\mathbb{R}$, where $T_{\mathbf{P}}(x,y) = xy$ denotes the product t-norm.

Based on a general representation theorem for strongly complete fuzzy orderings [4], we can further prove that

$$L_r(x,y) = \min(1, \max(0, 1 - \tfrac{1}{r}(x-y)))$$

is a strongly complete $T_{\mathbf{L}}$-E_r-ordering on $\mathbb{R}$ and that

$$L'_r(x,y) = \min(1, \exp(-\tfrac{1}{r}(x-y))$$

is a strongly complete $T_{\mathbf{P}}$-E'_r-ordering on $\mathbb{R}$. As $T_{\mathbf{L}} \leq T_{\mathbf{P}}$, we can trivially conclude that L'_r is also a strongly complete $T_{\mathbf{L}}$-E'_r-ordering.

In order to generalize the notion of concordant and discordant pairs, we need the notion of a strict fuzzy ordering. We call a binary fuzzy relation R a *strict fuzzy ordering* with respect to T and a T-equivalence E, for brevity *strict T-E-ordering*, if it is irreflexive (i.e. $R(x,x) = 0$ for all $x \in X$), T-transitive, and E-extensional, that is,

$$T(E(x,x'), E(y,y'), R(x,y)) \leq R(x',y')$$

for all $x, x', y, y', z \in X$ [5].

Given a T-E-ordering L,

$$R(x,y) = \min(L(x,y), N_T(L(y,x))), \tag{2}$$

where $N_T(x) = \sup\{y \in [0,1] \mid T(x,y) = 0\}$ is the residual negation of T, is the most appropriate choice for extracting a strict fuzzy ordering from a given fuzzy ordering L (for a detailed argumentation, see [5]). From this construction, we can infer that the fuzzy relation

$$R_r(x,y) = \min(1, \max(0, \tfrac{1}{r}(y-x)))$$

is a strict $T_{\mathbf{L}}$-E_r-ordering and that

$$R'_r(x,y) = \max(0, 1 - \exp(-\tfrac{1}{r}(y-x)))$$

is a strict $T_{\mathbf{L}}$-E'_r-ordering.

If a given $T_{\mathbf{L}}$-E-ordering L is strongly complete, it can be proved that the fuzzy relation R defined as in (2) simplifies to

$$R(x,y) = 1 - L(y,x)$$

and that the following holds:

$$R(x,y) + E(x,y) + R(y,x) = 1 \tag{3}$$

$$\min(R(x,y), R(y,x)) = 0 \tag{4}$$

5 A Fuzzy Ordering-Based Rank Correlation Coefficient

The previous section has provided us with the apparatus that is necessary to define a generalized rank correlation measure. Assume that the data are given as in (1) again (with $x_i \in X$ and $y_i \in Y$ for all $i = 1, \ldots, n$). Further assume that we are given two $T_{\mathbf{L}}$-equivalences $E_X : X^2 \to [0,1]$ and $E_Y : Y^2 \to [0,1]$, a strongly complete $T_{\mathbf{L}}$-E_X-ordering $L_X : X^2 \to [0,1]$ and a strongly complete $T_{\mathbf{L}}$-E_Y-ordering $L_Y : Y^2 \to [0,1]$. Therefore, we can define a strict $T_{\mathbf{L}}$-E_X-ordering on X as $R_X(x_1,x_2) = 1 - L_X(x_2,x_1)$ and a strict $T_{\mathbf{L}}$-E_Y-ordering on Y as $R_Y(y_1,y_2) = 1 - L_Y(y_2,y_1)$.

Spearman's rho is based on rankings. Rankings are crisp concepts in which it is not easy to accommodate degrees of relationship in a straightforward way. Thus it is more meaningful to use pairwise comparisons to

define a concept of rank correlation, just like Kendall's tau and the gamma measure do.

Given an index pair (i,j), we can compute the degree to which $((x_i,y_i),(x_j,y_j))$ is a concordant pair as

$$\tilde{C}(i,j) = \min(R_X(x_i,x_j), R_Y(y_i,y_j))$$

and the degree to which $((x_i,y_i),(x_j,y_j))$ is a discordant pair as

$$\tilde{D}(i,j) = \min(R_X(x_i,x_j), R_Y(y_j,y_i)).$$

If we adopt the simple sigma count idea to measure the cardinality of a fuzzy set [8], we can compute the numbers of concordant pairs $\tilde{C}$ and discordant pairs $\tilde{D}$, respectively, as

$$\tilde{C} = \sum_{i=1}^{n} \sum_{j \neq i} \tilde{C}(i,j),$$
$$\tilde{D} = \sum_{i=1}^{n} \sum_{j \neq i} \tilde{D}(i,j).$$

The question arises whether we should attempt to generalize τ_a, τ_b or γ. As the main motivation is to get rid of the influence of close-to-ties pairs in the presence of noise, it is immediate that the idea behind γ is the most promising one. So, with the assumptions from above, we define our *fuzzy ordering-based rank correlation measure* $\tilde{\gamma}$ as

$$\tilde{\gamma} = \frac{\tilde{C} - \tilde{D}}{\tilde{C} + \tilde{D}}.$$

To interpret the meaning of $\tilde{\gamma}$, we note that, for all index pairs (i,j), the equality

$$\tilde{C}(i,j) + \tilde{C}(j,i) + \tilde{D}(i,j) + \tilde{D}(j,i) + \tilde{T}(i,j) = 1 \quad (5)$$

holds, where $\tilde{T}(i,j)$ denotes the degree to which (i,j) is a tie in either variable:

$$\tilde{T}(i,j) = \max(E_X(x_i,x_j), E_Y(y_i,y_j))$$

Moreover, we can infer the following:

$$\tilde{C} = \sum_{i=1}^{n} \sum_{j>i} (\tilde{C}(i,j) + \tilde{C}(j,i))$$
$$\tilde{D} = \sum_{i=1}^{n} \sum_{j>i} (\tilde{D}(i,j) + \tilde{D}(i,j))$$

Thus, by (5), $\tilde{C}+\tilde{D}$ equals the number of non-tie pairs if we consider each choice of indices i,j only once (in contrast to considering (i,j) *and* (j,i) independently for each i and j). So $\tilde{\gamma}$ measures the difference of concordant and discordant pairs relative to the number of non-tie pairs; the concept of "tiedness" is a fuzzy one, however.

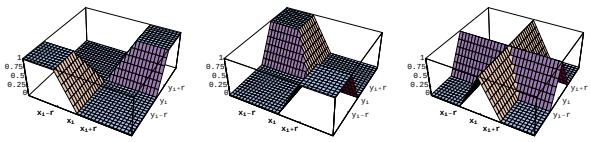


Figure 2: $\tilde{C}(i,j)+\tilde{C}(j,i)$ (left), $\tilde{D}(i,j)+\tilde{D}(j,i)$ (middle), and $\tilde{T}(i,j)$ (right) plotted as functions of x_j and y_j for fixed x_i and y_i (using the relations E_r and R_r).

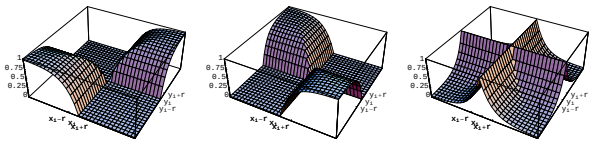


Figure 3: $\tilde{C}(i,j)+\tilde{C}(j,i)$ (left), $\tilde{D}(i,j)+\tilde{D}(j,i)$ (middle), and $\tilde{T}(i,j)$ (right) plotted as functions of x_j and y_j for fixed x_i and y_i (using the relations E'_r and R'_r).

It is obvious that, in case that E_X and E_Y are crisp equalities and that R_X and R_Y are crisp linear strict orderings, that $\tilde{\gamma}$ coincides with γ. So what is the difference if R_X and R_Y are non-trivial fuzzy relations? The above interpretation shows that concordant/discordant pairs are counted more if they are dissimilar and less if they are similar—which perfectly corresponds to our intention. Let us demonstrate this fact with an example.

Assume $X = Y = \mathbb{R}$, $E_X = E_Y = E_r$, and $R_X = R_Y = R_r$ for some $r > 0$. Fixing some x_i and y_i and considering $\tilde{C}(i,j)+\tilde{C}(j,i)$, $\tilde{D}(i,j)+\tilde{D}(j,i)$, and $\tilde{T}(i,j)$ as functions of the two variables x_j and y_j, the graphs shown in Figure 2 can be obtained. It can be seen that pairs are counted fully if $|x_i - x_j| > r$ and $|y_i - y_j| > r$ (i.e. like in the classical γ measure). If one of the two distances is smaller than r, the pair is considered as a tie to the corresponding degree $\tilde{T}(i,j)$ and only counted to a degree of $1 - \tilde{T}(i,j)$. One also sees that, if r is chosen so large that $|x_i - x_j| \leq r$ and $|y_i - y_j| \leq r$ for all pairs, all pairs are counted to a degree proportionally to the minimum of these two distances. If the relations $E_X = E_Y = E'_r$, and $R_X = R_Y = R'_r$ are used, the effect is qualitatively similar, r also controls to which degree a close-to-tie pair is counted, also in a monotonic, yet asymptotic fashion (see Figure 3).

It is clear from the above examples that, the smaller r, the more $\tilde{\gamma}$ resembles to γ. For both, the variant based on E_r/R_r and the variant based on E'_r/R'_r, it can be proved that $\tilde{\gamma}$ converges to γ for $r \to 0$.

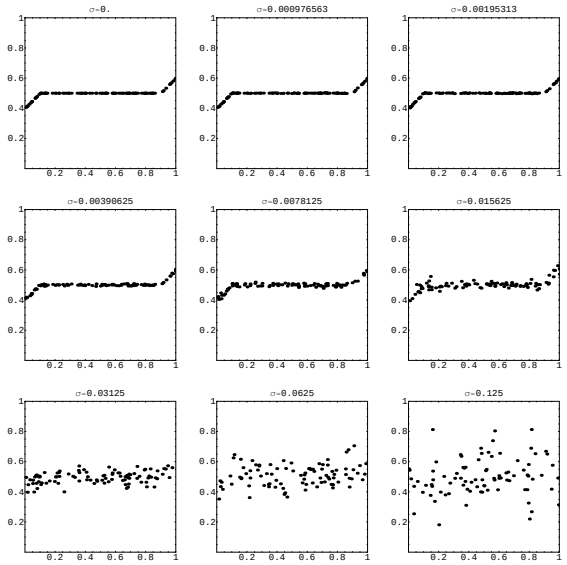


Figure 4: Different data sets obtained from contaminating a non-decreasing relationship by normally distributed noise with different standard deviations.

Another property of $\tilde{\gamma}$ is immediate to see: if the fuzzy relations R_X and R_Y are continuous (assuming that this notion makes sense on X and Y), then $\tilde{\gamma}$ depends continuously on the data set $(x_i, y_i)_{i=1}^n$.

6 Experiments

Let us first reconsider the example from Section 3. More specifically, we are given 100 uniformly distributed random values $(x_1, \ldots, x_{100})$ from the unit interval. The list $(y_1, \ldots, y_{100})$ is computed as $y_i = f(x_i)$, where f is a simple, piecewise linear, non-decreasing function that has a relatively large flat area. In order to study how different rank correlation measures react to noise, we contaminated the data points with additive, independent, normally distributed noise with 0 mean and standard deviation σ. Figure 4 shows these data sets. Figure 6 displays the results that we obtained for different rank correlation measures. We compared ρ, τ_b, γ and different variants of $\tilde{\gamma}$. Every line in Figure 6 corresponds to the results obtained by one rank correlation measure depending on the noise level σ. The two lines for τ_b (dotted, black) and γ (dotted, light gray) coincide except for no noise ($\sigma = 0$). Both lines reveal that these two measures react to noise in an non-robust way. More or less the same is true for ρ (dotted, medium gray). The other lines correspond to different variants of $\tilde{\gamma}$. Solid lines correspond to $\tilde{\gamma}$ using R_r and dashed lines denote the results for $\tilde{\gamma}$ using R'_r (where we use the same r for both components). We used $r = 0.05$ (black), $r = 0.2$ (medium gray), and $r = 0.5$ (light gray). We see that all six different variants react to the noise in a more robust way than the three crisp measures. Clearly, the higher r, the more noise is neglected. Note, however, that, the larger r, the more difficult it is for $\tilde{\gamma}$ to find out whether there are slightly non-monotonic parts in the data.

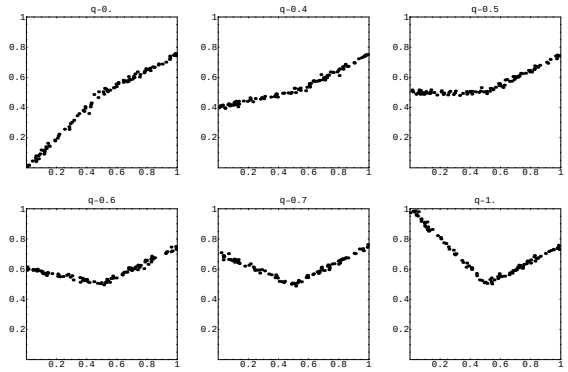


Figure 5: Noisy data sets that correspond to monotonic ($q \leq 0.5$) and non-monotonic relationships ($q > 0.5$).

So let us consider a different setting. Now we fix the noise level $\sigma = 0.01$ and use different functions to create the second list $(y_1, \ldots, y_{100})$. Right of $x = 0.5$, we use $f(x) = \frac{x}{2} + \frac{1}{4}$ and to the left or $x = 0.5$, we linearly interpolate between $(0, q)$ and $(0.5, 0.5)$. It is clear, that this relationship is monotonic if and only if $q \leq 0.5$. The data sets are displayed in Figure 5 and the results are presented in Figure 7, where we use the same conventions to distinguish the lines as in Figure 6. We see that all variants of $\tilde{\gamma}$ show acceptable results for $q \leq 0.5$, whereas ρ, τ_b and γ again have problems to handle the noise in case of the large proportion of ties that occurs for $q = 0.5$. We also see that $\tilde{\gamma}$ already yields significantly lower values for $q = 0.6$ in the case $r = 0.05$ (no matter which of the two variants is considered). For larger r, however, we see that $\tilde{\gamma}$ cannot detect the slight non-monotonicity for $q = 0.6$ that well. These two examples demonstrate that, when choosing r, there is a trade-off between robustness (the larger r, the better) and sensitivity (the smaller r, the better).

As a third set of experiments, we have tried to figure out the variance of $\tilde{\gamma}$. For this study, we have computed all rank correlation measures used in the above experiments for different test data several times and computed the variance of the results. In all experiments, we have encountered that τ_b and γ had higher variances than all variants of $\tilde{\gamma}$. The variances we obtained for different variants of $\tilde{\gamma}$ obeyed a simple

and unsurprising rule: the larger r, the smaller the variance. Interestingly, the variances we obtained for Spearman's ρ were also very low, comparable to the lower values for $\tilde{\gamma}$ with a large r.

Note that the authors have carried out numerous experiments to solidify the above claims. As the space in this paper is limited, we just quoted the most interesting and demonstrative results.

7 Concluding Remarks

This paper, as the appellative term "towards" in the title suggests, attempts to present first ideas that the authors consider promising. The examples of the previous section are intended to support this viewpoint. They are illustrative and indicative, but they cannot replace a formal investigation of the properties of $\tilde{\gamma}$. As it has been done exhaustively for Spearman's rho and Kendall's tau, a significance analysis and a variance analysis have to be carried out. Note, however, that this cannot be done analogously for $\tilde{\gamma}$. Both Spearman's rho and Kendall's tau are fully determined by the ranking of the lists $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$. Thus, combinatorial techniques can be used to study variances and significance levels [11]—not so for $\tilde{\gamma}$ that always depends on the distance relationships of the values, too, so this analysis can only be done by some distribution assumptions. These studies are left to future research.

To determine the right choice for the parameter r is another open question. As we have noted above, there is a trade-off between robustness on the one side and sensitivity/significance on the other side. So this topic goes hand in hand with a more formal statistical analysis. Profound results concerning the choice of r, again, can only be expected with specific distribution assumptions. In any case, we want to note in advance that $\tilde{\gamma}$ depends continuously on r, so at least we can be sure that $\tilde{\gamma}$ will react robust to slightly sub-optimal choices of r.

Finally, we would like to remark that this investigation was inspired by a problem in bioinformatics: how to infer sets of co-transcribed genes in procaryotic genomes (so-called *operons*) from the gene expression levels measured by microarray experiments [3, 15, **?**]. It will also be subject of future research to evaluate the rank correlation measures introduced in this paper in this domain.

References

[1] H. Abdi. Coefficients of correlation, alienation and determination. In N. J. Salkind, editor, *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA, 2007.

[2] H. Abdi. The Kendall rank correlation coefficient. In N. J. Salkind, editor, *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA, 2007.

[3] J. Bockhorst, Y. Qiu, J. Glasner, M. Liu, F. Blattner, and M. Craven. Predicting bacterial transcription units using sequence and expression data. *Bioinformatics*, 19(Suppl. 1):i34–i43, 2003.

[4] U. Bodenhofer. A similarity-based generalization of fuzzy orderings preserving the classical axioms. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 8(5):593–610, 2000.

[5] U. Bodenhofer and M. Demirci. Strict fuzzy orderings in a similarity-based setting. In *Proc. Joint 4th Conf. of the European Society for Fuzzy Logic and Technology and 11 Recontres Francophones sur la Logique Floue et ses Applications*, pages 297–302, Barcelona, September 2005.

[6] B. De Baets and R. Mesiar. Pseudo-metrics and T-equivalences. *J. Fuzzy Math.*, 5(2):471–481, 1997.

[7] B. De Baets and R. Mesiar. Metrics and T-equalities. *J. Math. Anal. Appl.*, 267:331–347, 2002.

[8] A. DeLuca and S. Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Inf. Control*, 20:301–312, 1972.

[9] L. A. Goodman and W. H. Kruskal. Measures of association for cross classifications. *J. Amer. Statist. Assoc.*, 49(268):732–764, 1954.

[10] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.

[11] M. G. Kendall. *Rank Correlation Methods*. Charles Griffin & Co., London, third edition, 1962.

[12] F. Klawonn. Fuzzy sets and vague environments. *Fuzzy Sets and Systems*, 66:207–221, 1994.

[13] W. H. Kruskal. Ordinal measures of association. *J. Amer. Statist. Assoc.*, 53(284):814–861, 1958.

[14] K. Pearson. Notes on the history of correlation. *Biometrika*, 13:25–45, 1920.

[15] C. Sabatti, L. Rohlin, M.-K. Oh, and J. C. Liao. Co-expression pattern from DNA microarray experiments as a tool for operon prediction. *Nucleic Acids Res.*, 30(13):2886–2893, 2002.

[16] C. Spearman. The proof and measurement of association between two things. *Am. J. Psychol.*, 15(1):72–101, 1904.

[17] C. Spearman. Demonstration of formulae for true measurement of correlation. *Am. J. Psychol.*, 18(2):161–169, 1907.

[18] E. Trillas and L. Valverde. An inquiry into indistinguishability operators. In H. J. Skala, S. Termini, and E. Trillas, editors, *Aspects of Vagueness*, pages 231–256. Reidel, Dordrecht, 1984.

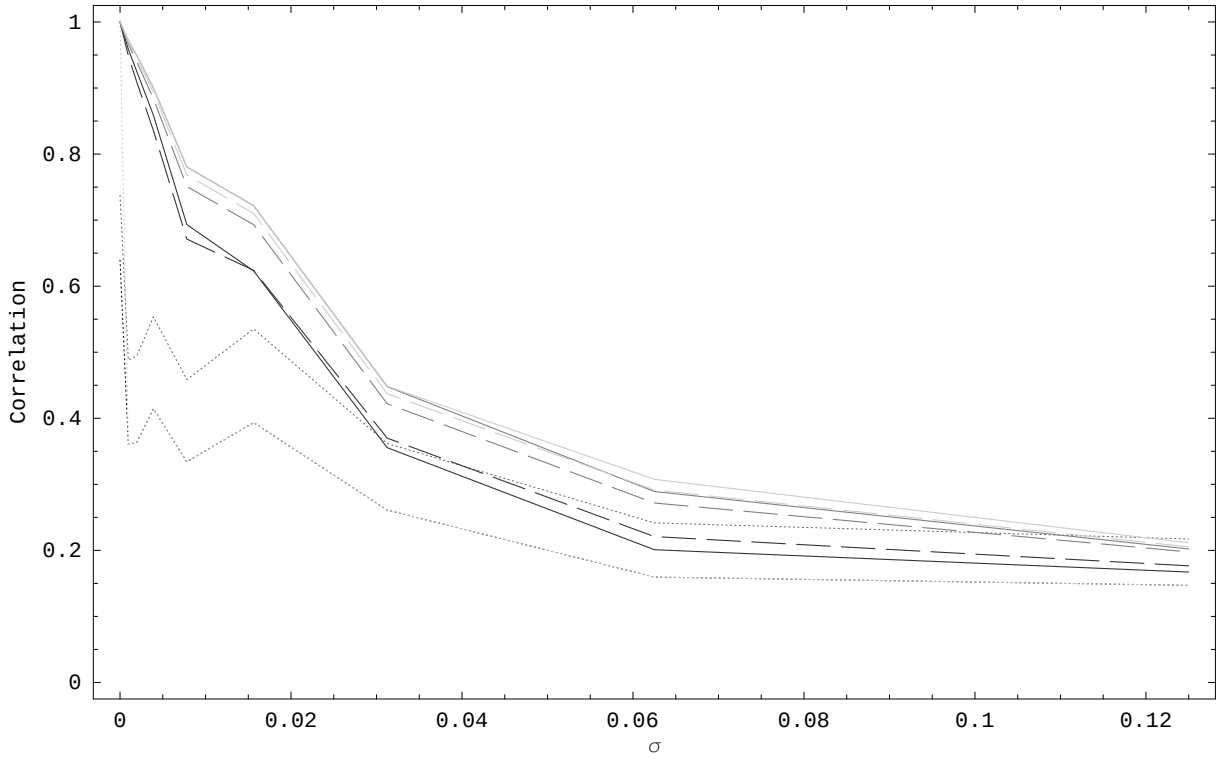


Figure 6: Results obtained by applying different rank correlation measures to the data sets shown in Figure 4.

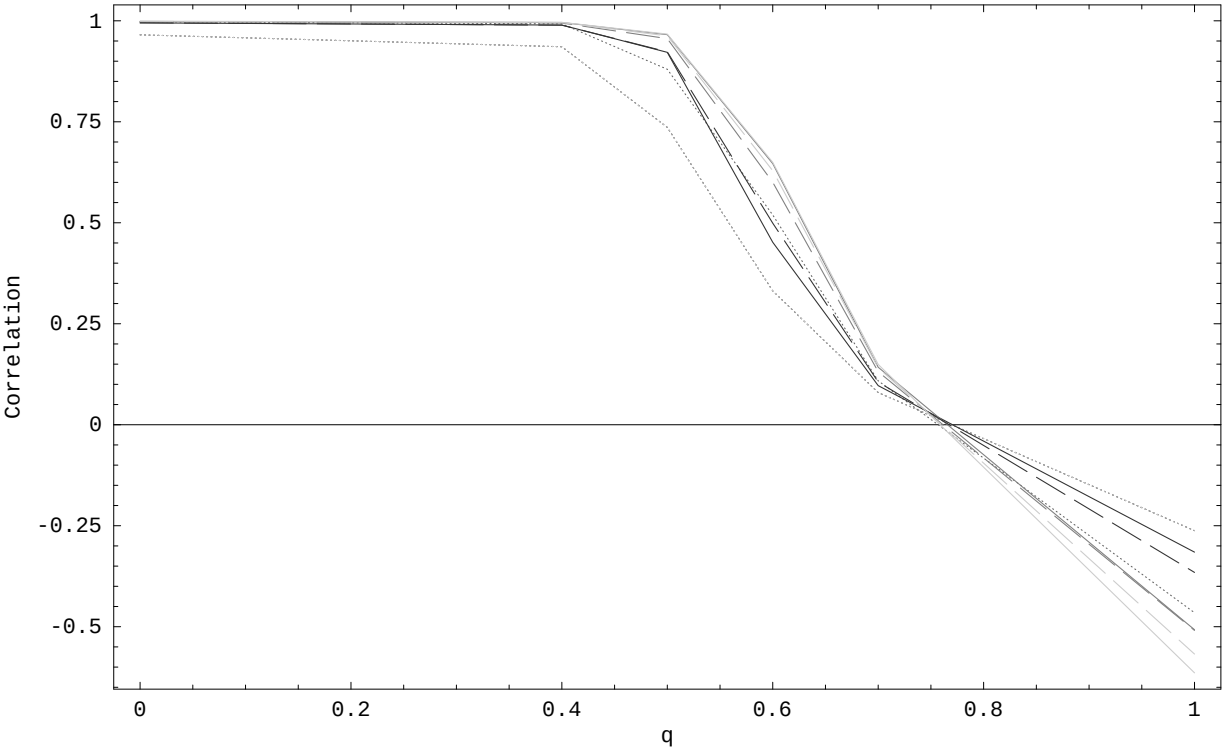


Figure 7: Results obtained by applying different rank correlation measures to the data sets shown in Figure 5.

SPoID: Do Not Throw Meaningful Incomplete Sequences Away!

Céline Fiot
LIRMM - CNRS
161 rue Ada
34392 Montpellier, France
fiot@lirmm.fr

Anne Laurent
LIRMM - CNRS
161 rue Ada
34392 Montpellier, France
laurent@lirmm.fr

Maguelonne Teisseire
LIRMM - CNRS
161 rue Ada
34392 Montpellier, France
teisseire@lirmm.fr

Abstract

Industrial databases often contain a large amount of unfilled information. During the knowledge discovery process one processing step is often necessary in order to remove these incomplete data either by deleting or assessing them. When the data mining task consists in mining for frequent sequences, incomplete data are, most of the time, deleted, which leads to an important loss of information. Extracted knowledge then becomes less representative of the database. Therefore we propose a method that uses the partial information contained in incomplete records, only temporary ignoring the missing part of the record. Experiments run on various synthetic datasets show the validity of our proposal as well in terms of quality as in terms of the robustness to the rate of missing values.

Keywords: Data mining, sequential patterns, missing values, incomplete data.

1 Introduction

For the last decade, data mining application fields have widened. Approaches which were designed for specific applications are now applied to other ones. Especially it is the case of sequential patterns. This data mining technique aims at discovering knowlegde from temporal databases. First designed to analyse customer behavior, they are now used in various industrial or biological fields. The searched datasets are sets of time-stamped records. Each of them is constituted of a set of values.

But sequential pattern approaches must be adapted, these kinds of data having different format or some of them being imperfect. In particular, industrial databases often contain incomplete data (records containing unfilled attributes) due to breakdowns or errors, for instance. However existing methods only allow the analysis of complete data, without considering incomplete records, which may represent an important loss of information. Besides, replacement of missing values or their estimation are often either too simplistic to perform unbiased results, or too time-consuming to be applied on large datasets. For this reason it could be useful to implement a method for mining sequential patterns within incomplete databases.

In this paper we propose an extension of the principles originally developed by [2] in order to discover frequent sequences within databases containing values missing at random. This approach was designed on the basis of an association rule method for mining incomplete databases [14] and of a technique often used in machine learning [8]: ignoring missing values without ignoring the whole involved record. This principle consists in only making use of available information (i.e. filled-in attributes) and ignoring missing information. Thus only partial complete databases are mined for each pattern and the whole dataset is used to discover all the patterns.
We here redefine this principle for sequential pattern mining, thus proposing to discover maximal frequent sequences within time-stamped incomplete databases. For this reason, we adapt the concepts linked to sequential pattern discovery, leading to the SPoID (Sequential Patterns over Incomplete Data) algorithm. This algorithm uses a well-known and efficient sequential pattern mining algorithm, PSP [9]. It has been tested on synthetic datasets to show the validity of our approach.

The remainder of the paper is organized as follows. Section 2 introduces the methods for association rule mining in the presence of missing values, and the concepts linked to sequential pattern discovery. Section 3 details our approach to mine for sequential patterns within incomplete data, and we describe our algorithm in Section 4. Section 5 is then dedicated to experiments that show the validity of our approach. Finally,

we discuss in Section 6 further work opened by this proposal and we conclude in Section 7.

2 Sequential Patterns and Incompleteness

Sequential patterns are often introduced as an extension of association rules, initially proposed in [1]. They highlight correlations between database records as well as their temporal relationship. Even so these algorithms do not mine incomplete records contained in the database. These missing values must then be removed either by deletion or replacement. Quality of results then depends on this preprocess. Moreover this step is often time-consuming. In order to reduce the preprocessing due to missing values and to improve the sequential pattern quality, we propose a method for sequential pattern mining within incomplete databases. This method is based on association rules approaches. In this section we first define the concepts linked to sequential pattern mining, then we detail our motivations before introducing techniques that allow association rule mining handling missing values.

2.1 Sequential Patterns

Sequential patterns are based on the idea of *maximal frequent sequences*.
Let $\mathcal{R}$ be a set of objects records where each record R consists of three information elements: an object-id, a record timestamp and a set of attributes/items in the record. Let $I = \{i_1, i_2, ..., i_m\}$ be a set of items or attributes. An *itemset* is a non-empty set of attributes i_k, denoted by $(i_1 i_2 \dots i_k)$. It is a non-ordered representation. A *sequence* s is a non-empty ordered list of itemsets s_p, denoted by $< s_1 s_2 ... s_p >$. A *n-sequence* is a sequence of n items (or of size n).

Example 1. *Let us consider an example of market basket analysis. The object is a customer, and records are the transactions made by this customer. Timestamps are the date of transactions. If a customer purchases products e, a, k, u, and f according to the sequence $s =< (e)\ (a\ k)\ (u)\ (f) >$, then all items of the sequence were bought separately, except products a and k which were purchased at the same time. In this example, s is a 5-sequence.*

One sequence $S =< s_1\ s_2 ... s_p >$ is a *subsequence* of another one $S' =< s'_1\ s'_2\ ... s'_m >$ if there are integers $l_1 < l_2 < ... < l_p$ such that $s_1 \subseteq s'_{l_1}$, $s_2 \subseteq s'_{l_2}$, ..., $s_p \subseteq s'_{l_p}$.

Example 2. *The sequence $s' =<(a)\ (f)>$ is a subsequence of s because $(a) \subseteq (a\ k)$ and $(f) \subseteq (f)$. However, $<(a)\ (k)>$ is not a subsequence of s.*

All records from the same object o are grouped together and sorted in increasing order of their timestamp. They are called a data sequence. An object *supports* a sequence s if it is included within the data sequence of this object (s is a subsequence of the data sequence). The *frequency* of a sequence ($freq(s)$) is defined as the percentage of objects supporting s in the whole set of objects $\mathcal{O}$. In order to decide whether a sequence is frequent or not, a minimum frequency value ($minFreq$) is specified by the user and the sequence is said to be frequent if the condition $freq(s) \geq minFreq$ holds. A sequence that may be frequent is a *candidate sequence*. Given a database of object records, the problem of sequential pattern mining is to find all maximal sequences of which the frequency is greater than a specified threshold ($minFreq$) [2]. Each of these sequences represents a sequential pattern, also called a maximal frequent sequence.

Several extensions were proposed to consider incremental mining for sequential patterns [10], to handle numerical and quantitative values [7, 3, 6] or to generalize sequential patterns with respect to various temporal parameters (time-interval between events of a sequence, grouping several records into a single itemset...) [15, 11, 5]. However, no technique was proposed to deal with missing values while sequential pattern mining. For this reason, in the following sections, we propose an approach that can mine maximal frequent sequences from an incomplete sequence database.

2.2 Motivations

Let us consider the database given by Table 1. The goal is to extract the sequential patterns with a minimum frequency equal to 50%.

Table 1: Complete database.

O.	Seq.
O1	(a b) (b c d) (b c e)
O2	(a) (b c) (b d)
O3	(a b) (b c) (b c d)

Table 2: Incomplete database.

O.	Seq.
O1	(a b) (? ? c) (? b c)
O2	(a) (? c) (b d)
O3	(a b) (? c) (? ? c)

The sequential patterns obtained are: <(a b)(b c d)>, <(a b)(b c)(b c)> and <(a)(b c)(b d)>. Now, let us consider the incomplete database given by Table 2.

Some information in the data sequences has not been filled in and these values are missing. Let us consider that these values are identified as missing and unfilled. In order to mine for patterns, previous approaches require the suppression of missing values.

During the pre-processing step, incomplete records may either be completely deleted, which leads to the dataset in Table 3, or be partially but definitively deleted (only missing values are removed), as shown

by Table 4.

Table 3: After incomplete record deletion.

O.	Seq.
O1	$(a\ b)$
O2	$(a)\ (b\ d)$
O3	$(a\ b)$

Table 4: After missing value deletion.

O.	Seq.
O1	$(a\ b)\ (c)\ (b\ c)$
O2	$(a)\ (c)\ (b\ d)$
O3	$(a\ b)\ (c)\ (c)$

Considering the first pre-processing, discovered pattern $minFreq = 50\%$ is: <(a b)>. In the best case, Table 4, resulting patterns for $minFreq = 50\%$ are: <(a b)(c)(c)> and <(a)(c)(b)>.

Note that only a short part of the database is used to discover information and only part of frequent patterns and frequent items of the complete dataset are found. In particular, item d is not frequent in the incomplete database.
As deleting missing values or incomplete records leads to an important loss of information, using the whole dataset without deletion appears essential. In Section 3, we thus propose an approach that uses partial and temporary disabling of incomplete data. All the records will then be used for discovering all the sequential patterns, but each of them is extracted from a partial dataset. This method has been instigated from an association rule approach.

2.3 Association Rules and Missing Values

Some works were proposed to discover association rules within incomplete databases. Especially, [12, 13] implement an assessment systems based on a probabilistic distribution. With this approach one missing value can represent various values while mining for association rules. These methods are particularly well-suited to classical non-historized relational databases, but they are not easily adaptable to the specific format of data sequence detailed in section 2.1.

That is the reason why we chose to start our approach from the RAR algorithm (Robust Association Rules), proposed by [14]. This method, fully compatible with the original proposal [1], allow the user to consider incomplete data while association rule mining within incomplete relational databases, thanks to partial and temporary omission of such incomplete records. The main idea consists in only taking into account filled-in attributes in incomplete records. The whole database is not used to discover each rule but to generate the whole set of rules. This technique is based on the valid database concept, which is a complete dataset for a given itemset. The remaining part of the database is temporary ignored.
In order to consider this dataset partitioning, definitions of support (percentage of records in database that include the rule items) and confidence (probability for a record to contain the right part of the rule knowing it contains the left part) were reformulated. Furthermore, a new concept was introduced to take into account the size of the complete sample used to compute the rule support. This representativity measure also allow to prune rules that are slightly significant with respect to the initial database.

From this principle we redefine in the next section the definitions linked to sequential pattern mining previously detailed. These new definitions are then used by our algorithm to mine incomplete sequence database.

3 SPoID: Dealing with Incomplete Data

3.1 Sequential Patterns over Incomplete Data (SPoID)

As deleting incomplete records leads to an important loss of information we adapted an association rule mining approach robust to missing values. In this section we describe our method, SPoID (Sequential Patterns over Incomplete Data), based on the principles of the RAR algorithm proposed by [14].

The main idea of our approach, as the one of the RAR method, is incomplete elements disabling, within our context, incomplete sequences. While the RAR algorithm only regards complete records for association rules mining, we propose to only take into account the complete data sequences for each candidate sequence. In other words, when an incomplete data sequence is scanned, only filled-in time-stamped attributes will be considered for frequency calculation. Thus each candidate sequence will be considered as a frequent sequence on a partial database, but the whole dataset will be used to find the whole set of frequent sequences.

Let us consider a candidate sequence S, the set $\mathcal{O}$ of objects in the database can be divided into three disjoint subsets (Figure 1):the set of data sequences supporting S, denoted by $\mathcal{O}_S$, the set of data sequences not supporting S, denoted by $\mathcal{O}_{\overline{S}}$, and the set of data sequences that we do not know if they support S or not, denoted by $\mathcal{O}^*_S$.

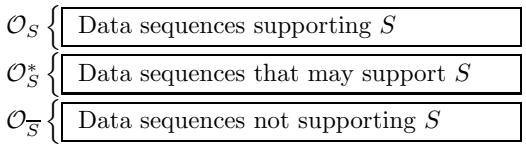


Figure 1: Partition of the database depending on S inclusion.

3.2 Definitions

For each candidate sequence S, only the subsets $\mathcal{O}_{\overline{S}} \cup \mathcal{O}_S$ will be kept to determine wether the sequence S is frequent or not. This data sequence set represents the *valid database* for S.

Definition 1. *A* valid database *is a database only containing complete data sequences for a given candidate sequence, i.e. each value of each record in the data sequence corresponds to an identified item i of $\mathcal{I}$, the set of items in the database.*

Constitution of a valid database leans on temporary disabling data sequences that contain missing values for items in the candidate sequence. This implies to redefine the frequency calculation to take into account the database partial deactivation.

Definition 2. *A data sequence is* disabled *for a candidate sequence S if it is incomplete for S (i.e. we cannot decide whether it supports S or not). The set of data sequence disabled for a candidate sequence S is denoted by $Dis(S)$.*

The frequency definition given in section 2.1 is then modified in order to consider the valid database concept, and thus that only one part of the dataset is used for frequency calculation.

Definition 3. *The* frequency *of a sequence S is the appearance rate of this sequence among the data sequences that can support it. It is defined as the ratio of the number of data sequences supporting S by the number of data sequences that surely include S or not (complete data sequences for S). It is given by:*

$$Supp(S) = \frac{|\mathcal{O}_S|}{|\mathcal{O}| - |Dis(S)|}$$

Property 1. *Considering minor restrictions, this definition holds the antimonotonicity property of the support definition enonciated by [2].*

Proof.

$$Supp(S) = \frac{|\mathcal{O}_S|}{|\mathcal{O}| - |Dis(S)|} = \frac{|\mathcal{O}_S|}{|\mathcal{O}_S| + |\mathcal{O}_{\overline{S}}|} = \frac{1}{1 + \frac{|\mathcal{O}_{\overline{S}}|}{|\mathcal{O}_S|}}$$

However, if $S' \subseteq S$, then $|\mathcal{O}_{S'}| \geq |\mathcal{O}_S|$. Indeed, if a sequence S' is supported by the data sequence of an object o, then either this object supports its supersequence S or it does not support it, or we do not know. But in any case, any object that does not support S' cannot include its supersequence S. Moreover, if $S' \subseteq S$, then $|\mathcal{O}_{\overline{S}}| \geq |\mathcal{O}_{\overline{S'}}|$. Indeed, $o \in \mathcal{O}_{\overline{S'}} \Rightarrow \exists i \backslash s'_i \not\subset o$, however $S' \subseteq S \Rightarrow \forall i, \exists s_k \backslash s'_i \subseteq s_k$, then $\exists k \backslash s_k \not\subset o$ and, in that case, o does not include S.
Considering that none of these cardinalities is null, both inequalities 3.2 and 3.2 can be multiplied member to member. That leading to:

$$\begin{aligned} |\mathcal{O}_{\overline{S}}||\mathcal{O}_{S'}| \geq |\mathcal{O}_S||\mathcal{O}_{\overline{S'}}| &\Rightarrow 1 + \frac{|\mathcal{O}_{\overline{S}}|}{|\mathcal{O}_S|} \geq 1 + \frac{|\mathcal{O}_{\overline{S'}}|}{|\mathcal{O}_{S'}|} \\ &\Rightarrow \frac{1}{1+\frac{|\mathcal{O}_{\overline{S'}}|}{|\mathcal{O}_{S'}|}} \geq \frac{1}{1+\frac{|\mathcal{O}_{\overline{S}}|}{|\mathcal{O}_S|}} \\ &\Rightarrow Supp(S') \geq Supp(S) \end{aligned}$$

Then, the support definition for valid databases holds the antimonotony property. □

As the new support definition is antimonotonic, we can use the various properties described in [2] in order to implement the sequential mining algorithm within incomplete databases. However, the *frequency* concept must also be regarded taking into account the size of the valid database used to compute it. Therefore we define a *representativity* criteria and a minimum representativity threshold *minRep*, that must be satisfied: a valid database must be a significative sample of the whole dataset for a sequence S to be frequent, even if the condition $freq(S) \geq minFreq$ holds.

Definition 4. *The* representativity *$Rep(S)$ of a sequence S is defined as the ratio of the number of data sequences including S or that cannot include it by the total number of data sequences in the whole dataset. It is given by:*

$$Rep(S) = \frac{|\mathcal{O}| - |Dis(S)|}{|\mathcal{O}|}$$

Definition 5. *A sequence is said to be* representative *if its representativity is greater than a minimum representativity value minRep.*

In other words, to be kept as frequent, a candidate sequence must have a representativity greater than the minimum representativity threshold *minRep* and its frequency must be no less than the user-defined minimum threshold $minFreq$.

3.3 Representativity Threshold and Margin of Error

Statistics use sampling techniques that allow to only consider a population subset to assess a proportion, satisfying an error interval with a sufficient confidence. These tools help to determine the optimal sampling size depending on the data distribution. Thus considering a random data distribution, [16] uses the Chernoff bound to set the minimal size of a random sample for association rule mining. This result was also proved theoretically and experimentally by [17].

We thus propose to use two kinds of representativity depending on the user needs: the minimum representativity threshold can be defined either by the

user as a percentage of the dataset size, or it can be an absolute number of data sequences computed from statistics formula related to the data distribution and user-defined parameters for error and confidence level. However our experiments show that the optimal representativity threshold is not an absolute value but rather depends on the missing value rate of datasets.

4 Implementation

4.1 An Example

The incomplete database used in this example is given by Table 2. Let $minFreq$ be 50%. First support and representativity are computed for each item to determine which one are frequent. Item a is certainly supported by the three objects, then its frequency is $freq(a) = 3/3 = 100\%$ and its represetativity is equal to 1. It is the same for items b and c.
For item d, $\mathcal{O}_{<d>} = \{O2\}$ and $Dis(< d >) = \{O1, O3\}$, then $freq(d) = 1/(3\text{-}2) = 1$ and $rep(d) = (3\text{-}2)/3 = 0.33$. If $minRep$ is 0.3, then $rep(d) > minRep$ and d is a frequent item. On the other hand, if $minRep = 0.4$, then $rep(d) < minRep$ and d is not a frequent item because the valid database regarded to compute its frequency is not significant enough.

Let us consider $minRep$=0.3. Now we consider the candidate sequence $S =< (a\ b)(a\ b\ c\ d) >$. This sequence cannot be supported by one of the data sequence, because none of them contains an itemset composed of 4 items, either complete or not. Then $\mathcal{O}_S = \{\}$, $Dis(S) = \{\}$ and $\mathcal{O}_{\overline{S}} = \{O1, O2, O3\}$ and $freq(S)$=0. S is not frequent. Now we consider $S' =< (a\ b)(b\ c) >$. It is supported by $O1$, it cannot be supported by $O2$ but maybe by $O3$. Then, $\mathcal{O}_{S'} = \{O1\}$, $Dis(S') = \{O3\}$ and $\mathcal{O}_{\overline{S'}} = \{O2\}$. That leads to $freq(S')$=1/(3-1)=50% and $rep(S')$=(3-1)/3 = 0.67. S' is then both representative and frequent.

Applying this method, discovered patterns for $minFreq$=50% and $minRep$ = 0.3 are: <(a b)(c)(b c)> and <(a)(c)(b d)>. Even if these patterns are not exactly the one obtained on the complete database, they are closer to the one we should get than the one discovered using the preprocessed dataset. Experiments detailed in section 5 show that there exists a value of the minimum representativity for which the algorithm SPoID extracts the whole set of sequential patterns of the complete database from an incomplete one.

4.2 Algorithm

The algorithm SPoID runs similarly to the generate-prune sequential pattern mining algorithms. It consists in generating all the candidate k-sequences from the frequent (k-1)-sequences. Then the database is scanned to count the number of data sequences that support each candidate sequence. The main difference stands in the counting of incomplete data sequences. This counting step is described by the algorithm Alg. 1: for each candidate sequence, for each object,

- if the candidate sequence is found, the absolute value of the frequency is incremented,
- if the candidate sequence is not found nor a sequence with missing values that could be replaced to complete the candidate sequence, then the object does not support the candidate sequence. The absolute frequency is not incremented.
- an incomplete data sequence in which missing values could be replaced by items of the candidate sequence is found. In that case, the object is added to the disabled object set.

Once the whole dataset scanned, the absolute value of the frequency is divided by the substraction of the number of disabled objects to the number of objects in the database. The representativity is also computed. Then the pruning step is run to delete candidate sequences that are neither frequent nor representative.

SPoID - **Input:** $|\mathcal{O}|$, sequence database,
$minSup$, minimum support
$minRep$, minimum representativity
(user-defined or computed)
Ouput: $SPList$, frequent sequence list

$C \leftarrow \{i \in \mathcal{I}\}$; $k = 1$;
$F \leftarrow$ getFrepnRep(C,$minFreq$,$minRep$) ;
$SPList$.add(F) ;
While ($C \neq \varnothing$) **do**
  k++ ; $C \leftarrow$ generate(F,k) ;
  For each candidate sequence $s \in C$ **do**
    For each object $o \in \mathcal{O}$ **do**
      [Search for s within S_o]
      If ($s \in S_o$) **Then**
        $support(s)$++ ;
        $Dis(s) \leftarrow Dis(s) \backslash$o ;
      Else
        If ($\tilde{s} \in S_o / \tilde{s}$ may be s) **Then**
          $Dis(s) \leftarrow Dis(s) \cup o$;
        End If
      End If
    End For
    $Sup(s) \leftarrow support(s)/(|\mathcal{O}| - |Dis(s)|)$;
    $Rep(s) \leftarrow |\mathcal{O}| - |Dis(s)|/|\mathcal{O}|$;
    If (($Sup(s) < minSup$)||
      ($Rep(s) < minRep$)) **Then**
      prune(s) ;
    End If
  End For
End While
return SPList ;

Algorithm 1 – *SPoID - main algorithm.*

The temporal complexity of this algorithm is, in the worse case, the same as the one of the algorithm TOTALLYFUZZY presented by [6]. We use the same kind of optimizations to reduce the number of database scans. On the other hand, the space complexity is much less than the one of TOTALLYFUZZY, as it is similar to the one of PSP.

5 Experiments

These experiments were carried out on a PC - Linux 2.6.7 OS, CPU 2.8 GHz with 512 MB of memory. The algorithm was implemented in Java on the PSP principle. In particular, the Prefix-Tree structure was used to store the candidate and frequent sequences.
We used synthetic datasets randomly generated. Then some items were randomly replaced by missing values. Sequential patterns were extracted from the complete database and from the preprocessed incomplete ones (i.e. incomplete databases in which incomplete records have been deleted). Then those patterns are compared to the one discovered by our algorithm SPoID. Results here detailed were obtained from several synthetic datasets containing around 2000 sequences of 20 transactions in average. Each transaction contains around 10 items chosen among 100.

Our analysis is based on the several counting:

- the total number of sequential patterns discovered by SPoID,
- the number of sequential patterns discovered by SPoID, that are discovered in the complete database,
- the number of wrong sequential patterns discovered by SPoID (that groups together the patterns that are not discovered in the complete database and the one not found by SPoID but should be).

Table 5 sums up these notations.

Table 5: Notations for the different kinds of patterns.

β	# sequential patterns discovered by SPoID, also contained in the complete dataset
δ	# different sequential patterns
θ	# sequential patterns discovered by SPoID in the incomplete database
τ	# sequential patterns discovered in the complete dataset

First, Figure 2 shows the evolution of the ratio β/θ, with respect to the minimum representativity threshold. It can be noted that this rate increases according to *minRep*. It means that among sequential patterns discovered by SPoID, the proportion of sequential patterns obtained on the complete dataset increases with the *minRep* threshold.

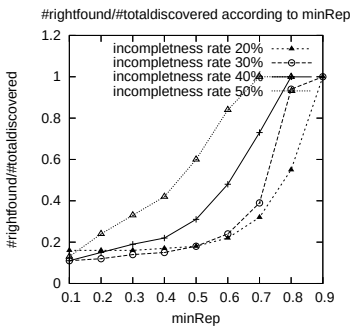


Figure 2: β/θ rate according to *minRep*.

This observation can be completed by analysing Figure 3, which represents evolution of the ratio β/τ (number of discovered sequential patterns with respect to sequential patterns that should be discovered) according to the minimum representativity threshold. This ratio decreases while *minRep* increases. It means that the minimum representativity threshold should be low enough to allow the discovery of all the sequential patterns obtained in the complete database.

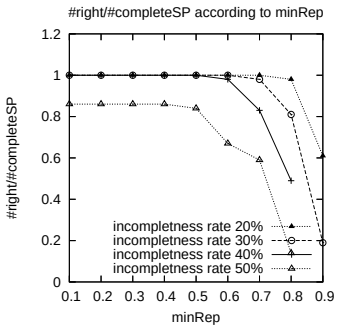


Figure 3: β/τ rate according to *minRep*.

Then we show that there exists an optimal value of the representativity threshold for which the ratio β/θ and β/τ are the closest to 1. This value is the threshold at which the right patterns discovered in the incomplete database are the most numerous compared to the number of wrong patterns. Figure 4 focuses on this optimal value *minRep*. This graph describes the evolution of the ratio β/δ according to the minimum representativity. It can be noted that there is not an absolute value for the minimum representativity, that would be common to every database independently from the incompleteness rate and only depending on an error margin. From these results, the minimum representativity threshold only depends on the incompleteness rate of the database.
Whatever the proportion of missing values in the incomplete database, the overall behavior of the ratio

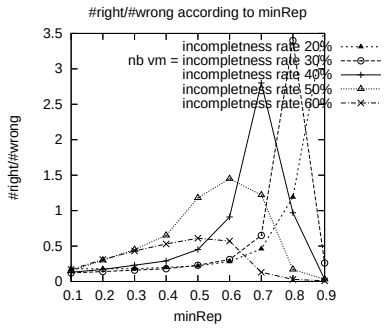


Figure 4: β/δ rate according to *minRep*.

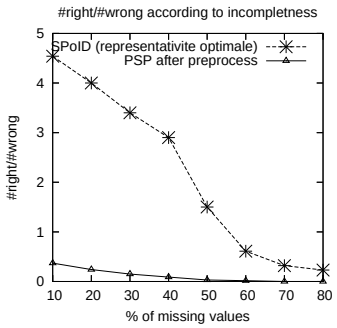


Figure 5: β/δ rate according to the missing value proportion.

β/δ is similar: it increases until it reaches a maximum before decreasing. This maximal point corresponds to the average optimal representativity, for which the number of right patterns discoverd by SPoID is the highest and the number of wrong patterns is the lowest. Table 6 gives the value of optimal representativity empirically found, for each incompleteness rate in datasets.

Table 6: Average optimal representativity according to the missing value proportion in the database.

% of missing values	**optimal** *minRep*
10%	0.97
20%	0.9
30%	0.81
40%	0.74
50%	0.6
60%	0.48
70%	0.39
80%	0.22

Figure 4 also shows the SPoID algorithm performances depending on the missing value proportion in the database. A difference can be noted between the overall evolution of the ratio β/δ for databases containing less than 40% of missing values and the one containing 50% of incomplete records or more.
Thus Figure 5 gives the comparison of SPoID success rate with results obtained on a preprocessed incomplete database. This figure shows that some right patterns, discovered by SPoID are not found using a deletion preprocessing step. It can be noted that the ratio of right patterns strongly decreases between 40 and 50% of missing values: the number of wrong patterns becomes proportionnaly slightly higher compared to the number of right patterns.
It can also be noted that this ratio becomes less than 1 when the missing value percentage exceeds 50%. SPoID can then discover sequential patterns within incomplete database if at least half of the records in the dataset are complete, while the former methods requiring a preprocessing step do not find all the frequent patterns since 10% of missing values.

Lastly, the analysis of runtime performances shows that at constant incompleteness rate, runtime of SPoID is slightly constant while the *minRep* threshold decreases.The qualitative analysis of corresponding candidate and frequent sequences has shown that the number of candidate sequences increases but, as the minimum representativity is lower, the time spent for scanning the database decreases. We also noted that runtime increases with the incompleteness rate.

6 Further Work

Results detailed in these experiments show that the method SPoID is robust until around 40% of incompleteness. But these interesting results could be improved using another approach. Indeed, as we introduced it in section 2.3, some work were done for mining association rules using probabilities to assess missing values. These proposals [12, 13] cannot easily be adapted for mining sequences. However we are currently working on using a frequency distribution for modelizing incomplete data.
This frequency distribution is not a probability distribution and does not have the properties of a possibility distribution. We rather designed it as a fuzzy set, of which the membership function represents the level of certainty of a missing value to be for each possible value present in the database. Some experiments are currently carried out. First results are available in [4]. Next step will then consist in detecting the different kinds of incomplete information including the attributes that should not be considered as incomplete even if unfilled. Lastly we think about regarding noise in a further version of our algorithm, as it is a common imperfection in real-life databases.

7 Conclusion

Temporal databases available from many fields such as biological data or industrial process data most of the time contain imperfect data. More especially they may contain a lot of incomplete data. But the most adapted data mining technique to analyse such time-stamped datasets, i.e. sequential pattern discovery, cannot be easily applied to incomplete data. There is indeed no mining technique for discovering frequent sequences from incomplete databases. Therefore in this paper we proposed new definitions for sequential pattern mining in order to handle random incompleteness in data sequences. These new definitions enable the user to manage missing values directly during the mining task, then avoiding a heavy preprocessing step. Our method and algorithm SPoID has been implemented and tested on synthetic datasets. We have thus shown the robustness of our approach until an incompleteness rate around 40% , while the existing approaches give erroneous results since 10% of missing values. We now work on extending this approach in order to take into account other types of missing values such as data not randomly spread.

References

[1] R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216. Peter Buneman and Sushil Jajodia, 26–28 1993.

[2] R. Agrawal and R. Srikant. Mining sequential patterns. In *Eleventh International Conference on Data Engineering*, pages 3–14, Taipei, Taiwan, 1995. IEEE Computer Society Press.

[3] R.-S. Chen, G.-H. Tzeng, C.-C. Chen, and Y.-C. Hu. Discovery of Fuzzy Sequential Patterns for Fuzzy Partitions in Quantitative Attributes. In *ACS/IEEE Int. Conf. on Computer Systems and Applications*, pages 144–150, 2001.

[4] C. Fiot. Approximate sequential patterns for incomplete databases. Technical Report RR-07003, LIRMM, France, 02 2007.

[5] C. Fiot, A. Laurent, and M. Teisseire. Des motifs séquentiels généralisés aux contraintes de temps étendues. In *6èmes journées d'Extraction et Gestion des Connaissances*, pages 603–614, 2006.

[6] C. Fiot, A. Laurent, and M. Teisseire. From crispness to fuzziness: Three algorithms for soft sequential pattern mining. *IEEE Transactions on Fuzzy Systems*, to appear.

[7] T.P. Hong, K.Y. Lin, and S.L. Wang. Mining Fuzzy Sequential Patterns from Multiple-Items Transactions. In *Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf.*, pages 1317–1321, 2001.

[8] W. Z. Liu, A. P. White, S. G. Thompson, and M. A. Bramer. Techniques for dealing with missing values in classification. *Lecture Notes in Computer Science*, 1280, 1997.

[9] F. Masseglia, F. Cathala, and P. Poncelet. The PSP Approach for Mining Sequential Patterns. In *Principles of Data Mining and Knowledge Discovery*, pages 176–184, 1998.

[10] F. Masseglia, P. Poncelet, and M. Teisseire. Incremental mining of sequential patterns in large databases. Technical report, LIRMM, France, 01 2000.

[11] F. Masseglia, P. Poncelet, and M. Teisseire. Pre-Processing Time Constraints for Efficiently Mining Generalized Sequential Patterns. In *11th Int. Symp. on Temporal Representation and Reasoning (TIME '04)*, pages 87–95, 2004.

[12] J. Nayak and D. Cook. Approximate association rule mining. In *Florida Artificial Intelligence Research Symposium*, 2001.

[13] V. Ng and J. Lee. Quantitative association rules over incomplete data. In *IEEE International Conference*, pages 2821–2826, 1998.

[14] A. Ragel and B. Cremilleux. Treatment of missing values for association rules. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 258–270, 1998.

[15] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. In *5th International Conference on Extending Database Technology (EDBT '96)*, pages 3–17, London, UK, 1996. Springer-Verlag.

[16] H. Toivonen. Sampling large databases for association rules. In *VLDB '96: Proceedings of the 22th International Conference on Very Large Data Bases*, pages 134–145, 1996.

[17] M. J. Zaki, S. Parthasarathy, W. Li, and M. Ogihara. Evaluation of sampling for data mining of association rules. Technical report, Rochester, NY, USA, 1996.

Measuring Variation Strength in Gradual Dependencies

C. Molina, J.M. Serrano
Department of Informatics
University of Jaén, Spain
e-mail {carlosmo, jschica}@ujaen.es

D. Sánchez[1], **M.A. Vila**
Department of Computer Science and A.I
University of Granada, Spain
e-mail {daniel, vila}@decsai.ugr.es

Abstract

In this paper we extend a previous definition of gradual dependence as a special kind of (crisp) association rule, in order to measure not only the existence of a tendency, but its strength. The new proposal is based on the idea of fuzzy association rule and the definition of variation strength in the degree of fulfilment of an imprecise property by different objects. We study the new semantics and properties of the resulting fuzzy gradual dependence, and we propose a way to adapt existing fuzzy association rule mining algorithms for the new task of mining such dependencies.

1 Introduction

Gradual dependencies are rules that represent a relation among the variation in the degree of fulfilment of imprecise properties by different objects [10]. This kind of rules express a *tendency*. Consider for instance a database containing data about weight and speed of a set of trucks, and consider the restrictions *high* related to weight and *slow* related to speed, represented by means of suitable fuzzy sets on the domains of the attributes. An example of gradual dependence is *the higher the weight, the lower the speed*, meaning that as the weight of a truck increases, its speed tends to decrease.

The variations in the membership degree considered in gradual dependencies can be of two types: *the more* and *the less*, meaning that the membership degree of the first object to the considered fuzzy set is greater or lower than the membership of the second one, respectively. Hence we can consider four types of gradual dependencies: *the more X is A, the more Y is B* (expressed as $(>, X, A) \rightarrow (>, Y, B)$), *the more X is A, the less Y is B* (expressed as $(>, X, A) \rightarrow (<, Y, B)$), and so on.

In [10], the evaluation and representation of these gradual associations is based on linear regression analysis. The starting point of this approach is the idea of *contingency diagram*. Given two attributes X and Y, fuzzy sets A and B defined on X and Y, respectively, and a database $\mathcal{D}$ containing pairs of values $(x, y) \in X \times Y$, a contingency diagram is a two-dimensional plot of points $(A(x), B(y))$ such that $A(x) > 0$. A gradual dependence, represented as a *tendency rule* $A \rightharpoonup^t B$, means that " ... an increase in $A(x)$ comes along with an increase in $B(y)$". The validity of the rule is assessed on the basis of the regression coefficients $[\alpha, \beta]$ of the line that approximates the points in the contingency diagram (α being the slope of the line) and the quality of the regression as given by the R^2 coefficient.

In [4] we introduced an alternative approach, in which a gradual dependence is a rule of the form $(*_1, X, A) \rightarrow (*_2, Y, B)$, with $*_1, *_2 \in \{<, >\}$. The dependence holds in $\mathcal{D}$ iff $\forall (x, y), (x', y') \in \mathcal{D}$, $A(x) *_1 A(x')$ implies $B(y) *_2 B(y')$. The discovery of such dependencies is based on mining for association rules in a suitable set of transactions obtained from the database. For that purpose we define items of the form $[>, X, A]$ and $[<, X, A]$, expressing the two possible tendencies of attribute X with respect to the restriction A, and one transaction associated to every pair of objects. An item of the form $[<, X, A]$ (resp. $[<, X, A]$) is in the transaction associated to the pair of objects (o, o') (with values x and x' of X respectively) iff $A(x) < A(x')$ (resp. $A(x) > A(x')$). This way, a gradual dependence $(*_1, X, A) \rightarrow (*_2, Y, B)$ in a database $\mathcal{D}$ corresponds to an association rule of the form $[*_1, X, A] \Rightarrow [*_2, Y, B]$ in the corresponding set of transactions (one for each pair of objects in $\mathcal{D}$). For example, *the higher the weight, the lower the speed* can be expressed by the association rule $[>, Weight, High] \Rightarrow [>, Speed, Low]$. Support and

[1]Corresponding author

accuracy of the rule are employed in order to measure the importance and accuracy of the gradual dependence.

The latter has the advantage that algorithms to discover gradual rules can be obtained by a simple modification of any (crisp) association rule discovery algorithm. However, the semantics of both approaches are different since in [10] the relation between the magnitude of variation in both variables is taken into account, whilst in [4] only the fulfilment of the variation is considered.

In order to illustrate the difference, let us come back to our first example. Let us suppose we have three trucks whose fulfilment of the restrictions *high weight*, *slow speed*, and *big size* is shown in table 1. Let us assess the two gradual dependencies *the higher the weight, the lower the speed* and *the higher the weight, the bigger the size* using the approaches in [10] and [4]. Using [4], both dependencies hold with total accuracy since every time a truck is heavier than another, it is slower and bigger. For this approach, both dependencies hold to the same degree. However, if we look at the contingency diagrams for both dependencies (figure 1), it can be seen that the slope of the regression line for the dependence *the higher the weight, the lower the speed* (the parameters of the regression line are approximately $[0.167, 0.167]$) is smaller than for the dependence *the higher the weight, the bigger the size* (approximately $[1.3, -0.67]$). In both cases, clearly, the regression line fits perfectly the points in the contingency diagrams, so the quality of the regression is $R^2 = 1$. Hence the second dependence is stronger than the first one.

In this paper we propose an extension to the approach in [4] that incorporates the magnitude of variation in the degree of fulfilment of the restrictions in both variables, with the objective of detecting the strength of the dependence in cases like the example above. The new approach is based on the concept of fuzzy association rule, and it is related to previous work about the discovery of fuzzy approximate dependencies [2].

The paper is organized as follows: in section 2 we briefly recall a previous approach to gradual dependencies and we extend it by considering membership variation and fuzzy association rules. In section 3 we introduce the particular case of fuzzy gradual dependencies generated by the approach to fuzzy association rules in [6]. Section 4 is devoted to mining issues and to show some experiments. Finally, section 5 contains our conclusions and future research.

Truck	High weight	Slow speed	Big size
t_1	0.2	0.2	0.2
t_2	0.5	0.25	0.6
t_3	0.8	0.3	1

Table 1: Membership degrees of *high weight*, *slow speed*, and *big size* for three trucks

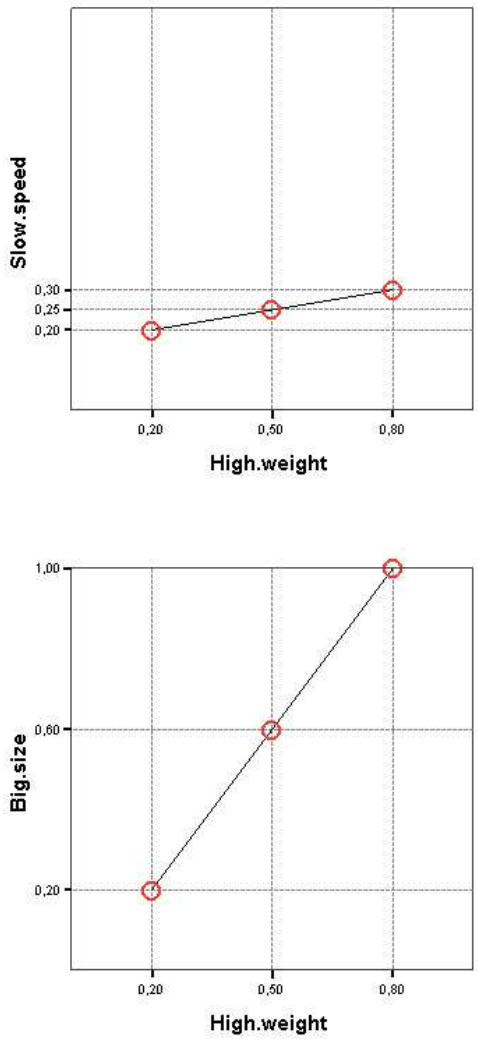


Figure 1: Contingency diagrams for the gradual dependencies *the higher the weight, the lower the speed* and *the higher the weight, the bigger the size* from the data in table 1

$$CF(I_1 \Rightarrow I_2) = \begin{cases} \frac{conf(I_1 \Rightarrow I_2) - supp(I_2)}{1 - supp(I_2)} & conf(I_1 \Rightarrow I_2) > supp(I_2) \\ \frac{conf(I_1 \Rightarrow I_2) - supp(I_2)}{supp(I_2)} & conf(I_1 \Rightarrow I_2) \leq supp(I_2) \end{cases} \quad (1)$$

2 Gradual dependencies with variation strength

In this section we extend our definition of gradual dependence [4] in order to incorporate variation strength in the assessment. First we briefly recall the definition in [4]. Then we define the concept of variation, and use it to extend the definition in [4] by using fuzzy association rules.

2.1 Our previous approach

In [4], a gradual dependence is defined as follows: let X and Y be two attributes, A and B fuzzy sets defined on the domains of X and Y, respectively, and a database $\mathcal{D}$ containing pairs of values $(x, y) \in X \times Y$. Let $*_1, *_2 \in \{<, >\}$. A gradual dependence of the form $(*_1, X, A) \rightarrow (*_2, Y, B)$ holds in $\mathcal{D}$ iff $\forall (x, y), (x', y') \in \mathcal{D}$, $A(x) *_1 A(x')$ implies $B(y) *_2 B(y')$.

This way, a gradual dependence is seen as a rule on a dataset consisting of pairs of objects of the original database. Hence, we use association rules in order to assess gradual dependencies in a database. As it is well known, given a set I of items and a bag T of transactions with $t \subseteq I \ \forall t \in T$, an association rule is an expression of the form $I_1 \Rightarrow I_2$ with $I_1, I_2 \subset I$, $I_1 \cap I_2 = \emptyset$ [1]. This rule is said to hold in T iff every transaction that contains I_1 contains also I_2. The usual measures are *support* and *confidence*, the former being the number or percentage of transactions containing $I_1 \cup I_2$, and the latter being the percentage of transactions containing I_1 that contain I_2. Many other measures have been proposed, see for example [5, 12, 3, 9]. In this paper we shall employ Shortliffe and Buchanan's *certainty factors*, as proposed in [3]. Let $supp(I_j)$ be the support of the itemset I_j and let $supp(I_1 \Rightarrow I_2) = supp(I_1 \cup I_2)$ be the support of the rule. Let $conf(I_1 \Rightarrow I_2) = supp(I_1 \Rightarrow I_2)/supp(I_1)$ be the confidence. The certainty factor of the rule, $CF(I_1 \Rightarrow I_2)$, is defined in equation 1.

The certainty factor yields a value in $[-1, 1]$ and measures how our belief that I_2 is in a transaction changes when we are told that I_1 is in that transaction. Positive values indicate our belief increases, negative values mean our belief decreases, and 0 means no change. Certainty factors have better properties than confidence, and help to solve some of its inconveniences. In particular, it helps to reduce the number of rules obtained by eliminating those rules that correspond in fact to statistical independence or negative dependence (up to 80 % in some of our experiments). This is shown, among other properties of certainty factors as accuracy measures for association rules, in [3]. Finally, let us remark that the calculation of the certainty factor in the final step of any association rule mining algorithm is straightforward and does not modify the time complexity of the algorithm, since support of the consequent and support and confidence of the rule are all available in this step.

We employ association rules in order to mine for gradual dependencies as follows: let $GI^{\mathcal{D}} = \{[>, X, A], [<, X, A], [>, Y, B], [<, Y, B]\}$ be a set of items and $GT^{\mathcal{D}}$ be a set of transactions containing items from $GI^{\mathcal{D}}$. $GT^{\mathcal{D}}$ is obtained from $\mathcal{D}$ as follows: $\forall\ o = (x, y), o' = (x', y') \in \mathcal{D}$ there is one transaction $gt_{oo'} \in GT^{\mathcal{D}}$ such that $[*, X, A] \in gt_{oo'}$ iff $A(x) * A(x')$ and $[*, Y, B] \in gt_{oo'}$ iff $B(y) * B(y')$, with $* \in \{<, >\}$. Let us remark that $GT^{\mathcal{D}}$ is a *crisp* set of transactions. Then, the gradual dependence $(*_1, X, A) \rightarrow (*_2, Y, B)$ holds in $\mathcal{D}$ iff the (crisp) association rule $[*_1, X, A] \Rightarrow [*_2, Y, B]$ holds in $GT^{\mathcal{D}}$. The support and confidence of the association rule $[*_1, X, A] \Rightarrow [*_2, Y, B]$ can be employed to assess the gradual dependence $(*_1, X, A) \rightarrow (*_2, Y, B)$. We usually employ support and certainty factor.

Let us remark that with this approach, the support of an item of the form $[*, X, A]$ is

$$supp([*, X, A]) = \frac{\left|\left\{gt_{oo'} \in GT^{\mathcal{D}} \mid A(x) * A(x')\right\}\right|}{|GT^{\mathcal{D}}|} \quad (2)$$

and hence the support of a dependence $(*_1, X, A) \rightarrow (*_2, Y, B)$ is the support of the itemset $\{[*_1, X, A], [*_2, Y, B]\}$, as equation 3 shows.

Some important and intuitive properties of this approach are the following: let c be an operator in $\{>, <\}$ such that $c(>) =<$ and $c(<) =>$. Then $supp(\{[*_1, X_1, A_1], \ldots, [*_k, X_k, A_k]\}) = supp(\{[c(*_1), X_1, A_1], \ldots, [c(*_k), X_k, A_k]\})$ (in particular $supp([*, X, A]) = supp([c(*), X, A])$). As a consequence, $supp((*_1, X, A) \rightarrow (*_2, Y, B)) = supp((c(*_1), X, A) \rightarrow (c(*_2), Y, B))$, and the same happens with confidence and certainty factor.

$$supp((*_1, X, A) \rightarrow (*_2, Y, B)) = \frac{\left|\left\{gt_{oo'} \in GT^{\mathcal{D}} \mid A(x) *_1 A(x') \wedge B(y) *_2 B(y')\right\}\right|}{|GT^{\mathcal{D}}|} \quad (3)$$

2.2 Membership variation

In the previous approach, only the fact that the membership degree is greater (or lesser) is taken into account. This way, the membership of the item $[*, X, A]$ in a transaction $gt_{oo'} \in GT^{\mathcal{D}}$ corresponding to a pair $o = (x, y), o' = (x', y') \in \mathcal{D}$ is defined as in equation 4.

$$gt_{oo'}([*, X, A]) = \begin{cases} 1 & A(x) * A(x') \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

With this definition, $A(x) = 0$ and $A(x') = 0.1$ yield the same result than $A(x) = 0$ and $A(x') = 1$. However, as we saw in the introduction, this can lead to obtain the same accuracy for dependencies that are intuitively different.

In order to avoid this problem, we propose to replace equation 4 by another expression that provides a degree in $[0, 1]$. We call this a *variation degree*. This way, $gt_{oo'}([*, X, A]) \in [0, 1]$.

There are different possibilities to obtain the degree $gt_{oo'}([*, X, A])$. In this paper we propose to employ that of equation 5:

$$gt_{oo'}([*, X, A]) = v_*(A(x), A(x')) \quad (5)$$

where

$$v_*(a, b) = \begin{cases} |a - b| & a * b \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

As an example, let $o = (0, y)$, $o' = (0.1, y')$, and $o'' = (1, y'')$. Then, $gt_{oo'}([<, X, A]) = 0.1$, $gt_{oo''}([<, X, A]) = 1$, $gt_{o'o''}([<, X, A]) = 0.9$, $gt_{o'o}([<, X, A]) = gt_{o''o}([<, X, A]) = gt_{o''o'}([<, X, A]) = 0$.

The following proposition holds:

Proposition 2.1 *Equation 5 verifies*

1. $gt_{oo'}([*, X, A]) \in [0, 1]$
2. *Suppouse* $A(x) * A(x')$ *and* $A(x) * A(x'')$. *Then* $|A(x) - A(x')| > |A(x) - A(x'')|$ *implies* $gt_{oo'}([*, X, A]) > gt_{oo''}([*, X, A])$
3. $gt_{oo'}([*, X, A]) = gt_{o'o}([c(*), X, A])$

Proof: Trivial. □

We consider that the properties in proposition 2.1 must be verified by the variation degree, despite the way it is calculated.

2.3 A new approach to gradual dependencies

Taking variation degrees into account, we propose a new definition of gradual dependence as a modification of our definition in [4], as follows:

Definition 2.1 *Let X and Y be two attributes, A and B fuzzy sets defined on the domains of X and Y, respectively, and a database $\mathcal{D}$ containing pairs of values $(x, y) \in X \times Y$. Let $*_1, *_2 \in \{<, >\}$. A gradual dependence of the form $(*_1, X, A) \rightarrow (*_2, Y, B)$ holds in $\mathcal{D}$ iff $\forall o, o' \in \mathcal{D}$ with $o = (x, y)$ and $o' = (x', y')$, $v_{*_1}(A(x), A(x'))$ implies $v_{*_2}(B(y), B(y'))$.*

where v_* is that of equation 6. Let us remark that the implication that appears in this definition is a *fuzzy implication*. This has two main consequences: first, there are in fact different definitions of gradual dependence, depending on the implication considered. Second, a gradual dependence holds to a certain degree. Hence, we are working in fact with *fuzzy gradual dependencies*.

Now, we can extend our interpretation of gradual dependencies as association rules in [4] in order to consider the variation degree of items. A natural way to extend our first approach is to consider fuzzy association rules. There are many different approaches to the definition and assessment of fuzzy association rules. In general, the different extensions take as starting point, in one way or another, a generalization of transactions to fuzzy transactions as fuzzy subsets of items. The main difference between the different existing approaches is the way they assess the rules (see among others [11, 6, 13, 8]).

Using fuzzy association rules is natural in our case since each item has a membership degree to each transaction, so we have in fact a set of *fuzzy transactions*, i.e., fuzzy subsets of items. However, let us remark that since there is no a single definition of fuzzy gradual dependence, the approach employed for mining the fuzzy rules will define, in practice, a particular type of fuzzy gradual dependence.

Let $GI^{\mathcal{D}} = \{[>, X, A], [<, X, A], [>, Y, B], [<, Y, B]\}$ be a set of items and $\tilde{GT}^{\mathcal{D}}$ be a set of fuzzy transactions containing items from $GI^{\mathcal{D}}$. $\tilde{GT}^{\mathcal{D}}$ is obtained from $\mathcal{D}$ as follows: $\forall\ o = (x, y), o' = (x', y') \in \mathcal{D}$ there is one fuzzy transaction $gt_{oo'} \in \tilde{GT}^{\mathcal{D}}$ such that $gt_{oo'}([*, X, A]) = v_*(A(x), A(x'))$ and $gt_{oo'}([*, Y, B]) = v_*(B(y), B(y'))$, with $* \in \{<, >\}$.

Since a fuzzy association rule defines a special kind of fuzzy implication between the degrees of antecedent and consequent, we can conclude the following:

Proposition 2.2 *A fuzzy association rule* $[*_1, X, A] \Rightarrow [*_2, Y, B]$ *in* $\tilde{G}T^{\mathcal{D}}$ *defines a fuzzy gradual dependence* $(*_1, X, A) \rightarrow (*_2, Y, B)$ *in* $\mathcal{D}$.

i.e., fuzzy association rules in $\tilde{G}T^{\mathcal{D}}$ define some particular types of fuzzy gradual dependencies in $\mathcal{D}$.

Following proposition 2.2, the support and confidence (or other accuracy measures) of the fuzzy association rule $[*_1, X, A] \Rightarrow [*_2, Y, B]$ can be employed to assess a particular type of fuzzy gradual dependence $(*_1, X, A) \rightarrow (*_2, Y, B)$.

3 A particular definition of fuzzy gradual dependence

As we have seen, there are many possible ways to define fuzzy gradual dependencies, in particular starting from an specific approach to fuzzy association rules. In this paper we shall employ the approach to fuzzy association rules introduced in [6] to obtain a particular definition of fuzzy gradual dependence.

3.1 Our approach to fuzzy association rules

In [6], fuzzy association rules are defined and assessed as follows: let $I = \{i_1, \ldots, i_m\}$ be a set of items and $\tilde{T}$ be a set of fuzzy transactions, where each fuzzy transaction is a fuzzy subset of I. For every fuzzy transaction $\tilde{\tau} \in \tilde{T}$ we note $\tilde{\tau}(i_k)$ the membership degree of i_k in $\tilde{\tau}$. For an itemset I_0 we note $\tilde{\tau}(I_0) = \min_{i_k \in I_0} \tilde{\tau}(i_k)$ the degree to which I_0 is in a transaction $\tilde{\tau}$. A fuzzy association rule is an implication of the form $I_1 \Rightarrow I_2$ such that $I_1, I_2 \subset I$ and $I_1 \cap I_2 = \emptyset$. Notice that this is the same definition of a crisp association rule since, from the structural point of view, there is no difference. The difference is that for fuzzy rules the starting point is a set of fuzzy transactions, and the problem is how to assess the support and accuracy. Strictly speaking, what we call fuzzy association rules are association rules assessed on fuzzy transactions.

We call *representation* of the item i_k, noted $\tilde{\Gamma}_{i_k}$, to the (fuzzy) set of transactions where i_k appears, defined as in equation 7. This representation can be extended to itemsets as in equation 8.

$$\tilde{\Gamma}_{i_k}(\tilde{\tau}) = \tilde{\tau}(i_k) \tag{7}$$

$$\tilde{\Gamma}_{I_0}(\tilde{\tau}) = \min_{i_k \in I_0} \tilde{\Gamma}_{i_k}(\tilde{\tau}) = \min_{i_k \in I_0} \tilde{\tau}(i_k) = \tilde{\tau}(I_0) \tag{8}$$

In order to measure the interest and accuracy of a fuzzy association rule, we employ a semantic approach based on the evaluation of quantified sentences, using the fuzzy quantifier $Q_M(x) = x$, as follows:

- The support of an itemset I_0 is the evaluation of the quantified sentence Q_M of $\tilde{T}$ are $\tilde{\Gamma}_{I_0}$.
- The support of the fuzzy association rule $I_1 \Rightarrow I_2$ in $\tilde{T}$, $Supp(I_1 \Rightarrow I_2)$, is the evaluation of the quantified sentence Q_M of T are $\tilde{\Gamma}_{I_1 \cup I_2} = Q$ of T are $(\tilde{\Gamma}_{I_1} \cap \tilde{\Gamma}_{I_2})$.
- The confidence of the fuzzy association rule $I_1 \Rightarrow I_2$ in $\tilde{T}$, $Conf(I_1 \Rightarrow I_2)$, is the evaluation of the quantified sentence Q of $\tilde{\Gamma}_{I_1}$ are $\tilde{\Gamma}_{I_2}$.
- The certainty factor is obtained from support and confidence using equation 1.

We evaluate a quantified sentence of the form Q of F are G by means of method GD, defined in [7] as

$$GD_Q(G/F) = \sum_{\alpha_i \in \Lambda(G/F)} (\alpha_i - \alpha_{i+1}) Q\left(\frac{|(G \cap F)_{\alpha_i}|}{|F_{\alpha_i}|}\right) \tag{9}$$

where $\triangle(G/F) = \Lambda(G \cap F) \cup \Lambda(F)$, $\Lambda(F)$ being the level set of F, and $\Lambda(G/F) = \{\alpha_1, ..., \alpha_p\}$ with $\alpha_i > \alpha_{i+1}$ for every $i \in \{1, ..., p-1\}$, and considering $\alpha_{p+1} = 0$. The set F is assumed to be normalized. If not, F is normalized and the same normalization factor is applied to $G \cap F$.

It is possible to employ different fuzzy quantifiers, provided they verify certain properties [6]. We employ the quantifier Q_M since the resulting approach is a generalization of the ordinary association rule assessment framework in the crisp case (i.e., if the set of transactions is crisp, the measures described above yield the ordinary measures for support, confidence, and certainty factor). This is true only for Q_M. Other important properties defining the semantics of this proposal are those of equations 10 and 11.

$$Conf(I_1 \Rightarrow I_2) = 1 \text{ iff } \tilde{\tau}(I_1) \leq \tilde{\tau}(I_2) \quad \forall \tilde{\tau} \in \tilde{T} \tag{10}$$

$$CF(I_1 \Rightarrow I_2) = 1 \text{ iff } Conf(I_1 \Rightarrow I_2) = 1 \tag{11}$$

3.2 Fuzzy gradual dependence

Following the approach in the previous section, and using the quantifier $Q_M(x) = x$, a fuzzy gradual dependence $(*_1, X, A) \rightarrow (*_2, Y, B)$ in $\mathcal{D}$ is a fuzzy association rule $[*_1, X, A] \Rightarrow [*_2, Y, B]$ in $\tilde{G}T^{\mathcal{D}}$ that holds

$$supp([*_1, X, A] \Rightarrow [*_2, Y, B]) = \sum_{\alpha_i \in \Lambda\left(\left(\tilde{\Gamma}_{[*_1,X,A]} \cap \tilde{\Gamma}_{[*_2,Y,B]}\right)/\tilde{GT}^{\mathcal{D}}\right)} (\alpha_i - \alpha_{i+1}) \left(\frac{\left| \left(\tilde{\Gamma}_{[*_1,X,A]} \cap \tilde{\Gamma}_{[*_2,Y,B]}\right)_{\alpha_i} \right|}{|\tilde{GT}^{\mathcal{D}}_{\alpha_i}|} \right) \quad (12)$$

$$conf([*_1, X, A] \Rightarrow [*_2, Y, B]) = \sum_{\alpha_i \in \Lambda\left(\tilde{\Gamma}_{[*_1,X,A]}/\tilde{\Gamma}_{[*_2,Y,B]}\right)} (\alpha_i - \alpha_{i+1}) \left(\frac{\left| \left(\tilde{\Gamma}_{[*_1,X,A]} \cap \tilde{\Gamma}_{[*_2,Y,B]}\right)_{\alpha_i} \right|}{\left| \left(\tilde{\Gamma}_{[*_1,X,A]}\right)_{\alpha_i} \right|} \right) \quad (13)$$

with support and confidence given by equations 12 and 13, where $\tilde{\Gamma}_{[*_1,X,A]}$ is a fuzzy subset of transactions such that $\tilde{\Gamma}_{[*_1,X,A]}(gt_{oo'}) = gt_{oo'}([*_1, X, A])$ (similar for $\tilde{\Gamma}_{[*_2,Y,B]}$) and the α_i correspond to the union of the level sets of the fuzzy sets involved, arranged in decreasing order (in equation 13, $\tilde{\Gamma}_{[*_1,X,A]}$ must be normalized, otherwise we should normalize it first and apply the same factor to the intersection $\tilde{\Gamma}_{[*_1,X,A]} \cap \tilde{\Gamma}_{[*_2,Y,B]}$). The certainty factor is obtained as in equation 1.

The following properties from the approach in [4] keep holding:

Proposition 3.1 *Let c be an operator in $\{>, <\}$ such that $c(>) =<$ and $c(<) =>$. Then, $supp([*, X, A]) = supp([c(*), X, A])$.*

Proof: By proposition 2.1, $gt_{oo'}([*, X, A]) = gt_{o'o}([c(*), X, A])$ for every pair $o = (x, y)$, $o' = (x', y')$. Hence,

$$\Lambda\left(\tilde{\Gamma}_{[*,X,A]}/\tilde{GT}^{\mathcal{D}}\right) = \Lambda\left(\tilde{\Gamma}_{[c(*),X,A]}/\tilde{GT}^{\mathcal{D}}\right)$$

and $\forall \alpha_i$

$$\left| \left(\tilde{\Gamma}_{[*,X,A]}\right)_{\alpha_i} \right| = \left| \left(\tilde{\Gamma}_{[c(*),X,A]}\right)_{\alpha_i} \right|$$

so $supp([*, X, A]) = supp([c(*), X, A])$. □

Proposition 3.2 *The generalization to itemsets hold as well, so $supp(\{[*_1, X_1, A_1], \ldots, [*_k, X_k, A_k]\}) = supp(\{[c(*_1), X_1, A_1], \ldots, [c(*_k), X_k, A_k]\})$*

Proof: Same as proposition 3.1. □

Corollary 3.1 *It follows that:*

$supp([*_1, X, A] \Rightarrow [*_2, Y, B]) = supp([c(*_1), X, A] \Rightarrow [c(*_2), Y, B])$,

$conf([*_1, X, A] \Rightarrow [*_2, Y, B]) = conf([c(*_1), X, A] \Rightarrow [c(*_2), Y, B])$,

$CF([*_1, X, A] \Rightarrow [*_2, Y, B]) = CF([c(*_1), X, A] \Rightarrow [c(*_2), Y, B])$.

This last corollary implies that in order to assess all the possible gradual dependencies involving only items of the form $[*_1, X, A]$ and $[*_2, Y, B]$ it is enough to measure support and accuracy for $[<, X, A] \Rightarrow [<, Y, B]$ and $[<, X, A] \Rightarrow [>, Y, B]$.

The following propositions allow us to provide an interpretation of the semantics of our fuzzy gradual dependence and some relation to the approach in [10]:

Proposition 3.3 *$conf([*_1, X, A] \Rightarrow [*_2, Y, B]) = 1$ iff $v_{*_1}(A(x), A(x')) \leq v_{*_2}(B(y), B(y'))$ $\forall o, o' \in \mathcal{D}$*

Proof: $v_{*_1}(A(x), A(x')) \leq v_{*_2}(B(y), B(y'))$ $\forall o, o' \in \mathcal{D}$ iff $gt_{oo'}([*_1, X, A]) \leq gt_{oo'}([*_2, Y, B])$ $\forall gt_{oo'} \tilde{GT}^{\mathcal{D}}$. By equation 10, this is true iff $conf([*_1, X, A] \Rightarrow [*_2, Y, B]) = 1$. □

Proposition 3.4 *$CF([*_1, X, A] \Rightarrow [*_2, Y, B]) = 1$ iff $v_{*_1}(A(x), A(x')) \leq v_{*_2}(B(y), B(y'))$ $\forall o, o' \in \mathcal{D}$*

Proof: Immediate by proposition 3.3 and equation 11. □

Proposition 3.5 *If $CF([*_1, X, A] \Rightarrow [*_2, Y, B]) = 1$ then $A \rightharpoonup^t B[\alpha, \beta]$ holds with $|\alpha| \geq 1$.*

Proof: Let us consider first the dependence $[<, X, A] \Rightarrow [<, Y, B]$. If $CF([<, X, A] \Rightarrow [<, Y, B]) = 1$ then by proposition 3.4, $v_<(A(x), A(x')) \leq v_<(B(y), B(y'))$ $\forall o, o' \in \mathcal{D}$. As a consequence, $A(x) < A(x')$ implies $A(x') - A(x) \leq B(y') - B(y)$, i.e., $(B(y') - B(y))/(A(x') - A(x)) \geq 1$. Therefore, the slope of all the lines linking pairs of points in the contingency diagram is greater or equal than 1 (no points with membership 0 are considered in the diagram). Hence, the slope of the regression line for all the points is greater or equal than 1.

The proof is similar for the rule $[<, X, A] \Rightarrow [>, Y, B]$, but yielding $(B(y') - B(y))/(A(x') - A(x)) \leq -1$ and hence a slope for the regression line less or equal than

Rule	supp	conf	CF	α	β	R^2
$(>, Weight, High) \rightarrow (>, Speed, Low)$	0.33	0,1	0,06	0.167	0.167	1
$(>, Weight, High) \rightarrow (>, Size, Big)$	0.2	1	1	1.3	-0.67	1

Table 2: Assessment of the gradual dependencies of the example in the introduction using our new approach and that in [10] (values are approximate)

-1. Hence, we have covered all the possibilities and $|\alpha| \geq 1$. □

It is easy to show that the reciprocal of proposition 3.5 holds when $R^2 = 1$, but cannot be guaranteed otherwise.

In order to illustrate these results, let us come back to the example in the introduction. The assessment of the rules (approximate values) is shown in table 2. As expected, the new approach takes into account the variation membership and, instead of yielding two dependencies with confidence and certainty factor equal to one, only in the second case this happens. In fact, the first one has a very low accuracy. Let us remark also that for the second dependence, confidence and certainty factor are one and, at the same time, the slope of the corresponding regression line is greater than 1 (as expected since $R^2 = 1$).

4 Mining gradual dependencies

4.1 Algorithm

In general, the problem we face is that of mining gradual dependencies as association rules in a database $\mathcal{D}$ containing a description of a set of objects in terms of a set of attributes $\{X_1, \ldots, X_m\}$. For each attribute X_i we have a set of n_i fuzzy restrictions defined by fuzzy sets $\{A_{i1}, \ldots, A_{in_i}\}$. We consider a set of items $GI^{\mathcal{D}} = \{[*, X_i, A_{ij}]\}$ with $* \in \{<, >\}$, $i \in \{1, \ldots, m\}$, and $j \in \{1, \ldots, n_i\}$. We shall also consider a bag of fuzzy transactions $\tilde{GT}^{\mathcal{D}}$ containing items of $GI^{\mathcal{D}}$, and obtained from $\mathcal{D}$ as explained in previous sections. Finally, we impose an usual restriction on the rules: no pair of items appearing in the left or right part of a rule can share the same attribute.

A first approach to solve the problem of mining gradual dependencies would be simply to build the set $\tilde{GT}^{\mathcal{D}}$ of transactions and to apply any of the existing algorithms for mining fuzzy association rules. As it is well known, most of the existing algorithms work in two steps: the first one (the most computationally expensive) is to discover the frequent itemsets, i.e., those with support above a minimum user-defined threshold. In the second one, and starting from the frequent itemsets, those rules with enough accuracy are obtained.

The complexity of the second step is not modified as it depends on the number of frequent itemsets, and is not affected by the calculation of the certainty factor. However, the main inconvenience of this approach in our problem is the complexity of discovering the frequent itemsets with respect to the number of objects: while finding frequent itemsets in $\mathcal{D}$ has a complexity $O(n)$ in the number of objects (multiplied by another factors related to number of items and other, depending on the algorithm), finding frequent itemsets in $\tilde{GT}^{\mathcal{D}}$ has a complexity $O(n^2)$.

This problem can be solved to an extent by considering a fixed number of equidistributed levels (degrees) in the definition of the fuzzy sets. In [4] we proposed a solution for the approach presented in that paper (using crisp association rules). With that solution, the complexity of finding the support of itemsets of size p is $n + k^p$. The extent to which this solution is good depends on the relation between n and k^p.

We have developed algorithms based on similar principles for the discovery of fuzzy association rules [6] and fuzzy approximate dependencies [2]. At this moment we are working in an algorithm whose complexity will be $n + k^{p+1}$ for itemsets of size p. We shall describe it in a future paper.

4.2 Experiments

In [4] we performed a small experiment on a real database with data about soils and weather in Granada (southeast of Spain). The data collected in a set of farms includes among others attributes about average temperature, raining and altitude, ph, and percentages of clay and sand in the soil. Fuzzy sets *High*, *Medium* and *Low* have been defined on the domain of each attribute. Our intention is not to discuss the dependencies obtained, but to show the variation between the crisp and fuzzy approaches in a real dataset.

Table 3 shows a set of gradual dependencies obtained that were found interesting by experts, with their support and certainty factor, and the variation in support and certainty factor when using fuzzy rules instead of crisp ones. As expected, in all the cases, support and certainty factor diminished. It is remarkable the case of dependence 5, that is detected using the crisp approach but not by the fuzzy one.

		Crisp		Fuzzy	
#	**Dependence**	**supp**	**CF**	**supp**	**CF**
1	$(>, ATEMP, High) \rightarrow (>, ALT, Low)$	0,14	0,93	0,015	0,53
2	$(>, ATEMP, Low) \rightarrow (>, ALT, High)$	0,13	0,93	0.087	0,9
3	$(>, ARAIN, Medium) \rightarrow (>, ATEMP, Medium)$	0,24	0,9	0,1	0,68
4	$(>, ARAIN, Low) \rightarrow (>, ALT, Low)$	0,3	0,85	0,18	0.65
5	$(>, ATEMP, Low) \rightarrow (>, ARAIN, Low)$	0,12	0,84	-	-
6	$(>, CLAY, High) \rightarrow (>, SAND, Low)$	0,09	0,84	0,04	0,64

Table 3: Some gradual dependencies obtained from a real database about soil characteristics and weather in olive cultivation farms, using crisp and fuzzy association rules.

5 Conclusions

We have extended our definition of gradual dependence in [4] in order to incorporate variation strength. For that purpose we have introduced the new notion of degree of variation associated to a pair of objets. We have provided a definition of gradual dependence on the basis of fuzzy association rules over a set of fuzzy transactions obtained from the original dataset by using the degree of variation. We have shown that the new approach is better in capturing the variation strength in gradual dependencies, and we have shown some properties that explain the semantics of the new approach, as well as some results that relate the new approach to the approaches in [10] and [4].

Several research avenues remain open. First, we want to investigate the semantics of fuzzy gradual dependencies obtained by using other approaches to fuzzy association rules, like the measures introduced in [8]. Second, we are working in an algorithm able to reduce the complexity of the mining process when employing existing algorithms for mining fuzzy association rules. Finally, we will apply our techniques to mine for fuzzy gradual dependencies in real databases.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. Of the 1993 ACM SIGMOD Conference*, pages 207–216, 1993.

[2] F. Berzal, I. Blanco, D. Sánchez, J.M. Serrano, and M.A. Vila. A definition for fuzzy approximate dependencies. *Fuzzy Sets and Systems*, 149(1):105–129, 2005.

[3] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila. Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis*, 6:221–235, 2002.

[4] F. Berzal, J.C. Cubero, D. Sánchez, J.M. Serrano, and M.A. Vila. An alternative approach to discover gradual dependencies. *Int. Journal of Uncertainty, Fuzziness and Knowledge-based Systems*. Submitted.

[5] S. Brin, R. Motwani, J.D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. *SIGMOD Record*, 26(2):255–264, 1997.

[6] M. Delgado, N. Marín, D. Sánchez, and M.A. Vila. Fuzzy association rules: General model and applications. *IEEE Transactions on Fuzzy Systems*, 11(2):214–225, 2003.

[7] M. Delgado, D. Sánchez, and M.A. Vila. Fuzzy cardinality based evaluation of quantified sentences. *International Journal of Approximate Reasoning*, 23:23–66, 2000.

[8] D. Dubois, E. Hüllermeier, and H. Prade. A systematic approach to the assessment of fuzzy association rules. *Data Mining and Knowledge Discovery*, 13(2):167–192, 2006.

[9] M. Hahsler and K. Hornik. New probabilistic interest measures for association rules. Technical report, Department of Statistics and Mathematics, Vienna University of Economics and Business Administration, 2006.

[10] E. Hüllermeier. Association rules for expressing gradual dependencies. In *Proceedings PKDD 2002 Lecture Notes in Computer Science 2431*, pages 200–211. 2002.

[11] Chan-Man Kuok, Ada Fu, and Man Hon Wong. Mining fuzzy association rules in databases. *SIGMOD Record*, 27(1):41–46, 1998.

[12] C. Silverstein, S. Brin, and R. Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data Mining and Knowledge Discovery*, 2:39–68, 1998.

[13] T. Sudkamp. Examples, counterexamples, and measuring fuzzy associations. *Fuzzy Sets and Systems*, 149(1):57–71, 2005.

Forest of Fuzzy Decision Trees and their Application in Video Mining

Marcin Detyniecki
Université Pierre et Marie Curie-Paris6
CNRS UMR 7606, LIP6,
104 av. du Président Kennedy
Paris, F-75016, France
Marcin.Detyniecki@lip6.fr

Christophe Marsala
Université Pierre et Marie Curie-Paris6
CNRS UMR 7606, LIP6,
104 av. du Président Kennedy
Paris, F-75016, France
Christophe.Marsala@lip6.fr

Abstract

One of the great challenges today is to index videos with high-level semantic concepts or features. The basis of our approach is to use a fuzzy decision trees (FDT) to construct the heart of the system in order to reduce the need of human usage in the process of indexation. But when we address large, unbalanced, multiclass data sets, a single classifier - such as the FDT - is insufficient. Therefore we study the use of forests of fuzzy decision trees (FFDT): (a) its effectiveness for a high level feature detection task and (b) the effect on performance from number of classifiers point of view.

Keywords: Video mining, High level Features, Fuzzy Decisions Trees.

1 Introduction

The growth of multimedia data and in particular of video data has caused a corresponding growth in the need to analyze and to exploit them. One of the great challenges today is to be able to index these data with high-level semantic concepts (or features) such as "indoor/outdoor", "people", "maps", "military personnel", etc. Today, we are able to calculate accurately low level features as for instance colour or texture histograms. Unfortunately, between these two levels there is an unbridgeable gap.

One solution to bridge the semantic gap is, based on a set of examples, to learn or to extract a general rule that will allow classifying new examples. This type of approach is known as inductive reasoning. Inductive machine learning is a well-known research topic with a large set of methods, the most common being the decision trees (DT). However, robustness and threshold problems appear when considering classical DTs. The introduction of fuzzy set theory in a learning method enables us to smooth out these negative effects. Thus, at the basis of our approach we use the Fuzzy Decision Trees (FDT).

Since 2001, each year, the National Institute of Standards and Technology (NIST) organizes the TREC Video Retrieval Evaluation. We will use this framework as basis for our comparison. Although decision trees are widely used as core technology in application using machine learning techniques, no other team participating to the challenge uses them. In fact, simple trees are not as effective when dealing with large unbalanced multiclass data sets [9]. Here, we not only propose to use them, but also to go farther by combining them in a boosting [5] manner, thus obtaining forests of fuzzy decision trees (FFDT).

In this paper, we describe the FFDT algorithm and we study (a) its effectiveness for a high level feature detection task and (b) the effect on performance from size point of view. Since today's approaches are evaluated based on the average precision, measure based on a ranking, we shortly discuss the fundamental difference between classification – what most of the learning algorithm do – and ranking optimization.

The paper is structured as follows: in Section 3, we describe our video processing method, which provides as result low level features. In Section 4 and Section 5 we present Fuzzy Decision Trees and Forests. Finally in Section 6 we detail the high level feature detection experiments and discuss the results.

2 Overview

To use the Fuzzy Decision Trees (FDT) learning method, a training set must be provided in which there are examples of keyframes *with* the high-level feature to be recognized and examples that do *not* possess that feature. For the sake of simplicity, presence of the feature is the *class descriptor*.

The general schema of the application is illustrated in

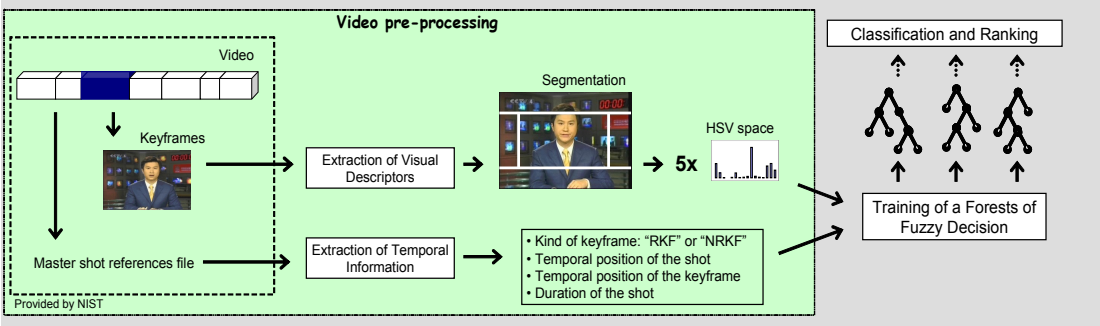


Figure 1: General schema

Figure 1. Our approach is decomposed in three main stages. First, the video is pre-processed in order to obtain a set of descriptors to feed the machine learning algorithm. Afterwards, in the training step, the system learns how to recognize the presence of a high-level feature in a shot using manually indexed videos (the so-called *development data* set). Finally, in the classification step, the system is able to recognize the presence of a high-level feature in a shot for any additional video.

The training stage is composed of several steps (see Figure 3):

Step 1 the video is segmented into temporal shots; each shot is associated with a representative keyframe;

Step 2 a set of descriptors is extracted from each keyframe of the whole set of keyframes; keyframes from the development data set are also associated with some high-level features (classes) obtained by means of the manual indexation of the videos;

Step 3 several training sets of keyframe descriptors (and the associated class) are built;

Step 4 each of these training sets is used to construct a fuzzy decision tree (FDT) and thus obtaining a forest.

The classification stage is decomposed into the following steps (see Figure 4):

Step 1 each shot to be classified is associated with a representative keyframe;

Step 2 a set of descriptors is extracted from that keyframe;

Step 3 the set is classified by means of each FDT of the forest;

Step 4 the presence of a high level feature is valued by aggregating the classification results given by all the trees of the forest.

3 Video pre-processing

In order to feed the data mining algorithm, the video (MPEG file) has to be pre-processed. A set of numerical descriptors, on which the algorithm will learn, has to be defined to characterize the video.

3.1 Video segmentation

First, the video is segmented into a set of shots. We used the results of the segmentation of a video into shots provided by [12] for the TRECVID challenge. Each video of the corpus is described by:

- a keyframe for each detected shot.
- an XML file that provides the time codes of the detected shots, their duration, and the time codes of the keyframes.

To reduce the number of the shots, all shots shorter than 2 seconds are automatically merged with their neighbours to provide a unique shot. A representative keyframe (RKF) of this new shot is selected among the extracted keyframes. The other keyframes are kept and are called non representative keyframe (NRKF). Thus, at the end, each shot is always associated with a unique RKF, but can also be associated with several NRKF.

3.2 Visual Information Descriptors

Visual Information Descriptors are at the basis of the learning process. They are obtained directly and exclusively from the keyframes.

In order to obtain spatial-related information, the image is segmented into 5 regions (see Figure 2) in order to isolate important descriptors, thus helping the learning algorithm to focus on the discriminative variables.

Each of the regions corresponds to a spatial part of the keyframe: top, bottom, left, right, and middle. The five regions are not of the same size, thus reflecting a visual importance of the contained information based on its position.

Afterwards, for each region, the associated histogram in the HSV space is computed. Based on the importance of the region, the histogram is computed in a more or less precise way, by varying the number of bins: 6x3x3 for Middle, and Bottom, 4x3x3 for Right, and 4x2x2 for Left, and Top.

At the end of this procedure, the visual information descriptors are a set of numerical values (belonging to $[0, 1]$) that characterizes every keyframe.

The choice of the number of regions and the number of bins for the histogram deserve further optimization. Moreover, more complementary visual descriptors could be added in order to enhance the possibilities of choice of the learning algorithm (FDT) for its decisions.



Figure 2: Spatial segmentation of a Keyframe

3.3 Temporal Information Descriptors

The *temporal information descriptors* are extracted from the shot detection process [12] and are linked to the sequence of frames in the video. Every keyframe is associated with the following information:

- the kind of the keyframe: representative (RKF) or non representative (NRKF);
- the temporal position of the shot containing the keyframe: time code of the beginning;
- the temporal position of the keyframe in the shot: relative time code since the beginning of the shot;
- the duration of the shot containing the keyframe.

3.4 Class Descriptor

The *class descriptor* is obtained from the manual indexation of the video. It corresponds to the feature(s) to be detected in a shot, as for instance sports, weather, desert, corporate leader, US flag, chart, etc.

The class descriptor is extracted from the file obtained from the collaborative work of indexation of the development video set. Note that a keyframe can be associated with more than one class descriptor depending on the result of the indexation process. We just choose the first class descriptor available, but clearly more refined selection should be done at this stage. Further research will address this point.

4 Fuzzy Decision Trees

Classical decision tree algorithms [3, 13] are one of the inductive learning algorithms that are most intensively used in data mining. Unfortunately they encounter technical problems when dealing with numerical attributes. That leads to the introduction of the fuzzy decision tree construction algorithms enabling the use of fuzzy values in the decision tree [10]. "Fuzziness" allows the decisions to be smoother, avoiding sharp thresholds. It also enables to have degrees of decision and of membership to a certain class.

Inductive learning rises from the *particular* to the *general*. A tree is built from its root to its leaves, by successive partitioning the training set into subsets. Each partition is done by means of a test on an attribute, which leads to the definition of a node of the tree.

Let us assume that a set of classes $C = \{c_1, ..., c_K\}$, representing a physical or a conceptual phenomenon, is considered. And that this phenomenon is described by means of a set of attributes $\mathcal{A} = \{A_1, ..., A_N\}$. In that case, a *description* is a N-tuple of attribute-value pairs (A_j, v_{jl}). Each description is linked with a particular class c_k from C to make up an *instance* (or *example*, or *case*) e_i of the phenomenon. Finally, the inductive learning is the process that generalizes from a *training set* $\mathcal{E} = \{e_1, ..., e_n\}$ of examples to a general law to bring out relations between descriptions and classes in C. In our case, each attribute A_j can take a fuzzy, numerical, or symbolic value v_{jl} in the set $\{v_{j1}, ..., v_{jm_j}\}$ of all possible values. We suppose that v_{jl} is associated with a membership function $\mu_{v_{jl}}$. Similarly, each c_k is supposed to be associated with a membership function μ_{c_k}.

4.1 Attribute Selection

Most algorithms designed for constructing decision trees proceed in the same way: the so-called *Top Down Induction of Decision Tree* (TDIDT) method. They build a tree from the root to the leaves, by successive partitioning the training set into subsets. Each partition is done by means of a test on one attribute and leads to the definition of a node of the tree. The attribute is selected by means of a *measure of discrimination* H. Such a measure enables us to order

the attributes according to increasing discrimination-accuracy when splitting the training set. The discrimination power of each attribute in $\mathcal{A}$ is valued with regard to the classes. The attribute with the highest discriminating power is selected to construct a node.

4.2 Construction of Fuzzy Partitions

The process of construction of FDT is based on the fuzzy partition for each numerical attribute. However, it is rare to know, a priori, such a fuzzy partition. Thus an automatic method of construction such a partition from a set of precise values was implemented. In this way we obtain a set of fuzzy values for each numerical attribute.

The algorithm [8] is based on the utilization of the mathematical morphology theory. Kernels of concordant values of a numerical attribute related to the values of the class can be found. Fuzzy values induced from a set of numerical values of an attribute are linked with the repartition of the values of the class related to the numerical attribute. Thus a contextual partitioning of an attribute is performed, enabling us to obtain the best partition related to that attribute with respect to the class.

4.3 Classification using Fuzzy Decision Tree

It is well-known that the path from the root to a leaf in a decision tree is equivalent to a production rule [7]. The premises for such a rule r are tests on attributes values, and the conclusion is the value of the class that labels the leaf of the path:

$$\texttt{if } A_{l_1} = v_{l_1} \texttt{ and } \ldots \texttt{and } A_{l_p} = v_{l_p} \texttt{ then } C = c_k$$

In a FDT, a leaf can be labelled by a set of values $\{c_1, ..., c_K\}$ for the class, each value c_j associated with a weight computed during the learning phase. Thus, a path of a fuzzy decision tree is equivalent to the following rule:

$$\begin{array}{c}\texttt{if } A_{l_1} = v_{l_1} \texttt{ and } \ldots \texttt{and } A_{l_p} = v_{l_p} \texttt{ then} \\ C = c_1 \texttt{ with the degree } P^*(c_1|(v_{l_1}, v_{l_2}, ...v_{l_p})) \\ \texttt{and } \ldots \texttt{ and } C = c_K \texttt{ with the degree} \\ P^*(c_K|(v_{l_1}, v_{l_2}, ...v_{l_p}))\end{array}$$

In a FDT, each value v_i can be either precise or fuzzy, and is described by means of a membership function μ_{v_i}. When a keyframe k, described by means of a set of values $\{A_1 = w_1; ...; A_n = w_n\}$, must be classified, its description is compared with the premises of the rule r, by looking to the degree with which the observed value w is near to the edge value v. This proximity is valued as $\texttt{Deg}(w, v)$. In our case, the value w is a precise value and we have $\texttt{Deg}(w, v) = \mu_v(w)$.

For each premise, $\texttt{Deg}(w_{l_i}, v_{l_i})$ is valued for the corresponding value w_{l_i}. Finally, given the rule r, the keyframe k is associated with the class c_j with a *final degree* $\texttt{Fdeg}_r(c_j)$ that is valued as the aggregation of all the degrees $\texttt{Deg}(w_{l_i}, v_{l_i})$ by means of the *minimum*:

$$\texttt{Fdeg}_r(c_j) = \min_{i\,=\,1...p} \texttt{Deg}(w_{l_i}, v_{l_i}).P^*(c_j|(v_{l_1}, v_{l_2}, ...v_{l_p}))$$

For each class c_j, the keyframe k is associated with the membership degree $\texttt{Fdeg}(c_j)$, from $[0, 1]$, computed based on the whole set of rules. If n_ρ is the number of rules given by the fuzzy decision tree:

$$\texttt{Fdeg}(c_j) = \max_{r\,=\,1...n_\rho} \texttt{Fdeg}_r(c_j)$$

The predicted class c_k associated with k can be chosen as the class with the highest $\texttt{Fdeg}(c_k)$.

4.4 The Salammbô Software

The construction and the use of the FDT was done by means of the Salammbô software. This software was developed for building FDT efficiently and it enables to test several kinds of parameters of the FDT [10]. Moreover, the automatic method to build a fuzzy partition on the set of values of a numerical attribute, mentioned above, was implemented enabling us to avoid the prior definition of fuzzy values of attributes.

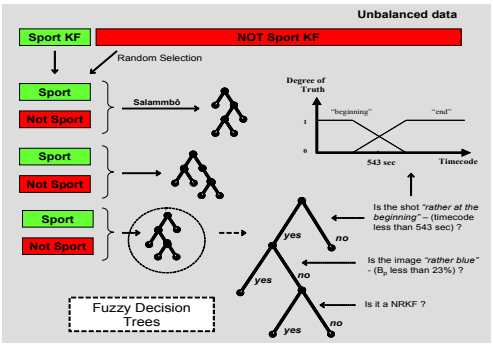


Figure 3: Growing a Forest of Fuzzy Decision Trees

5 Forests of Fuzzy Decision Trees

The use of Forests of Fuzzy Decision Trees (FFDT) is crucial when we have large, unbalanced, multiclass data sets [9]. In fact, in these cases a unique decision tree is extremely dependent on the data chosen for its construction and thus the result can be unstable and the tree generalises poorly.

Several previous works have explored the use of forests of decision trees. Generally, the decision trees of the forest are constructed classically, and they are used

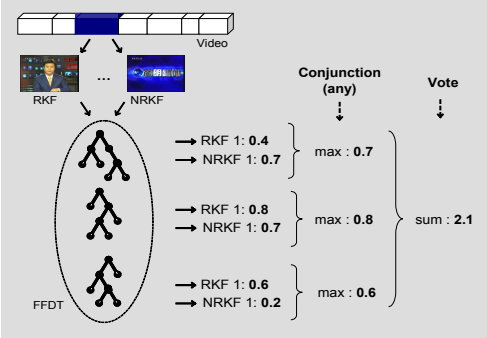


Figure 4: Fuzzy Classification and ranking using a Forest of FDT

to classify cases either classically [2, 6], or by means of the fuzzy set theory [4]. Forests of fuzzy decision trees with a fuzzy-based construction of the trees and a fuzzy classification of new cases have been proposed in [9].

A forest is composed of a given number n of Fuzzy Decision Trees. Each FDT F_i of the forest is constructed from a training set T_i. Each training set T_i is a random sample of the whole training set, as described hereafter.

Although the FDT can handle several classes simultaneously, in domains where a great number of classes exist, we decompose the problem by constructing a forest for each single class. The number of forest is thus the number of classes. For instance, if the aim is to associate each case e with one of the three classes c_1, c_2, and c_3, we construct a forest to recognize if e can be associated with c_1 or not, another forest to recognize if e can be associated with c_2 or not, an a last one to recognize if e can be associated with c_3 or not. Thus, here a forest is dedicated to the recognition of a single high level feature class and is composed of fuzzy decision trees that classify into a binary class (yes or no).

In order to use the Fuzzy Decision Trees learning method we need two training sets, one with keyframes that contain the feature to be recognized and another one with keyframes that do *not* possess that feature (see Figure 3). Since the FDTs are based on a measure estimating the quality of a decision, the two classes need to be equal in number of cases. For example, if one class outnumbers the other one, the best decision will be to always classify an example as being part of the majority class.

Thus, to have a valid training set for the construction of a FDT, we have to balance the number of keyframes of each class by (randomly) selecting a subset of the whole development data set with an equal number of cases in each class.

By repeating the random selection of examples and each time building a FDT, we obtain a robust forest classifier that are able to cover the description space of all training examples.

5.1 Classification with a forest

After the construction of the FDT as explained previously (Section 4.3), each FDT is used to classify the whole test set of keyframes. Then, by means of the classification, each keyframe k from the test set is associated with a membership degree $\texttt{FDeg}(c)$ to each class c.

With a forest of n FDTs, corresponding to a single class to be recognized, the classification of a keyframe k is performed in two steps (see Figure 4):

1. classification of k by means of the n FDT of the forest: k is classified with each FDT F_i in order to obtain a degree $\texttt{Fdeg}(c)$ of k to belong to the class c. We denote $d_i(k) = \texttt{Fdeg}(c)$ the degree given by F_i for k.
2. sum (or average) of the $d_i(k)$, $i = 1, ..., n$ degrees for each k in order to obtain a single value $d(k) = \sum_{i=1}^{n} d_i(k)$, which corresponds to the degree for the forest to believe that the k contains the feature. The higher $d(k)$, the higher it is believed that k contains the corresponding feature.

5.2 Ranking of the test shots

Finally, shots are ranked according to a degree $D(S)$: the higher $D(S)$, the higher the FFDT believed that S contains the corresponding feature.

Since only high-level features in shot are considered, the degrees of all the keyframes (RKF and NRKF) that pertained to the same shot S are aggregated to obtain the degree $D(S)$. We considers that a shot contains a given feature if at least one of its keyframes contains the feature. Therefore, the degree $D(S)$ for the shot S containing the feature is $D(S) = \max_{k \in S}(d(k))$.

6 Learning High-level Features

In other to compare our approach to others we participated to the high-level feature extraction task at TRECVID 2006 [11]. The aim of that task is to propose, for each high-level feature, a ranking of at most 2000 shots that contain it. The addressed features (and their identification number) are: sports (1), weather (3), office (5), meeting (6), desert (10), mountain (12), waterscape/waterfront (17), corporate leader (22), police security (23), military personnel

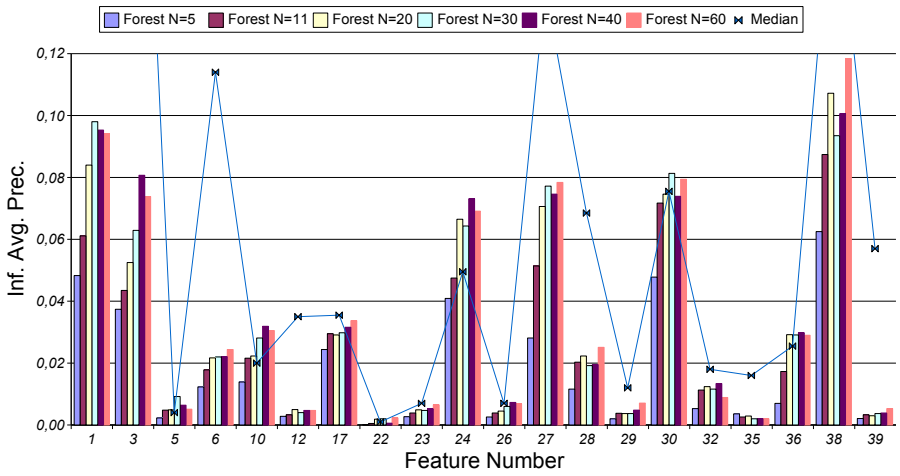


Figure 5: Importance of the size of the forest

(24), animal (26), computer TV screen (27), US flag (28), airplane (29), car (30), truck (32), people marching (35), explosion fire (36), maps (38), and charts (39).

For this study forests of 5, 11, 20, 30, 40, and 60 Fuzzy Decision Trees were built and compared between each other for each high-level feature.

6.1 Experimental roadmap

The TRECVID video corpus [11] is composed of:

- the *development data*: used to construct the system. It is composed of 137 video news (recorded in November 2004), 30mn length in average. These videos were segmented into shots and are associated with an XML file that gives information about the shots (time code, duration, etc.). Each shot is associated with a representative keyframe and, possibly, a set of non representative keyframes. Each keyframe has been (manually) indexed by one or more high level feature. The development data is composed of around 74500 keyframes (*devel* keyframes).

- the *test data*: used to evaluate the system. It is composed of 259 video news (recorded in November and December 2005), 30mn length in average. As in the development data, these videos have also been segmented into shots and are associated with an XML file that gives information about the shots (timecode, duration, etc.). Each shot is associated with a representative keyframe and, possibly, a set of non representative keyframes. None of the keyframes is indexed. The test data is constituted by around 146000 keyframes (*test* keyframes).

The set of training keyframes is highly unbalanced. For instance, for the Sports feature there are 60066 keyframes without the Sports feature, while only 1570 keyframes with it. As stated in Section 5 this problem can be solved by sampling the data set. The size of the training set was limited to 5000 keyframes at most, with as many keyframes in each class.

All runs are evaluated by means of the Inferred Average Precision (InfAP) [11], and the number of hits at several depths of the ranking (first 100, first 1000, all 2000). These values were computed by means of the software provide by the NIST and the reference file published after the competition. In order to provide an idea of the degree of complexity for the detection of each feature, the median InfAP from the 88 submitted results that were sent for evaluation to TRECVID 2006 is also shown.

6.2 Fuzzy Decision Forest vs other methods

Even though our descriptors are rather simple and incomplete, by looking Figure 5, we observe that the performance in average of the FFDT is around the median of all submitted runs.

We observe that the features that relatively to others seem to work better, in decreasing order, are: maps (38), sports (1), car (30), computer TV screen (27), military personnel (24), and weather (3). It is interesting to notice that (a) while map detection performs well, chart detection performs poorly and (b) that besides military personnel detection these features seem to be easily characterized by visual descriptors, which is coherent with our approach.

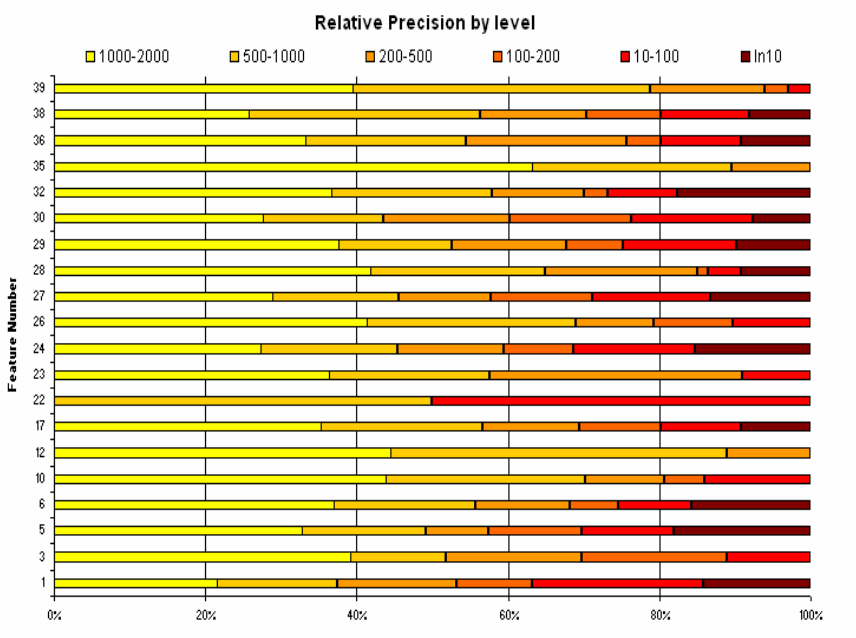


Figure 6: Classification vs ranking

If we compare our approach to (the median of) the others, and only by looking the features where some reasonable results were obtained by all the participants, we observe that our approach is particularly interesting for the detection of the features military personnel (24), desert (10), car (30) and waterscape/waterfront (17). And that our approach is relatively weak for weather (3), meeting (6) and sport (1). We guess that the reason for this is, in the one hand the existence of specialized systems for these features and, in the other hand, the simplicity of our visual descriptors. Further works will address these precises questions.

6.3 Size of the Forests

By looking on the overall performance (Figure 5) and to the per feature results, we clearly remark that by increasing the number of FDTs of a forest we improve the results. This confirms the hypothesis that FFDT is a suitable technique for covering large, unbalanced, multiclass data sets.

However, we observe that there is a limit in the number of FDT for a forest. In fact, too many classifiers lead to an over specialization to the development data, implying a loss in the generalization power.

6.4 Classification vs Ranking

Our approach performs better (relatively to others), when looking at large recall list (e.g. all the 2000 ranked shots), than when looking at the top of the list (e.g. shots ranked within the first 100). In Figure 6, we show the distribution of good classified shots in relation with their ranking.

Although the use of a forest increases the accuracy of the values and the overall ranking with respect to a single FDT classifier, it appears that the improvement is still not sufficient and uniform enough. In fact, classification algorithms like decision trees do not optimize the ranking, but only the decision concerning the class.

In other words, a classification algorithm like decision trees will try to keep all degrees above a certain (decision) threshold rather that focusing on the fact that the higher the degrees have to be more certainly in the class. In terms of error, in classification, it should be avoided that correct examples are classified under a certain threshold, while in ranking optimisation and error at high a the degree of certainty should cost more than at a lower one.

The presence of false positives in the classification (examples wrongly classified as being in the class) will degrade the ranking because both true positives (examples perfectly classified in the class) and false positives will have the same degree of classification, the sorting will not garantee that true positives will be on the top of the ranking. However, the classification garantees that “in a sufficiently large amount” of examples, the precision will be good.

However, in Figure 6, for each feature, the relative precision by level is presented. Ideally, a good rank-

ing system should provide 100% of the true positives within the top positions of the ranking. With a forest, it can be seen that in general, the global precision for a feature is given mostly by the true positives ranked after the 1000^{th} rank in the ordered shots. Supporting the hypothesis that a classifier is able to detect the presence of a feature in keyframes, does not automatically imply that it can sort them on the presence criteria.

For some rare features (truck (32), military personnel (24), meeting (6), office (5), sports (1)), the number of relative true positives found within the first 10 ranked is very high. On the contrary, for the features people marching (35) and mountain (12), the true positives are never ranked within the 500 first of the list.

The direct learning of ranking is an interesting alternative and is today a new field of research [1].

7 Conclusion

Our aim is to build a system that can automatically recognize a set of high level features for each shot in a video. The basis of our approach is to use a forest of fuzzy decision trees to construct the heart of an automatic system, in order to reduce the need of human usage in the process of indexation.

Although decision trees are very popular as key technology in several application domains, video indexation seems to be an exception. And even if our low level features are rather simple, when comparing our approach to others on the basis of the TRECVID 2006 challenge, we notice that the FFDTs have performance close to the median.

Combining several classifiers (here FDTs into a FFDT) improve the results and this especially since we address a large, unbalanced, multiclass video data collection. There is a performance limit due to overfitting.

Some high level features are "easier" to learn than others. It seems to be a strong correlation between the quality of the results and how well a feature can be described visually. But further research should be done in this direction, in particular by studying the homogeneity of the visual classes.

On the final analysis, there is strong difference between classifications – what most machine learning algorithms do – and ranking optimization. In fact, in the classification case we maximise the accuracy of the decision and not the degree of confidence of this decision. While combining several classifiers through a vote (sum of the degrees) seems to improve the exactness of the confidence degree, the direct learning of the ranking is an alternative promising research field.

References

[1] M. R. Amini, A. Tombros, N. Usunier, and M. Lalmas. Learning-based summarisation of xml documents. *Jour. of Information Retrieval*, 2007, *to appear.*

[2] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[3] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification And Regression Trees.* Chapman and Hall, New York, 1984.

[4] K. Crockett, Z. Bandar, and D. McLean. Growing a fuzzy decision forest. In *Proc. of the 10th IEEE Int. Conf. on Fuzzy Systems*, pages 614–617, 2001.

[5] Y. Freund, R. Iyer, R. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *Jour. of Machine Learning Research*, 4:933–969, 2003.

[6] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, 2006.

[7] C. Marsala. Fuzzy decision trees to help flexible querying. *Kybernetika*, 36(6):689–705, 2000.

[8] C. Marsala and B. Bouchon-Meunier. Fuzzy partioning using mathematical morphology in a learning scheme. In *Proc. of the 5th IEEE Int. Conf. on Fuzzy Systems*, pages 1512–1517, New Orleans, USA, September 1996.

[9] C. Marsala and B. Bouchon-Meunier. Forest of fuzzy decision trees. In M. Mareš, R. Mesiar, V. Novák, J. Ramik, and A. Stupňanová, editors, *Proc. of the 7th IFSA World Congress*, pages 369–374, Prague, Czech Republic, June 1997.

[10] C. Marsala and B. Bouchon-Meunier. An adaptable system to construct fuzzy decision trees. In *Proc. of the NAFIPS'99*, pages 223–227, New York, USA, 1999.

[11] NIST. Guidelines for the TRECVID 2006 evaluation - National Institute of Standards and Technology, 2006. http://www-nlpir.nist.gov/projects/tv2006/tv2006.html.

[12] C. Petersohn. Fraunhofer HHI at TRECVID 2004: Shot boundary detection system. Technical report, TREC Video Retrieval Evaluation Online Proceedings, TRECVID, 2004. http://www-nlpir.nist.gov/projects/tvpubs/tvpapers04/fraunhofer.pdf.

[13] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):86–106, 1986.

Fuzzy-Relational Classification: Combining Pairwise Decomposition Techniques with Fuzzy Preference Modeling

Eyke Hüllermeier and Klaus Brinker
Philipps-Universität Marburg
Department of Mathematics and Computer Science
{eyke,brinker}@informatik.uni-marburg.de

Abstract

This paper introduces a new approach to classification which combines pairwise decomposition techniques from machine learning with ideas and tools from fuzzy preference modeling. The approach, called fuzzy relational classification, effectively reduces the problem of classification to a problem of decision making based on a fuzzy preference relation. It will be shown that, by decomposing such a relation into a strict preference, an indifference, and an incomparability relation, it becomes possible to quantify different types of uncertainty in classification, and thereby to support more sophisticated classification and postprocessing strategies.
Keywords: Machine learning, classification, fuzzy preference relations, decision analysis.

1 Introduction

As one of the standard problems of supervised learning, the performance task of classification has been studied intensively in the field of machine learning. The arguably simplest type of classification problems are *dichotomous* (*binary*, *two-class*) problems for which a multitude of efficient and theoretically well-founded classification methods exists. Needless to say, however, practically relevant problems are rarely restricted to the binary case. One approach for tackling *polychotomous* problems is to use model classes that are able to represent a multi-class classifier, i.e., an $\mathcal{X} \rightarrow \mathcal{L}$ mapping for $|\mathcal{L}| > 2$, directly. An alternative strategy to approach such problems is to transform the original problem into several binary problems via a *class binarization* technique. The most popular class binarization technique is the unordered or one-against-rest binarization, where one takes each class in turn and learns a binary concept that discriminates this class from all other classes.

The key idea of the alternative *learning by pairwise comparison* (LPC) approach (aka pairwise classification, round robin learning, one-vs-one) is to transform an m-class problem into $m(m-1)/2$ binary problems, one for each pair of classes.[1] At classification time, a query instance is submitted to all binary models, and the predictions of these models are combined into an overall classification. In [5, 6], it was shown that pairwise classification is not only more accurate than the one-against-rest technique but that, despite the fact that the number of models that have to be learned is quadratic in the number of classes, pairwise classification is also more efficient (at least in the training phase) than one-against-rest classification.

This paper elaborates on another interesting aspect of the LPC approach: Assuming that every binary learner outputs a score in the unit interval (or, more generally, an ordered scale), and that this score can reasonably be interpreted as a "fuzzy preference" for the first in comparison with the second class, the complete ensemble of pairwise learners produces a *fuzzy preference relation*. The final classification decision is then made on the basis of this relation. In other words, the problem of classification has been reduced, in a first step, to a problem of decision making based on a fuzzy preference relation.

The novel aspect here is to look at the ensemble of predictions as a fuzzy preference relation. This perspective establishes a close connection between (pairwise) learning and fuzzy preference modeling, and therefore allows for applying techniques from the former field in the context of machine learning. In this paper, we are especially interested in exploiting techniques for decomposing a fuzzy (weak) preference relation into a preference structure consisting of a strict preference, an indifference, and an incomparability relation.

[1] Alternatively, one can consider a binary problem for every *ordered* pair of classes, in which case the total number of such problems is doubled. We shall come back to this point later on.

As will be argued in more detail later on, the latter two relations have a quite interesting interpretation and important meaning in the context of classification, where they represent two types of uncertainty: ambiguity and ignorance. Consequently, these relations can support more sophisticated classification strategies, including those that allow for partial reject options.

The remainder of the paper is organized as follows. Section 2 details the LPC approach to classification, and section 3 recalls the basics of fuzzy preference structures. The idea of classification based on fuzzy preference relations is outlined in section 4. Section 5 elaborates on an important element of this approach, namely learning weak preferences between class labels. First empirical results are presented in section 6, and section 7 concludes the paper.

2 Learning by Pairwise Comparison

As mentioned earlier, learning by pairwise comparison (LPC) transforms a multi-class classification problem, i.e., a problem involving $m > 2$ classes (labels) $\mathcal{L} = \{\lambda_1 \dots \lambda_m\}$, into a number of *binary* problems. To this end, a separate model (base learner) $\mathcal{M}_{i,j}$ is trained for each *pair* of labels $(\lambda_i, \lambda_j) \in \mathcal{L}$. $\mathcal{M}_{i,j}$ is intended to separate the objects with label λ_i from those having label λ_j. If $(x, \lambda_a) \in \mathcal{X} \times \mathcal{L}$ is an original training example (revealing that instance x has label λ_a), then x is considered as a *positive* example for all learners $\mathcal{M}_{a,j}$ and as a *negative* example for the learners $\mathcal{M}_{j,a}$ $(j \neq a)$; those models $\mathcal{M}_{i,j}$ with $a \notin \{i, j\}$ simply ignore this example.

At classification time, a query x is submitted to all learners, and each prediction $\mathcal{M}_{i,j}(x)$ is interpreted as a vote for a label. In particular, if $\mathcal{M}_{i,j}$ is a $\{0, 1\}$-valued classifier, $\mathcal{M}_{i,j}(x) = 1$ is counted as a vote for λ_i, while $\mathcal{M}_{i,j}(x) = 0$ would be considered as a vote for λ_j. Given these outputs, the simplest classification strategy is to predict the class label with the highest number of votes. A straightforward extension of the above voting scheme to the case of $[0, 1]$-valued (scoring) classifiers yields a *weighted voting procedure*: The score for label λ_i is computed by

$$r_i \stackrel{\text{df}}{=} \sum_{1 \leq j \neq i \leq m} r_{i,j}, \tag{1}$$

where $r_{i,j} = \mathcal{M}_{i,j}(x)$, and again the label with the highest score is predicted.

The votes $r_{i,j}$ in (1) and, hence, the learners $\mathcal{M}_{i,j}$ are usually assumed to be (additively) *reciprocal*, that is,

$$r_{j,i} \equiv 1 - r_{i,j} \tag{2}$$

and correspondingly $\mathcal{M}_{i,j}(x) \equiv 1 - \mathcal{M}_{j,i}(x)$. Practically, this means that only one half of the $m(m-1)$ classifiers $\mathcal{M}_{i,j}$ needs to be trained, for example those for $i < j$. As will be explained in more detail below, this restriction is not very useful in our approach. Therefore, we will train the whole set of classifiers $\mathcal{M}_{i,j}$, $1 \leq i \neq j \leq m$, which means that no particular relation between $r_{i,j}$ and $r_{j,i}$ will be assumed.

3 Fuzzy Preference Structures

Considering the classification problem as a *decision problem*, namely a problem of deciding on a class label for a query input x, an output $r_{i,j} = \mathcal{M}_{i,j}(x)$ can be interpreted as a *preference* for label λ_i in comparison with label λ_j: The higher $r_{i,j}$, the more preferred is λ_i as a classification for x, i.e., the more likely λ_i appears in comparison with label λ_j. Correspondingly, the matrix

$$\mathcal{R} = \begin{bmatrix} - & r_{1,2} & \dots & r_{1,m} \\ r_{2,1} & - & \dots & r_{2,m} \\ \vdots & & & \vdots \\ r_{m,1} & r_{m,2} & \dots & - \end{bmatrix} \tag{3}$$

obtained by collecting the outputs of the whole classifier ensemble can be interpreted as a *fuzzy* or *valued preference relation*. A classification decision can then be made on the basis of the relation (3). To this end, one can resort to corresponding techniques that have been developed and investigated quite thoroughly in fuzzy preference modeling and decision making [4]. In principle, the simple voting scheme (1) outlined in section 2 can be seen as a special case of such a decision making technique.

In this paper, our interest concerns the application of techniques for decomposing the relation $\mathcal{R}$ into three associated relations with different meaning. Suppose that $\mathcal{R}$ can be considered as a *weak preference relation*, which means that $r_{i,j} = \mathcal{R}(\lambda_i, \lambda_j)$ is interpreted as $\lambda_i \succeq \lambda_j$, that is, "label λ_i is at least as likely as label λ_j". From this relation, one can derive a *fuzzy preference structure* consisting of a *strict preference relation* $\mathcal{P}$, an *indifference relation* $\mathcal{I}$, and an *incomparability relation* $\mathcal{J}$. Referring to the class of t-norms [9] to operate on fuzzy preference degrees, a fuzzy preference structure can be defined as follows: Let (T, S, N) be a continuous De Morgan triplet consisting of a strong negation N, a t-norm T, and its N-dual t-conorm S; moreover, denote the T-intersection of two sets A and B by $A \cap_T B$ and the S-union by $A \cup_S B$. A fuzzy preference structure on $\mathcal{L}$ is a triplet $(\mathcal{P}, \mathcal{I}, \mathcal{J})$ of fuzzy relations satisfying

- $\mathcal{P}$ and $\mathcal{J}$ are irreflexive, $\mathcal{I}$ is reflexive;
- $\mathcal{P}$ is T-asymmetrical ($\mathcal{P} \cap_T \mathcal{P}^t = \emptyset$), $\mathcal{I}$ and $\mathcal{J}$ are symmetrical;

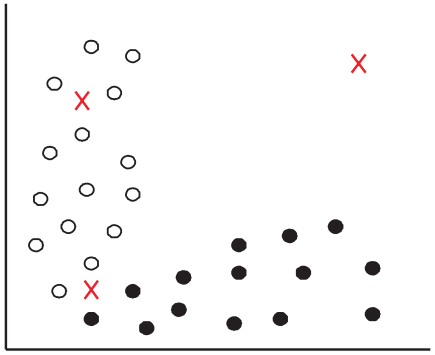


Figure 1: Classification scenario: Observations from two classes (points) and new query instances (crosses).

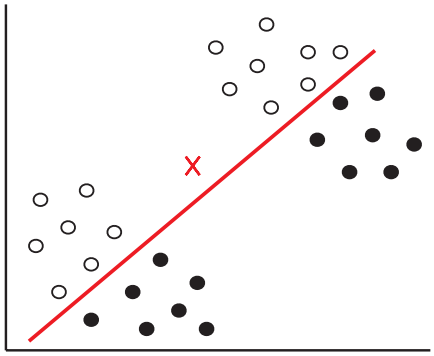


Figure 2: Given the assumption of linear separability, the query instance can be classified quite safely, even though it is spatially isolated from all other examples.

- $\mathcal{P} \cap_T \mathcal{I} = \emptyset$, $\mathcal{P} \cap_T \mathcal{J} = \emptyset$, $\mathcal{I} \cap_T \mathcal{J} = \emptyset$;
- $\mathcal{P} \cup_S \mathcal{P}^t \cup_S \mathcal{I} \cup_S \mathcal{J} = \mathcal{L} \times \mathcal{L}$.

The question of how to decompose a weak (valued) preference relation $\mathcal{R} \in [0,1]^{m \times m}$ into a strict preference relation $\mathcal{P}$, an indifference relation $\mathcal{I}$, and an incomparability relation $\mathcal{J}$ such that $(\mathcal{P}, \mathcal{I}, \mathcal{J})$ is a fuzzy preference structure have been studied extensively in the literature (e.g. [4, 1]). Without going into technical detail, we only give an example of a commonly employed decomposition scheme (again, we denote $r_{i,j} = \mathcal{R}(\lambda_i, \lambda_j)$):

$$\begin{aligned} \mathcal{P}(\lambda_i, \lambda_j) &= r_{i,j} \times (1 - r_{j,i}) \\ \mathcal{I}(\lambda_i, \lambda_j) &= r_{i,j} \times r_{j,i} \\ \mathcal{J}(\lambda_i, \lambda_j) &= (1 - r_{i,j}) \times (1 - r_{j,i}) \end{aligned} \tag{4}$$

A related decomposition scheme will also be used in the experimental part below.

4 Fuzzy Modeling of Classification Knowledge

The relations $\mathcal{I}$ and $\mathcal{J}$ have a very interesting meaning in the context of classification: Indifference corresponds to the *ambiguity* of a classification decision, while incompatibility reflects the corresponding degree of *ignorance*. To illustrate what we mean, respectively, by ambiguity and ignorance, consider the simple classification scenario shown in Fig. 1: Given observations from two classes, `black` and `white`, three new instances marked by a cross need to be classified. Obviously, given the current observations, the upper left instance can quite safely be classified as `white`. The case of the lower left instance, however, involves a high level of ambiguity, since both classes, `black` and `white`, appear plausible. The third situation is an example of ignorance: The upper right instance is located in a region of the instance space in which no observations have been made so far. Consequently, there is neither evidence in favor of class `black` nor in favor of class `white`.

In the above example, the meaning of and difference between ambiguity and ignorance is intuitively quite obvious. Upon closer examination, however, these concepts turn out to be more intricate. In particular, one should realize that ignorance is not immediately linked with sparseness of the input space. This is due to the fact that generalization in machine learning is not only based on the observed data but also involves a model class with associated model assumptions. In fact, a direct connection between ignorance and sparely populated regions of the input space can only be established for instance-based (prototype-based) classifiers, since these classifiers are explicitly based on the assumption that closely neighbored instances belong to the same class.

The situation is different, however, for other types of models. For example, Fig. 2 shows a scenario in which a query point in a sparse input region can be classified quite safely, given the observed data *in conjunction with the assumption of a linear model*. In other words, given the correctness of the inductive bias of the learner (linearity assumption), the current observations allow for quite confident conclusions about the label of the query, even though the latter does not have any close neighbors.

The above considerations give rise to the following conception of ambiguity and ignorance in the context of classification: Let $\mathfrak{M}$ denote the model class underlying the classification problem, and let $\mathcal{V} = \mathcal{V}(\mathcal{D})$ be the set of models which are compatible with the examples given, i.e., the set of models which can still be regarded as possible candidates given the data $\mathcal{D}$; in the machine learning literature, $\mathcal{V}$ is called the *version*

space. Now, given a query $x_0 \in \mathcal{X}$, the set of possible predictions is

$$Y_0 = \{\mathcal{M}(x) \,|\, \mathcal{M} \in \mathcal{V}(\mathcal{D}) \subseteq \mathfrak{M}\} \tag{5}$$

If the output of a model $\mathcal{M} \in \mathfrak{M}$ is a (deterministic) class label, then Y_0 is a subset of class labels ($Y_0 \subseteq \mathcal{L}$). Otherwise, if $\mathfrak{M}$ is a class of probabilistic classifiers, then Y_0 is a class of probability distributions over $\mathcal{L}$. In any case, it seems reasonable to define the degree of ignorance of a prediction in terms of the *diversity* of Y_0: The more predictions appear possible, i.e., the higher the diversity of predictions, the higher is the degree of ignorance.

According to this view, ignorance (incomparability) corresponds to that part of the (total) uncertainty about a prediction which can potentially be reduced by gathering more examples and thereby shrinking the version space. As opposed to this, the degree of ambiguity (indifference) corresponds to that part of the uncertainty which is due to a known conflict and which cannot be reduced any further.

The general idea of our method is to learn the weak preference relation (3), using an LPC approach, and to decompose this relation into a preference structure $(\mathcal{P}, \mathcal{I}, \mathcal{J})$ such that $\mathcal{J}$ characterizes the ignorance involved in a prediction, in the sense as outlined above, and $\mathcal{I}$ the ambiguity of the classification. In this context, two important questions have to be answered: Firstly, how to learn a suitable weak preference relation $\mathcal{R}$, and secondly, how to decompose $\mathcal{R}$ into a structure $(\mathcal{P}, \mathcal{I}, \mathcal{J})$. These problems will be discussed in more detail in the following section.

5 Learning Weak Preference Relations

As mentioned above, the first step of our method consists of learning the weak preference relation $\mathcal{R}$. More specifically, for every pair of labels (λ_i, λ_j), we have to induce models $\mathcal{M}_{i,j}$ and $\mathcal{M}_{j,i}$ such that, for a given query input x, $\mathcal{M}_{i,j}(x)$ corresponds to the degree of weak preference $\lambda_i \succeq \lambda_j$ and, vice versa, $\mathcal{M}_{j,i}(x)$ to the degree of weak preference $\lambda_j \succeq \lambda_i$.

The models $\mathcal{M}_{i,j}$ are of special importance as they directly determine the degrees of ambiguity and ignorance associated with a comparison between λ_i and λ_j. This fact is also crucial for the properties that the models $\mathcal{M}_{i,j}$ should obey.

According to the idea outlined in the previous section, a weak preference in favor of a class label should be derived from the set (5) of possible predictions. As this set in turn depends on the version space $\mathcal{V}$, the problem comes down to computing or at least approximating this space. In this connection, it deserves mentioning that an exact representation of the version space will usually not be possible for reasons of complexity. Apart from that, however, a representation of that kind would not be very useful either. In fact, despite the theoretical appeal of the version space concept, a considerable practical drawback concerns its extreme sensitivity toward noise and inconsistencies in the data.

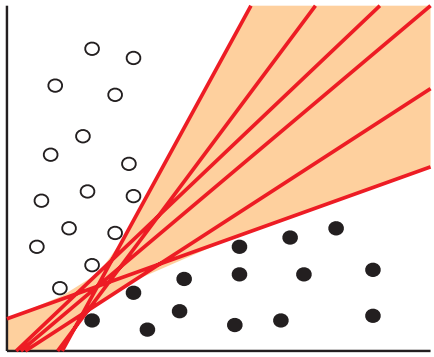


Figure 3: Illustration of the version space (class of hyperplanes that classify the training data correctly) and the "region of ignorance" (shaded in light color).

To overcome these problems, our idea is to approximate a version space in terms of a finite number of representative models. More specifically, consider the problem of learning a binary model $\mathcal{M}_{i,j}$ from an underlying model class $\mathfrak{M}$. To approximate the version space associated with $\mathcal{M}_{i,j}$, we induce a finite set of models

$$\mathbb{M}_{i,j} = \{\, \mathcal{M}_{i,j}^1, \mathcal{M}_{i,j}^2 \ldots \mathcal{M}_{i,j}^K \,\} \subseteq \mathfrak{M} \tag{6}$$

The set of possible predictions (5) is approximated correspondingly by

$$\widehat{Y}_0 = \mathbb{M}_{i,j}(x) = \bigcup_{k=1\ldots K} \mathcal{M}_{i,j}^k(x).$$

The way in which the models in (6) are obtained depends on the model class $\mathfrak{M}$. The basic idea is to apply randomization techniques as they are typically employed in ensemble learning methods. In the experiments below, we shall use ensembles of linear perceptrons, each of which is trained on a random permutation of the whole data.

An illustration is given in Fig. 3. Assuming that the two classes `black` and `white` can be separated in terms of a linear hyperplane, the version space consists of all those hyperplanes that classify the training data correctly. Given a new query instance, a unique class label can be assigned only if that instance lies on the same side of *all* hyperplanes (this situation is sometimes called "unanimous voting" [11]). Otherwise,

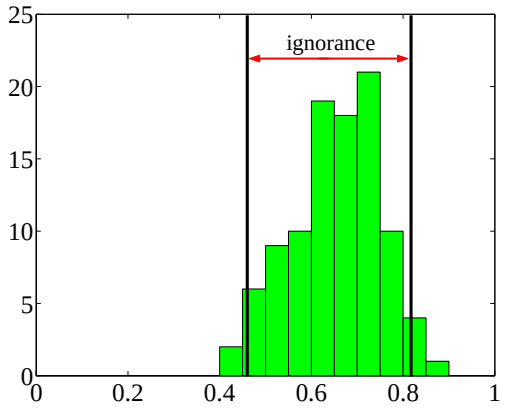


Figure 4: Distribution of the scores output by an ensemble $\mathbb{M}_{i,j}$. The degree of ignorance corresponds to the imprecision (width) of the distribution (here measured in a robust way in terms of the distance between the α- and $(1-\alpha)$-quantile).

both predictions are possible; the corresponding set of instances constitutes the "region of ignorance" which is shaded in light color.

In the above example, $\{0,1\}$-valued classifiers were used for the sake of simplicity. In the context of fuzzy classification, however, scoring classifiers with outputs in the unit interval are more reasonable. Suppose that each ensemble member $\mathcal{M}_{i,j}^k$ in (6) outputs a score $s_{i,j}^k \in [0,1]$. The minimum of these scores would in principle be suitable as a degree of (weak) preference for λ_i in comparison with λ_j:

$$r_{i,j} = \min_{k=1\ldots K} s_{i,j}^k.$$

As this order statistic is quite sensitive toward noise and outliers, however, we propose to replace it by the empirical α-quantile of the distribution of the $s_{i,j}^k$ (a reasonable choice is $\alpha = 0.1$).

Note that, in case the models in $\mathfrak{M}$ are reciprocal, only $\mathbb{M}_{i,j}$ or $\mathbb{M}_{j,i}$ needs to be trained, but not both. We then have $s_{i,j}^k = 1 - s_{j,i}^k$, and the α-quantile for $\mathcal{M}_{i,j}$ is given by 1 minus the $(1-\alpha)$-quantile for $\mathcal{M}_{j,i}$. In other words, the degree of ignorance is directly reflected by the distribution of the scores $s_{i,j}^k = 1 - s_{j,i}^k$ and corresponds to the length of the interval between the α-quantile and the $(1-\alpha)$-quantile of this distribution. Thus, the more precise this distribution, the smaller the degree of ignorance. In particular, if all models $\mathcal{M}_{i,j}^k$ output the same score s, the ignorance component shrinks to 0. An illustration is given in Fig. 4.

Our approach of *fuzzy relational classification* (FRC) as outlined above can be seen as a technique for deriving a condensed representation of the classification-relevant information contained in the version space. Once a preference structure $(\mathcal{P}, \mathcal{I}, \mathcal{J})$ has been induced, it can be taken as a point of departure for sophisticated decision strategies which go beyond simple voting procedures. This approach becomes especially interesting in extended classification scenarios, that is, generalizations of the conventional setting in which a single decision in favor of a unique class label is requested. For example, it might be allowed to predict several class labels instead of single one in cases of ambiguity, or to defer an immediate decision in cases of ignorance (or ambiguity). The latter scenario is known as *classification with reject option* in the literature, where one often distinguishes between *ambiguity rejection* [2, 7] and *distance rejection* [3]. Interestingly, this corresponds roughly to our distinction between ambiguity and ignorance. As we explained above, however, our conception of ignorance is more general and arguably more faithful, as it takes the underlying model assumptions into account: equating distance (between the query and observed examples) with ignorance does make sense for instance-based classifiers but not necessarily for other approaches with different model assumptions.

Of course, the design of suitable decision policies is highly application-specific and beyond the scope of this paper. In the next section, we therefore restrict ourselves to a simple experimental setup which is suitable for testing a key feature of FRL, namely its ability to represent the amount of uncertainty associated with a classification. More specifically, we used FRL as a means for implementing a reject option in the context of binary classification.

	name	# features	# examples
1	australian_scale	14	690
2	breast-cancer_scale	10	683
3	fourclass_scale	2	862
4	german	24	1000
5	heart_scale	13	270
6	splice_scale	60	1000
7	sonar_scale	60	208
8	w1a	300	2477

Figure 5: Data sets used in the experiments.

6 Experimental Results

We conducted an experimental study on 8 binary classification data sets from the Statlog and UCI repositories (cf. Fig. 5).[2] Each of the data sets was randomly

[2]These are preprocessed versions from the LIBSVM-website.

split into a training and test set of (roughly) equal size. As model classes $\mathbb{M}_{i,j}$, we used ensembles of 100 perceptrons with linear kernels and the default additive diagonal constant 1 (to account for non-separable problems), which were induced on the training data. Each perceptron was provided with a random permutation of the training set in order to obtain a diverse ensemble [8]. This process was repeated 10 times to reduce the bias induced by the random splitting procedure, and the results were averaged.

On the test sets, the real-valued classification outputs of the perceptrons were converted into normalized scores using a common logistic regression approach by Platt [10]. For a given test instance, the weak preference component $r_{i,j} = \mathcal{R}(\lambda_i, \lambda_j)$ was derived by the 0.1-quantile of the distribution of the scores from the ensemble $\mathbb{M}_{i,j}$ (see section 5). Moreover, as a decomposition scheme we used a slight modification of (4):

$$\begin{aligned} \mathcal{P}(\lambda_i, \lambda_j) &= r_{i,j}\,(1 - r_{j,i}) \\ \mathcal{I}(\lambda_i, \lambda_j) &= 2\,r_{i,j}\,r_{j,i} \\ \mathcal{J}(\lambda_i, \lambda_j) &= 1 - (r_{i,j} + r_{j,i}) \end{aligned} \tag{7}$$

The reason for the modification is that in (7), the ignorance component nicely agrees with our derivation of weak preference degrees: It just corresponds to the width of the distribution of the scores generated by $\mathbb{M}_{i,j}$ (or, more precisely, the length of the interval between the quantiles of this distribution); therefore, it reflects the diversity of the predictions and becomes 0 if all ensemble members $\mathcal{M}_{i,j}^k$ agree on exactly the same score.

Finally, all test instances were ordered with respect to the associated degrees of indifference (ignorance), and corresponding accuracy-rejection diagrams were derived. These diagrams provide a visual representation of the accuracy levels α as a function of the rejection rate ρ: If the ρ% test instances with the highest degrees of indifference (ignorance) are refused, then the classification rate on the remaining test instances is α. Obviously, the effectiveness of FRL in representing uncertainty is in direct correspondence with the shape of the accuracy-rejection curve: If the degree of indifference (ignorance) produced by FRL is a good indicator of the reliability of a classification, then the ordering of instances according to indifference (ignorance) is in agreement with their respective degree of reliability (chance of misclassification), which in turn means that the accuracy-rejection curve is increasing. The presumption that FRL is indeed effective in this sense is perfectly confirmed by the experimental results, as can be seen in Fig. 6–7.

7 Conclusions

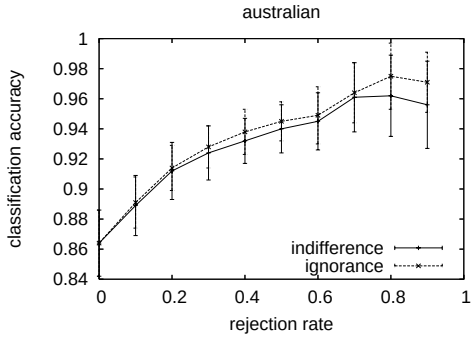
In this paper, we have introduced a new approach to classification learning which refers to the concept of fuzzy preference structures. This approach is intimately related with learning by pairwise comparison (LPC), a well-known machine learning technique for reducing multi-class to binary problems. The key idea of our approach, called *fuzzy relational classification* (FRC), is to use LPC in order to learn a fuzzy (weak) preference relation among the potential class labels. The original classification problem thus becomes a problem of decision making, namely of taking a course of action on the basis of this fuzzy preference relation. This way, our approach makes machine learning amenable to techniques and decision making strategies that have been studied intensively in the literature on fuzzy preferences.

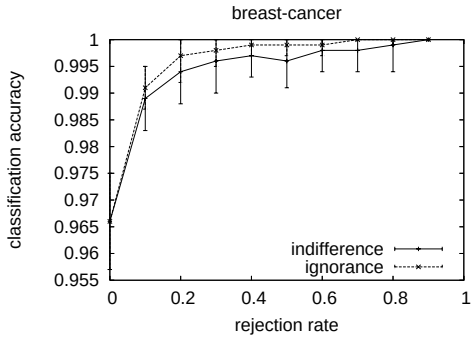
An interesting example of corresponding techniques has been considered in more detail in this paper, namely the decomposition of a weak preference relation into a strict preference, an indifference, and an incomparability relation. We have argued that, in a classification context, indifference can be interpreted as the *ambiguity* of a prediction while indifference represents the level of *ignorance*. These concepts can be extremely useful, especially in extended classification scenarios which go beyond the prediction of a single label or do offer the option to abstain from a immediate classification decision.

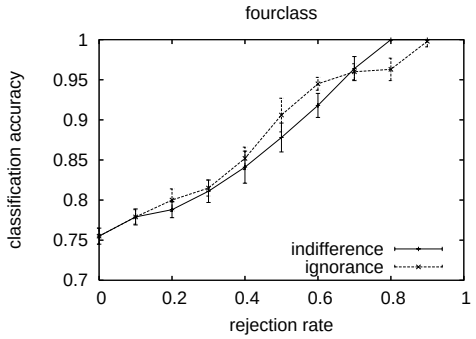
First empirical studies have shown that FRC is indeed able to represent the uncertainty related to a classification decision: The implementation of a reject options turned out to be highly effective, regardless of whether the decision to abstain is made on the basis of the degree of ambiguity or the degree of ignorance.

The main contribution of this paper is a basic conceptual framework of fuzzy relational classification, including first empirical evidence in favor of its usefulness. Nevertheless, this framework is far from being complete and still leaves much scope for further developments. This concerns almost all steps of the approach and includes both aspects of learning and decision making. Just to give an example, our approach outlined in section 5 is of course not the only way to learn a weak preference relation. Moreover, the aspect of optimal decision making on the basis of pairwise preferences has not yet been addressed (as it strongly depends on the classification scenario). Issues of that kind will therefore be explored in future work.

Acknowledgements. This research was supported by the German Research Foundation (DFG).







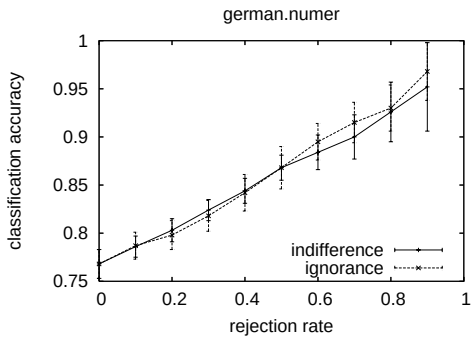
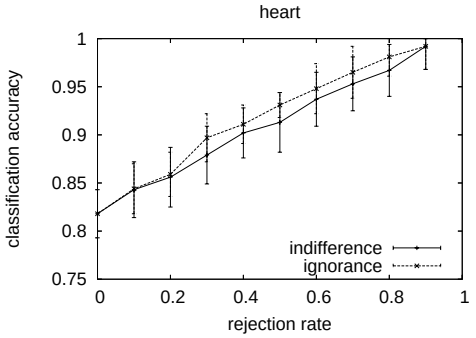
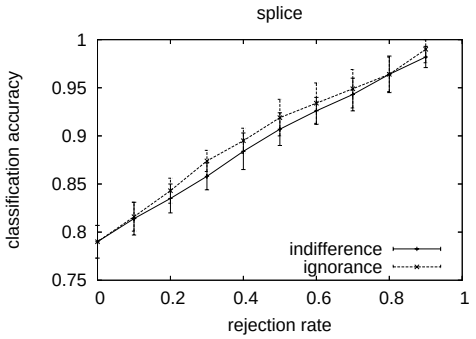
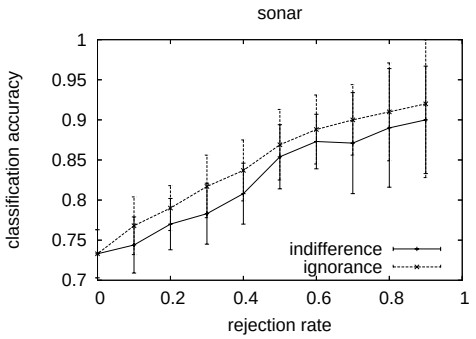


Figure 6: Accuracy-rejection curves for the data sets 1–4.







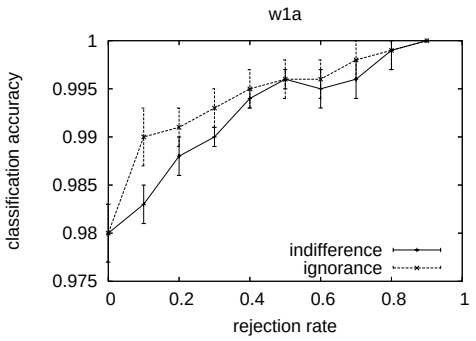


Figure 7: Accuracy-rejection curves for the data sets 5–8.

References

[1] B. De Baets, B. Van de Walle, and E. Kerre. Fuzzy preference structures and their characterization. *The Journal of Fuzzy Mathematics*, 3:373–381, 1995.

[2] CK. Chow. An optimum character recognition system using decision functions. *Trans. on Electronic Computers*, 6:247–253, 1957.

[3] B. Dubuisson and MH. Masson. A statistical decision rule with incomplete knowledge about classes. *Pattern Recognition*, 26(1):155–165, 1993.

[4] J. Fodor and M. Roubens. *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, 1994.

[5] J. Fürnkranz. Round robin rule learning. In *ICML-2001, Proc. 18th International Conference on Machine Learning*, pages 146–153, Williamstown, MA, 2001.

[6] J. Fürnkranz. Round robin classification. *Journal of Machine Learning Research*, 2:721–747, 2002.

[7] TM. Ha. The optimum class-selective decision rule. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):608–615, 1997.

[8] Ralf Herbrich. *Learning Kernel Classifiers*. MIT Press, 2002.

[9] EP. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, 2002.

[10] John Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In A.J. Smola, P. Bartlett, B. Schoelkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 61–74, Cambridge, MA, 1999. MIT Press.

[11] E. Smirnov, I. Sprinkhuizen-Kuyper, G. Nalbantov, and S. Vanderlooy. Version space support vector machines. In *Proc. ECAI-06, 17th European Conference on Artificial Intelligence*, pages 809–810, Riva del Garda, Spain, 2006.

Selecting the Optimal Rule Set Using a Bacterial Evolutionary Algorithm

Mario Drobics
eHealth systems
Biomedical Engineering
Austrian Research Centers GmbH
A-1220 Vienna, Austria
mario.drobics@arcsmed.at

János Botzheim
Department of Telecommunications
and Media Informatics
Budapest University of
Technology and Economics
H-1117 Budapest, Hungary
botzheim@tmit.bme.hu

Klaus-Peter Adlassnig
Section on Medical Expert
and Knowledge-Based Systems
Core Unit for Medical Statistics
and Informatics
Medical University Vienna
A-1090 Vienna, Austria
klaus-peter.adlassnig@meduniwien.ac.at

Abstract

In many regression learning algorithms for fuzzy rule bases it is not possible to define the error measure to be optimized freely. A possible alternative is the usage of global optimization algorithms like genetic programming approaches. These approaches, however, are very slow because of the high complexity of the search space. In this paper we present a novel approach where we first create a large set of (possibly) redundant rules using inductive rule learning and where we use a bacterial evolutionary algorithm to identify the best subset of rules in a subsequent step. The evolutionary algorithm tries to find an optimal rule set with respect to a freely definable goal function.

Keywords: Inductive Learning, Rule Selection, Interpretability, Bacterial Evolutionary Algorithm.

1 Introduction

Regression learning is concerned with finding a function $f(\mathbf{x}), \mathbb{R}^n \mapsto \mathbb{R}$ which best fits a given data set X. As fuzzy rule bases are capable of fulfilling requirements regarding interpretability and accuracy, they are often used in control applications, where expert knowledge is not available, but knowledge of the resulting system is essential [6].

Most methods for learning fuzzy rule bases (or fuzzy regression trees), however, choose a stepwise approach to construct the rule base. Therefore, decisions are based on local criteria like entropy gain, confidence and support, or improvement in goodness of fit (e.g. mean squared error). This approach has two shortcomings: Firstly, selecting accurate rules individually is not sufficient, as the interaction of the rules is very important for the overall performance of the rule base in the fuzzy case. Secondly, it is usually not possible to define the goal function freely. This becomes crucial, when global criteria—like interpretability measures [12]—are involved.

Recent approaches which try to use more problem specific selection criteria have been presented e.g. for regression trees, where a rough approximation of the expected output is used. As at the time a node is split no results from the subtrees are available, the quality of the split is measured by interpolating between the mean values of each subgroup [9, 14].

$$\mathrm{MSE}_{\mathrm{DT}}(P, z, X) = \frac{\sum_{\mathbf{x}\in X} \mu_X(\mathbf{x})(\tilde{z}_P(\mathbf{x}) - z(\mathbf{x}))^2}{|X|},$$

$$\tilde{z}_P(\mathbf{x}) = t(P(\mathbf{x}))\bar{z}(X|P) + t(\neg P(\mathbf{x}))\bar{z}(X|\neg P),$$

where X is the data set, P the predicate to be applied, z the actual goal function and $\bar{z}(X|P)$ the average of z in X, weighted according to P. $t(.)$ denotes the truth evaluation function. Although this approach is an improvement over traditional approaches and it might be adopted to other error measures easily, it still uses a step wise approach and is therefore restricted to an approximation of the final tree structure. Other approaches use subsequent optimization techniques like pruning [14] or meta-optimization techniques like boosting [16].

Currently, only a few approaches like genetic programming [17] are capable of optimizing a complete rule base. These approaches, however, are usually very complex and time consuming, as the search space is extremely large.

We overcome these limitations by splitting the construction of the rules and the construction of the final rule base. Namely we construct a large set of rules first, where all rules fulfill only minimal requirements in terms of confidence and support. Then we select a much smaller subset of these rules using a bacterial evolutionary algorithm (BEA). As in the BEA we can define the goal function freely, we finally obtain a rule set which perfectly fits the requirements. Comparable

approaches for classification problems using genetic algorithms have been presented in [11] and [19]. The main disadvantage of these approaches is their complexity, caused by the use of GAs and a binary coding. Furthermore, the key advantage of this approach—its ability to find a good *combination* of rules—is much more powerful when applied to regression learning.

Bacterial evolutionary algorithms are simpler then genetic algorithms and it is possible to reach lower error levels within a short time. They comprise of two operations inspired by the microbial evolution phenomenon. The bacterial mutation operation which optimizes the chromosome of one bacterium, and the gene transfer operation which transfers information between different bacteria within the population. BEA have already been successfully applied to rule learning [5] and feature selection [4].

In this paper we will first introduce the bacterial evolutionary algorithm and we show how this method can be applied to the problem of rule selection. Then some simulation results are presented, to illustrate the potential of this new approach. Finally, an outlook to future work is given.

2 Bacterial Evolutionary Algorithm

There are several optimization algorithms which were inspired by the processes of evolution. These processes can be easily applied in optimization problems where one individual corresponds to one solution of the problem. An individual can be represented by a sequence of numbers that can be bits as well. This sequence is called *chromosome*, which is nothing else than the individual itself. Bacterial evolutionary algorithms are a recent variant of genetic algorithms based on bacterial evolution rather than eukaryotic. Bacteria share chunks of their genes rather than perform neat crossover in chromosomes, which means bacterial genomes can grow or shrink. This mechanism is used both in the *bacterial mutation* and the *gene transfer* operations. The latter substitutes the genetic algorithms crossover operation, so information can be transferred between different individuals. As in this approach many operations can be performed in parallel, it can be adopted to a parallel computing environment in a straightforward manner.

2.1 Generating the initial rule set

In our approach we use a method which finds all rules fulfilling minimal requirements in terms of confidence and support called *FS-Miner* [8]. Although it might be possible to remove rules covering the same range of the data space using a partial ordering structure, we do not use this mechanism as we want to obtain the most comprehensive set of rules. Of course, other rule learning methods might be used as well (e.g. association rule miners [1, 10]). The underlying set of predicates was defined using *CompFS* [7]. For each attribute, a partition into five fuzzy sets was created automatically. Furthermore, ordering based predicates were defined, too [3].

2.2 The encoding method

In bacterial evolutionary algorithms, one bacterium $\xi_i, i \in I$ corresponds to one solution of the problem under investigation. For the task of selecting m_i rules from a set of n rules ($m_i \leq n$), the bacterium consists of a vector of rule indices $\xi_i = \{\xi_i^1, \ldots, \xi_i^{m_i}\}, 1 \leq \xi_i^k \leq n$ with ξ_i^k being the index of the k-th rule and $\xi_i^k \neq \xi_i^l$ for $k \neq l$.

This encoding method, although more complex than a simple binary coding, has strong benefits. First of all this encoding supports the implicit definition of subgroups from not consecutive rules. When using a binary coding, subgroups can only evolve amongst neighboring rules. Having subgroups of rules is, however, very important as these subgroups may contain interacting rules with a good overall performance. Furthermore, the evolutionary operations perform block operations which preserve these subgroups. Secondly, we have total control on the number of rules in the rule base. By specifying the length of elements inserted or deleted from the bacterium, we determine the overall number of rules involved. When using a binary coding, the number of rules is equivalent to the number of 1's, making it much harder to control the overall number of rules in a single step.

2.3 The evaluation function

Similar to genetic algorithms the fitness of a bacterium ξ_i is evaluated using an *evaluation function* $\phi(\xi_i)$. As this evaluation function is computed for all bacteria after each mutation, its efficiency has a major influence on the overall runtime performance of the algorithm.

When a regression problem is only optimized with respect to the overall error the problem occurs, that the number of rules will increase rapidly. Although this effect can be reduced by using a separate test data set, we might want to obtain the best result involving a certain number of rules. As, however, defining the size of the final rule base a-priory is difficult, we use a fuzzy predicate which is incorporated in the target function to limit the size of the rule base. Doing so, we can compute the best result for a given threshold size [19].

For a given data set $S^\star \subset S$ we compute the error estimate of a rule set ξ_i as the normalized mean squared

error of the corresponding output function f. To ensure that we do not obtain infinitely large rule sets we define a fuzzy predicate $\mathrm{LE}(s, m)$ according to:

$$\mathrm{LE}(s, \overline{m}) = \begin{cases} 1 & s <= \overline{m} \\ e^{-\frac{1}{2}\left(\frac{\overline{m}-s}{\overline{m}/2}\right)^2} & \text{otherwise} \end{cases},$$

with $s = \mathrm{MS}(f)$ being the actual model size, and $\overline{m}$ the desired maximum number of rules. The overall error measure $\phi(f, S^\star)$ is then computed as:

$$\phi(f, S^\star) = \frac{\mathrm{mseN}(f, S^\star)}{(1 - \mathrm{RoNP}(f, S^\star))^4 * \mathrm{LE}(\mathrm{MS}(f), \overline{m})}, \quad (1)$$

where

$$\mathrm{mseN}(f, S^\star) = \frac{\sum_{\mathbf{x} \in S^\star} (z(\mathbf{x}) - f(\mathbf{x}))^2}{(\max_{\mathbf{x} \in S^\star} z(\mathbf{x}) - \min_{\mathbf{x} \in S^\star} z(\mathbf{x}))^2},$$

and $\mathrm{RoNP}(f)$ is the ratio of null predictions.

2.4 The evolutionary process

The basic algorithm consists of three steps [5, 13]. First, an initial population has to be created randomly. Then, bacterial mutation and gene transfer are applied, until a stopping criteria is fulfilled. The evolution cycle is summarized below:

Bacterial Evolutionary Algorithm

1. create initial population
2. apply bacterial mutation
3. apply gene transfer
4. until stopping condition is not fulfilled, go to 2.
5. return best bacterium

2.5 Generating the initial population

First, an initial bacterium population of N_{ind} bacteria $\{\xi_i, i \in I\}$ is created randomly ($I = \{1, \ldots, N_{\mathrm{ind}}\}$). Figure 1 shows a bacterium ξ_i with $n = 50$ and $m = 5$.

44	17	36	2	7
ξ_i^1	ξ_i^2	ξ_i^3	ξ_i^4	ξ_i^5

Figure 1: A single bacterium

2.6 Bacterial mutation

To find a global optimum, it is necessary to explore new regions of the search space, not yet covered by the current population. This is achieved by adding new, randomly generated information to the bacteria using bacterial mutation.

Bacterial mutation is applied to all bacteria $\xi_i, i \in I$. First, N_{clones} copies (clones) of the bacterium are created. Then, a random segment of length l is mutated in each clone. After mutating the same segment in all clones, all the clones and the original bacterium are evaluated using the evaluation function ϕ. The bacterium with the best evaluation result transfers the mutated segment to the other individuals. This step is repeated until each segment of the bacterium has been mutated once. The mutation may not only change the content, but also the length. The length of the new elements is chosen randomly as $l \pm l^\star$, where $l^\star$ is a parameter specifying the maximal change in length. When changing a segment of a bacterium, we must take care that the new segment is unique within the selected bacterium. At the end, the best bacterium is kept and the clones are discharged. Figure 2 shows an example mutation for $N_{\mathrm{clones}} = 3$ and $l = 1$.

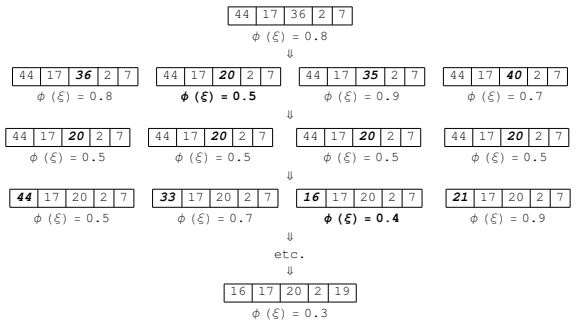


Figure 2: Bacterial mutation

2.7 Gene transfer

The bacterial mutation operator optimizes the bacteria in the population individually. To ensure that information from effective bacteria spreads over the whole population, gene transfer is applied.

First, the population must be sorted and divided into two halves according to their evaluation results. The bacteria with a higher evaluation are called superior half, the bacteria with a lower evaluation are referred to as inferior half. Then, one bacterium is randomly chosen from the superior half and another from the inferior half. These two bacteria are called the source bacterium, and the destination bacterium, respectively. A segment from the source bacterium is randomly chosen and this segment is used to overwrite a random segment of the destination bacterium, if the source segment is not already in the destination bacterium. These two segments may vary in size up to a given length. This ensures—together with the variable length in the bacterial mutation step—that the bacteria are automatically adjusted to the optimal length.

Gene transfer is repeated N_{inf} times, where N_{inf} is the number of "infections" per generation. Figure 3 shows an example for the gene transfer operations ($N_{\text{ind}} = 4, N_{\text{inf}} = 3$).

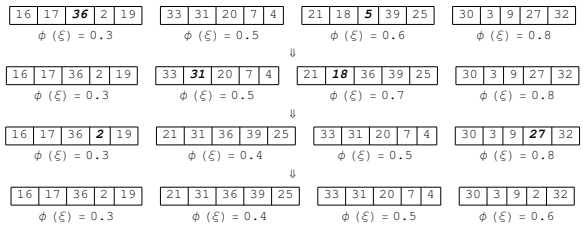


Figure 3: Gene transfer

2.8 Stopping condition

If all individuals in the population are equal or the maximum number of generations N_{gen} is reached, the algorithm ends, otherwise it returns to the bacterial mutation step. Typically, a small number of generations (below 10) already leads to good results. If a target value for evaluation function exists, a threshold value might be defined alternatively.

3 Simulation results

3.1 Performance Comparison

To compare the performance of our approach with other methods, we used seven data sets from the UCI repository [2]. We compared the results obtained with a fuzzy rule base learner *FS-FOIL*, a fuzzy decision tree learner *FS-ID3*, and a fuzzy regression tree learner *FS-LiRT*—all using the same sets of predicates as used in the BEA rules. Furthermore, we used two methods from the WEKA 3-4 toolkit [18], namely *M5-Prime* [15], and *M5-Rules*. *M5-Prime* and *M5-Rules* generate decision trees and decision rules to solve the regression learning problem. The latter two methods do not use a predefined set of predicates/decisions but compute the decision boundaries problem specific. Furthermore, *FS-FOIL* and *FS-ID3* use predefined linguistic output classes, while all other methods compute individual numeric output values for each rule/leaf. For all these methods we disabled local linear models to ensure equal expressiveness of the underlying language. All tests have been carried out using 5-fold cross validation with identical subsets for all test runs. The results of these tests are shown in Fig. 4 where the average normalized mean error, the average ratio of null predictions, and the average number of rules are printed for each test run. For the BEA we used the parameter setting shown in Table 1.

no. of generations	5
no. of individuals	4
no. of clones	5
mutation length	3 ± 2
no. of gene transfers	20
gene transfer length	2 ± 2
max. length	20

Table 1: Parameter setting for the BEA

The results obtained using the BEA are in five of seven cases better than those of the other methods. The trade-off, however, is a decreased coverage (i.e. a higher ratio of null predictions). This is caused by the design of the evaluation function ϕ in Equ. 1, where the ratio of null predictions is a divisor of the normalized mean squared error. This means, that a decrease of x% in coverage is equivalent to an increase of x% of the MSE in terms of the evaluation measure. A high ratio of null predictions as for the servo data set indicates, that a large portion of the data (approx. 30%), shows an "untypical" behaviour and should be analyzed further. Using a different design of the evaluation function could, however, be used to obtain a rule base with a higher coverage, but also a higher MSE.

NMSE RoNP Size	*FS-FOIL*	*FS-ID3*	*FS-LiRT*	*M5P*	*M5Ru*	*BEA*
autoMpg	0.019 0.357 16.6	0.017 0 36.	0.014 0 22.8	**0.012** 0 16.	0.013 0 11.8	0.016 0.096 21.2
autoPrice	0.018 0.071 20.8	0.021 0 34.	0.017 0 12.	0.023 0 8.6	0.023 0 7.4	**0.014** 0.0131 20.2
bodyFat	**0.003** 0.012 7.	0.006 0 5.	0.008 0 3.6	0.007 0 18.6	0.01 0 17.6	0.008 0.032 18.6
cloud	0.042 0.038 28.8	0.031 0 25.4	0.044 0 8.2	0.039 0 7.2	0.044 0 6.2	**0.025** 0.048 17.6
housing	0.015 0.154 21.8	0.014 0 21.	0.012 0 17.	0.012 0 18.2	0.016 0 13.	**0.011** 0.048 19.8
servo	0.017 0.364 19.2	0.023 0 44.	0.012 0 9.6	0.028 0 10.6	0.042 0 6.8	**0.006** 0.309 23.
veteran	0.116 0.185 22.	0.084 0 65.4	0.161 0 19.6	0.053 0 1.8	0.055 0 1.8	**0.051** 0.185 20.6

Figure 4: Comparison of results for 7 UCI data sets

3.2 Housing data

Let us take a closer look to the results obtained for the housing data set. Initially 120, 98, 429, 232, and 50 rules (929 in total) were created for each of the five goal classes using *FS-Miner*. The partitioning of the input

domains and the definition of the corresponding fuzzy predicates were done using *CompFS*. The partitioning of the goal attribute is shown in Fig. 5.

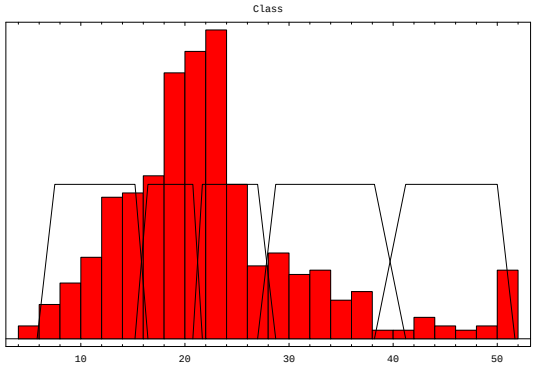


Figure 5: Partitioning of the class parameter in the housing data set

Afterward we applied the BEA to obtain the final rule set. In the first of five trial runs 4, 4, 9, 2, and 1 rules (total 20 rules) involving only two or three predicates have been selected. The resulting rule base is shown in Fig. 6. The additional columns show the number of correctly classified samples (tt), the number of misclassified samples (tf), the number of unclassified samples of the goal class (rt), and the corresponding confidence and support.

4 Conclusions

In this paper we have shown how bacterial evolutionary algorithms can be applied to identify the optimal subset of rules for a given regression learning problem. Bacterial evolutionary algorithm seems to be more efficient than the standard genetic algorithms. The reason for that is the different nature of the operations in the BEA ([13]). Bacterial mutation is more effective than classical mutation in GA because of the cloning procedure. Every clone brings a new chance to find a better solution anywhere in the search space, thus wider space can be explored. In the gene transfer we do not lose good individuals, because the information flow is directed from the superior sub-population to the inferior one. The algorithm presented is capable of optimizing freely definable goal functions which enables the explicit formulation of interpretability quality measures. We have shown, that although the underlying language (i.e. the predicates used) is very simple, we are often capable of finding better solutions than by traditional top-down approaches.

Future work will be concerned with implementing a parallel version of the BEA. We hope, that using a parallel implementation we can also solve large real-world applications within reasonable time. Furthermore, we want to study different evaluation functions which allow the user to specify his needs more easily.

References

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, and C. Zaniolo, editors, *Proc. 20th Int. Conf. on Very Large Data Bases*, pages 487–499. Morgan Kaufmann, 1994.

[2] C. L. Blake and C. J. Merz. UCI repository of machine learning databases. Univ. of California, Irvine, Dept. of Information and Computer Sciences, 1998. http://www.ics.uci.edu/∼mlearn/MLRepository.html.

[3] U. Bodenhofer. *Fuzzy Orderings — Theory and Application Perspectives*. Studies in Fuzziness and Soft Computing. Physica-Verlag, Heidelberg.

[4] J. Botzheim, M. Drobics, and L. T. Kóczy. Feature selection using bacterial optimization. In *Proc. 10th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 797–804, Perugia, July 2004.

[5] J. Botzheim, B. Hámori, L. T. Kóczy, and A. E. Ruano. Bacterial algorithm applied for fuzzy rule extraction. In *Proc. Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 1021–1026, Annecy, FR, 2002.

[6] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors. *Interpretability Issues in Fuzzy Modeling*, volume 128 of *Studies in Fuzziness and Soft Computing*. Springer, Berlin, 2003.

[7] M. Drobics. Choosing the best predicates for data-driven fuzzy modeling. In *Proc. 13th IEEE Int. Conf. on Fuzzy Systems*, pages 245–249, Budapest, July 2004.

[8] M. Drobics. *Data Analysis Using Fuzzy Expressions*, volume C 48 of *Schriftenreihe der Johannes-Kepler-Universität Linz*. Universitätsverlag Rudolf Trauner, 2005.

[9] M. Drobics and J. Himmelbauer. Creating comprehensible regression models—inductive learning and optimization of fuzzy regression trees using comprehensible fuzzy predicates. *Soft Computing*, 11:421–438, 2006.

[10] T.-P. Hong, K.-Y. Lin, and S.-L. Wang. Fuzzy data mining for interesting generalized association rules. *Fuzzy Sets and Systems*, 138(2):255–269, 2003.

Class	tt	tf	rt	conf	supp	Condition
class_Is_0_VL	73.38	23.26	31.54	0.76	0.15	LSTAT_IsAtLeast_H && CRIM_IsAtLeast_L
	82.24	28.48	22.78	0.75	0.16	LSTAT_IsAtLeast_H && PTRATIO_IsAtLeast_H
	9.36	2.32	95.34	0.82	0.02	CRIM_IsAtLeast_H && RM_IsAtLeast_H
	26.98	4.97	77.55	0.85	0.05	TAX_Is_VH && B_Is_VL
class_Is_1_L	69.5	43.28	76.85	0.63	0.14	RM_IsAtMost_L && NOX_IsAtMost_M
	26.72	10.26	118.2	0.73	0.05	RM_Is_L && DIS_IsAtLeast_H
	22.43	10.03	122.66	0.71	0.04	LSTAT_Is_M && TAX_Is_H
	61.41	39.59	85.01	0.62	0.12	LSTAT_Is_M && INDUS_IsAtLeast_L
class_Is_2_M	58.82	32.67	90.87	0.66	0.12	RM_Is_M && NOX_IsAtMost_M
	60.26	36.72	89.35	0.63	0.12	RM_Is_M && CRIM_Is_VL
	23.3	11.28	124.78	0.68	0.05	LSTAT_IsAtMost_L && RM_IsAtMost_H && ZN_Is_L
	8.72	4.03	139.09	0.7	0.02	TAX_Is_L && CHAS_Is_1
	26.52	11.25	122.11	0.72	0.05	LSTAT_Is_L && TAX_Is_L && NOX_IsAtMost_L
	5.62	2.65	142.57	0.73	0.01	LSTAT_Is_L && CRIM_Is_L
	8.81	6.08	139.48	0.62	0.02	LSTAT_IsAtLeast_M && AGE_Is_L && B_Is_VH
	54.85	28.96	93.47	0.66	0.11	TAX_Is_L && AGE_IsAtMost_M && RM_IsAtMost_H
	50.	19.18	99.22	0.74	0.1	RM_Is_M && PTRATIO_IsAtMost_M && DIS_IsAtLeast_L
class_Is_3_H	13.31	2.07	60.12	0.89	0.03	ZN_IsAtLeast_L && RM_Is_H && TAX_Is_VL
	31.61	16.75	42.21	0.66	0.06	LSTAT_Is_VL && RM_Is_H && RAD_IsAtMost_H
class_Is_4_VH	9.33	5.69	22.37	0.62	0.02	RM_IsAtLeast_H && PTRATIO_Is_VL && DIS_IsAtMost_L

Figure 6: Result obtained for the housing data set

[11] H. Ishibuchi and T. Yamamoto. Fuzzy rule selection by data mining criteria and genetic algorithms. In *Proc. of Genetic and Evolutionary Computation Conference*, pages 399–406, 2002.

[12] D. Nauck. Measuring interpretability in rule-based classification systems. In *Proc. 3rd IEEE Int. Conf. on Fuzzy Systems*, pages 196–201, Tokyo, October 2003.

[13] N. E. Nawa and T. Furuhashi. Fuzzy system parameters discovery by bacterial evolutionary algorithm. *IEEE Trans. Fuzzy Syst.*, 7:608–616, 1999.

[14] C. Olaru and L. Wehenkel. A complete fuzzy decision tree technique. *Fuzzy Sets and Systems*, 138(2):221–254, 2003.

[15] J. R. Quinlan. Learning with Continuous Classes. In *Proc. 5th Austr. Joint Conf. on Artificial Intelligence*, pages 343–348, 1992.

[16] G. Ridgeway, D. Madigan, and T. Richardson. Boosting methodology for regression problems. In *Proc. Artificial Intelligence and Statistics*, 1999.

[17] M. Setnes and H. Roubos. GA-fuzzy modeling and classification: Complexity and performance. *IEEE Trans. Fuzzy Systems*, 8(5):509–522, October 2000.

[18] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools with Java implementations.* Morgan Kaufmann, 2000.

[19] Y. Yi and E. Hüllermeier. Learning complexity-bounded rule-based classifiers by combining association analysis and genetic algorithms. In *Proc. Joint 4th Conf. of the European Society for Fuzzy Logic and Technology and 11 Recontres Francophones sur la Logique Floue et ses Applications*, Barcelona, September 2005.

Session 7

Fuzzy Sets in Distributive Artificial Intelligence – A. Averkin

Fuzzy multiagent distributed assembly chart planning in agriculture*

Tikhonov E.
Russia, Msk. reg., Dmitrov,
Vnukovsky mkr., 13 45
yevgesh@mail.ru

Abstract

Agriculture in Russia has some specifics such as large distribution, inclement climate and big market competition. Appearing of resources-economy, precision and management technologies demanded from agriculture companies new planning and control methods. The paper is dedicated to assembly charts planning in agriculture with multi-agent approach. For resources distribution we used method of virtual auction and fuzzy logic controller. JACK Intelligent Agents platform is used for implementation.

Keywords: multi-agent technology, production planning, agriculture, fuzzy logic, JACK Intelligent Agents.

1 Introduction

For agriculture production in inclement and unstable climate conditions it's very important to react intelligent on different unexpected situations. As usual season plans of cultivating crops differs from results at the end of the year. Plans can change some times a day because of climate changes, breakdowns, supply delays or just a human factor.

The main agriculture instrument for planning is assembly chart which helps economists find out expenses and profits of nurturing specific crop. This chart includes land, aggregates, machines, tractors, personal and technological operations. Working the assembly chart up is serious problem especially in large agro-holdings and when taking into account quality figures of aggregates, time component, land condition and changing environment. These requirements make a further stress to the agriculture production planning system, which must be dynamically adaptable to both local and distributed utilization of production resources and materials.

In this work the multi-agent approach applied for agriculture assembly chart planning discussed in [1-3]. Planning system is implementing with Java based JACK Intelligent Agents platform.

2 Model

Agriculture production to stay competitive faces the problem of prime cost reduction using various information about soil, machines, fertilizers. The effects of these trends can be summarized as increasing complexity and the need to respond to continual change under decreasing costs. To meet these new business challenges, manufacturing operations require additional functionality, like robustness, scalability or reconfigurability, while maintaining simple and transparent processes [3].

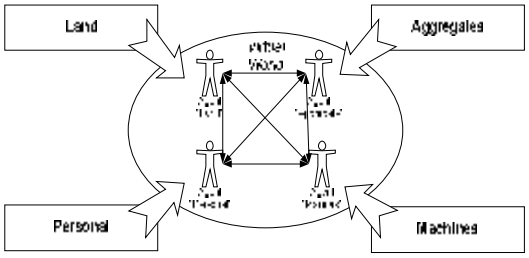


Figure. 1. Presentation entities of model as communicating agents.

Such models could be applied as self-organized intelligent agents in virtual worlds [4] where each agent is an entity of the model with its own properties. And it could cooperate with other agents to achieve own goals.

Let's represent main objects of our assembly chart model with intelligent agents which are able to communicate with each other. For assembly chart main objects are: lands, aggregates, machines and personal (Fig. 1). There could be unlimited number of agents of these types in the model. Main properties of land, aggregate and machine agents shown in figure 2.

Each operation takes special aggregates, machines and personal in order to grow the crop. But the quality and cost of carrying the task out is different on various lands. And the task handling may be done with different quality and cost as well. So we need to plan choosing aggregates and machines for land processing in the way to maximize quality and minimize prime-costs.

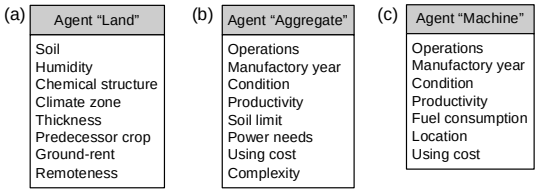


Figure 2. Model objects properties: (a) Land, (b) Aggregate, (b) Machine

This approach can help easily renegotiate, re-plan and make right and coordinated decision.

We included "Expert" and different crops (wheat, potato, grass and etc.) agents as well representing experts in the system.

3 Agent architecture

Among the main requirements of an autonomous agent is the ability to perform means-end reasoning, i.e. the ability to select a course of actions that ultimately achieves the goals of the agent. There has long been a strive in the Artificial Intelligence community for an agent architecture, that enables selection of a course of actions that ultimately achieves the goals of the agent. This has led to many approaches to practical reasoning, where the most notorious and respected of which is the agent model known as Beliefs, Desires and Intentions [5].

The BDI kernel, seen in figure 3, is the pivot and functions as the interpreter. Its execution cycle is as follows: At time t: certain goals are established and certain beliefs are held. An event occurs that alters the beliefs or modify the goals, and the new combination of goals and beliefs trigger plans. If the required capabilities exist, the goal that was triggered by the event, is placed in the intentions structure.

As the main objective of agent is planning we use fuzzy module to determine if the model object represented by agent capable to do necessary tasks based on current conditions. For this purpose the fuzzy logic controller is used.

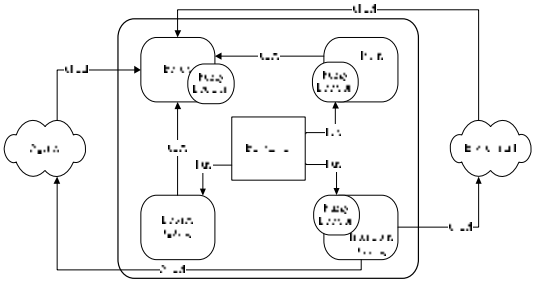


Figure 3. The BDI model with the addition of making fuzzy decision support, interacting agents and the environment.

An executable plan is found and executed, whereby the agent performs the action. Performing the action may change the environment and establish new goals and beliefs, and so the BDI kernel's execution cycle starts again.

3.1 Fuzzy controller

The main purpose of fuzzy logic controller use is to increase flexibility of choosing contractor for agriculture operations maintaining. It is also used to enlarge agent beliefset utilization quality as shown in fig. 3 and fig. 4.

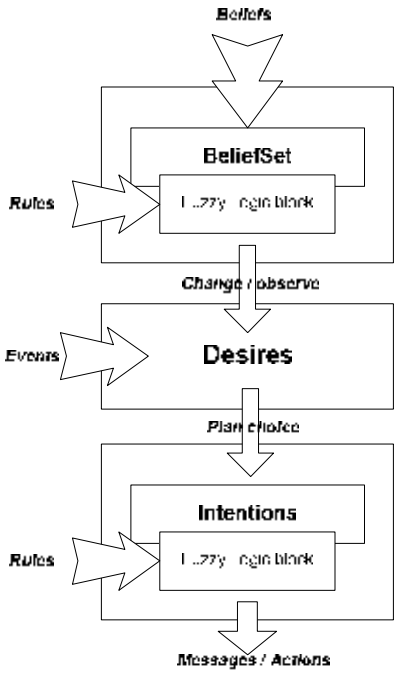


Figure 4. The BDI model with fuzzy logic blocks, introspection steps.

Taking into account the fact that each agricultural technologic operation needs fixed soil condition it is very convenient to use fuzzy logic block with the

support of passing needed knowledge as rules. The same situation we observe on aggregates and machines which have common properties but show them in different ways.

Why we use fuzzy logic controller? The answer to this question comes from the fuzziness paradigm where one or more properties have fuzzy limits. This problem corresponds with agriculture. There is also a lot of indeterminism in agriculture mostly because of natural production conditions. Also fuzzy logic tool is very convenient as it allows reducing (some kind of data fusion) to one grade diversified factors (qualitative and quantitative) what is very valuable around agriculture experts (agronomists).

The main principles of fuzzy logic is perfectly described in [6],[7].

Figure 5 shows structure of fuzzy control module used in our agents. It consists of rule base, fuzzification part, inference block, defuzzification part. But using communicating agents allow us on-line download or tune parameters of fuzzy controller. That also allows to take opportunity of available knowledge to improve processes and real-time decision making.

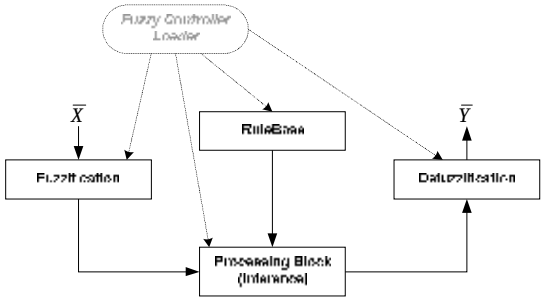


Figure 5. Fuzzy controller.

Fuzzy control engine is build into agent's architecture but setting of system parameters mainly happens during work process depending on current internal state, beliefs and environment condition.

For example, agent "Field *i*" starting agricultural production process with current soil parameters (predecessor, chemical structure, humidity and etc.). It requires some knowledge to decide if soil parameters favorable for starting technological process operations for specific crop. Our agent system divide this problem into five steps as shown in figure 6.

Main steps are:

1. Sending request to expert agent ("Expert") ,
2. Agent "Expert" requesting all known crop agents (wheat, potato, etc),
3. Crop agent estimates suit coefficient based on information attached to request message. This coefficient also may be evaluated using fuzzy logic controller. It is supposed that crop agent has appropriate knowledge about suitable conditions,
4. Sending evaluated coefficients to "Field *i*" agent with necessary knowledge about operation starting parameters for soil,
5. Receiving estimated coefficients and choosing the best one (highest possible). Loading knowledge into fuzzy controller about "crop' and beliefset monitoring for operation starting.

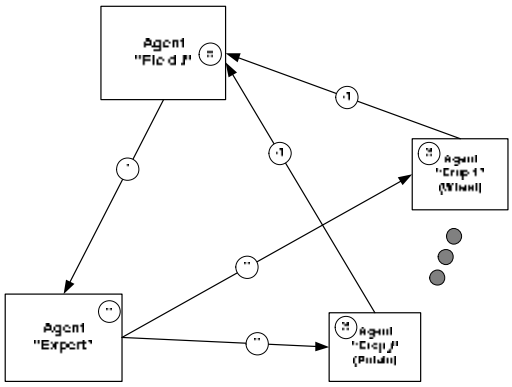


Figure 6. Agent "Field *i*" knowledge obtaining process.

Knowledge passes through agent's network as a set of linguistic terms and variables, membership functions and collection of linguistic rules to attain certain objectives in the form of IF-THEN rule with condition and conclusion.

We used *jFuzzyLogic* package for embedding fuzzy control block into agent. *jFuzzyLogic* is a fuzzy logic package written in java [8]. *jFuzzyLogic* is a java implementation of a Fuzzy Logic software package. It implements a complete Fuzzy inference system (FIS) as well as Fuzzy Control Logic compliance (FCL) according to IEC 1131 [9]. It will allow for real-time packet analysis and can be integrated into other Web-based or Agent-based Network Tools.

The Fuzzy Control applications programmed in Fuzzy Control Language FCL are encapsulated in Function Blocks (or Programs). The Function Block Types defined in Fuzzy Control Language (FCL) shall specify the input and output parameters and the Fuzzy Control specific rules and declarations. The corresponding Functions Block Instances shall contain the specific data of the Fuzzy Control applications.

All descriptions in FCL are enclosed between FUNCTION_BLOCK, END_FUNCTION_BLOCK statements. For example, fuzzification of "humidity" parameter with FCL may be like

```
VAR_INPUT                // Define input variables
    humidity : REAL;
    ...
END_VAR

FUZZIFY humidity            // Fuzzify input variable 'humidity': {'dry' , 'average' , 'damp'}
    TERM dry := (0, 1) (40, 0) ;
    TERM average := (30, 0) (50,1) (60,1) (80,0);
    TERM damp := (70, 0) (90, 1) (100, 1);
END_FUZZIFY.
```

Defuzzification part in FCL also has its method and looks like

```
DEFUZZIFY tip                    // Defzzzify output variable 'tip' : {'poor', 'middling', 'best' }
    TERM poor := (0,0) (20,1) (40,0);
    TERM middling := (30,0) (55,1) (80,0);
    TERM best := (70,0) (90,1) (100,0);
    ACCU : MAX;              // Use 'max' accumulation method
    METHOD : COG;            // Use 'Center Of Gravity' defuzzification method
    DEFAULT := 0;         // Default value is 0 (if no rule activates defuzzifier)
END_DEFUZZIFY.
```

Production rules contain into section "RULEBLOCK" and include directions for activation and accumulation.

```
RULEBLOCK No1
    AND : MIN;               // Use 'min' for 'and' (also implicit use 'max' for 'or' to fulfill DeMorgan's Law)
    ACT : MIN;               // Use 'min' activation method
    RULE 1 : IF humidity IS dry AND thickness IS solid THEN tip IS poor;
    RULE 2 : IF humidity IS dry AND thickness IS average THEN tip IS poor;
    RULE 3 : IF humidity IS dry AND thickness IS soft THEN tip IS middling;
    ...
END_RULEBLOCK
```

Actually we can implement much more features as it provided in jFuzzyLogic package: weighting factor, subconditions, various membership functions, defuzzification, accumulation and aggregation methods and etc.

4 Planning process

The plan is created collectively by a community of simple planning agents that use a sophisticated auction-based negotiation, supported by use of the social knowledge and acquaintance models. The core of our system is a community of planning agents which making production plans for individual orders, taking care of conflicts and managing re-planning and plan reconfiguration.

A stable and industry accepted approach to the coordination of agents' joint activity is based on clear cut roles (even temporary) in the multi-agent community. Let us have a coordinator ("field" agent) who is in charge of proper task decomposition and subcontracting contractors for implementing components of the tasks. A classical and industry accepted negotiation algorithm is contract-net-protocol. There is used the simplified version of contract-net in our system.

Any agent (will become a coordinator) can initiate the contract net by requesting some contractors for specific services. Each contractor carries out its own internal reasoning and suggests a collaboration proposal.

Planning process starts at the problem of choosing appropriate crop and sort of the crop according soil conditions on specific lands. It's proposed that land agent knows its main parameters. They can be achieved from experts (agronomists), extracted from previous land use or obtained from special sensors [10]. But land agent can't initiate production process in this situation as it doesn't know anything about what are to grow. So the land agent has to resort to the help of other agents through the communication process.

At figure 6 it's shown how knowledge can be obtained as a set of production rules, linguistic terms and methods for use with fuzzy controller. Fuzzy

support in this case is very valuable and helps us to make a reasonable decision taking into account economical, agricultural and ecological factors.

When essential knowledge has loaded agent can proceed to soil condition controlling. This operation lean on data gathered from sensors or received with messages (fig. 7). When auspicious conditions take place event is send and agent starts aggregates selection. Conditions are considered favorable when estimated with fuzzy controller grade after defuzzification reaches given threshold or gets into interval (fig. 8,9).

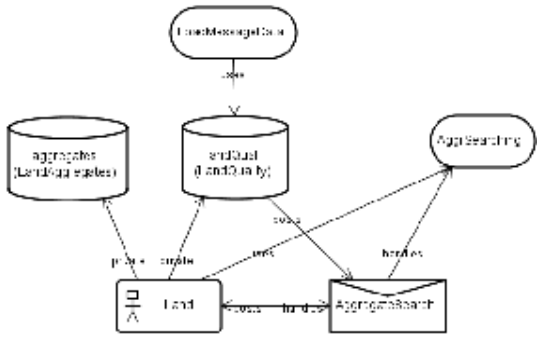


Figure 7. Soil controlling and operations initiation diagram.

Figures 8 and 9 demonstrate our approach for the case when two parameters are controlled (“humidity” and “thickness”) and threshold estimates (“tip”).

Input values are normalized to [1..100] interval and generated randomly. There is defuzzifier chart for the case of values 65 (“humidity”) and 75 (“thickness”) in figure 9.

Classically operation of crop and crops’ sort choosing is done by experts before. As well all operation periods are selected by agronomists.

After choosing the sort and, of the crop and coming appropriate conditions agent starts looking for resources necessary for undertaking the operation. In this case agent initiate choosing plan (“AggrSearching”) through posting event (“AggregateSearch”). The plan is responsible for negotiating all known aggregates (“aggregates”) by sending bids for carrying current operation out (collaboration).

Received message initiates estimate computation based on its own beliefs, obtained operation data and appropriate knowledge (fig. 10) with the help of fuzzy logic controller. With this estimate we can measure the extent of convenience using current aggregate.

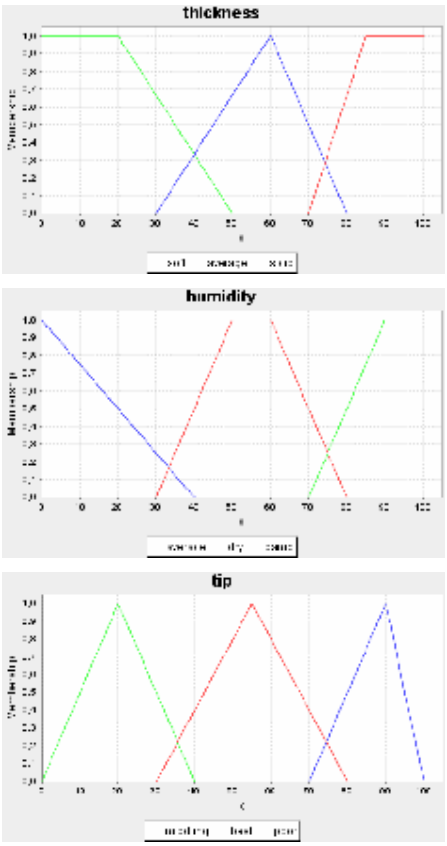


Figure 8. Charts for linguistic variables in the ruleSet: (top) input var. “thickness”, (middle) input var. “humidity”, (bottom) output var. “tip”.

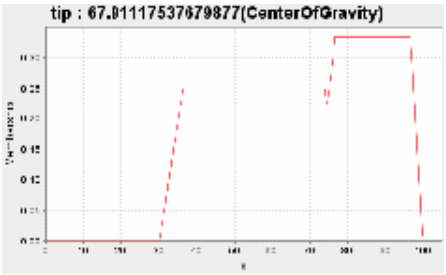


Figure 9. Defuzzifier chart.

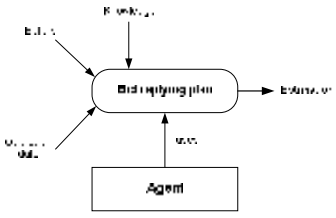


Figure 10. Estimation process diagram.

In picture 11 estimating plan is called “Choose” and after computation it becomes the contactor for sub collaborator finding. It is necessary for the best tractor selection in addition to aggregate (fig. 12).

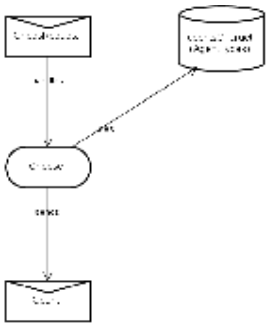


Figure 11. "Data-event-plan" diagram for aggregate agent.

While standard resource distribution approach is based on market mechanism, our estimations point of view allows take qualitative characteristics into consideration.

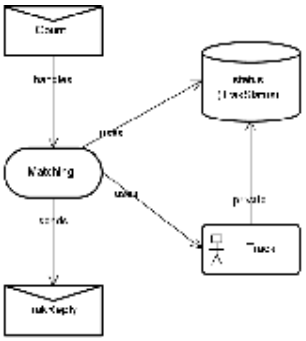


Figure 12. "Data-event-plan" diagram for tractor agent.

Upon receiving proposals for collaboration, the coordinator carries out a computational process by which it selects the best possible collaborator(s) – see Figure 13. The contract net protocol can be also multi-staged. For each single-staged communication within a community of n field agents, it is needed to send $2(n + 1)$ messages in the worst case.

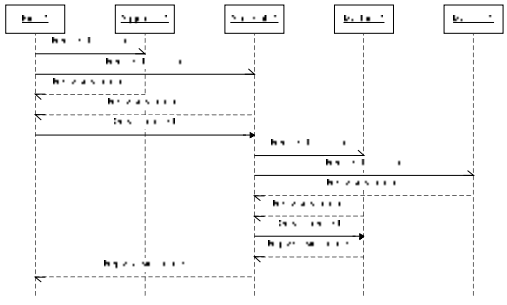


Fig. 13. Contraction based on a simplified Contract Net.

When the requesting messages are sent by agents, information about sending agent and current task conditions enclosed into message. This received information with agents' beliefs and conditions is used in fuzzy module to determine quality of undertaking the task.

This can help us eliminate from using low quality aggregates and machines even if theirs prime-costs are lower than others. This follows the fact that yield loss could be more than expensive on qualitative cultivation.

When the replies with quality mark and using cost are received contactor chooses the best variant for adding to overall plan and "sign a contract" with all collaborators.

For implementation we are using Jack Intelligent Agents framework [11],[12]. JACK is an important and novel contribution to the field of agent-oriented software engineering. Rather than invent a new language, an existing popular language (Java) has been augmented with constructs for agent communication, knowledge representation, and for both deliberative (goal-based) and reactive (event-based) programming. This has been achieved in a way that allows the programmer to mix familiar Java statements with agent programming constructs. Although JACK is strongly oriented toward the BDI paradigm, its component-based architecture supports a wide variety of agent programming styles. Also JACK architecture allows embedding necessary java modules for extension agent's possibilities.

5 Related work

It's supposed to progress development of planning agents in agricultural sector. In particular agent training ability, classification and wide past experience use are in our sight to implement for more intelligent decision making support. Also it is planned to add into system the ability of expert consulting on agricultural machinery selection problem according to local production conditions and market limitations.

Conclusions

The research described in this paper contributes to the multi-agent planning in agriculture sector with distributed knowledge using. This approach opens for farming production new horizons in basic technological planning that was mainly in experts interests.

Our research also has been driven by the idea of embedding the fuzzy logic controller into agent. Transferring knowledge as a set of production rules helps agriculture agents make adequate and soft decisions in changing environment. For this case FCL was adopted with agents.

**The paper is supported of RFBI grants 06-01-00576 and 07-01-00782*

References

[1] Michal Pechoucek, Agent Based Production Planning, PLANET News, 2003.

[2] Rui M. Lima, Rui M. Sousa, Paulo J. Martins, Distributed Production Planning and Control Agent Based System, International Journal of Production Research, 2006, 44(18 - 19): 3693-3709.

[3] Duncan C. McFarlane, Stefan Bussmann Developments in Holonic Production Planning and Control, Int. Journal of Production Planning and Control, Vol. 11, No. 6, 2000, pp. 522 –536.

[4] Skobelev P.O., Multi-agent approach to intelligent system building of coordinated agriculture business managing. 2000.

[5] Jakob Appel, Kurt Nielsen, Applied Multi-Agent Systems Principles to Cooperating Robots in Tomorrow's Agricultural Environment, PhD Thesis, 2005.

[6] V. Novak, I. Perfiliefa, J. Mockor, Mathematical principles of fuzzy logic, Kluwer Academic Publishers, 1999.

[7] D. Rutkovsky, M. Pilnsky, L. Rutkovsky, Neural networks, genetic algorithms and fuzzy systems, Naukowe PWN, 2004.

[8] http://jfuzzylogic.sourceforge.net/

[9] IEC 1131 - PROGRAMMABLE CONTROLLERS Part 7 - Fuzzy Control Programming Committee Draft CD 1.0 (Rel. 19 Jan 97).

[10] Spyridon Fountas, Systems Analysis of Precision Agriculture, PhD thesis, 2004.

[11] Rick Evertsz, Martyn Fletcher, Richard Jones, Jacquie Jarvis, James Brusey, and Sandy Dance. Implementing Industrial Multi-Agent Systems using JACK, 2005.

[12] http://www.agent-software.com/

SYNTHESIS OF DISTRIBUTED FUZZY HIERARCHICAL MODEL IN DECISION SUPPORT SYSTEMS IN FUZZY ENVIRONMENT*

Averkin A.N.
Computer center of the Russian Academy of Science
averkin@ccas.ru

Titova N.V.
Computer center of the Russian Academy of Science
titova_nina@inbox.ru

Agrafonova T.V.
Computer center of the Russian Academy of Science
solnishko3@yandex.ru

Abstract

As the result of the development of Fuzzy Multiple Criteria Decision Making (FMCDM) with the help of fuzzy set theory a number of innovations have been made possible. The new approach of FMCDM – Fuzzy Hierarchical Modeling - is introduced. It is shown in the article, how to use the fuzzy hierarchical model with other methods of FMCDM. Advantages of the method are described also. We propose a novel approach to overcome the inherent limitations of Hierarchical Methods by exploiting multiple distributed information repositories.

Keywords: Fuzzy multiple criteria decision making, fuzzy hierarchical modeling.

1 Distributed Fuzzy Hierarchical Model

The basis of the model is the hierarchical structure of the factors, which was received as a result of function-structured decomposition of the data domain [1, 4-6]. The meta-levels of the structure are the following: first level is the level of the global aims, the second level is the level of the rival's aims, and the third level is the level of the measures for the achievement of the global aims and rivals aims removal. The last level is the level of the concrete actions. The links in the hierarchy define the dependency of the upper level element realization from the corresponding underlying level element. Thereby, the realization of possible measures for the achievement of the aim depends on some concrete undertaken actions. This hierarchy allows evaluating the importance of all the elements of the level taking into consideration their contribution in the top levels elements realization. The hierarchical structure analysis model allows to process local factors estimations. These estimations have, as a rule, fuzzy and inconsistent nature, got from sources of different reliability (from expert with different competence level). This hierarchical model also allows to get total global consistent and reliable in the sense of theories of the fuzzy sets estimations. Thereby, each decision will be characterized by its importance taking into consideration its role in the factors structure. But, such a decision characteristic is insufficient for all-round estimation. The additional characteristics of the decision, such as its realization in economic, social, politic etc senses, must be considered. But, the hierarchical model can find the most needed decision in the current situation.

We propose a novel approach to overcome the inherent limitations of Hierarchical Methods by exploiting multiple distributed information repositories. The construction of fuzzy hierarchical model can be distributed between a number of experts. They may work to a single domain or to different domains. Also the distributed computational methods are used for making the expert estimation and for receiving the result.

2. Hierarchy Analysis Methods

The special role in the complex object analysis plays the analysis of the factors links graph's structure (the graph has the form of ordered hierarchical ranked structure). Directly influenced factors are situated on the graph's last level. The realization of these factors (as a rule they represent the concrete actions), spreading upwards on consecutively located levels of the factors hierarchical structure, will bring into the realization of all above located factors and, finally, - to the achievement of the global aims of the considered complex object development. At the moment, the strict statement of the hierarchy multilevel factors structure building problem doesn't exist. But, it is possible to indicate the principles of its practice construction. These principles are formulated in the form of six necessary conditions, which must satisfy considered hierarchical structures. It is naturally, that real hierarchical structures will satisfy these conditions only in certain measure, which depends on the used methods and algorithms of their formation.

1. Hierarchical factors structures are built on the base of the profound sense of used fac-

tors; the factors in the underlying level reveal the sense of the upper level factors, or the underlying level factors represent the events, which realization promotes the realization of upper level factors.

2. The realization of some of the factors, lying on the same level, must not influence the realization of the other factors of this level. In other words, the factors of the same level must be independent from each other.
3. Factors on the considered level directly depends only on the factors of the nearest underlying level of the hierarchy.
4. Fullness of the factors uncovering: factor on the considered hierarchy level is completely realized, if all the influencing its realization factors of the next underlying level are also realized.
5. Positive relationship between the upper level factors and underlying level factors: the realization of the underlying level factors must not provide the reduction of the realization possibility of the upper level factors.
6. Linearity of the functional links between the adjacent levels factors.

3 Analysis of a Hierarchy with Fuzzy Estimations

First of all, we should build the hierarchy. On the objects set $Z = \{1, 2, \ldots, N\}$ is defined the oriented graph $G_r = (Z, W)$ without cycles with the vertexes set coinciding with the objects set, and the arcs set W. The presence of the arc $(i, j) \in W$ means that the weight z_i of the object (vertex) i directly depends on the weight z_j of the object j. The graph G_r has the structure of the purposes and tasks graph of some complex system, if all the vertexes of this graph can be located on non crossing levels $V_1, \ldots, V_M$ in such a way, that the graph's arcs connect only the vertexes of the adjacent levels and these arcs lead from top to bottom, from the level V_i to the level V_{i+1}, i = 1, …, M-1; the vertexes, from which arcs don't leave, are located on the level V_M; all the vertexes, in which arcs do not enter, are located on the level V_1. The construction of the hierarchy is one of the most difficult stage because of the difficult formalization of the used objects, such as aims, rivals aims etc. After hierarchy construction, the elementary estimations should be made by experts. The elementary estimation consists on the getting for certain vertex $i \in V_m$ paired estimations $r_{jk}^{(i)}$ of the arcs weights $(i, j) \in W$, $j \in \Gamma_i = \{k \mid (i, k) \in W\}$. Paired estimations show, in how many times the contribution of the object j is more than the contribution of the object k in the achievement of the object i aim; $j, k \in \Gamma_i$. These estimations can be exact ($r_{jk}^{(i)} \in R_+$ - nonnegative numbers), interval ($r_{jk}^{(i)} = [a_{jk}^{(i)}, b_{jk}^{(i)}] \subset R$ - intervals) or fuzzy numbers ($r_{jk}^{(i)} = \{(t, \mu_{jk}^{(i)}(t)) \mid t \in R_+\}$ – closed convex fuzzy sets on R_+). The last case includes the linguistic estimates and two previous cases. Thereby, we get as a result of an elementary estimation an weighted binary relation $R^{(i)} = \{((j, k), r_{jk}^{(i)}) \mid j, k \in \Gamma_i\}$ on the objects set Γ_i, which gives the intensity of the objects superiority. After getting the estimations, we must average them. In each of the elementary estimations several experts can participate, so for some pairs (j, k) of the objects $j, k \in \Gamma_i$ different experts s can assign different estimations $r_{jk}^{(i)s}$ (s – expert's number). The procedure of the expert estimation averaging consists in the determination of the mean geometric estimation.

4 Hierarchic Structure Arcs Weights Determination

The result of the pairs estimations average in the i elementary estimation – exact, interval or fuzzy relation $R_{(i)}$ – is used in the determination of the weights $y_{i,j}$, of all the arcs $(i, j) \in W$, coming out of the vertex i. The arcs weights satisfy the following condition:

$$\sum_{j \in \Gamma_i} y_{ij} = 1; \quad y_{ij} \geq 0, \quad \forall i \in \Gamma_i .$$

If there are several objects on the first level y_1, then the "zero" elementary estimation is made, it means, that the pair comparison of the objects importance coefficients must be made. As a result of the "zero" estimation, the importance coefficients of the first level objects are determined.

5. The Importance Coefficients Determination

After the elementary estimations results processing, the importance coefficients z_j of the objects $j \in V_1$ of the first level of the hierarchic structure are determined. And also the weights y_{ji} of all the arcs $(i, j) \in W$ are determined (the coefficients of the relative importance of the vertex $Y_{ji}^{(s)}$ for the vertex $Y_i^{(s-1)}$ of the nearest upper level, where s – is a level number. The weights of the underlying level objects are determined by the recurrence from top to bottom recalculation of the objects weights (objects importance coefficients):

$$z_i = \sum_{j \in \Gamma_i^{-1}} y_{ji} z_j, \; i \in V_2,$$

……

$$z_i = \sum_{j \in \Gamma_i^{-1}} y_{ji} z_j, \; i \in V_M$$

($\Gamma_i^{-1} = \{j \mid (j, i) \in W\}$).

6 The Different Experts Estimations Consensus Analysis

The coefficients importance validity is determined by the elementary estimations results validity. In the case, then the initial pairs estimations are fuzzy or mixed, the results validity is equal to the consensus degree of the initial fuzzy relation $R^{(i)}$ and the resulting over transitive matrix, which is determined as a result of a special estimations approximation problem solution. The solution of the estimation approximation problem is made, using a modified method of Makeev and Shahnov [2-3]. In the case, then the estimations are exact or interval, the results validity is characterized by the degree of the intervals bounds changes, which are assigned by the experts.

7 Conclusion

Decomposition in Hierarchic Model is made until the level which contains factors with qualitative or quantitative scale of values. To apply the FMCDM methods the construction of every scale reflection to [0,1] is needed. It means that it necessary to create membership function that will convert every value from the scale to real number from [0,1]. The number is interpreted as preference of selected factor value for the main hierarchical goal (factor of the upper level) achievement. Zero is interpreted as index of minimum preference than One is interpreted as index of maximum preference.

On the basis of relative importance weights it is possible to construct unified scale for the scale gradations of the factors. Using FMCDM methods allows as getting the preference coefficient of the alternative, it means the preference coefficients of last level factors value collection.

**The paper is supported of RFBI grants 06-01-00576 and 07-01-00782*

References

[1] Saathy T, Kerns K. Analytic Planning. Systems - M: Radio i Svyaz. 1991.

[2] Makeev S.P., Shahnov I.F. Alterna tives sequencing on the base of fuzzy estimations. - M.: CC AS USSR, 1989.

[3] Makeev S.P., Shahnov I.F. Objects sequencing in the hierarchical systems // Izvestia AS USSR. Technicheskaya Cybernetika, 1991, N 3, c. 29 – 46.

[4] Jensen R.E., 1984, Comparison of eigenvector, least squares and logarithmic least squares methods, EJOR, 127 p.

[5] Saathy, T.L., Vargas, L.G. Inconsistency and rank preservation // Journal of Mathematical Psychology, vol. 28, pp. 205 - 214, 1984.

[6] Saaty, T.L., Multicriteria Decision Making - The Analytic Hierarchy Process, RWS Publications, 2nd edition, 1990.

Fuzzy calculating and fuzzy control in wireless sensor network

Irina Kalganova
Russia, Moscow, 127434, Krasnostudencheskiy proezd, 4-110
irinafedina@mail.ru

Abstract

At present time configurable wireless sensor networks are given special consideration. Wireless sensor networks are now a static elements union. To make such a network a powerful system that is able to replace traditional computers, research is required of the application of different theories and technologies for sensor network development. However, building models of approximate reasoning and its usage in computer systems presents an important scientific problem. Fuzzy logic provides effective representation of real world information. The mathematical method of fuzzy information representation allows the building an adequate model of the reality. The article describes an algorithm of a fuzzy controller used to make soft computing and soft managements of a wireless network. Such a network consists of devices with limited resources which are able to work during many months and possibly years.

It has been suggested that AI methods can be used to benefit the technological development of wireless sensor networks (WSN). Now, WSN conception changes the human role, as its elements - sensor microcontrollers - become more independent and underlying of human tasks. The Homocentric model of network calculating, with humans as the main network unit becomes a thing of the past. Humans shift from calculating central to its periphery, concentrating on process control, becoming the mediator between real world and computers. Actually a new class of distributed computers has been created which will open new outlooks. WSN has made its first step on the road to the next epoch when computers will be connected with the physical world. At this point they will be able to pre-determine users wishes, and will be able to make independent decisions.

In last time, fuzzy controllers are widely adopted for practical issues. They allow managing of complex and ill formalized processes basing on linguistic information. In this paper, research of fuzzy logic algorithm application has been carried out for optimization of WSM work. One of the application examples can be optimization device work frequency. WSN consist of tiny devices with limited resources but they have to work without recharging over periods of many years. So issues of energy saving occur. To prolong the devices "life" methods of fuzzy analyses are suggested for more flexible management of device work. For example, if the network is intended for environment observation and has to report about unusual cases then the fuzzy controller algorithm is convenient to optimize device work frequency. If the situation around is normal, network nodes work less frequently to lose less energy. As soon as the situation oversteps the limits network the device begins to work more frequently to store all information. Stored data can be used to analyze by device itself, network or network administrator.

Fuzzy controller block is placed in every network device. In our example, frequency of sensor data reading, the Fuzzy controller input is sensor data and output is a required value. If we take for example a network which watches the temperature conditions i.e. the network device periodically reads the temperature value. In this case the fuzzy control input is temperature value and the output is the temperature reading frequency value. A Network device containing such a fuzzy controller module can react with more flexibility to changes in its environment. This makes it possible to save a great deal of energy. This module works without the assistance of human interaction and so therefore provides a device for independency.

Fuzzy controller input can be heterogeneous data, for example temperature and light or temperature and humidity. Thus a device can analyze different parameters depending on the task at hand.

There are 2 main methods to design a fuzzy controller module to application work based on hardware capabilities.

The first method is the use of a special device for fuzzy management applications. They are divided into 2 types:

- Fuzzy coprocessors. These are special devices for execution of input fuzzy operations. They are controlled by universal microprocessors or microcontrollers, for example VY86C570 (Togai InfraLogic, Inc);
- Microcontrollers, which have special commands and registers to execute

fuzzy input, for example Motorola 68HC12;

The Second method is the use of universal microprocessors or microcontrollers. Fuzzy controller module is implemented as software in this case. To do it there are several variants:

- Using of fuzzy frameworks. They collect variables information, fuzzy sets and fuzzy operators which are selected by the user and generate program code afterwards;
- Using of finished libraries. They are designed especially to develop fuzzy controller applications;
- Developer designs whole system. Such a fuzzy controller module has a slower response rate than special fuzzy hardware. In spite of this, this particular way of fuzzy module development is more flexible in making some changes in fuzzy application. Such fuzzy controllers allow humans to try several various variants of applications without the devices changing.

To design the fuzzy controller module it is necessary to create a Knowledge Base which should include concept information divided into Data Base and Rules Base. Knowledge Base is a static data structure which has predefined size and installed contents. Knowledge Base structure should be defined before starting application work. Data Base contains information about fuzzy sets which are linguistic terms used in fuzzy rules. To simplify a calculation, each fuzzy set membership function is presented as a piecewise-linear function and has three or four points. Points count depends on the membership function form: triangular or trapezoidal.

These point values in a fuzzy set match up places which they have in variable definitional domain. An Example of variable definitional domain division into five trapezoidal fuzzy sets is shown on the picture below.

Rules Base is rule set. Each rule uses matched fuzzy set from the Data Base. There are two possible data structure designs to store Knowledge Base. The decision depends on hardware capability.

If computer is power optimal, structure design can consist of entries array where each entry is a rule from Rules Base and has variable fields (antecedents and consequents) of integer or point type. These integers or points are reference to fuzzy sets array. Sometimes it is necessary to use dynamic count of rules and fuzzy sets for several Knowledge Bases. In this case, static arrays should be replaced by lists in dynamic memory.

Microcontrollers can use only simple data structures consisting of integer or float arrays. Structure bases on as many arrays as antecedents and consequents and defining points of fuzzy sets. For example for trapezoidal fuzzy sets Data Base is float arrays. Each element of these arrays contains information about antecedent i, fuzzy set j and defining fuzzy set point k. Rules Base consist of three arrays. These arrays size are rules count. Rules link to fuzzy sets which Data Base keeps.

Each node of network has a fuzzy controller module which has a set of input and output linguistic variables. Thus this type of sensor node can be called a fuzzy sensor. Fuzzy sensor contains linguistic variables called also attributes, Knowledge Base consisted of fuzzy rules and машина вывода which carries out calculation of output variables.

In order for the fuzzy system to be more flexible, capability of linguistic rules and Knowledge Base remote setting has been added. Thus it is possible for the fuzzy sensor to change its behavior if the user wishes, whether it is human or a computer program. Input variables are linguistic mapping of physical sensors placed on the device. Output variables values are results of fuzzy controller work.

A set of such fuzzy sensors which are wireless network nodes can organize hierarchy. Thus, it is possible to form a distributed system of information processing and storing. The hierarchy is created in the following way, output variables of some fuzzy sensor can be input variables of other fuzzy sensor or sensors. Multilevel hierarchies of fuzzy sensors are created using this principle. This distributed system allows tasks of any complexity to be implemented.

There is capability to build distributed Data Base to process fuzzy queries in wireless sensor network. This capability is easily realized, but is not expedient, as queries to Data Base and fuzzy modifiers presented in the form of ordinary relational Data Base are limited. Also the Data Base has a large memory size. Therefore the use of SQL extended to fuzzy case to process fuzzy queries is not effective. Following this a special fuzzy query language, and communications in distributed systems were developed. This language can be used for

- sensor network requests
- fuzzy triggers creating
- fuzzy active Data Bases creating
- fuzzy sensors communications

It is based on this approach that the distributed Data Base was created. Its data is stored in various network nodes. The Client uses such distributed fuzzy Data Base as ordinary Data Base. The query can be both simple, for example a query of attribute values for a given period, and complex containing aggregation functions. Also it can be fuzzy which asks output attributes fuzzy calculation. If the node does not have enough

information to execute a query it creates and sends its own queries to other nodes.

Five general, goal-oriented, data fusion methods are in use today in WSN (ordered by data complexity) - data association, identity fusion, effect estimation, pattern recognition and artificial intelligence. Ten discrete data fusion techniques can be identified within these five general categories: figure of merit and gating technique in the data association, Kalman filters in the identity fusion, Bayesian decision theory and Dempster-Shafer evidentional reasoning in the effect estimation, adaptive neural networks and cluster methods in the pattern recognition, expert systems, blackboard architecture and fuzzy logic in the artificial intelligence.

The sensor data fusion technology focuses on the acquisition of high-level information (artificial intelligence level), i.e. information that is related to many conventional physical quantities in a non-analytical way. In these complex cases, fuzzy production systems and fuzzy neural networks are more effective and they compute and report linguistic assessments of numerically acquired values. Two methods are proposed to realize the aggregation from basic measurements. The first one performs a combination of the relevant features by means of a rule-based description of the relations between them. With the second, the aggregation is realized through an interpolation mechanism that creates a fuzzy partition of the numeric multi-dimensional space of the basic features. This partition can be realized with fuzzy neural networks. But fuzzy sensor can also can be used on low levels of data fusion, e.g. for filtering and for pattern recognition.

Aggregation functions can be modeled using T-norm and T-conorm. To do it only one membership function of line form should be built. But rules are different for various aggregation functions. For example it is required to calculate maximum of two temperature values which are measured different nodes. Two linguistic variables, *Temp1* and *Temp2*, should be created and membership functions, *f1* and *f2*, should be built for these variables. So every variable has one membership variable of line form. Result is output linguistic variable *Max* which has line function *fmax* as *Max* membership function. In this case rules have to be following:

IF Temp1 =f1 OR Temp2 = f2 THEN Max =fmax

To calculate minimum of *Temp1* and *Temp2* rules have to be following:

IF Temp1 = line AND Temp2 = line THEN Max = line

Factors which can influence on exactness of fuzzy logic system has to be take into account. For example:

- selecting of fuzzy sets kind and number
- selecting of defuzzyfication method
- selecting of operators which are used in rules. Operators selecting is very important and directly influences on result exactness.

In this fuzzy system following T-norm and T_conorm are used:

- Zade's T-norm and T_conorm
- probabilistic T-norm and T_conorm
- Lukasevich's T-norm and T_conorm
- parametric Franc's nd T-conorm
- parametric Sugeno's d T-conorm

The same system can use various types of T-norm and T-conorm. Setting can be made in any moment during system working.

Such module can be separate into following parts:

- fuzzy controller algorithm extended to using of T-norm and T-conorm
- initial data download
- result storing
- fuzzy model setting
- dynamic changing of T-norm and T-conorm

In subsystem of dynamic changing of T-norm and T-conorm their queue task is executed. They change at all time during managing process. It allows to do the system management is more appropriate of expert opinion about its work.

Data fusion algorithms in production form can be easily decomposed and they have hierarchical form by nature. So sensor nodes hierarchy can realize data fusion inside WSN. The knowledge bases for processing of data in given node is distributed inside WSN.

WSNs are huge dynamic databases but for more effective using of information we need more effective organization. The most interesting approach is to use WSN as distributed computing environment for intelligent data processing methods and as storehouse of this methods and not only tools for data measuring and transmitting.. Thus methods are to provide distributed accumulation, transmitting and using of these knowledge. One of approaches is to use expert system with knowledge base distributed among fuzzy sensor nodes in WSN. The physical data attributes are processed by fuzzy sensor node knowledge base and by knowledge base of neighbor fuzzy sensor nodes.

But the main problem is the cost and the complexity of data delivery in data fusion fuzzy sensor because the position of this fuzzy sensor has to be fixed and closed to the user. So the assignment of WSN as point for data fusion must be dynamic

procedure and the fuzzy sensor position should be optimized in regarding of the query, WSN and environment status. Together with fuzzy sensor's assignment its knowledge base should be changed. When cluster head function is delivered inside cluster of nodes from one fuzzy sensor to other fuzzy sensor than knowledge base with cluster head functions of first fuzzy sensor should be send to second fuzzy sensor.

When user requests about some attribute from particular fuzzy sensor and this fuzzy sensor has no rules to compute it, then the request should be send to the nearest fuzzy sensor where necessary knowledge base was located. E.g., for monitoring of dynamic object (goal) moving or other changes of goal parameters the positions of fuzzy sensors responsible for this monitoring can change and data fusion methods must be transmit to the fuzzy sensor responsible for this monitoring and located near the goal at the moment of data request.

The suggested approach consists in moving up of fuzzy sensor with data fusion rules to the event point and in hierarchical data processing inside this fuzzy sensor. At first fuzzy sensor should get the necessary physical data and then compute the users' inquiry answer. Computing procedure for inquiry answer starts from fuzzy sensor with data fusion rules, than distributes across WSN. The universal character of these algorithms by production rules allows to simplify these procedure by transmission only rule's parameters instead of program code.

Thus fuzzy distributed knowledge base for distributed data base query processing has the following properties:

- It functions as distributed expert system.
- Knowledge in production form can be transmitted between nodes.
- Knowledge base for inquiry answer can be send in fuzzy sensor together with the inquiry
- Special language is used for knowledge base transmission between nodes.

Knowledge-based program of data fusion uses parallel computing algorithms and destines for the whole WSN and not only for certain fuzzy sensor.

Until now middleware was rarely designed for wireless networks support. Most middleware systems were for enterprise networks. Some of then focuses on how to control quality of service adaptation in middleware architecture, and the quality of service is specified by fuzzy rules and membership functions [15] an can be realized by fuzzy sensor. But there are already some projects underway that aim to develop middleware for WSN, such as [1, 3, 4, 5, 6, 7, 8, 9, 16]. Cougar [1], for example, adopts a database approach where sensor readings are treated like "virtual" relational database tables. An SQL-like query language is used to issue tasks to the WSN. The Smart Messages Project [7] is based on agent-like messages containing code and data, which migrate through the sensor network. NEST [8] provides so-called microcells as a basic abstraction. They are similar to operating system tasks with support for migration, replication, and grouping. SCADDS [4] is based on a paradigm called Directed Diffusion, which supports robust and energy-efficient delivery and in-network aggregation of sensor events. Project AGILA [16] envisions a new paradigm for programming and using sensor networks where applications consist of special programs called mobile agents that can migrate their code and state from one node to another as they execute. Mobile agents offer an unprecedented level of flexibility by allowing fluid applications to spread throughout the network and to intelligently position themselves in the optimal location for performing their task, whether it be detecting an intruder or tracking a wildfire. By allowing new agents to be dynamically injected, a pre-existing network can be re-tasked.

However, most of the projects are in an early stage focusing on developing algorithms and components for WSN [2], which might later serve as a foundation for middleware. Moreover, most of the current results are based on simulations or small-scale experiments in laboratory settings. The suitability for large-scale networks still has to be proven.

There are several models of fuzzy databases, which can be easily, generalized as WSN fuzzy database in the fuzzy sensor friendly domains.

Applying fuzzy logic to databases has been an active research area since the 80's. The most important issues are the enhancement of existing data models for representing uncertain and/or imprecise data (fuzzy data), the extension of current database languages to handle fuzzy queries, and the use of fuzzy inference to deduce answers to questions in fuzzy expert database systems [11, 12].

Active databases, which incorporate Event-Condition-Action rules into the conventional (passive) databases, have been investigated by many researchers over the past decade [13, 14]. They provide the capability to react to database (and possibly external) stimuli, called events, without user intervention. Fuzzy triggers model combines fuzzy logic features with active database capabilities to provide a high-level view of data stored in a database was proposed in [10].

In spite of the fact that knowledge representation with fuzzy production rules is quite natural and simple procedure, it suppose rather slow interface with the user for rules and membership function acquisition. Besides the resulted knowledge base usually needs validation and verification. A

learning pattern that corresponds to expert opinion about desired measurement process can greatly accelerate this process. So the nearest goal is using beside fuzzy sensor neuro-fuzzy networks (embedded in one node or distributed among several nodes of WSN), genetic algorithms and artificial immune systems. These models can control by adaptation and learning of fuzzy sensor for optimization such hierarchical processes as data clusterization, filtration, aggregation, association and fusion. These possibilities will increase the effectiveness in static and dynamic object monitoring, monitoring of environments, buildings and industry processes. Besides they will be able self-learning simultaneously with control of basic processes – data and knowledge transfer inside WSN, energy saving, defense of WSN from attacks. For the case of heterogonous WSN functions dedicated nodes should realize the most part of soft computing methods.

References

[1] Cougar Project. http://www.cs.cornell.edu/database/cougar.

[2] Mobile Information and Communication Systems. http://www.nccr-mics.ch./

[3] Power Aware Distributed Systems. http://www.pads.east.isi.edu./

[4] Scalable Coordination Architectures for Deeply Distributed Systems. http://www.isi.edu/div7/scadds.

[5] Sensorwebs Project. http://www.basics.eecs.berkeley.edu/sensorwebs.

[6] Smart-Its Project. http://www.smart-its.org./

[7] Smart Messages Project. http://www.rutgers.edu/sm.

[8] A Network Virtual Machine for Real-Time Coordination Services. http://www.cs.virginia.edu/nest.

[9] Webdust Project. http://www.cs.rutgers.edu/dataman/webdust.

[10] Bouaziz T., Karvonen J., Pesonen, A. and Wolski A. Design and Implementation of TEMPO Fuzzy Triggers. Research Report No TTE1-2-97, VTT Information Technology, Finland, March 1997. Also at ftp://ftp.vtt.fi/pub/projects/rapid/tempo-design.ps

[11] Zemankova M. FILIP: a Fuzzy Intelligent Information System with Learning Capabilities. In [IS89], pp. 473-486.

[12] Leung K. S., Wong M.H. and Lam W. A Fuzzy Expert Database System. In Data & Knowledge Engineering , Vol. 4, 1989, pp. 287-304.

[13] Paton, N.W., Diaz, O., Williams, M.H., Campin, J., Dinn, A., and Jaim, A. Dimension of Active Behavior. In Proceedings of the 1st Int. Workshop on Rules in Database Systems, Edinburg (Scotland), August 1993, pp. 40-57

[14] Widom J. and, Ceri S. (Eds.). Active Database Systems: Triggers and Rules For Advanced Data-base Processing. Morgan Kaufmann, 1996.

[15] B. Li and K. Nahrstedt, “A Control-based Middleware Framework for Quality of Service Adaptations,” IEEE Jour. of Selected Areas in Communications, Special Issue on Service Enabling Platforms, vol. 17, no. 9, Sept. 1999, pp. 1632–1650.

[16] Chien-Liang Fok, Gruia-Catalin Roman, and Chenyang Lu, "Rapid Development and Flexible Deployment of Adaptive Wireless Sensor Network Applications," International Conference on Distributed Computing Systems (ICDCS'05), Columbus, OH, June 2005 also in http://mobilab.wustl.edu/projects/agilla.

[17] O. Cordón, F. Herrera, A. Peregrín, Practical Study on the Implementation of Fuzzy Logic Controllers, 1998

[18] Ren´e Jager, Fuzzy Logic in Control, 1995

[19] O. Cordón, F. Herrera, A. Peregrín, Applicability of the Fuzzy Operators in the Design of Fuzzy Logic Controllers, 1995

[20] Paolo Dadone, Design Optimization of Fuzzy Logic Systems, 2001

SOFT COMPUTING IN WIRELESS SENSORS NETWORKS

Averkin A.N.
Dorodnicyn Computing Center of the Russian Academy of Sciences
averkin@ccas.ru

Belenki A.G.
Lomonosov Moscow State University
abelenkiy@gmail.com

Abstract

The embedded soft computing approach in wireless sensor networks is suggested. This approach means a combination of embedded fuzzy logic and neural networks models for information processing in complex environment with uncertain, imprecise, fuzzy measuring data. It is generalization of soft computing concept for the embedded, distributed, adaptive systems.

Keywords: embedded fuzzy logic, data fusion, clusterization, aggregation, fuzzy distributed knowledge base.

1 Introduction

The technology of fuzzy and neuro-fuzzy systems in WSN uses soft computing approach to increase performance of wireless sensors networks (WSN) and to make them more intelligent. WSN is one of the most promising technologies of the 21st century. For the first time «smart» sensors were implemented by Berkeley University of California together with INTEL corporation.

The prototype of WSN node is a software-hardware platform for deployment of several specialized sensors on the base on an autonomous wireless controller. WSN consists of a large number of tiny devices, which are deployed in real environment and function as a united network. To provide sensor nodes with a possibility of self-organization the specialized software was designed together with IDE for application development. The specialized software implements the possibilities of communication, routing and application support for WSN.

The increasing of WSN performance means preliminary processing of raw data, data fusion, clusterization and aggregation. Intelligent WSN provides also distributive decision-making and queries processing, knowledge-based routing and power consumption. Methods of decision-making and information processing based on symbol models of classic artificial intelligence are too complex for embedded realization in WSN due to limited communication and power resources of WSN.

Only few companies in the world solve this problem by embedded soft computing approaches. These approaches for WSN mean a combination of embedded fuzzy logic and neural networks models for information processing in complex environment with uncertain, imprecise, fuzzy measuring data. It is generalization of soft computing concept for the embedded, distributed, adaptive systems. These approaches were suggested by Russian scientists in 1997 [2], when hybridization of soft computing and mathematical statistics was used to process the results of heterogeneous measurements for environment monitoring applications.

The first realization of these approaches was made in 2004 [1,2,4]. The main part of our embedded soft computing and soft computing approaches is Smart Node (SN) model for WSN [3]. The core of SN is Fuzzy Engine, which consists of three modules: knowledge base (a set of fuzzy production rules), fuzzification and defuzzification modules, which transform numerical measurements in linguistic form and vice-versa. The output of SN can approximate of any function of input parameters, e.g. when it is impossible or difficult to measure certain parameter, it can be computed by SN with the use of special rules from knowledge base. Similarly the special rules can be created for data fusion, clusterization, aggregation, routing and power consumption. The knowledge base of SN can be created as a result of knowledge acquisition from exert or by supervised neural network learning. Application knowledge in nodes can significantly improve the resource and energy efficiency, for example by application-specific data caching and aggregation in intermediate node

SN are realized inside MeshNetics™ platform [4]. MeshNetics™ is a family of software components, algorithms, hardware designs and solutions that enable next generation M2M applications [3]. By adding expert system to monitoring, controlling, tracking applications, it enables new generation of solutions optimized for the needs of end-users. With MeshNetics™ businesses profit from reduced pricing structure, life-cycle time and enhanced competitiveness of their new and existing products.

MeshNetics™ enables expert remote monitoring and control of a wide range of processes, assets, systems, and facilities. Built-in expert system with advanced algorithms, neural networks and fuzzy logic creates a new generation of WSN's and frees businesses from

being trapped into costly and complex wired "static" systems and enhances the possibilities by listening actively to your environment.
Built-in expert system on the base of SN, that allows to support hybrid distributive expert system on the nodes of WSN, which can realize, together with data collection and communication, a large class of existing algorithms of data fusion and aggregation.
Meshnetics™ SN is the node of Meshnetics™ platform, which include Smart Engine. Smart Engine is given by a number of its parameters, i.e. rules patterns, variables, terms, membership functions, triangular norms. These parameters can transmit across WSN. These transmissions can be defined by user, or by Smart Node (SN) itself. E.g., in goal tracking process these transmissions can be realized by SN in dependence of mutual smart node and goal positions.
Software environment of SN is universal tool for intellectual decision support in WSN and it is strongly connected with following power-aware&networked embedded Computer Systems research areas: application-driven network architectures, emerging platforms and technology, resource constrained real-time OS's, distributed algorithms (broadcast, anycast, multicast, convergecast) in lossy wireless networks, ad hoc multi-hop routing, , in-network aggregation and processing, coverage and density, ranging and localization, resilient aggregators, distributed feature extraction, tracking, and collaborative signal processing.

2 Smart Nodes Challenges

2.1 Data Processing Challenges

The main goal of WSN activity is collecting a tremendous amount of data and transmitting them to the user. Collected data can be interpreted as distributed knowledge base. In this approach both system user and system designer have the problems of control for distributed data processing processes of uncertain, incomplete or redundant sensor measurements.
The main factors, that influence on WSM effectiveness and make problems for designers are the following:

- WSN nodes have very restricted computational and storage power.
- Node communication range is limited. In most cases nodes can directly communicate with immediate neighbors only.
- WSN consists of a large number of unreliable nodes, that produce measurement or transmission errors
- WSN must continue to operate at all times even when some of it nodes get physically destroyed at unpredictable times.
- WSN must continue to operate without interruption when new nodes are added to the network in order to replace the failed ones or extend the network.
- As a result, node communication may require different paths at different times depending on the state of end-to-end link between communicating parts of the network.
- The decrease of battery power is different in different nodes.
- Data transmission time and power losses increase with the size of WSN.

Two last problems are of great importance to the user. The carrying capacity of network only slightly depends on number of nodes (it increases as log N, where N – number of nodes). The most rational output is decreasing of traffic by adding distributive hierarchical data processing inside network and by providing the user with relevant answers only. There are a number of various algorithms for this processing realization. These algorithms depends on data types and data generalizations levels. But in traditional models of distributive hierarchical data processing each algorithm is strictly connected with certain node for given network topology. Changing of network topology implies reboot of nodes and this process needs transmitting of large pieces of code. To solve the problem in SN transmitting of code units has changed by transmitting of knowledge units. These knowledge units are used for reboot local note with tuning parameters. The last problem (irregularity of power consumption) is solved by embedding power consumption rules in given sensor node. On the base of these rules the sensor node can make autonomous decision about utility of participation in data collection and data transmission for given states of the environments, neighbor nodes and decision-making node.
As software tool for this purpose we have realized universal tool of fuzzy sensor shell with production knowledge model, embedded in all nodes of SN. This shell may approximate a large class of existing algorithms of data fusion and aggregation. In this approach each sensor becomes intelligent agent with knowledge about himself and its' environment and it is able to autonomous decision-making. Sensor may control this knowledge and send it to other node. In this case WSN can be interpreted as distributed data base and knowledge base with the possibilities of mobility and adaptability. The fact allows using multi-agent technologies and distributive intelligent decision support systems.

2.2 Middleware Challenges

SN technology sits between the operating system and the application and thus belongs to middleware. Thus the main purpose of middleware for sensor networks is to support the development, maintenance, deployment, and execution of sensing-based applications. This includes mechanisms for formulating complex high level sensing tasks, communicating this task to the WSN, coordination of sensor nodes to split the task and distribute it to the individual sensor nodes, data fusion for merging the sensor readings of the individual sensor nodes into a high-level result, and reporting the result back to the task issuer. The most part of these mechanisms were successfully realized in SN.

Unique property of SN embedded middleware for WSN is imposed by the design principle application knowledge in nodes. Traditional middleware is designed to accommodate a wide variety of applications without necessarily needing application knowledge. SN embedded middleware for WSN has to provide mechanisms for injecting application knowledge into the infrastructure and the WSN.

For this purpose SN embedded middleware for WSN has to provide special knowledge representation language, special query language, special protocols of query forwarding, special methods of data fusion and aggregation and special methods of software update management. These new technologies have been realized on the base of fuzzy systems technology.

3 Possible Application Fields of SN

3.1 Smart Node in WSN Control.

When we use knowledge (meta-rules) to control WSS we have analogy with active network paradigm. The similarity is in sending together using together with each request special block of rules (capsule for active networks) to process the request by SN (server for active networks). The difference is that SN suppose two types of traffic – knowledge traffic and data traffic and in active networks there no knowledge traffic. For knowledge traffic and for communication with other subsystems (e.g. neuro-fuzzy systems) we can use FULL-like special language for fuzzy knowledge representation.

Using SN we can realize:

- Routing algorithms
- Optimal control of power consumption in WSN
- Data traffic control
- Q&S control

3.2 Using SM for Fuzzy Data Base Designing

From one perspective sensor networks are similar to distributed database systems. They store environmental data on distributed nodes and respond to aperiodic and long-lived periodic queries. Data interest can be pre-registered to the sensor network so that the corresponding data is collected and transmitted only when needed. These specified interests are similar to views in traditional databases because they filter the data according to the application's data semantics and shield the overwhelming volume of raw data from applications. Fuzzy query approach can be used to reduce this volume of raw data.

The extension of TinyDB by fuzzy attributes can be interpreted as fuzzy TinyDB and fuzzy active TinyDB. This possibility can be easy realized but does not seems very useful because there are only few possible classes of requests to WSN and representation of fuzzy modifier in TinyDB are rather restricted. Besides volume of DB is too large for SN memory. So using of extended SQL for fuzzy requests processing is not effective.

Thus in Fuzzy Meshnetics™ expert WSN on the base of SN we can completely substitute functions of fuzzy data base for WSN (possible TinyDB fuzzy extensions) by SSM data bases functions with special query processing language for SN. The language can be used for:

- Fuzzy queries to WSN (fuzziness can be in query only and also in WSN);
- For fuzzy triggers designing;
- For active data base designing;
- For communication between SN.

Traditional Event-Condition-Action triggers (active database rules) include a Boolean predicate as a trigger condition. As far as WSN can be considered as distributive database, we can see that SN with Event-Condition-Actions in KB realize embedded fuzzy triggers for this distributive database.

3.3 Using of SN for Data Aggregation and Fusion.

If all raw data is sent to base stations for further processing, the volume and burst ness of the traffic may cause many collisions and contribute to significant power loss. To minimize unnecessary data transmission, intermediate nodes or nearby nodes work together to filter and aggregate data before the data arrives at the destination.

Five general, goal-oriented, data fusion methods are in use today in WSN (ordered by data complexity) - data association, identity fusion, effect estimation, pattern recognition and artificial intelligence. Ten discrete data fusion techniques can be identified within these five general categories: figure of merit and gating technique in the data association, Kalman

filters in the identity fusion, Bayesian decision theory and Dempster-Shafer evidentional reasoning in the effect estimation, adaptive neural networks and cluster methods in the pattern recognition, expert systems, blackboard architecture and fuzzy logic in the artificial intelligence.

The SN data fusion technology focuses on the acquisition of high-level information (artificial intelligence level), i.e. information that is related to many conventional physical quantities in a non-analytical way. In these complex cases, fuzzy production systems and fuzzy neural networks are more effective and they compute and report linguistic assessments of numerically acquired values. Two methods are proposed to realize the aggregation from basic measurements. The first one performs a combination of the relevant features by means of a rule-based description of the relations between them. With the second, the aggregation is realized through an interpolation mechanism that creates a fuzzy partition of the numeric multi-dimensional space of the basic features. This partition can be realized with fuzzy neural networks. But SN can also can be used on low levels of data fusion, e.g. for filtering and for pattern recognition. Data fusion algorithms in production form can be easily decomposed and they have hierarchical form by nature. So sensor nodes hierarchy can realize data fusion inside WSN. The knowledge bases for processing of data in given node is distributed inside WSN.

But a naive placement of the fusion functions on the network nodes will diminish the usefulness of in-network fusion, and reduce the longevity of the network (and hence the application). Thus, managing the placement (and dynamic relocation) of the fusion functions on the network nodes with a view to saving power becomes an additional responsibility of the application programmer. Dynamic relocation may be required either because the remaining power level at the current node is going below threshold, or to save the power consumed in the network as a whole by reducing the total data transmission. Supporting the relocation of fusion functions at run-time has all the traditional challenges of process migration.

3.4 Using of SN for Fuzzy Distributed Expert System Designing

WSNs are huge dynamic databases but for more effective using of information we need more effective organization. The most interesting approach is to use WSN as distributed computing environment for intelligent data processing methods and as storehouse of this methods and not only tools for data measuring and transmitting.. Thus methods are to provide distributed accumulation, transmitting and using of these knowledge. One of approaches is to use expert system with knowledge base distributed among SNs in WSN. The real data attributes (IPA) are processed by SN knowledge base and by knowledge bases of neighbor SNs.

But the main problem is the cost and the complexity of data delivery in data fusion SN, because this data fusion SN, because this position of this SN must be fixed and close to the user. So the assignment of WSN as point for data fusion must be dynamic procedure and the SN position should be optimized in regarding of the query, WSN and environment status. Together with SN's assignment its knowledge base should be changed. When cluster head function is delivered inside cluster of nodes from one SN_1 to SN_2, than knowledge base with cluster head functions of SN_1 should be send to SN_2.

Thus fuzzy distributed knowledge base for distributed data base query processing has the following properties:

- It functions as distributed expert system.
- Knowledge in production form can be transmitted between nodes.
- Knowledge base for inquiry answer can be send in SN together with the inquiry
- Special language is used for knowledge base transmission between nodes.
- Knowledge-based program of data fusion uses parallel computing algorithms and destines for the whole WSM and not only for certain SN.

References:

[1] Suvorov V., Zubkov G., Averkin A. In formation Technologies on the Base of Wireless Sensors Networks // Proceedings of the International conference on Information problems in the 3rd millennium (Kazan, 2004).

[2] Averkin A., Belenki, Suvorov V., Zubkov G. Soft Computing in Wireless Sensors Networks, International Conference on Soft Computing and Measurements SCM'2005, St.Petersburg, 2005.

[3] Averkin A.N., Prokopchina S.V. Survey of Soft Measurement Conception, St.Peterburg: Gydrometeoisdat, 1997.

[4] http://meshnetics.com

EVALUATION OF ALTERNATIVES FOR THE DISPOSITION OF SURPLUS WEAPONS-USABLE PLUTONIUM ON THE BASIS OF FUZZY SETS*

Averkin A.N.
Computer center of the Russian Academy of Science
averkin@ccas.ru

Kosterev V.V.
Moscow Engineering Physics Institute
vvkosterev@mephi.ru

Stankevich K.V.
Moscow Engineering Physics Institute
S_KV@bk.ru

Abstract

Multiattribute evaluation of alternatives for the disposition of surplus weapons-usable plutonium on the base of fuzzy sets has been made.

Keywords: Multiattribute evaluation of alternatives, surplus weapons-usable plutonium, fuzzy sets.

1 Introduction

One of the major problems of nuclear power is the problem of recycling of nuclear materials: their safe long-term storage and management.

Process of the Russian-American reduction of strategic nuclear arms has put on the agenda a problem of proliferation of nuclear weapons - highly enriched uranium and plutonium.

The problem of recycling of weapon plutonium is the most actual today and attracts attention of politicians, scientific and engineers in Russia and abroad. The main objective of the decision of this problem is to prevent the proliferation of nuclear weapon. Besides it is necessary to provide efficiency of using of plutonium and safety of man and environment.

The problem of disposition of surplus plutonium can be formalized as a multi-attribute problem of a choice of alternatives from a set of possible alternatives.

The Department of Energy of USA - Office of Fissile Materials Disposition has announced a Record of Decision selecting alternatives for disposition of surplus plutonium. There are thirteen alternatives which fall into reactor, immobilization and direct disposal alternatives. Thirteen alternatives of recycling surpluses weapons-usable plutonium are considered in report [1]. Five reactor's alternatives envisage that surplus weapons-usable plutonium is used for production of MOX fuel for nuclear reactors. Other six alternatives include different ways to immobilize surplus plutonium with radioactive glass-bonded zeolite, or mix with borosilicate glass, ceramic and radioactive materials. The last two alternatives present pellets, immobilized with ceramic and placed in a borehole. It means using inert matrix or directly deep borehole.

Because the decision for plutonium disposition involves multiple criteria it is appropriate to use multi-attribute utility model for this study [2]. It bases on the calculation a multi-attribute utility $U(x_1, x_2,\ldots, x_n)$, where x_i represents the level of performance on measure *i*. $U(x_1, x_2,\ldots, x_n)$ can be decomposed into an additive, multiplicative or other form to simplify assessment.

An additive multi-attribute utility model can be represented as follows:

$$u(x_1, x_2,..., x_n) = \sum_{i=1}^{n} k_i f_i(x_i), \qquad (1)$$

where $f_i(x_i)$ - is a single-attribute value function over measure i that is scaled from 0 to 1, k_i - is the weight for measure i and $\sum_{i=1}^{n} k_i = 1$.

The purpose of this report is development an additive multi-attribute utility model on the basis of fuzzy sets [3] and evaluation of alternatives for the disposition of surplus weapons-usable plutonium.

2 Evaluation and ordering of alternatives on the basis of an additive multi-attribute utility model and fuzzy sets.

The method of ordering of alternatives on the basis of the additive convolution expanded on a case of fuzzy information is used. This approach bases on the assumption of additive independence of criteria.

Let it is required to order **m** alternatives: $a_1, a_2, \ldots, a_m$, which are estimated on **n** criteria. We shall designate an estimation of **i**th alternatives on **j**th criterion as $\tilde{A}_{ij}$, i=1, m; j=1, n. Relative importance of criterion is determined by weight multiplier K_j. The weighed estimation of **i**th alternatives is determined under the formula

$$\tilde{A}_i = \sum_{j=1}^{n} K_j \times \tilde{A}_{ij}, \qquad (2)$$

where $\sum_{j=1}^{n} K_i = 1.$

In the assumption of fuzziness of assessment of alternatives by corresponding criteria $\tilde{A}_{ij}$ and relative importance of alternatives K_j they are fuzzy numbers and might be represented by membership functions $m_{\tilde{A}_{ij}}(a_{ij})$ and $m_{K_j}(k_j)$ accordingly.

Membership function of the weighed estimation of **i**th alternative $m_{\tilde{A}_i}(a_i)$ is determined according to the formula (2) on the basis of the extension principle [4].

$$m_Z(z) = \max_{z=x*y}\{\min[m_X(x), m_Y(y)]\}, \quad (3)$$

where symbol * means algebraic operation.

The membership functions $m_{\tilde{A}_{ij}}(a_{ij})$ and $m_{K_j}(k_j)$ were set as symmetric triangular functions. Their tops correspond to assessment of alternatives and relative importance of criteria, and the left and right borders characterize the fuzziness of assessments.

The alternatives and the five goals or objectives (criteria) are shown in the table 1:
The alternatives can be categorized into 3 parts:

I. Reactor alternatives:
1 - Existing Light Water Reactors, Existing Facilities; 2-Existing Light Water Reactors, Greenfield Facilities; 3-Partially Completed Light Water Reactors; 4-Evolutionary Light Water Reactors; 5-CANDU Reactors;

II. Immobilization alternatives:
6 - Vitrification Greenfield; 7-Vitrification Can-in-Canister; 8-Vitrification Adjunct Melter; 9-Ceramic Greenfield; 10-Ceramic Can-in-Canister; 11-Electrometallurgical Treatment;

III. Direct disposal alternatives:
12 - Deep Borehole (Immobilization); 13-Deep Borehole (Direct Emplacement).

The membership functions of the weighed assessment of **i**th alternative $m_{\tilde{A}_i}(a_i)$ are calculated for assessment uncertainty from 1% up to 100 Figure 1 represents the membership functions $m_{\tilde{A}_i}(a_i)$ of alternatives with 5% of uncertainty.

Thus, the best is the thirteenth alternative, namely a burial place of plutonium in boreholes.

To compare the fuzzy numbers M and N we must evaluate the degree of possibility of $M \geq N$ according to extension principle [4]

$$v(M \geq N) = \sup_{x,\, y\,:\, x \geq y}\{\min[m_M(x), m_N(y)]\}. \quad (4)$$

This formula (4) is an extension of the inequality $x \geq y$ according to the extension principle.

According to the formula (4) we received the degree of possibility that **i**th alternative more then **13**th alternative:
v (1) = 0.8863, v (2) = 0.4185, v (3) = 0.3361, v (4) = 0.3245, v (5) = 0.5470, v (6) = 0.4333, v (7) = 0.9588, v (8) = 0.4349, v (9) = 0.4349, v (10) = 0.9621, v (11) = 0.3591, v (12) = 0.5305, v (13) = 1.0.

According this example the degree of that the tenth, or, for example, the second alternative is the best, is equal to 0.9621 and 0.4185 accordingly. It is similarly possible to carry out the comparison of alternatives with **10**th alternative and with others.

The received results allow carrying out evaluation and ordering of alternatives for the disposition of surplus weapons-usable plutonium in view of real uncertainty of input data.

Table 1

Objectives / № of alternative (a_i)	Theft	Diversion	Irreversibility	International Cooperation	Timeliness
1	0.2281	0.2879	0.7000	0.5278	0.6577
2	0.2311	0.2879	0.7000	0.5278	0.4414
3	0.2378	0.2879	0.7000	0.5278	0.4036
4	0.2323	0.2877	0.7000	0.5278	0.4000
5	0.2278	0.2879	0.7000	0.4496	0.6162
6	0.2941	0.2584	0.7000	0.4483	0.5712
7	0.2941	0.2584	0.7000	0.4483	0.8162

8	0.2954	0.2584	0.7000	0.4483	0.5712
9	0.2941	0.2599	0.7000	0.4483	0.5712
10	0.2867	0.2635	0.7000	0.4483	0.8162
11	0.3055	0.2093	0.7000	0.4483	0.5711
12	0.3166	0.3017	0.6000	0.4548	0.6667
13	0.3116	0.2664	0.8500	0.4548	0.6667

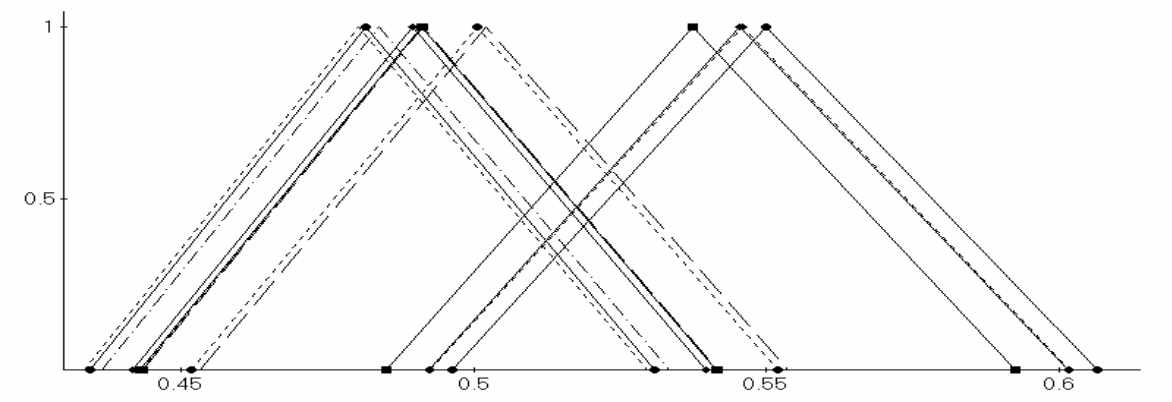


Figure 1: Membership functions of alternatives from the left to the right: 4,3,11,2,6,8,9,12,5,1,7,10,13.

Acknowledgement

The paper is supported by RFBI grants 06-01-00576 and 07-01-00782

References

[1] Evaluation of Alternatives for the Disposition of Surplus weapons-Usable Plutonium – ANRCP-1997-1, 1997.
[2] Keeney, R.L. and Raiffa, H. Decisions with Multiple Objectives. New York: Wiley, 1976.
[3] Borisov A.H., Krumberg O.A., Fedorov I.P. Decision-Making on the Basis of Fuzzy Models. Examples of Utilisation, Riga: "Zinatne", 1990.
[4] Dubois, D., Prade, H. Fuzzy sets and systems, theory and Applications, Academic Press, New York, 1980.

TOWARDS TO AUTOMATIC TACTICS' ANALYSIS IN SOCCER*

Alexey Averkin
Dorodnicyn Computing Center of the Russian Academy of Sciences
averkin@ccas.ru

Anton Gusev
Dorodnicyn Computing Center of the Russian Academy of Sciences
gusev.anton@gmail.com

Mikhail Shestakov
Russian State University of Physical Education
shtv@infosport.ru

Abstract

One of the main problems in soccer is to coordinate in real time scale collective behavior of team members, who solve a common task by solving individual tasks. The specificity of the problem consists in the fact that soccer players with their collective and individual skills, individual behavior and incomplete knowledge of the environment and limited resources should succeed in reaching the common goal in dynamically changing game.

.

Keywords: spatial reasoning, formal spatial situation description, intelligent soccer tactics' analysis

1 Introduction

Soccer players act in time and space inside an unpredictable environment that often complicates team work. When organizing a team play, the main task is to provide coordinated actions of players, all trying to realize their individual actions. Thus, behavior of soccer players as a team is a more complex phenomenon than coordinated individual actions of separate athletes.

At present we are building a digitized soccer match analysis tool and implementing situation patterns in a real soccer match. Our research started from interviewing soccer coaches in order to understand their interpretation of detailed spatial situation description. So we determined fuzzy linguistic variables set and defined formal description of a spatial situation.

The objective of this study was to create an Intelligent System of Tactics' Analysis (ISTA) in a group of interacting players in game situations and during tactical training exercises

2 Methods

To create ISTA, we used multi-agent technologies and fuzzy semiotic models to control individual agent. According to the adopted terminology we use the word "agent" instead of "soccer player". Three layers were distinguished in a cognitive agent: physical actions, individual behavior and coordinated behavior.

When analyzing team work of a multi-agent system, one should take into account that it is organized with the help of the group (team) plan of actions of the agents. This plan has the following peculiarities.

1. The group plan requires the consent of the group of agents to follow some instructions in the group actions.

2. The agents should take responsibilities in respect of their individual actions and the activity of the group as a whole (individual intentions about how to act).

3. At the same time, the agent should take responsibility in respect of the actions of the other agents (coordinated intention).

4. The plan of the group activity may include plans of individual agents for the assigned actions and plans of the subgroups.

Spatial situation on a soccer field is described by determination of spatial relations between players, ball and goal. We might not calculate dimensions of these objects, but accept that a soccer player is <middle-sized>, the goal is <big> and the ball is <small>. Determination of static spatial relations (e. g. relations of directions and distance) is shown on picture below (Figure 1; the same approach as used in [1], [2]).

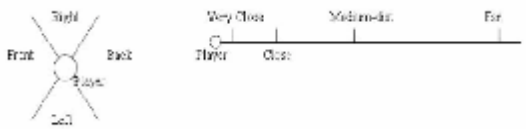


Figure 1: Spatial relations (distance, directions).

Then we've added the speed of each player and the ball. Possible speeds are <standing> (0 m/s), <walking> (<2 m/s),<jogging> (from 2 to 4 m/s),<running>

(from 4 to 5,8 m/s),<high speed running> (from 5,8 to 7 m/s) and <sprinting> (>7 m/s). This information is taken from previous time step (while building a spatial situation description we need current and previous coordinates of the players and the ball).

There are many typical spatial situations on a soccer field. Formal Spatial situation description defines a "pattern" abstracted from objects coordinates using qualitative spatial data. We are trying to recognize 'local' situation double-pass etc., and predict movements of each player during realization of this game pattern. Two players can play together having determined the same spatial situation pattern, without sending any messages (as players in real soccer). So if a player "knows" that his role is "assistance" in double-pass pattern, he will pass back to first player after catching the ball. In other case he will do something ignoring his partner

We are building formal situation description for each 0.1 second of a soccer match (digitized record of a real human match). Then we are using qualitative spatial reasoning to determine the game pattern, which exists on the field at the moment.

Hierarchy of tactical group actions in soccer consists of tactical variant, tactical scheme, and tactical combination.

In our work we recognized the following combinations: in attack - wall pass, pass to a third player, scissor movement; in defense – counteraction to wall pass, counteraction to the combination "pass to a third player", covering. Tactical schemes consisting of several tactical combinations were registered: in attack – run and center, run through the center, flank run with pass to the other flank; in defense – man-to-man marking, zonal marking, horizontal and vertical displacement.

The system ISTA was adapted in order to work with Qualisys motion capture camera with high-speed video "Oqus" (Qualisys AB, Sweden), that permitted to create a database of game combinations for the learning sample of the artificial neural network. To complete computer training support tool, which deals with digitized match record database and automatically recognizes pre-defined situation pattern, now we use fully-defined spatial situation description, while human players do their spatial reasoning using some dominant objects. We are to provide an algorithm to cut unnecessary spatial relations and objects from agent spatial situation description.

3 Results

Using this approach for processing video records of matches played by teams of different qualification permitted to identify some patterns in tactical actions of the teams used in a match. We have analyzed 4 games of the leading European club teams (group A), 5 games of the Russian national youth team (16 yrs) in the European championship of 2006 (group B), 5 games of Second Division teams participating in the regular Russian championship (group C).

Our study permitted to reveal some specific features of tactics organization in teams of different qualification during official games. Data presented in Table 1 and Table2 demonstrates that teams of the group A used less diverse tactical schemes ($P < 0.05$), but repeated the same scheme more often ($P < 0.01$) in a game, than teams of the groups B and C. Team-work in defense was more often used in the group A, than in the other two groups ($P < 0.05$). It should be noted that reliability of team actions in defense was considerably lower in the groups B and C, than in the group A.

Table 1: Group tactical attacking actions with % success rate for different qualification teams

	Number of schemes	**Number of repetitions**	**%**
Group A	6±1,1	31±3,5	20
Group B	9±2,1	18±2,1	40
Group C	8±3,1	12±1,3	50

Table 2: Group tactical actions in defence with % success rate for different qualification teams

	Number of schemes	**Number of repetitions**	**%**
Group A	4±0,2	20±3,1	80
Group B	3±0,2	15±4,3	60
Group C	3±0,7	12±2,4	50

Data displayed in Table 3 revealed formation of tactical schemes in attack and distribution of tactical combinations by 4 zones of the field.

Table 3: Number of group combinations in a single scheme of attack and their distribution by field zones

	Number	**Zone**			
	of	**1**	**2**	**3**	**4**

	combi-nations				
Group A	4±0,2	5	25	30	40
Group B	3±0,2	10	40	35	15
Group C	3±0,7	30	50	20	10

Zone 1 – one quarter of the field around the penalty area of attacking team, zone 4 - one quarter of the field around the penalty area of defending team, zones 2 and 3 – in the center of the soccer field on the sides of attacking and defending teams correspondingly. Players of the group A used more combinations during the realization of a single scheme ($P < 0.05$), than players of lower qualification (groups B and C). It was found out that teams of higher qualification improved interaction in immediate proximity of the opposing goal (group A - 40% in the zone 4), while less qualified teams organized most of collective actions on their own half of the field (group B - 40% in the zone 2; group C – 50% in the zone 2). The lack of sufficient technique preparedness of players of the groups B and C must have prevented them from reliable performance on the opposing half of the field in case of active defense of the opponent.

4 Conclusions

Digitized data (e.g. coordinates of players and ball) can easily be used for automatic tactics' review of match or training process. Usage of ISTA coach assistant is the way to raise effectiveness of training.

Given results was highly commended by real soccer coaches, since ISTA is the way to reduce time spend for match review and it is the way to equitably analyze tactics of both teams. Furthermore we have revealed the following tendencies in organization of team and group actions in soccer. Less qualified teams improvise more during the game that interferes the use of collective actions. To organize team work effectively, tactics of a team must be determined by the chosen style of playing, based on minimal number of tactical variants and schemes. Nowadays we see perspectives in greater universality of players from the point of view of organization of group actions when tackling on the whole field.

Acknowledgement

The paper is supported by RFBI grants 06-01-00576 and 07-01-00782

References

[1] Kandrashina E., Spatial and temporal knowledge representation in artificial intelligence, Radio I Svjaz, Moscow, Russia, 1989.

[2] Musto, A., On Spatial Reference Frames in Qualitative Motion Representation, FKI-Bericht Nr. FKI-230-99, Technische Universitat Munchen, Germany, 1999.

Session 8

Fuzzy Sets – Philosphy and Criticism – R. Seising

Scientific Theories and the Computational Theory of Perceptions – A Structuralis View Including Fuzzy Sets

Rudolf Seising
Medical University of Vienna
Core Unit for Medical Statistics and Informatics
Spitalgasse 23, A - 1090 Vienna, Austria
Rudolf.Seising@meduniwien.ac.at

Abstract

Fuzzy Logic, Computing with words, and the Computational Theory of Preceptions build a stack of methodologies to help bridge the gap between systems and phenomena in reality and scientific theories. Bridging this gap has been a problem in all philosophical approaches, and the so-called structuralist view of theories in one that was established in the 20th century. In this paper we present a "fuzzy extension" of this view on scientific research. For this scientific work in progress we show that the "fuzzy structuralist approach" can be fruitful to interpret quantum mechanics and evolutionary biology.

Keywords: Fuzzy Sets, Computational Theory of Perceptions, Philosophy of Science.

1 Introduction

In this Introduction, to deal with basic definitions of basic terms, why not use WIKIPEDIA? – What is Science? – WIKIPEDIA's article "history of science" explains: "Science is a body of empirical and theoretical knowledge, produced by a global community of researchers, making use of specific techniques for the observation and explanation of real phenomena, this *techne* summed up under the banner of *scientific method*." [1] Here we notice that science and technology have a close connection and we can find many examples for interactions of science and technology in its history.

The mathematical theory of fuzzy sets is very young compared to others e.g. calculus, group theory, topology, and probability theory, and we know that fuzzy sets became well-known because of the enormous large number of application systems the field of applications in industry, economy and other areas since the first fuzzy steam engine that was conceptualized by Sedrak Assilian and Ebrahim Mamdani in the 1970s [2]. Nevertheless, the theory of fuzzy sets has also a great theoretical potential: this consists – in my point of view – a new interpretation of the scientific process using fuzzy sets. In the paper at hand we will have a look on both sciences – the body of empirical and theoretical knowledge – and scientists – members of the community of researchers who observe real phenomena and produce scientific theories – the protagonists in the process of knowledge production.

Since Galileo and Descartes science is extensively mathematical. Results of measurements, natural laws and scientific theories have got mathematical expressions, "the book of nature is written in the language of mathematics" and the scientific method incorporates its rigor.

One of the earliest areas of mathematics is computing. – What is computing? – WIKIPEDIA explains: "Originally, the word *computing* was synonymous with counting and calculating, and a science and technology that deal with the original sense of computing mathematical calculations." [3]

"Calculations" explains WIKIPEDIA as follows: "A *calculation* is a deliberate process for transforming one or more inputs into one or more results." [4]

In most cases inputs and outputs of mathematical computing are numbers: "A *number* is an abstract idea used in counting and measuring" as WIKIPEDIA explains. [5]

Computing with numbers was and is very successful in and also for the development of modern science but at the end of the 20th century Lotfi A. Zadeh proposed to compute with very different objects and in his first *AI Magazine*-article in the year 2001, *A New Direction in AI. Toward a Computational Theory of Perceptions* [6] he assumed "that progress has been, and continues to be, slow in those areas where a methodology is needed in which the objects of computation are perceptions—perceptions of time, distance, form, and other attributes of physical and mental objects." ([6], p. 73)

Already in 1996 he had published the article *Fuzzy Logic = Computing with Words* [8] where he proposed a method for reasoning and computing with words based on the theory of fuzzy sets instead of exact computing with numbers. He argued that "the main contribution of fuzzy logic is a methodology for computing with words. No other methodology serves this purpose." ([7], p. 103.)

"A *word* is a unit of language that carries meaning and consists of one or more morphemes which are linked more or less tightly together, and has a phonetical value." That's the explanation of WIKIPEDIA [8] and here is, what we reads about "meaning": "In linguistics, *meaning* is the content carried by the words or signs exchanged by people when communicating through language. Restated, the communication of meaning is the purpose and function of language. A communicated meaning will (more or less accurately) replicate between individuals either a direct perception or some sentient derivation thereof. Meanings may take many forms, such as evoking a certain idea, or denoting a certain real-world entity." [10]

About the concept of perceptions WIKIPEDIA says: "In psychology and the cognitive sciences, *perception* is the process of acquiring, interpreting, selecting, and organizing sensory information." [11]

Sensory information flows in our brain. We do not understand totally how the brain can analyze these sensations that may come from our five traditional senses – sight, audition, touch, taste, and smell – or others e.g. pain, heat, cold, gravitation, etc. and we do not know how the brain transforms this information into perceptions – it seems to be an active and constructive transformation, i.e. the brain is building and interpreting the sensory inputs and perceptions are this transformation's output. There are many examples that identical sensory inputs can result in different outputs. In these well-known examples we can perceive two perceptions that are contained in the painting and then we can swap between them.

"Humans have a remarkable capability to perform a wide variety of physical and mental tasks without any measurements and any computations. Everyday examples of such tasks are parking a car, driving in city traffic, playing golf, cooking a meal, and summarizing a story. In performing such tasks, for example, driving in city traffic, humans base whatever decisions have to be made on information that, for the most part, is perception, rather than measurement, based. The computational theory of perceptions (CTP), which is outlined in this article, is inspired by the remarkable human capability to operate on, and reason with, perception-based information." This is a section in Zadeh's article [6]. Here and in his 1999 article *From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions* [11], he pointed out that 1) measurements can be represented or manipulated by numbers, 2) we are able to represent or manipulate perceptions with words.

As we saw already, fuzzy logic (FL) is a methodology for CW and now Zadeh stated "CW provides a methodology for what may be called a computational theory of perceptions (CTP)." ([11], p. 108.) Therefore we have a methodology hierarchy (fig. 1):

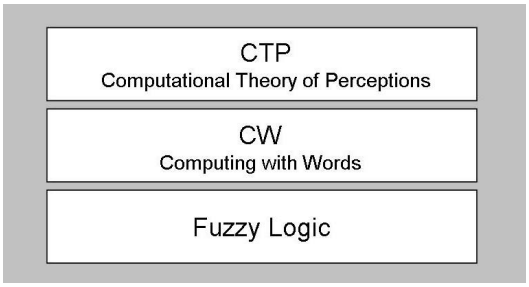


Figure 1: Zadeh's "methodology hierarchy".

Let's abandon WIKIPEDIA. Let's use Zadeh's methodology hierarchy in philosophy of science.

2 Philosophy of Science

Philosophy of science is the branch in philosophy that reflects the basis of science, their assumptions and implications, their methods and results, their theories and experiments. There are various scientific disciplines and therefore it is understood that we can distinguish between the philosophy of astronomy and physics, chemistry, and other empirical sciences, and we can also be interested in philosophies of social sciences and the humanities. However, these differing philosophies of scientific disciplines arose in differing historical periods and the earliest philosophical reflections on modern science started with theories and experiments in mechanics in the 17th century. Two main views in philosophy of science arose in about the same period: The philosophical view of *rationalism* came to fundamental, logical and theoretical investigations using logics and mathematics to formulate axioms and laws whereas the view of *empiricism* was an is to have experiments to find or prove or refute natural laws. In both directions – from experimental results to theoretical laws or from theoretical laws to experimental proves or refutations – scientists have to bridge the gap that separates theory and practice in science.

From the empiricist point of view the source of our knowledge is sense experience. The English philosopher John Locke (1632-1704) used the analogy of the mind of a newborn as a "tabula rasa" that will be written by the sensual perceptions the baby has later. In Locke's opinion this perceptions provide information about the physical world. Locke's view is called "material empiricism" whereas the so called idealistic empiricism was hold by George Berkeley (1684-1753) and David Hume (1711-1776), an Irish and a Scottish philosopher: there exists no material world, only the perceptions are real.

This epistemological dispute is of great interest for historians of science but it is ongoing till this day and therefore it is of great interest for today's philosophers of science, too. Searching a bridge over the gap between rationalism and empiricism is a slow-burning stove in the history of philosophy of science.

Two trends in obtaining systematic rational reconstructions of empirical theories can be found in the philosophy of science in the latter half of the 20th century: the *Carnap approach* and the *Suppes approach* (named after the German-US-American philosopher Rudolf Carnap (1891-1970) and the US-American philosopher Patrick Suppes (born 1922)). In both approaches, the first step consists of an axiomatization that intends to determine the mathematical structure of the theory in question. The difference of these views can be found in the way this task is performed. Rudolf Carnap was firmly convinced that only formal languages can provide the suitable tools to achieve the desired precision. Consequently the Carnap approach says that a theory has to be axiomatized within a formal language.

On the other hand, the Suppes approach uses informal logic and informal set theory. Thus, in this approach, one is able to axiomatize physical theories in a precise way without recourse to formal languages. This approach traces back to the proposal of Suppes in the 1950s to include the axiomatization of empirical theories of science in the metamathematical program of the French group "Bourbaki" [13]. The Suppes approach is the basis of what is called today the *structuralist view* in philosophy of science.

In the 1970s, the US-American physicist and philosopher Joseph D. Sneed developed informal semantics meant to consider not only mathematical aspects, but also application subjects of scientific theories in this framework, based on this method. In his book *The Logical Structure of Mathematical Physics* [14], Sneed presents this view as stating that all empirical claims of physical theories have the form "x is an S" where "is an S" is a set-theoretical predicate (e.g., "x is a classical particle mechanics"). Every physical system that fulfills this predicate is called a model of the theory. To give concrete examples, the class M of a theory's models is characterized by empirical laws that consist of conditions governing the connection of the components of physical systems. Therefore, we have models of a scientific theory, and by removing their empirical laws, we get the class M_p of so-called potential models of the theory. Potential models of an empirical theory consist of theoretical terms, i.e., observables with values that can be measured in accordance with the theory. This connection between theory and empiricism is the basis of the philosophical "problem of theoretical terms".

If we remove the theoretical terms of a theory in its potential models, we get structures that are to be treated on a purely empirical layer; we call the class M_{pp} of these structures of a scientific theory its "par-

tial potential models." Finally, every physical theory has a class I of intended systems (or applications) and, of course, different intended systems of a theory may partially overlap. This means that there is a class C of constraints that produces cross connections between the overlapping intended systems. In brief, this structuralist view of scientific theories regards the core K of a theory as a quadruple $K = \langle M_p, M_{pp}, M, C \rangle$. This core can be supplemented by the class I of intended applications of the theory $T = \langle K, I \rangle$. To make it clear that this concept reflects both sides of scientific theories, these classes of K and I are shown in figure 3. Thus we notice that M_{pp} and I are entities of an empirical layer whereas M_p and M_{pp} are structures in a theoretical layer oft the schema. In the next section we will extend this structuralist view of theories by fuzzy sets and fuzzy relations to represent perceptions as important components in the interpretation of scientific theories.

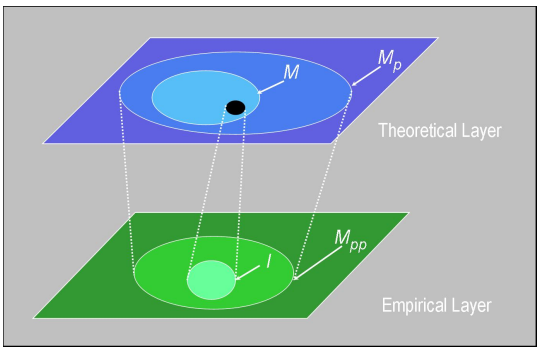


Figure 2: Empirical and theoretical structure layers to consider scientific theories.

3 Fuzzy Structuralism

Our modification of this approach pertains to the empirical layer in figure 2. Now, we will distinguish between real systems and phenomena at the one hand and perceptions of these entities at the other hand. Thus we introduce a lower layer – the real layer – and we rename our former empirical layer into "fuzzy layer" because the partial potential models and intended systems are not real systems because they have got a minimal structure by the scientist's observation. They are perception-based systems and therefore we have to distinguish them from real systems and phenomena that have no structure before someone imposes it tom them.

The layer of perceptions lies between the layer of real systems and phenomena and the layer of theoretical structures. According to Zadeh's Computational Theory of Perceptions (CTP) we represent perceptions in this intermediate layer as fuzzy sets. Whereas measurements are crisp perceptions are fuzzy, and because of the resolutions of our sense organs (e.g. aligning discrimination of the eye) perceptions are also granular, as Zadeh wrote in the *AI Magazine* in 2001: „perceptions, in general, are both fuzzy and granular or, for short *f*-granular [6]. Figure 3 shows Zadehs confrontation of crisp (*C*) and fuzzy (*F*) granulation of a linguistic variable.

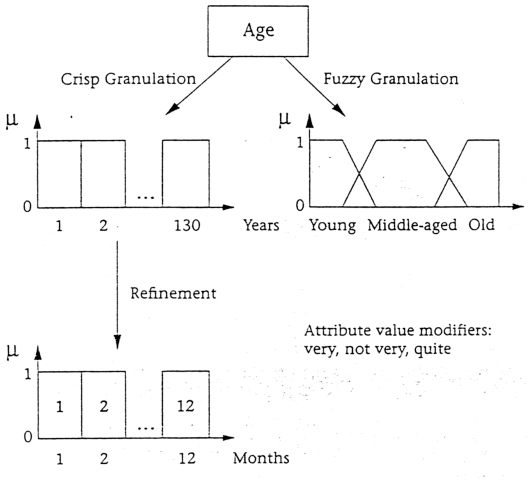


Figure 3: F- and C-Granulation of *Age*. [11].

When Zadeh established the CTP on the basis of Computing with Words (CW) that in turn bases on his theory of fuzzy sets and systems, he believed devoutly that these methodologies will get a certain place in science: "In coming years, computing with words and perceptions is likely to emerge as an important direction in science and technology" [7]. To take Zadeh at his word we establish his methodologies of fuzzy sets, computing with words and perceptions in our structuralist approach in philosophy of science. As aforementioned we introduce a fuzzy-layer of perceptions between the empirical layer of real systems and phenomena and the theoretical layer where we have the structures of models and potential models. Thus the relationship from real systems and theoretical structures has two parts: fuzzification and defuzzification.

- *Fuzzification: From Phenomena to Perceptions:*

Measurements are crisp and perceptions are fuzzy and granular. To represent perceptions we use fuzzy

sets, e. g. A^F, B^F, C^F, ... It is also possible that a scientist perceives not only single but interlinked phenomena, e.g. two entities move similarly or inversly, or something is faster or slower than the other, or it is more bright or dark, or it smells analogue etc. Such relationships can be charakterized by fuzzy-relations f^F, g^F, h^F, ….

- *Defuzzification: From Perceptions to Models:*

"Measure what is measurable and make measurable what is not so." is a sentence imputed to Galileo. In modern scientific theories this is the way to come from perceptions to measurements resp. quantities to be measured. We interpret this transfer as a defuzzification from perceptions represented by fuzzy sets A^F, B^F, C^F, … and relations between perceptions represented by fuzzy relations f^F, g^F, h^F, … to ordinary (crisp) sets A^C, B^C, C^C, … and relations f^C, g^C, h^C, … These sets and relations are basic entities to construct (potential) models of a scientific theory in the theoretical layer.

- *Theoretization: From Phenomena to Perceptions*

The composition of fuzzification and defuzzification yields the operation of a relationship T that can be named as *theoretization*, because it transfers phenomena and systems in the real (or empirical) layer into structures in the theoretical layer (fig. 4).

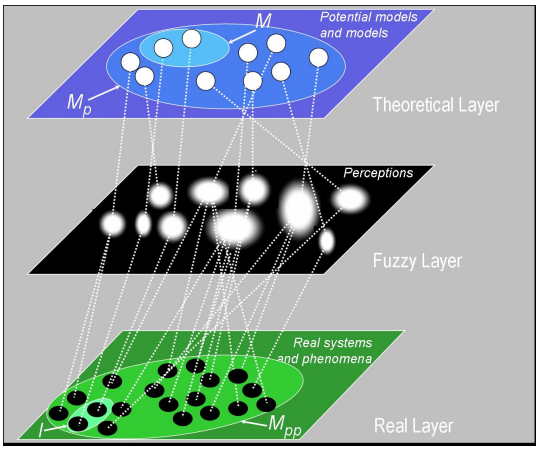


Figure 4: Empirical, fuzzy, and theoretical layer of crisp and fuzzy structures in scientific research.

In the structuralist view of theories the concept of theoretization is defined as an intertheoretic relation, i. e. a set theoretical relation between two theories T and T'. This theoretization relation exists if T' results from T by adding new theoretical terms und introduction of new laws that connect the former theoretical terms of theory T with this new theoretical terms in theory T'.

Sucessive adding of new theoretical terms establishes a hierarchy of theories and a comparative concept of theoriticity. In this manner the space-time theory arose from the euclidean geometry by adding the term „time" to the term „length", and from classical space-time-theory we get classical kinematics by adding the term „velocity". Classical kinematics turns to classical (Newtonian) mechanics by additional introduction of the terms „force" and „mass".

- The old theory T is covered with a new theoretical layer by the new theory T'.
- T-theoretical terms are not T'-theoretical but T'-non-theoretical terms and reciprocally they may not be some of the T-non-theoretical terms. The old theory must not change by the new theory in any way!
- In this hierarchy it holds that the higher in the hierarchy the more theoretical terms exist and the lower layers are characterized by the non-theoretical basic of the theory.

What happens in the lowest layer of this hierarchy? – Here exists a theory T with theoretical terms and relations but it is not a theoretization of another theory. This theory T covers phenomena and intended systems initially with theoretical terms. This is an initial theoretization because the T-theoretical terms are the only theoretical terms in this situation. They have been derived directly as measurements from observed phenomena. This derivation has got the name theoretization in the last section and it is a serial connection of fuzzification and defuzzification.

4 Fuzzy Structuralist Views on Scientific Disciplines

In this section we present two examples of new scientific theories in the 20th and in the 19th century in physics and in biology: quantum mechanics and evolutionism. Both theories have been established because previous theories did not provide the appropriate tool to explain new phenomena that have been observed in experiments or in nature.

4.1 The Case of Quantum Mechanics

The scientific revolution of quantum mechanics led to a new mathematical conceptual basis in physics concerning subatomic systems. Whereas in classical theories, the state of a system is represented by observables (e. g., position and momentum in the case of Newton's particle mechanics) that relate to human perception, states in quantum mechanics are not.

In classical physics we can observe these classical variables in experiments with classical systems. But quantum mechanical systems are not particles. The German physicist Werner Heisenberg (1901-1976), the Danish physicist Niels Bohr (1885-1962), and others introduced new theoretical systems (quanta) into the new quantum mechanics theory that differ significantly from those of classical physics. To determine the state of a quantum is much more difficult than to determine that of classical systems, as we cannot measure sharp values for all required variables simultaneously. This is the meaning of Heisenberg's uncertainty principle. We can experiment with a quantum in order to measure a position value, and we can also experiment with a quantum in order to measure a momentum value. However, we cannot conduct both experiments simultaneously and thus are not able to get both values for the same point in time respectively. Nonetheless, we can predict these values as outcomes of experiments at this point in time. Since predictions are targeted on future events, we cannot give them the logical values "true" or "false," but must assign them probabilities.

To determine the state of a quantum, we have to modify the classical concept: analogous to the state of classical systems, the state of a quantum mechanical system consists of all the probability distributions of all the system's properties that are formally possible in this physical theory. In classical physics the probability distributions for the observables position and momentum and all the other formally possible properties are marginal probability distributions of the unique probability distribution of the system's state. In quantum mechanics, however, there is no probability distribution that would be the meeting point of all probability distributions of all observables: in accordance with the uncertainty principle, not all of them are compatible.

In 1926, the German physicist Max Born (1882-1970) proposed an interpretation of the non-classical peculiarity of quantum mechanics, namely that the quantum mechanical wave function is a "probability-amplitude" [14]. This means that the absolute square of its value equals the probability of its having a certain position or a certain momentum if we measure the position or momentum respectively. Thus, the absolute square of the quantum mechanical state function equals the probability density function of its having a certain position or a certain momentum in the position or momentum representation of the wave function respectively. But there is no joint probability for the event in which both variables have a certain value, as there is no classical probability space (no Boolean algebra) that comprises these events.

In 1932, the Hungarian mathematician John von Neumann published the Mathematical Foundations of Quantum Mechanics [15], in which he defined the quantum mechanical probability amplitude as a one-dimensional subspace of an abstract Hilbert space, which is defined as the state function of a quantum. There are varying representations of a quantum's state, e.g., the "position picture" and the "momentum picture." These representations are complementary, which means that a subatomic system cannot be presented in both classical pictures at the same time.

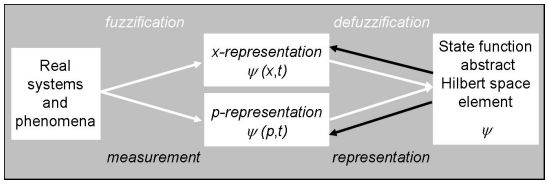


Figure 5: Quantum mechanical *theoretization* from real systems/ phenomena to a Hilbert space vector.

Figure 5 illustrates this case example of the fuzzy structuralist view in philosophy of science: A physician experiments with real subatomic systems or phenomena. He measures or predicts probability distributions of position- or momentum values. With these observable values he represents the quantum in the position or momentum picture with probability amplitudes $\psi(x,t)$ or $\psi(p,t)$ – we interpreted this process as *fuzzification* because quanta don't have position and momentum. The abstraction process from these classical pictures to get an abstract Hilbert space element ψ that representing the quantum theoretical *state* of the quantum, we interpret as *defuzzification*. The composition of these both processes is the quantum mechanical theoretization.

4.2 The Case of Evolutionary Biology

In the last third of the 20th century and in our first years of the 21st century biology is the leading scientific discipline. One of the most important researchers in this time and area was the evolutionary biologists and one of the architects of the synthetic theory of evolution, the German-Us-American Ernst Mayr (1904-2005) who was also an important historian and philosopher of biology. *What is biology?* was the title of his last book, and therein he distinguished between two rather different fields, mechanistic or functional biology and historical biology. "Functional biology deals with the physiology of all activities of living organisms, particularly with all cellular processes, including those of the genome. These functional processes ultimately can be explained purely mechanistically by chemistry and physics.

The other branch of biology is historical biology. A knowledge of history is not needed for the explanation of a purely functional process. However, it is indispensable for the explanation of all aspects of the living world that involve the dimension of historical time – in other words, as we now know, all aspects dealing with evolution. This field is evolutionary biology." ([16], p. 24)

Philosophy of science in the 20th century was based on exact sciences and especially on the new theories in physics (theory of relativity and quantum mechanics). Most philosophers of science in that century did have their background in physics but not in biology. There were only some scientists – and Mayr was one of them since the 1970s – who argued that we need a philosophy of modern biology that is different from philosophy of exact sciences and particularly Mayr emphasized this difference. E. g. in physics it is important to discover new facts or natural laws but in biology it is more important to develop new concepts and to round out concepts in being.

One reason for the missing of a philosophy of biology is that the basic principles of physics are simply not applicable to animate systems and another reason is that biology potentially bases on self-contained principles that are inapplicable to inanimate systems. The discovery of this difference in the basics of physics at the one hand and biology at the other hand was a fundamental intellectual revolution that began with Charles Darwin's *Origin of Species* in 1859. Thereupon modern biology emerged as an autonomous scientific discipline and a restructuring of the philosophy of science was prepared. To establish a philosophy of modern biology it was necessary 1) to eliminate and replace the principles of exact sciences by principles pertinent to biology and 2) to add new basic biological principles. Mayr "found that biology, even enough it is a genuine science, has certain characteristics not found in other sciences." ([16], p. 4) Among others he referred to the following:

… *unsharp separation of classes of phenomena*:
"The seemingly endless variety of phenomena, it was said, actually consisted of a limited number of natural kinds (essences or types), each forming a class. The members of each class were thought to be identical, constant, and sharply separated from the members of any other essence. Therefore variation was nonessential and accidental. […] Typological thinking, therefore, is unable to accommodate variation ad has given rise to a misleading conception of human race. Caucasians, Africans, Asians, and Inuits are types for a typologist that differ conspicuously from other human ethnic groups and are sharply separated from them. This mode of thinking leads to racism. Darwin completely rejected typological thinking and instead used an entirely different concept, now called *population thinking*" […]. ([16], p. 27)

… *variation or chance events*:
"One of the consequences of the acceptance of deterministic Newtonian laws was that it left no room for variation or chance events. […] The refutation of strict determinism and of the possibility of absolute prediction freed the way for the study of variation and of chance phenomena, so important in biology." ([16], p. 27)

… missing strict regularities:
"The philosophers of logical positivism, and indeed all philosophers with a background in physics and mathematics, base their theories on natural laws and such theories are therefore usually strictly deterministic. In biology there are also regularities, but various authors […] severely questions whether these are the same as the natural laws of the physical sciences. There is no consensus yet in the answer to this controversy. Laws certainly play a rather small role in theory construction in biology." ([16], p. 28)

In the times before the last decades in the 20th century these characteristics have got probabilistic for-

mulations, but we think that this is the wrong way to get fruitful solutions in the philosophy of biology. Physics concern the inanimate world with many indistinguishable objects and therefore it can be meaningful to argue with probabilities but: "In a biopopulation, by contrast, every individual is unique, while the statistical mean value of a population is an abstraction." ([16], p. 29) Because biological systems are high complex Mayr concluded: "Population thinking and populations are not laws but concepts. It is one of the most fundamental differences between biology and the so called exact sciences that in biology theories usually are based on concepts while in the physical sciences they are based on natural laws. Examples of concepts that became important bases of theories in various branches of biology are territory, female choice, sexual selection, resource, and geographic isolation. And even though, through appropriate rewording, some of these concepts can be phrased as laws, they are something entirely different from the Newtonian natural laws. ([16], p. 30)

Obviously we can consider these concepts with our fuzzy glasses and may be this is a good way to get interesting results in philosophy of biology. This means that the difference between the exact sciences and the life sciences is manifest in the missing exact structures of biological theories in the theoretical layer of sciences. Until today there is no part of defuzzification of a biological theoretization process and therefore no exact theory of evolution (fig. 6).

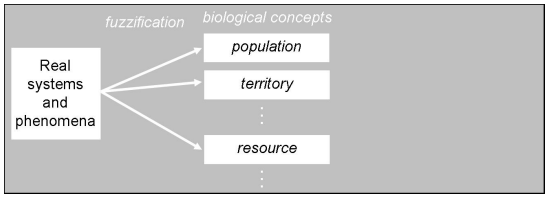


Figure 6: Fuzzification to evolutionary concepts.

5 Conclusion

The Computational Theory of Perceptions is an appropriate methodology to represent what happens in scientific research to bridge the gap between empirical observations and the abstract construction of theoretical structures. In the structuralist view of theories there is an empirical layer of real phenomena and systems that have some minimal structure and a theoretical layer of potential models and models that are full structured entities. But there is no representation of the scientist's perceptions. The modified view on the structuralist approach that is presented in this paper is scientific work in progress. This fuzzy structuralist view may open up a fruitful way to understand scientific research.

Acknowledgement

The author would like to thank Wolfgang Balzer (Munich) for his introduction to the structuralist view of theories. Many thanks go to Witold Pedrycz (Edmonton) for his encouragement to make a shift towards this interdisciplinary and philosophy of science-oriented research project. This work is partially supported by Grant # AB00158OFF from the Vienna Federal Data Center.

References

[1] http://en.wikipedia.org/wiki/History_of_science (24.4.2007).
[2] E. H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a Fuzzy logic controller, Int. J. Man-Machine Studies, 7, 1975, pp. 1-13.
[3] http://en.wikipedia.org/wiki/Computing (24.4.2007).
[4] http://en.wikipedia.org/wiki/Calculation (24.4.2007).
[5] http://en.wikipedia.org/wiki/Number (24.4.2007).
[6] L. A. Zadeh, A New Direction in AI. Toward a Computational Theory of Perceptions. AI-Magazine, 22, 1, 2001, pp. 73-84.
[7] L. A. Zadeh. Fuzzy Logic = Computing with Words, IEEE Transactions on Fuzzy Systems, Vol. 4, No. 2, 1996, pp. 103-111.
[8] http://en.wikipedia.org/wiki/Word (24.4.2007).
[9] http://en.wikipedia.org/wiki/Meaning_%28linguistic%29 (24.4.2007).
[10] http://en.wikipedia.org/wiki/Perception (24.4.2007).
[11] L. A. Zadeh, From Computing with Numbers to Computing with Words – From Manipulation of Measurements to Manipulation of Perceptions, IEEE Trans. Circuits and Syst.-I: Fund. Theory and Applications, 45, 1, 1999, pp. 105-119.
[12] N. Bourbaki, (pseudo.): *Elements of Mathematics: Theory of Sets*, Addison-Wesley, Reading, Mass. 1968.
[13] J. D. Sneed, *The Logical Structure of Mathematical Physics*, Dordrecht: Reidel 1971.
[14] M. Born, Zur Quantenmechanik der Stoßvorgänge, Z. für Physik, vol. 37, pp. 863-867, 1926.
[15] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton Univ. Press, Princeton, New Jersey, 1955.
[16] E. Mayr, *What Makes Biology Unique? Considerations on the autonomy of a scientific discipline*, Cambridge University Press 2004.

Fuzzy Set Theory and Philosophical Foundations of Medicine

Julia Limberg
Medical University of Vienna
Core Unit for Medical Statistics and Informatics
Spitalgasse 23, A - 1090 Vienna, Austria
e0107842@student.tuwien.ac.at

Rudolf Seising
Medical University of Vienna
Core Unit for Medical Statistics and Informatics
Spitalgasse 23, A - 1090 Vienna, Austria
Rudolf.Seising@meduniwien.ac.at

Abstract

Dealing with notions of health, illness and disease contains dealing with fuzziness. As the paper will demonstrate, states of these notions do not only exist or not exist. The medical philosopher and physician Sadegh-Zadeh introduced the notions of fuzzy health, fuzzy illness and fuzzy disease. A closer look will be taken on the concept of fuzzy disease. Because there are different possibilities to interpret the concept of disease, amongst others by linguistic and social backgrounds, Sadegh Zadeh indroduced potential candidates: complex „human conditions“. Afterwards, fuzzifications of life sciences will be extended and the fuzzification of the genome, including two approaches that deal with this theme, will be discussed.

Keywords: Fuzzy, Disease, Gene, Genomes, DNA.

1 Introduction

Health, Illness and Disease are notions originated in the theory of medicine that can't be defined in classical logic. Therefore, Kazem Sadegh-Zadeh, a doctor and philosopher in medicine, has been discussing the meaning of these notions since the 1980s in a new way. To illustrate his ideas, he chose the fuzzy-theoretical way. ([1], p. 607):

- "health is a matter of degree,
- illness is a matter of degree,
- and disease is a matter of degree."

In 1982, within the framework of a conference of medicine and philosophy, he blurred the notion of patienthood (being afflicted by a malady) as a new notion in the theory of medicine "of which the notion of health will be the additive inverse in the following sense:

Health = 1 – patienthood" [2].



Figure 1: Kazem Sadegh-Zadeh

In 2000 Sadegh-Zadeh presented a concept based on the Fuzzy Set Theory. In his article "Fuzzy Health, Illness and Disease" [1] he fuzzified the notions "Health", "Illness" and "Disease" and in doing so, he expanded their meanings. Particularly, he paid attention to the notion of fuzzy disease, of which the interpretation and definition proves to be very complex.

In this paper, after a short summary about this medical theoretical concept of fuzzy disease another approach of Sadegh-Zadeh, "Fuzzy genomes" [14], will be adduced. In this latter approach, the literally notions will be taken slightly into the background and the interpretation of diseases through their genomes will give the structural definition of disease. Because the definition of genes or genomes is not obvious, some basics will be described. Finally, another approach about the fuzzy concept of genes that bases in many aspects on the theory of Kazem Sadegh-Zadeh will be presented shortly.

2 Disease and Fuzzy Disease

Apparently, humans have thought about the existence of disease since humans are thinking. Many points in the bible refer to diseases that are linked with unbelief and punishment.

According to Karl Eduard Rothschuh (1908-1984), the well-known German physician and historian of medicine, this metaphysical interpretation of disease is still in the human's consciousness. Amongst this view, Rothschuh also mentioned a model of disease in association with philosophy. [3] Nevertheless, this view remained much generalized and interpreting the disease as a phenomenon or a "matter of evil" can't be considered satisfactory. So, Rothschuh introduced his third model of disease: The disease as a naturalistic model.

Thus, disease can be described in two ways. Firstly, disease is treated as a real entity. In compliance with this view, disease has an autonomous existence and has its roots in something like a seed. The disease grows like plants grow and this kind of growing can be regarded as the clinical tenor of a disease. The second naturalistic view regards the disease as a consequence of a disorganization of the organism and the organism's functional and structural components.

Similar to these models, respectively interpretations, Sadegh-Zadeh combined the notions of disease. By observing linguistic and social backgrounds, he introduced potential candidates for disease – complex "human conditions" like heart attack, apoplexia, cancer, etc.

A declaration like "heart attack is a disease" is well known in common language use. Other conditions, known in society as diseases, are now taken to approximate the notion of disease. These conditions do not merely arise from the biological state of the body. They may, as well, be described as large fuzzy sets which contain many different aspects of the sick person's environment, including religion and society.

Conditions which can be described as "pain", "distress" or "feeling of loneliness" may also be considered as aspects.

To define a definition as a whole one has to take a set of human conditions (D) with its corresponding criteria (C) into account. Following rules are established:

- Every element that is member of the basic set $\{D_1, D_2, \ldots, D_n\}$ is a disease and
- Every element that is similar to a disease with respect to the criteria $\{C_1, C_2, \ldots, C_n\}$ is a disease.

The first declaration seems to be clear. However, the second declaration poses a problem, because *similarity* has to be described. For this reason, the fuzzy set difference of two fuzzy sets A and B – *differ(A, B)* – is introduced as a starting point, which is calculated as follows:

$$differ\ (A,B) = \frac{\sum_i \max(0, \mu_A(x_i) - \mu_B(x_i)) + \sum_i \max(0, \mu_B(x_i) - \mu_A(x_i))}{c(A \cup B)}$$

C, given in the denominator, is the sum of the membership values of the corresponding fuzzy set (*fuzzy set count*). For instance, there is a fuzzy set X with $X = \{(x, 0.6), (y, 0.9)\}$, $c(X)$ will be calculated as: 0.6 + 0.9 = 1.5.

Back to fuzzy difference: Let's assume that there is a fuzzy set Y with $Y = \{(x, 0.7), (y, 0.4)\}$. Fuzzy-difference *differ(X, Y)* is calculated as

$$\frac{(0+0.5)+0.1+0)}{1.6} = 0.375$$

X differs from Y to a degree of 0.375.

Similarity of two fuzzy sets is resulted from the inversion of fuzzy difference. According to the example above, similarity would be given as: 1 – 0.375 = 0.625.

In order to avoid comparing apples and oranges, descriptions of similarity should be reduced to assimilable subsets. For example, one raises the question how similar are the two diseases D_i and D_j considering a few criteria $\{C_1, C_2, \ldots. C_m\}$.

Let's assume A to be a fuzzy set of arbitrary dimension and X as a part of this set; so $A \backslash X$. Human conditions, like heart attack and stomach ulcer, can be arranged according to their assimilable criteria $\{C_1, C_2, \ldots. C_m\}$:

- heart_ attack\{$(C_1, a_1), (C_2, a_2), \ldots, (C_m, a_m)$}
- stomach_ulcer\{$(C_1, b_1), (C_2, b_2), \ldots, (C_m, b_m)$}
- heart_ attack\{bodily_lesion, 1), (pain, 0.7), (distress, 0.8)}
- stomach_ulcer\{(bodily_lesion, 1), (pain, 0.3), (distress, 0.5)}

To calculate similarities between fuzzy sets, the following theorem is used:

$$Theorem: similar(A, B) = \frac{c(A \cap B)}{c(A \cup B)}$$

Similar comparisons include several degrees of partial (p) similarity, symbolized as p-similar($A\backslash X$, $B\backslash Y$), under the terms of the following definition:

$$p\text{-similar}\ (A\backslash X, B\backslash Y) = r, \text{ if similar } (X, Y) = r.$$

According to the example above and using the theorem above, this would mean:

$$p\text{-similar (heart_ attack}\backslash X, \text{stomach_ulcer } \backslash Y) = 0.72$$

Assuming that $\{D_1, ..., D_n\}$ would be a small set of human conditions, because of a set of criteria $\{C_1, \ldots, C_n\}$ which these conditions have in common. Each of these conditions is interpreted in a certain society as a disease.

For this society there is an agreement of degree ε of partial similarity. This degree is a pillar of this society's notion of disease:

- Every element in the basic set $\{D_1, \ldots, D_n\}$ is a disease
- A human condition H\X is a disease, if there is a disease $D_i\backslash Y \in \{D_1, \ldots, D_n\}$ and there is a $\varepsilon > 0$, so that p-similar $(H\backslash X, D_i\backslash Y) \geq \varepsilon$

Granted, that there is the criteria set heart_attack\{(C_1, 1), (C_2, 0.7), (C_3, 0.8)} as an element in basic set $\{D_1, \ldots, D_n\}$ and therefore a disease by definition.

The question of whether something that is not contained in basic set $\{D_1, \ldots, D_n\}$, like haemorrhoids, could be identified as a disease is decided by the degree ε of partial similarity.

For example, $\varepsilon = 0.6$ is asked and there is a human condition like: haemorrhoids\{(C_1, 0.9), (C_2, 0.2), (C_3, 0.55)}, the result is:

p-similar (haemorrhoids\$X$, heart_attack\$Y$) = 0.66.

Since 0.66 > 0.6 haemorrhoids can be described as disease.

According to this definition a proper choice of ε is essential: The smaller the ε that is chosen, the more diseases will exist and vice versa.

However, the value of ε is not chosen by the doctor, but by society.

Anyway, this notion of disease is a notion that can be comprised in binary logic, because there is made an explicit difference between states that are consistent with a disease and states that are not. Therefore, Sadegh-Zadeh expands this notion of disease to *a* notion of "disease to a certain degree". This can be achieved by the definition as follows:

Let's assume $\mathcal{H}$ to be a small set of human conditions. A fuzzy set $\mathcal{D}$ over $\mathcal{H}$ is considered as *a* set of diseases only if there is a subset $\{D_1, \ldots, D_n\}$ of $\mathcal{H}$ and there is a function $\mu_{\mathcal{D}}$, so that:

$\mu_{\mathcal{D}}$: $\mathcal{H}$-> [0,1] with $\mu_{\mathcal{D}}(H_i\backslash X)$ =

- 1, if $H_i\backslash X \in \{D_1, \ldots, D_n\}$, called prototype disease
- ε, if there is a prototype disease $Hj\backslash Y$ with p-similar $(H_i\backslash X, H_j\backslash Y) = \varepsilon$, and there is no prototype disease $H_k\backslash Z$ with p-similar $(H_i\backslash X, H_k\backslash Z) > \varepsilon$ and $\mathcal{D} = \{(H_i, \mu_{\mathcal{D}}(H_i)) \mid H_i \in \mathcal{H}\}$.

In this expanded definition a fuzzy set of following kind is created:

$\mathcal{D} = \{(D_1, \mu_{\mathcal{D}}(D_1)), \ldots, (D_q, \mu_{\mathcal{D}}(D_q))\}$, which consists of individual archetypes of diseases, which are all members of the set $\mathcal{D}$ to different degrees.

The membership-degree $\mu_{\mathcal{D}}(D_i)$ is of interval [0,1].

These new findings are now applied to the example of haemorrhoids:

- haemorrhoids\{(C_1, 0.9), (C_2, 0.2), (C_3, 0.55)}.

These criteria are compared with a prototype disease. The already known set heart_attack is called into the equation: heart_attack\{(C_1, 1), (C_2, 0.7, (C_3, 0.8)}.

Drawing a comparison shows that haemorrhoids may be considered as a disease to a degree of 0.66.

Accepting another individual with another set, haemorrhoids\{(C_1, 0.2), (C_2, 0.1), (C_3, 0.1)}, and taking this individual in comparison to heart_attack would result in a membership of degree 0.16 to the set of diseases. From this, we conclude that a person may have a disease to a certain degree and that this person may have no disease to a certain degree at the same time. [1]

To demonstrate Sadegh-Zadeh's ideas of the fuzzy-disease, we implemented a computer program that we already presented in an earlier paper. [4]

By making requests of membership degrees of certain symptoms, similarities to existing diseases that are stored in a database are calculated and decisions will be made on whether an entered group of symptoms provides an indication of the existence of a disease.

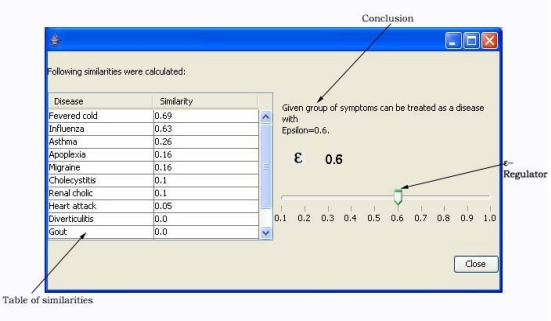


Figure 2: Form of calculation of similarities – analysis.

The preset ε-value is compared with the reference value, the highest value of similarity (listed at the top of the table of similarities). If the value of ε is higher than the reference value or is equal to this value, at least, the conclusion is drawn that the given group of symptoms can be treated as a disease to an extent of the degree ε.

An organisms attributes, that can be measured, e.g. by determine the temperature of the body, or viewed, e.g. "wheals", are called phenotypes. [5]

So far, by indicating different symptoms, we discussed the phenotype attributes of a disease. Now, our point of interest will be the cause and the roots of a disease.

According to the director of the National Institute of Environmental Health Sciences Kenneth Olden "diseases are caused by multi-factorial interactions of genes and environment". [6]

Let's conjure up the definition of diseases that Rothschuh claimed: Diseases are based on seeds. One may assume that seeds grow better if there is a fertilizer and if seeds are planted in a convenient kind of ground. So, the fertilizer can be compared to the environmental influence and the ground, as a living material, can be compared with the human body and therefore his genes. Now, it is time for a ground survey – or rather a definition of genes.

3 Genes and Fuzzy Genes

In June 27, 1994, Bill Gates quoted in *Business Week* "The gene is by far the most sophisticated program around" and in the *Stanford Encyclopedia of Philosophy* H.-J. Rheinberger declared inter alia: "There has never been a generally accepted definition of the 'gene' in genetics. There exist several, different accounts of the historical development and diversification of the gene concept as well. Today, along with the completion of the human genome sequence and the beginning of what has been called the era of postgenomics, genetics is again experiencing a time of conceptual change, voices even being raised to abandon the concept of the gene altogether." [7]

In this paper, the concept of genes will not be abandoned. Instead of that, we will keep in mind that the definition of the gene has always been changing according to newest findings in science and accept the definition of the gene, formulated by the Sequence Ontology Consortium as "a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions". According to Karen Eilbeck, a co-ordinator of the Sequence Ontology Consortium, it took 25 scientists nearly two days to reach this definition of a gene. [8]

The "locatable region of genomic sequence", respective all the genetic instructions for the development and functioning of living organisms are contained within the DNA[1]. The DNA is a double helix and made from many units of nucleotides. A nucleotide consists of a base, a sugar and one ore more phosphate groups.

In the DNA, the backbone consists of the phosphates and sugars and the purine and pyrimidine bases adenine (A), guanine (G), cytosine (C) and thymine (T) are inward-looking.

A segment of DNA may code a protein. The genetic code describes the relationship between the DNA sequence and the protein sequence. Only one of the two strands of the DNA codes the protein.

[1] As an exception, one has to mention that there is a group of viruses that have RNA-genomes.

A coded DNA sequence consists of many codons, which are read from a certain starting point.

Every codon consists of three nucleotides and accords to one amino acid. Indeed, the DNA of a gene contains all the Information that is needed for synthesis of protein, but DNA isn't the direct matrice for its creation.

In fact, the genetic information of the DNA first has to be transcribed in the base sequence of a single-stranded ribonucleic acid – the RNA, which contains of the sugar ribose and the base uracil (U) instead of 2-deoxyribose and thymine. After other transformations, a working copy of the gene is achieved, called messenger-RNA (mRNA). This mRNA describes a transportable information system for the synthesis of a specific protein [5][9][10].

So far, we've described the genesis, the location and the composition of a gene. Now, the reader might be interested in the questions of "How many genes are in the individual's organism?", respective "How to count bases?" and "How may one analyze given bases in order to recognize a disease?".

According to Ernst Peter Fisher, who is professor of the history of science at the university of Constance, the concept of genes is fuzzy. As if one only takes the sequences of the genome into account, one also has to recognize that genes are fuzzy entities. There are sequences that overlap and that recur – from those sequences the membership to a specific gene is unclear and counting is only possible if there is a definition [11]. Although there are problems in finding a precise definition and in giving an impression of the count of genes, definitions and the number of genes are still in the human's point of interest in order to verify the gene predictions.

In 1995, Victor Velculescu developed a new technique called SAGE (serial analysis of gene expression). First, RNA molecules are isolated and then their sequence is copied transcribed to DNA. Finally, one receives a piece of 20 pairs of bases from this copy of DNA; with this piece the gene that is considered may be identified [12][13]. To compare and to determine genes, one consults a database that stores information about already known sequences. The National Center for Biotechnology Information (NCBI)[2] offers a database that stores a collection of all publicly available DNA sequences that can be drawn for comparisons.

Comparing sequences is a difficult task and there are many different methods that describe possibilities and algorithms. In public databases, genes are stored as well-defined, crisp sets of bases.

As already adumbrated, there are good reasons for considering the gene in a fuzzy-theoretical way.

In the article "Fuzzy Genomes", Sadegh-Zadeh justifies the need of a fuzzy definition and shows a way of realization [14]. Furthermore, he developed a method to compare these fuzzy genomes.

The last section of this paper deals with a short summary of Sadegh-Zadeh's fuzzy genomes and another approach of sequence comparison by Angela Torres and Juan J. Nieto [15] will be taken into contrast.

4 Fuzzy genomes and comparisons on base sequences – two approaches

Both theories that will be presented are based on a fuzzy-theoretical definition of the gene and particular attention is paid on the comparison in order to identify diseases.

4.1 Approach 1: Fuzzy genomes by Sadegh-Zadeh

Having analysed a human's germplasm, one has to decide if a given section of RNA is a disease, respectively, a special form of disease, such as HIV.

Decisions on these cases are made by comparing known sequences of diseases with the section of RNA. Therefore, Sadegh-Zadeh transforms DNA and RNA to fuzzy sets.

According to the "RNA alphabet" of the bases <U, C, A, G>, U could be written as 1000, because appearance of U is true and there is no C, no A and no G. C could be written as 0100, A as 0010, G as 0001. So, an RNA sequence UACUGU can be transformed into the following bit sequence: 1000001001001000 00011000.

[2] http://www.ncbi.nlm.nih.gov/

This sequence has a length of 24 and can be represented in a 24-dimensional bit vector: (1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0).

To combine all possible appearances of a character in the alphabet, Sadegh-Zadeh builds up a fuzzy-matrix. This matrix contains all bases, or rather every character of the RNA alphabet and its membership to a given base sequence.

Considering every single base when building up a matrix, there are two points of interest:

1) Where is the position of the base and
2) To what extent and accordingly to which membership are the bases given.

For example, there is a RNA sequence UAC. Thus, the sequence consists of three bases.

The corresponding fuzzy-matrix would be:

Fuzzy_matrix (UAC) =

<(U in 1, 1), (C in 1, 0), (A in 1, 0), (G in 1, 0)

(U in 2, 0), (C in 2, 0), (A in 2, 1), (G in 2, 0)

(U in 3, 0), (C in 3, 1), (A in 3, 0), (G in 3, 0)>

An example, considering the first row: In UAC, U is at the first position. Therefore, U has membership 1, at position 1 (write U in 1, 1), there is no C at position 1, therefore C has membership 0 to the first position, the same with A and G.

One has to mention, that appearance or membership isn't obliged to be either 1 or 0. Membership may accept every value between 0 and 1. This could be the case, if a base is defective and can't be defined in one single class or there is uncertainty about the correct identification of a piece of a sequence.

Having a (m x n)-matrix, one may build up the corresponding (m x n)-vector with length n - thus, creating an n-dimensional vector.

Dealing with dimensions, Sadegh-Zadeh used the hypercube that was introduced into fuzzy set theory by Bart Kosko in [16].

For instance, there is a fuzzy set A with $\{(x_1, a_1), \ldots, (x_n, a_n)\}$. This set has an n-dimensional vector $(a_1, \ldots, a_n)$ and its members are in [0,1]. Consequently, it is a point in an n-dimensional unit hypercube $[0,1]^n$. The n appearances of x_i are assigned to coordinates in the hypercube.

According to the ground set Ω with $n \geq 1$ members, the fuzzy powerset is given as $F(2^{\Omega})$. Hence, one builds up a hypercube with 2^n corners.

If A consists of 3 members, one could display A in a 3-dimensional cube (with 2^3=8 corners).

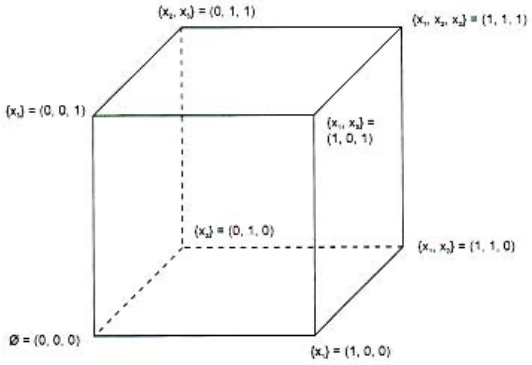


Figure 3: 3-dimensional cube [14].

For example, considering the fuzzy set A as $\{(x_1, 0.5), (x_2, 0.4), (x_3, 0.7)\}$. According to the coordinate axes x_1, x_2, x_3, A would be a point (0.5, 0.4, 0.7) in the 3-dimensional hypercube.

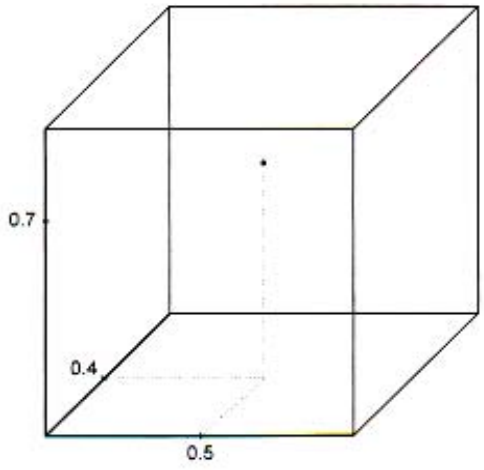


Figure 4: Point (0.5, 0.4, 0.7) in 3-dimensional cube [14].

A sequence of DNA or RNA is a point in a $4n$-dimensional hypercube. The $4n$ dimensions are reasoned by the fact that there are four bases, so m=4.

Displaying a sequence as a point in an n-dimensional hypercube depicts its order in comparison to any other sequence.

This suggests that one calculates differences or similarities between ordered fuzzy sets: Differences and, consequently, similarities of two polynucleotides A and B may be calculated after the definition of the difference, already mentioned as:

$$differ\ (A,B) = \frac{\sum_i \max(0, \mu_A(x_i) - \mu_B(x_i)) + \sum_i \max(0, \mu_B(x_i) - \mu_A(x_i))}{c(A \cup B)}$$

or as Hamming distances in the cube.

Analogous, similarity between sequences are calculated as the inverse of difference or as:

$$similar(A,B) = \frac{c(A \cap B)}{c(A \cup B)}$$

A degree that determines the vagueness of a set is referred to its fuzzy entropy, denoted as *ent*, so that the hypercube is mapped as follows:

$$ent : F(2^{\Omega}) \to [0, 1]$$

Considering a set's entropy, one is interested in determining the *nearest* and the *farthest* set.

Let's assume that there is a set A = (0.2, 0.8, 0.6). Then the nearest and farthest sets are given as: A_{near}= (0, 1, 1) and A_{far} = (1, 0, 0).

According to the hypercube, there is always a nearest and a farthest vertex to A. Fuzzy entropy of any set A is defined as the ratio of the Hamming distance from vertex A_{near} to A_{far}:

$$ent(A) = \frac{l^1(A, A_{near})}{l^1(A, A_{far})}$$

Clarity, denoted as *clar(A)*, is defined as the opposite to fuzzy entropy:

$$clar(A) = 1 - ent(A)$$

At the edges, *clar(A)* = 1 and in the centre of the hypercube *clar(A)* = 0.

From that, we can follow that real, what means existing, polynucleotides like UAC has entropy of 0 and therefore clarity of 1, whereas a fuzzy polynucleotide is near and far from a real polynucleotide to a certain degree [14].

4.2 Approach 2: The fuzzy polynucleotide space: basic properties by Torres and Nieto

As this approach is partly similar to the approach that has been presented firstly, only the common bases and the differences will be described.

Torres' and Nieto's approach bases on Sadegh-Zadeh's approach; a 12-dimensional hypercube and the RNA-Alphabet are also taken in consideration. The main difference between their theory and Sadegh-Zadeh's theory results from the fact, that Torres and Nieto do not generalize the 12 dimension to *n* dimensions. Instead of doing so, they leave at 12 dimensions and compare genes by frequency of occurrence of a certain base.

In a given sequence there are the four bases U, C, A, G and three of these bases build up a codon. A sequence is now characterized because of the frequency of every single base at any position in every codon.

For example, we consider a sequence as: UACUGA. The codons would be given as $codon_1$: UAC and $codon_2$: UGA. Considering U, we would conclude that U occurs at position 1 in $codon_1$ and also, U occurs at position 1 in $codon_2$. All in all, U occurs 2 times in the whole sequence at position 1. Thus, the fraction of U in the first base is calculated as 2/2 = 1 = 100%.

By applying this method to every base, 3*4 = 12 fractions will be calculated, as there are 3 positions in a codon and 4 possible bases.

The following table shows a table of fractions of a sequence S_1 that would be given for example as: S_1 = CAUUGU

Table 1: Table of fractions of sequence CAU UGU.

Position	**Count of nucleotides**					**Fractions**			
	U	C	A	G	*Sum*	U	C	A	G
1.	1	1	0	0	*2*	0.5	0.5	0	0
2.	0	0	1	1	*2*	0	0	0.5	0.5
3.	2	0	0	0	*2*	1	0	0	0

After calculating fractions of a base, a vector of fractions with length = 12 remains and stands for the whole sequence. In the example above, the sequence S_1 with CAU UGU would result in a vector V_1 = {0.5, 0.5, 0, 0, 0, 0, 0.5, 0.5, 1, 0, 0, 0}.

In order to compare a sequence with another sequence, the sequences' vectors of fractions are considered and the similarity between these sequences is calculated as:

$$sim(A,B) = c(A \vee B) / C$$

with

$$C = \left(\frac{a_1 + b_1}{2}, \ldots, \frac{a_n + b_n}{2} \right)$$

The difference between sequences is given as:

$$dif(A,B) = 1 - sim(A,B)$$

They conclude that:

- $sim(A,B) \neq similar(A,B)$ and
- $dif(A,B) \neq differ(A,B)$.

Every sequence of bases can be compared with every other sequence, by comparing the 12 fractions of the sequences, whether they are of the same length or not [15].

Acknowledgement

This work is partially supported by Grant # AB00158OFF from the Vienna Federal Data Center.

References

[1] Kazem Sadegh-Zadeh: Fuzzy Health, Illness, and Disease. Journal of Medicine and Philosophy, Vol. 25, No. 5 (2000), pp. 605-638: 612

[2] Kazem Sadegh-Zadeh: Fundamentals of clinical methodology: 3. Nosology. Artifical Intelligence in Medicine, 17 (1999) 81-108

[3] Karl Eduard Rothschuh: Der Krankheitsbegriff (Was ist Krankheit?), Hippokrates 43 (1972) 3-17

[4] Julia Limberg and Rudolf Seising: Fuzzy Health, Illness, and Disease Sadegh-Zadeh's framework and a program to identify diseases. Proceedings of the FUZZ-IEEE 2007, IEEE International Conference on Fuzzy Systems :: Intelligence is Fuzzy::, Imperial College, London, UK, 23-26 July, 2007 (in print).

[5] Benjamin Lewin: *Molekularbiologie der Gene*. Spektrum Akademischer Verlag. 1998.

[6] OLPA,107th Congress, Hearing I: Testimony on Environmental Health Problems. Available at: http://olpa.od.nih.gov/hearings/107/session2/testimonies/environmental.asp

[7] Hans-Jörg Rheinberger: Gene. Stanford Encyclopedia of Philosophy. Available at: http://plato.stanford.edu/entries/gene/

[8] Neurophilosophy. The Concept Of A Gene Has Envolved. Available at: http://neurophilosophy.wordpress.com/2006/05/30/the-concept-of-a-gene-has-evolved/

[9] Gerhard Thews, Ernst Mutschler, Peter Vaupel: *Anatomie, Physiologie, Pathophysiologie des Menschen*. Stuttgart: Wissenschaftliche Verlagsgesellschaft, Auflage: 5. 1999.

[10] DNA: http://en.wikipedia.org/wiki/DNA

[11] Ernst Peter Fisher: Gene sind wie Sand am Meer. Available at: http://www.netdoktor.de/feature/gene_fischer.htm

[12] SAGE: http://en.wikipedia.org/wiki/Serial_Analysis_of_Gene_Expression

[13] Ernst Peter Fischer: *Das Genom*. Frankfurt: Fischer (Tb.). Auflage: 2., 2002.

[14] Kazem Sadegh-Zadeh: Fuzzy Genomes. Artificial Intelligence in Medicine, 18 (2000) pp. 1-28.

[15] Angela Torres and Juan J. Nieto: The fuzzy polynucleotide space: basic properties. In: Bioinformatics, Vol.19, No.5 (2003), pp.587-592

[16] Bart Kosko: Fuzzy Entropy and Conditioning, Information Sciences, 40 (1986), pp. 165-174.

Fuzzy Logic as a Theory of Vagueness: 15 Conceptual Questions

Jeremy Bradley
Vienna University of Technology
jeremy.bradley@gmx.net

Abstract

In spite of its successes as a tool in the field of engineering, fuzzy set theory has yet to achieve the universal footing that probability theory has across the various fields of mathematics, technology, philosophy and psychology. This paper sets out points of critique brought up regarding the fuzzy approach and seeks to analyze them, focusing on the question of whether anything that can be done about these matters. Do these criticisms have any practical relevance or any relevance with respect to the intended fields of usage. Do they or do they not diminish fuzzy logic's suitability as a theory of vagueness?

Keywords: Fuzzy Logic, Theory of Vagueness.

1 Introduction

Since its initial presentation in 1965 by Lotfi A. Zadeh in his paper *Fuzzy Sets* [34], fuzzy set theory has successfully established itself as a useful tool in the field of engineering. Though its purpose and validity in any context were highly controversial in the early years, this initial criticism was defused by the practical success of fuzzy set theory, to a large degree under the name of "fuzzy logic." This began with Assilian's and Mamdani's steam engine in the 1970s [17] and has extended over an ever-expanding range of applications, from noodle cookers to washing machines, up to the present day. The history of fuzzy set theory's birth, development and progression has been documented by Rudolf Seising in his book *The Fuzzification of Systems* [26].

The acceptance of fuzzy logic as a technical tool, however, has not necessarily led to an acceptance of fuzzy set theory as a theory of vagueness, or as an instrument handling natural language - a matter over which there is a certain rift within the fuzzy community, which will be examined later.

In one of these groups - the so-called "technicians" - many might argue that these aims are not, never were, and never will be the intended domain of fuzzy logic. However, probability theory, dealing with a type of uncertainty orthogonal to fuzziness, has in its longer lifetime managed to become a principle that to a large degree is accepted as an integral part of almost every aspect of life. It has managed to establish itself as much more than "just a tool," while dealing with a concept (degrees of probability - "with a degree of likeliness of 0.5, Austria will beat Liechtenstein in the upcoming match") no less natural than that of fuzzy logic (degrees of applicability – "Italy played rather well in yesterday's match").

Many philosophers and psycho-linguists have argued and still argue that fuzzy logic must deal with some of these points in order to receive any recognition as a valuable theory in these respective fields. Among other things, this paper represents the starting point of an attempt to analyze how relevant the said points actually are, and whether it would be possible to overcome them or whether trying to solve these problems is basically an attempt to teach an elephant to fly.

The core of this project are fifteen questions formulated by Christian G. Fermüller of the Vienna University of Technology after he was involved in organizing and participating in, the *Prague International Colloquium: Uncertainty - Reasoning about probability and vagueness* in the Czech Republic in September 2006. The context of this conference was not that of a fuzzy conference, but of philosophy, and, in particular, the theory of vagueness. Consequently, these questions stem from a narrower context of fuzzy logic as a theory of vagueness - the decision to approach the fuzzy community with these questions led to some interesting and unexpected results.

The 15 questions attempt to summarize the points of

criticism and contempt encountered in and around the conference in Prague, among other places. Through international survey work and extensive participation from all over the world, it has been possible to formulate some examinations of the points brought up. Hopefully, further contributions will, in future, make a more extensive analysis of these points possible.

It should be noted that this paper does not attempt to question the feasibility of fuzzy logic in any of its present day applications or to disqualify fuzzy logic in general. If anything, it is an attempted defense of fuzzy logic against points brought up against it, as anything but a "useful tool", outside of the fuzzy community.

2 Fifteen Points of Critique - A Reference Table

The following table provides an overview of the questions that will be addressed individually in the next section.

A number of participants graded the severity of the points raised on a scale from 1 to 10, where 1 denoted that the respective question was a "very serious objection," and 10 denoted that the point raised was "not serious at all and/or completely irrelevant." An average of the evaluations of each point was computed.

Also, a count was kept of how many participants had something specific to say about each question and considered it important enough to justify further contemplation.

At this point, this table can only serve as a reference. Possibly, with an increasing number of participants, it will eventually evolve into something resembling statistics.

Table 1: Overview over the Questions

QUESTION	SIGNIFICANCE	ANSWERS
1	5.5	9
2	7	9
3	6.5	8
4	7	5
5	7	6
6	7	8
7	6	8
8	7	8
9	7	7
10	6.5	9
11	6	4
12	7	7
13	7	8
14	7	9
15	6	5

3 Fifteen Points of Critique – The Questions, Analyses, Attempts at Answering Them

3.1 Improper Precision

What do truth values like 0.5476324 *mean? How do we arrive at such values? Does FL provide any means to distinguish reasonable from unreasonable attributions of values? (A complete theory of vagueness should provide answers to such questions.)*

In a fuzzy environment one can, in fact, encounter truth values of which the interpretation can be difficult, if the granularity of the application permits. In some simple cases, a numerical value other than 0 and 1 can have a meaning that could be considered straightforward – for example, when establishing a degree of applicability of the term "luminous" to a pixel in a grayscale bitmap image, allowing 256 different gray values, distributed at equal distances over the spectrum, assigning a pixel with the luminosity value 128 the fuzzy value 0.5 seems intuitive. With this same system, it is also possible to get more "complex" fuzzy values that still remain intuitive. For example, a luminosity value of 129 would, under linear mapping, lead to a fuzzy value of 0.50390625 ($\frac{129}{256}$) – a fuzzy value no less applicable than a probability value of the same sort.

However, when venturing out of this very narrow realm of very specific technical applications, the problem remains. What does it mean to say that Ginger loves Fred to a value of 0.5476324? Intuitively, colleagues agree, absolutely nothing. It can be seen as an abstraction of an actual value, such as any exact number will always be – a person listing his weight as 72 kilograms will rarely have this weight at any level of exactness. However, in the realm of real numbers, one can still determine what one is creating an abstraction from.

3.2 Linear Ordering of Truth Values

*This seems to force one to judge the relative truth of intuitively incomparable statements, such as e.g., "John is tall" "Mary is rich," "Ginger loves Fred," etc. How can this be justified? (Note: it is insufficient to point out that algebraic models may also be nonlinear. The deeper worry here is that this does not *explicitly* reflect the "incomparability" of at least*

some vague propositions.)

One element of classical logics which is actually preserved in fuzzy logics is linearity – which can be described with the following axiom:

$$(a \rightarrow b) \vee (b \rightarrow a)$$

Given that the value of a is smaller or equal to the value of b, any t-norm based fuzzy logic will compute $(a \rightarrow b)$ as 1. Thus, the content of this statement can be summarized as:

$$(val(a) \leq val(b)) \vee (val(b) \leq val(a))$$

The validity of this axiom in a fuzzy environment states that a linear ordering of fuzzy values is always legit. However, a linear ordering of values might not be desired, though. Thus, this axiom does not seem to fit into a theory of vagueness in the eyes of many philosophers. If you cannot clearly accept or reject statements, how can you compare them? The issue is, thus, that the mathematical frameworks allow a comparison which does not seem natural or intuitive.

The consensus here seems to be that only due to both values in a comparison operating by the same system, and thus theoretically comparable, they are not necessarily comparable in context, even if the logical framework would allow it. Statements such as "John is taller than Bill is fat" are not answerable in terms of crisp numbers in natural language, even if both values are theoretically crisp.

However, within the realm of crisp numbers, comparison between incomparable statements would be comparing two values of different units with each other – something that does not happen in a fuzzy comparison. In a fuzzy comparison, both values have the same dimensionality of truth, which is noted as a dimensionless value on an interval stretching from 0 to 1. While one cannot compare inches with pounds, one should be able to compare two truth values, no matter where they came from. Under the same conditions, such comparisons are quite possible in probability – it is possible to evaluate statements such as "it is more likely that I will die in a plane crash than that I will win the lottery", also if the safety of my travels has no connection whatsoever with the results of a lottery. Nevertheless, the comparison is graspable here.

3.3 Truth Functionality

*This seems to clash with many intuitions (see, e.g., D. Edgington for very explicit arguments against truth functional connectives applied to vague propositions [5]). In particular it is forcefully argued (by many experts) that the semantic status (truth value) of B given that both A and A → B are "true to some intermediary degree" depends also on the *intentional* relation between A and B, not only on their respective truth values.*

This problem, though avoided in some modal logics, is not unique to fuzzy logic. It is quite possible to create bogus implications which will, in spite of their abstract nature, still hold true when evaluated, also in a logic employing only crisp numbers. Saying that the moon being made of green cheese would cause pigs to fly, though clearly nonsense, would still evaluate as true – simply because the moon is not made of green cheese, and thus it is impossible to negate this statement. Lack of consideration of conditionality is, however, a problem encountered in most logics. It is, of course, a problem relevant to fuzzy logic. But it is not a problem of fuzzy logics or created by fuzzy logic.

The author was also informed that there are, in fact, non-functionally expressible theories of fuzzy sets, though they have not been practically applied up to now.

3.4 Higher Order Vagueness

Even if the truth values themselves are replaced by "fuzzy values" or something similar the problem does not disappear: – at some level (order) "improper precisions" must creep in for any formal fuzzy logic, – at every level (order) it remains unclear how we arrive at the corresponding "fuzzy truth value". (How should we distinguish between an artifact of the model and a "genuinely representing" property of truth values?)

This consideration might be judged to be related to Zadeh's type 2 fuzzy sets, recently presented and elaborated in *Type-2 Fuzzy Sets Made Simple* [18] by Jerry M. Mendel and Robert I. Bob John and *Type-2 Fuzzy Sets: Some Questions and Answers* [19] by Jerry M. Mendel. However, type 2 fuzzy sets deal with handling statistical uncertainty and varying degrees of applicability at the same time, while some people argue that the dimensionality fuzziness deals with has an infinite number of dimensions, and not just the one a fuzzy set deals with.

Dealing with higher order vagueness seems highly relevant to the "objective" of fuzzy set theory. However, the higher one goes with the vagueness considered, the more complex a model gets. And practically speaking, a certain degree of imprecision is generally accepted in order to keep the modeling simple, intuitive and graspable. Though type 2 fuzzy sets are superior in power, their complexity and their non-intuitive nature

has probably contributed to their having yet become an accepted standard.

Similarly, one could model a system with various levels of overlying uncertainty. However, in practice an engineer will at some point choose to "cut off" his measurements at a certain threshold at which viewing deeper into the problem handled is no longer relevant or necessary.

3.5 Different Truth Functions for Connectives

Where are the criteria that allow us to pick the right or best one? There seems to be a lack of arguments from "first principles".

It is, indeed, possible to compute the same connectives in various ways in fuzzy logic – a concept which in classical mathematics or probability theory would seem alien. Finding various vastly different interpretations for "+" or for the joint probability of two statements happening in unison would not be acceptable.

In fuzzy logic, however, it is, for example, possible to describe the "AND" connective with three different t-norms and t-conorms, the results of which are only rarely equal.

$$T_{min}(x, y) = min\{x, y\}$$
$$T_{\text{Ł}}(x, y) = max\{0, a + b - 1\}$$
$$T_{prod}(x, y) = x * y$$

In probability, the combined probability of two independent incidents can be denoted quite easily. The "AND" connective, for example, can be denoted with:

$$P(x\&y) = P(x) * P(y)$$

It is also quite possible to validate formulas such as this one through mathematical deduction, or, alternatively, through empirical validation.

However, once again, these concepts are intuitive in their respective areas. Fuzzy connectives are not. It is, thus, difficult to imagine them as anything but approximations that are to be picked depending on practical needs in the respective situations.

3.6 Worries about "(too) Many Logics"

Correct reasoning should – like rationality in general – point to just one overall logic of which other logics can be (modal etc.) extensions or "limit cases". However (modern) FL is about an ever increasing range of logics...

This question, in particular seems to accent a split between two vastly different approaches to fuzzy logic within the fuzzy community, a phenomenon further explored in the conclusion. While many people argue that fuzzy logic creates models and should thus be treated correspondingly, others are interested in the practical validity of data handled by a fuzzy system, as they do not regard "it is only a model" as a satisfying answer.

If one was to accept this though, the endless realm of models describing a specific situation would become something not specific to fuzzy logic – it is encountered in reality too, as one can quite easily see when looking at five vastly different maps of one and the same city. The question remaining here is whether in a given situation there is an ideal model of the situation.

This problem is generally perceived as a problem of reality mapping, not with the logic itself. Mapping of reality can be critical in any theory, as the complexity of reality is impossible to grasp with simple models.

Some even argue that this is not a problem, but rather a blessing, as the multitude of logics allows various applications to pick and choose, depending on the specific needs of the given situation.

3.7 Hedging via Disjunctions

[cited here from Roy Sorenson: entry for "Vagueness" in the Stanford Encyclopedia of Philosophy [32]:] "Critics of the many-valued approach complain that it botches phenomena such as hedging. If I regard you as a borderline case of 'tall man', I cannot sincerely assert that you are tall and I cannot sincerely assert that you are of average height. But I can assert the hedged claim 'You are tall or of average height'. The many-valued rule for disjunction is to assign the whole statement the truth-value of its highest disjunct. Normally, the added disjunct in a hedged claim is not more than the other disjuncts.

Thus it cannot increase the degree of truth. Disappointingly, the proponent of many-valued logic cannot trace the increase of assertibility to an increase [in] the degree of truth."

Sorensen [obviously] refers to disjunction as maximum. But can "disjunction for hedging" really be explained by, e.g., Lukasiewicz, "strong disjunction"? Why should any particular truth function for disjunction adequately represent hedging in natural languages? Granted that also disjunction of minimum is a "real disjunction", how many "real disjunctions" are there in natural language? How do we get to know them? Can FL provide guidance for answers?

Fuzzy hedging does indeed often lead to insufficiencies, particularly in the field of natural language processing, as it does not consider the intentional relation between terms, but only their mathematical relation.

For example, if a person is a borderline case between tall and average height, he might have the fuzzy value of 0.5 for both "tall" and "average height."

Intuitively, if I was to ask if the said person is "tall or of average height", the answer would have to be yes – if he is a borderline case. Surely covering both possibilities must fully include him.

However, as disjunctions are generally implemented with a maximum, in a fuzzy discourse the said person would only achieve a value of 0.5 with this computation, as the mathematical model is not aware of the fact that the said groups are adjacent to each other, and that the said person, with these truth values, must be right in the crack between these groups and thus should be included.

3.8 Sacrificed Principles of Classical Logics

(Most, if not all) fuzzy logics sacrifice principles of classical logics that seem intuitively "correct" even from (e.g.) a constructive or "relevance" point of view (e.g. the law of contradiction $\neg(A\&\neg A)$ and idempotence of conjunction $A \rightarrow A\&A$ etc.) How can such radical deviations from traditional "laws" be justified?

Dropping the law of the excluded middle, as is necessary in a fuzzy context, is something that seems relatively acceptable to many – the repeated confusion throughout history about the statement "either you're with us or you're against us" shows how human thinking and natural language do not generally deal with the absolutes on which the law of the excluded middle is based, and that thus its principle is not completely natural to human thinking in the first place.

Dropping the law of contradiction, however, seems more critical. The law of contradiction states that statements such as "a man cannot be tall and not tall at the same time" must be true – a statement, which unlike the law of the excluded middle, seems intuitive in natural language as well.

However, the only t-norm to conserve this rule is the aforementioned Łukasiewicz-norm, in which also a borderline case would evaluate to the desired result. Let us consider that a person is tall to the value of 0.5 – and thus is not tall to exactly the same degree. The results given by the three t-norms would be:

$$T_{min}(0.5, 0.5) = 0.5$$
$$T_{\mathrm{L}}(0.5, 0.5) = 0$$
$$T_{prod}(0.5, 0.5) = 0.25$$

However, the two t-norms "causing problems" here have proven themselves to be highly valuable in practical applications. "Excluding" them on these grounds does not seem plausible, in reality.

Arguments are also found noting that Boolean concepts, such as this one, are, in fact, not to be considered mandatory in all logical systems, as they do not necessarily hold in reality either. This statement, though natural to humans, does not hold in the context of quantum statements, for example.

3.9 Epistemic, Ontic or Pragmatic Character?

It is left unclear whether the "degree of truth" has an epistemic, an "ontic" or a "purely pragmatic" character; different interpretations (Giles [8] [9] / Ruspini [25] / Mundici [4] / Behounek [1]'s resource interpretation/voting semantics, etc.) seem to imply different answers. (See, e.g., Jeff Paris [2] [22] for problems with some of these interpretations). A theory of vagueness should include clear answers to such questions.

Most seem to credit fuzzy logic with having an ontic character that may be used pragmatically. The question generally seems to be seen as a question of semantics though, and thus not particularly interesting or relevant within the context of this survey.

3.10 Surface Phenomena

Fuzzy logic is only an ad-hoc model for some "surface phenomena" that may be useful for engineering purposes, but does not help us (a lot) in answering "deep questions" about correct reasoning, the metaphysical or ontological status of vague predicates, epistemic and – probably most important – prescriptive (deontic) aspects of logic in general.

It is true that modeling in fuzzy logic is generally based on surface phenomena. However, most seem to consider this a modeling issue, not an issue of the logical system used. Determining the metaphysical origins of knowledge is difficult in any circumstances. Any kind of reasoning is often going to be hard to analyze to its deepest level in practicality. Even successful doctors are often credited for making good decisions in a hypothetical and conjectural fashion, and not in a deductive manner.

3.11 Penumbral Connections

Many philosophers follow Kit Fine in asserting that "penumbral connections" should be modeled directly in any logic reasoning with vagueness. (E.g. if it is indeterminate whether X is blue or green, it is still definitely true that it is mono-colored etc.) Can FL compete with supervaluation in accommodating penumbral connections?

Kit Fine explains what he regards as a penumbral connection in detail in his 1975 paper '*Vagueness, truth and logic* [7]. Technicians argument that fuzzy logic should not compete here, since penumbral connections do not lie within its modeling range.

3.12 It is Only a Model

*FL often insists on a kind of application-oriented point of view. However, it is not enough to reply "it is only a model" to worries about a particular logic or semantic machinery. This would beg the question of whether the model is adequately representing how we should reason correctly in various situations. In general it is doubtful whether an "engineering approach" can help us to create a full-fledged theory of vagueness. Mathematical models can only be *a part of* or *a tool within* a theory of vagueness.*

Lots of answers agreed that fuzzy logic is, indeed, only a tool for modeling that comes from the field of engineering, but that there is nothing "only" about this. Some claim that fuzzy logic never pretended to offer a foundation for theoretical understanding of vagueness, while others claim that fuzzy logic is on its way towards eventual success at this task through a process of abstraction.

3.13 Relation to Natural Language

FL has an uneasy relation to natural language. On the one hand, it is often claimed that FL is "close to natural language discourse". On the other, it does not respect the fact that in natural language we do not use (concrete, linearly ordered) intermediate truth values and (different) truth functional connectives.

Fuzzy logic is a precise tool dealing with imprecise data, while natural language is purely imprecise – or is based on a model so complex that it is impossible to determine the relation from the possibly precise causes leading to human thought, and thus, to natural language (on a very basic level, the human brain does function digitally – a neuron either transmits a signal, or it does not). The author of this paper has previously been involved in papers on this topic [27] [28] [29].

There are, thus, limits to how well classical fuzzy logic can approximate natural language. It is theoretically possible that if, at some point in the future, the functioning of the human brain is understood, it will be possible to model natural language adequately if one was to handle a great number of stages of fuzziness. However, a model describing such a situation would become complex beyond comprehension – and, thus, unusable. Thus, the fact remains that while humans would refer to a person as "rather tall" or "really tall," fuzzy logic states that a person is tall to a degree of 0.7 and 0.9.

Many argue that natural language seems like a fairly abstract field in which to try to implement fuzzy logic. Though this might be true when considering applications today, it should be noted that the term fuzzy logic was actually not coined by Lotfi Zadeh, who spoke of fuzzy sets only initially, but by his Berkeley colleague George P. Lakoff, a professor of linguistics, in his 1972 paper *Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts* [14], in which he explored the possibilities of applying fuzzy set theory to natural language.

3.14 Operational Deficiency

*FL does not compare favorably with probability theory (PT) as a theory of (another type!) of uncertainty. Granted that FL is about degrees of truth as opposed to degrees of belief, one may be disappointed about the lack of convincing and robust models in FL as compared to PT. There is nothing like the paradigmatic application of PT to (e.g.) games of chance, where it is universally agreed that highly non-trivial, uniquely determined computations give you *demonstrably* and *empirically well-corroborated* (unique) values corresponding to rational expectation. Will there ever be similarly robust, non-trivial guidelines for complex information processing coming from FL?*

The author has yet to find a satisfactory answer to this question, and would appreciate opinions. Most seem to see it as a matter which only time will answer.

3.15 Record of Discourse

*Many theoreticians agree that paying attention to the specific *context* ("record of discourse") of an assertion (by competent speakers) is of utmost importance in understanding what's going on in a "(forced march) sorites" situation (and probably in all situations, where vagueness is involved). FL does not pay sufficient attention to this and therefore cannot compete with (in particular) contextualist theories of*

*vagueness (Shapiro [30], Graff [10] [11], Raffman [23] [24], etc.) with respect to questions about the best/correct way of *actual reasoning* in concrete dialogue scenarios (about sorites, etc.).*

This issue is based on Stuart Shapiro's 2006 book *Vagueness in Context* [30] [31], and lies quite outside of the interest range of the technical community addressed with these questions until now.

4 Conclusions

Interestingly, the main conclusion of the data collection up until now has not been of a technical, mathematical or philosophical nature, but of a sociological nature. There seems to be a fairly clean cut – even within scientific communities – regarding attitudes towards questions and contemplations of this nature. The author experienced this rift at the NAFIPS 2006 conference in Montreal, Canada, on a personal level, but was not aware of the magnitude of the rift between these two "schools of thought."

Within the "technical half" of the community [3] [15] [16], only a few of the issues addressed in this paper are relevant. The general maxim seems to be that fuzzy logic is a valid tool because it works – its practical successes invalidate conceptual and philosophical questions about it. If there were conceptual issues with fuzzy logic, it just would not work in practice. Trying to solve some of the points addressed here is not relevant for representatives of this community, as fuzzy logic is not supposed to deal with these issues and has never pretended to have solved them.

On the one hand, there are mathematicians and logicians [13] [33] [20] [21] who often seem to see contemplations of this sort as highly relevant and would themselves be interested in knowing some answers to questions asked – not surprising, as these questions stem from a philosophical context. Within this group, the practical successes of the fuzzy logic generally imply that fuzzy logic is an excellent abstraction of reality, but do not necessarily imply that it is a valid representation of the many layers of vagueness encountered in reality. Its successes do not imply that it is anything more than just a model or that it is a valid theory of vagueness.

For the author, who lacks a background in the fields of mathematics and philosophy, it is hard to see what there is "only" about a model, since from a technical point of view, very few methods applied and tools used are not models, abstractions and simplifications of reality, which is not difficult to model in its completeness, but impossible, as the Heisenberg uncertainty principle [12], among other theories, states.

Acknowledgement

The author would like to thank everyone who participated in this project in any way. At present, this includes Mirko Navara of the Czech Technical University, Carles Noguera and Lluís Godo Lacasa of the Spanish National Research Council, Robert Kosik and Reinhard Viertl of the Vienna University of Technology, Enric Trillas, Thomas Vetterlein and Christian Borgelt of the European Centre for Soft Computing, Siegfried Gottwald of the University of Leipzig, Jerry M. Mendel of the University of Southern California, Vesa A. Niskanen of the University of Helsinki, Hans-Jürgen Zimmermann of INFORM, Vilém Novák of the University of Ostrava, Klaus-Peter Adlassnig of the Medical University of Vienna as well as Petr Hajek, Petr Cintula and Libor Behounek of the Academy of Sciences of the Czech Republic.

Special thanks go to Christian G. Fermüller of the Vienna University of Technology for the initial penning of the questions dealt with in this chapter, and to Rudolf Seising of the Medical University of Vienna for extensive assistance with the "social networking" related to this project.

This work is partially supported by Grant # AB00158OFF from the Vienna Federal Data Center.

References

[1] L. Behounek: *An alternative justification of the axioms of fuzzy logics*, The Bulletin of Symbolic Logic, to appear.

[2] A. D. C. Bennett, J. B. Paris, A. Vencovská: *A new criteria for comparing fuzzy logics for uncertain reasoning*, Journal of Logic, Language, and Information, vol 9, pp 31–63, 2000.

[3] J. C. Bezdek: *Fuzzy Mathematics in Pattern Classification*, Ph. D. Thesis, Center for Apllied Mathematics, Cornell University, Ithaka, New York, 1973.

[4] R. Cignoli, I.M.L. D'Ottaviano, D. Mundici: *Algebraic Foundations of Many-valued Reasoning*, Trends in Logic, Volume 7. Kluwer Academic Publishers, Dordrecht, 1999.

[5] D. Edgington: *Vagueness by Degrees*, in *Vagueness: A Reader*, edited by R. Keefe and P. Smith, MIT Press, Cambridge, Massachusetts, pp. 294–316, 1997.

[6] C. Fermüller: *Revisiting Giles: Connecting Bets, Dialogue Games, and Fuzzy Logics*, in *Logic,*

Games and Philosophy: Foundational Perspectives, volume in the series *Logic, Epistemology and the Unity of Sciences*, Springer, to appear.

[7] K. Fine: *Vagueness, Truth and Logic*, Synthese 30, pp. 265–300, 1975.

[8] R. Giles: *A non-classical logic for physics*, Studia Logica 33, vol. 4, 399–417, 1974.

[9] R. Giles: *A non-classical logic for physics*, in *Selected Papers on Łukasiewicz Sentential Calculi*, edited by R. Wojcicki, G. Malinkowski, Polish Academy of Sciences, 13-51, 1997.

[10] D. Graff: *Shifting Sands: An Interest-Relative Theory of Vagueness*, Philosophical Topics 28, pp. 45-81, 2002.

[11] D. Graff: *Gap Principles, Penumbral Consequence, and Infinitely Higher-Order Vagueness*, in *Liars and Heaps: New Essays on the Semantics of Paradox*, edited by J. C. Beall, Oxford University Press, 2003.

[12] W. Heisenberg: *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Zeitschrift für Physik, 43, pp. 172–198, 1927.

[13] E. P. Klement, R. Mesiar, E. Pap: *Bausteine der Fuzzy Logic: T-Normen - Eigenschaften und Darstellungsstze*, in *Fuzzy Theorie und Statistik - Modelle und Anwendungen in der Diskussion, Computational Intelligence*, edited by R. Seising, Braunschweig, Wiesbaden, Vieweg, pp. 127–225, 1999.

[14] G. Lakoff: *edges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts*, Journal of Philosophical Logic, 2, pp. 458–508, 1973.

[15] E. H. Mamdani: *Twenty Years of Fuzzy Control: Experience Gained and Lessons Learnt*, IEEE Transactions on Fuzzy Systems, 1, pp 19–24, 1993.

[16] E. H. Mamdani: *Advances in the Linguistic Synthesis of Fuzzy Controllers*, International Journal of Man-Machine Studies, 8, pp 669–678, 1976.

[17] E. H. Mamdani, S. Assilian: *An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller*, International Journal of Man-Machine Studies, Vol. 7, Nr. 1, pp 1–13, 1975.

[18] J. M. Mendel, R. I. John: *Type-2 Fuzzy Sets Made Simple*, IEEE Transactions on Fuzzy Systems, vol. 10, no. 2, pp. 117–127, 2002.

[19] J.M. Mendel: *Type-2 Fuzzy Sets: Some Questions and Answers*, IEEE Neural Networks Society, pp. 10–13, August 2003.

[20] V. Novák: *Which Logic is the Real Fuzzy Logic?*, Fuzzy Sets and Systems, 157, pp. 635–641, 2006.

[21] V. Novák: *Fuzzy Sets as a Special Mathematical Model of Vagueness Phenomenon*, in *Computational Intelligence, Theory and Applications*, edited by B. Reusch, Springer, Heidelberg, pp. 683–690, 2006.

[22] J. B. Paris: *A semantics for fuzzy logic*, Soft Computing 1, 143–147, 1997.

[23] D. Raffman: *Vagueness and Context-Relativity*, Philosophical Studies 81, pp. 175–92, 1996.

[24] D. Raffman: *Is Perceptual Indiscriminability Nontransitive?*, Philosophical Topics 28, pp. 153-76, 2002

[25] E. H. Ruspini: *On the semantics of fuzzy logic*, Intl. J. Approx. Reasoning 5, 45–88, 1991.

[26] R. Seising: *The Fuzzification of Systems: The Genesis of Fuzzy Set Theory and its Initial Applications – Developments up to the 1970s*, Berlin, Springer, 2007.

[27] R. Seising, J. Bradley: *The Gap Between Scientific Theory and Application: Black and Zadeh: Vagueness and Fuzzy Sets*, NAFIPS 2006, Montreal, June 2006.

[28] R. Seising, J. Bradley: *From Vague or Loose Concepts to Hazy and Fuzzy Sets – Human Understanding Versus Exact Science*, KES 2006, Bournemouth, October 2006.

[29] R. Seising, J. Bradley: *Are Soft Computing and Its Applications in Technology and Medicine Human-Friendly?*, KES 2006, Bournemouth, October 2006.

[30] S. Shapiro: *Vagueness in Context*, Oxford University Press, USA, 2006.

[31] S. Shapiro: *Vagueness and Converstion*, in *Liars and Heaps: New Essays on the Semantics of Paradox*, edited by J. C. Beall, Oxford University Press. 2003.

[32] R. Sorensen: *Vagueness*, in *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/, 2006.

[33] E. Trillas: *Menger's Trace in Fuzzy Logic*, Segunda Época, Vol. 11, No. 27, pp. 89–96, 1996.

[34] L. A. Zadeh: *Fuzzy Sets*, Information and Control, vol. 8, pp. 338–353, 1965.

POSTMODERNISM AND CONTROL ENGINEERING

Valentina E. Balas
Aurel Vlaicu University of Arad, B-dul
Revolutiei Nr. 77
balas@inext.ro

Marius M. Balas
Aurel Vlaicu University of Arad, B-dul
Revolutiei Nr. 77
Marius_m_balas@yahoo.com

Abstract

This article intents to discuss the relation between modernism and postmodernism as a reaction to modernism, from the point of view of the science and technology. It makes a parallel between the intelligent control and the new trends in intellectual thoughts. It also addresses the relationship between postmodernism and the fuzzy set theory.

Keywords: Modernism, postmodernism, intelligent control, heuristics, soft computing, fuzzy logic.

1 Modernism and Postmodernism

Modernism is defined by some important features: *rationalism* (the belief in knowledge through reason), *empiricism* (the belief in knowledge through experience) and *materialism* (the belief in a purely physical universe). Postmodernism is a recent movement and a reaction to modernism.

The term Postmodernism was coined in the early 60s to describe the dissatisfaction with the modern architecture and became than a term for reaction to modernism in other fields as well [1]. Postmodern ideas in arts have influenced the philosophy and the analysis of culture and society. As engineers, we are interested to find out if a cultural movement, namely the postmodernism, is able to mark the scientific and the technological visions of our society, or at least, if similarities caused by the same social environment can be revealed in both fields. Since even the architecture, the domain that generated the term has an inevitable technological component, positive answers to the above mentioned quests are natural. The vice versa question is also challenging: are science and technology able to initiate and to determine global trends of the intellectual thinking?

Positive answers have been already given to these questions; our paper is meaning just to bring some personal arguments and opinions.

2 Modernism and Science

Science and society are intimately linked, although the science is often proclaiming its perfect objectivity. Scientific adventures, as the achievement of the atomic bomb, are proving that the society is able to take over the scientists. In the same time any society is depending of its scientific and technological platform. The modernism was initially supported by the mechanization of the industry, fired by the invention of the steam engines. The portrait of the modernist science has some corresponding features: a rigorous mathematical (numeric) support and formalism, a continuous search for abstracting and precision, etc. Further on, the electrical engineering was able not only to value another form of energy, but also to control the industrial processes. The same person, J.K. Maxwell built the theoretical pedestal of the electromagnetism, as well as a milestone for the automate control systems. Electricity made also possible the telecommunications. The electronic computing, analog as well as digital (based on the Boolean logic) emerged. A more than 2000 years developing and growing civilization, seeded by the ancient Greek philosophers, was by now ruling the Earth. The time to look at the stars and to put the foot on the Moon, and why not on other planets, has come.

The modernism best times may be considered as centered on the so called “la belle époque”, around 1900. The technological progress in all domains was finally bringing a significant improvement of the quality of life for everybody (well … almost). The young generation was educated with the help of the science-fiction authors J. Verne and H.G. Wells and the perspectives were bright. One of the last modernist cultural items that is fully containing the modernist science vision is considered to be the well-known G. Roddenberry’s TV series “*Star Trek – The Next Generation*.” Unfortunately the human mentality couldn’t match the human intelligence, and the euphoria of the new acquired technological breakdown generated two disastrous world wars and a subsequent long term cold war. These tragic events scattered away the general trust in science and technology, that could be

put in the position to invent and produce global destroying devices. The reactions against modernism began to structure themselves. Writers such as John Ralson Saul among others, have argued that postmodernism represents an accumulated disillusionment with the promises of the Enlightenment project and its progress of science, so central to modernist thinking.

2 Which is the Postmodernist Vision

Constantin Virgil Negoita and others consider Paris as the Postmodernism's birth place, "bursting full-blown" from the brains of Jean Baudrillard and Jean-Francois Lyotard.

Jean-Francois Lyotard understood modernity as a cultural condition characterized by constant change in the pursuit of progress, and postmodernity as the culmination of this process, where constant change has become a *status quo* and the notion of progress, obsolete. Following Ludwig Wittgenstein's critique of the possibility of absolute and total knowledge, Lyotard also further argued that the various "master-narratives" of progress, such as positivist science, Marxism and Structuralism, were defunct as methods of achieving progress. One of the most significant differences between modernism and postmodernism is its interest in universality or totality. While modernist artists aimed to capture universality or totality in some sense, postmodernists have rejected these ambitions as "metanarratives". "Simplifying to the extreme," says Lyotard, "I define *postmodern* as incredulity toward metanarratives" [1].

Postmodernism has features as the tolerance of ambiguity and disorder, stressing on skepticism and nihilism, the mixing of styles and manners, rejection of ultimate reality and absolute truth, lack of determinism and dogmatism. These features are not representing a simple fashion, they are not just "a rage against the machine", their origin is rather linked to the increasing complexity of our perception of the world.

The complexity of some problems created and leaven by the modernist era is now so high, that simple yes/no solutions are not any more possible; for instance the global heating can not be handled with the simple removal of its causes, we simply can not suddenly stop burning fuels. The more we know about a certain subject, a yes/no decision is harder to take about it; in a postmodern society, with a higher rate of educated people, even the public debates are more nuanced, the *pros* and *cons* list is longer.

As a typical postmodernist cultural item we can point again "*Star Trek*", but in its later versions "*The Third Generation*", etc. The popular and impressing futurist technology that was the asset of the first series is now often beginning to fail, of course in the worst possible moments. The members of the Enterprise crew, that were originally pictured as classical stone curved characters, are now beginning to manifest occasional psychic alienation symptoms.

We think that the postmodernist vision can be naturally associated with futurist Alvin Toffler's "*Third wave*" – the post-industrial society, that was characterized by demassification, diversity, knowledge-based production, and the acceleration of change. In 2007, we can say that Toffler's score is 3-1. From the four claimed items, he was mistaking (partially) only the demassification, the other are perfectly matching the actual postmodernist vision. The diversity is now an obvious attribute of the globalization, the knowledge-based society is the postmodernist developed version of the previous modernist information based society and the changes continued to accelerate.

Although useful distinctions can be drawn between the modernist and postmodernist eras, this does not erase the many continuities present between them. As noticed by A. Toffler, the three waves (pre-industrial, industrial and post-industrial) are coexisting. In a certain sense postmodernism is not as much a choice as a conviction.

3 Postmodernism and Science

Some scientific discoveries undermined the very essence of the modernist ideology: the rationalism and the materialism. We will name only two such scientific shocks: *the Relativity Theories* of A. Einstein and the *Big Bang Theory* on the beginning of the universe, of G. Lemaitre. The relativity put in cause the classical mechanics, one of the poles of modernism, a typical yes/not scientific discipline. A. Einstein itself failed to offer a deterministic explanation of the material world. On its side, the big bang theory shacked the idyllically image of the classical materialism: the matter was not created and will never disappear, the time has no beginning and will last forever and the space is endless.

As an anecdote, some (many) years ago, when we asked our Marxist philosophy professor, who was torturing the theory of the expansion of the universe (which was inducing the idea of a possible Creation of the Universe), what explanation can be however be given to the Hubble's law by the Marxism, he answered approximately that "we didn't find yet an acceptable explanation, but we are sure to find it sometimes, in the future." Of course, after a deeper analysis of the big bang's consequences, the materialism of the postmodernist era accepted the big bang idea, because this is not necessarily a proof of the existence of God, as some Marxists were fearing.

The XXth century quantum physics and astronomy showed tensioned evolutions, where thesis and antithesis were constantly emerging. This is also true for anthropology, medicine and biology. All these facts build the belief that truth is more relative than the Enlightenment thinkers had believed [2].

3 Postmodernism and Control Engineering

"Whereas modern science had previously dealt with matter and energy, postmodern science focuses on form and pattern" [2]. This is leading us towards a new fundamental vision of the Universe, as a triad *matter – energy – information*, where the information has a leading role. This vision is much older, even the first words of the Gospel of St. John can be interpreted in this sense, but only Claude Shannon offered a scientific model of the Information.

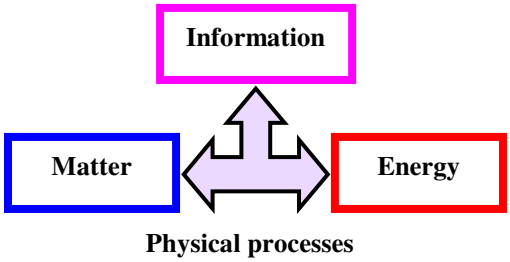


Figure 1. The *matter – energy – information* triad

We think that is not a coincidence that Postmodernism is contemporary to Electronics, the first technology that allow us to control as well energy and information. In industrial processes information is acting by means of the *intelligent control*. The intelligent control at its turn is powered by the Artificial Intelligence AI. Although the modernist shaped minds are objecting the approach, the only notable advances in AI are linked to a typical postmodernist concept, the *Soft Computing*, that is clustering fuzzy logic, neural networks, genetic algorithms and evolutionary computing [3], [4], etc.

The electrical engineering disciplines: electronics, computers, automate control, etc. can illustrate some effects with subversive influence on the modernist rationalistic scientific common sense, we will name only three:

- the Chua's circuit, essentially with only two capacities and one inductivity, is able to generate a chaotic dynamic;

- the precision of the electronic amplifiers is not depending at all of the components, excepting the feedback network; the precision is given essentially by the feedback reaction; a quite similar effect is characterizing the close loop control systems: the precision of the steady regimes is not depending of the components' precision, except the transducer; the simple presence of an integrative device in any point of the loop is eliminating the steady errors.

- the switching controllers' effect: a switching system can be potentially destabilized by an appropriate choice of the switching signal, even if the switching is between a number of Hurwitz-stable close loop systems. This phenomenon can produce catastrophes. For instance, in rare and unpredictable situations, the perturbations produced when switching between automate pilot and manual pilot may cause fatal airplane crashes. Reliable reports on such accidents are not easy to find, but it is unanimously accepted that the on-line switching of two different controllers may produce uncontrollable transient regimes and instability. This effect can not be explained by the conventional system theory in terms of frequency analysis, because its basic tool, the transfer function, is defined for null initial conditions, while the real applications has usually non-null initial conditions. Studying the systems by considering null initial conditions is simplifying a lot the manipulation of the equations, is revealing the specific behavior of the systems and it helps the comparisons between systems and the construction of the general theory of systems. But on the other hand, this quest for generalization can produce unexpected failures in specific conditions.

This is perhaps the most illustrative of the previous examples. The operational calculus (Laplace) is offering a comprehensive image for all the linear systems. The linear system theory established the conventional linear PID control and it can be easily associated to the modernist vision (universality, coherence, etc.) For nonlinear systems on the other hand, frequency analysis has few chances to produce satisfactory results, considering the huge diversification of the problems and the lack of a unified theory. In the case of barely controllable systems (highly nonlinear, time varying, etc.) the only unified approach is, for the time being, the heuristic one. Despite their inherent lack of rigor, the heuristic solutions can be always applied and may be very flexible and comprehensive. The theoretical tool that can help us in this matter is the phase trajectory, a time analysis [8], [9]. This kind of analyze is not able to reveal too much information on the internal structure of the systems, but is very helpful in applications, supporting the heuristic control decisions. The heuristic approach is totally opposite to the Cartesian rigorous modernist vision, but it is relevant for postmodernism. Although it can not be mathematically proved, the heuristics are bringing extremely positive results in most of the applications.

The classical mathematical approach: *hypothesis→ conclusion→demonstration* is now beginning to be replaced by a less elegant but more pragmatic methodology: *hypothesis→conclusion→computer simulation*. Instead of solving the differential equations, one let the computers to integrate them, numerically. As a result of this, most of the industrial products, including the 15 million items Airbus 380, are, in our postmodernist days, designed with the help of dedicated software, that are embedding general and specific knowledge of the domain, often acquired by simulations and represented linguistically by expert systems. The lack of a rigorous theory is compensated by serious experimental tests for validation.

4 Postmodernism and Fuzzy Logic

The Postmodern truth is fragmented, subjective and stemmed from approximate reasoning [2]. Epistemologically, this nuanciation of truth is unanimously associated with Lotfi A. Zadeh's fuzzy logic [2], [5], [6], [7], etc. The Aristotelian two valued logic, *true* and *false*, was dominating more than two millenniums the philosophy. After George Boole described it mathematically it get involved into technology, most of all, into control engineering, by the sequential control (relays, electronic digital circuits, PLCs, etc.). The climax of the Boolean logic was the conceiving and the development of the *digital computers*, the particular item that changed the world more that the landing on the Moon. The achievements of the digital technology are obvious and undeniable; however in certain situations it showed some limitations. These situations are generically characterized by the presence of different types and levels of *uncertainty*. If we are not able to classify a concept as true or false, the Boolean logic simply collapses. The uncertainty is anyway a constant of the human reasoning, which operates in a symbolic and qualitative manner. That is why before fuzzy logic, AI encountered enormous difficulties at the computer implementing stage.

Fuzzy logic is able to cope with uncertainty because it accepts not only two values 0 (false) and 1 (truth) for the membership functions, but all the interval bounded by 0 and 1. If the membership function of an element to a certain concept is 0.5, it means that we are not at all sure if the element is belonging to the concept or not, and the fuzziness is maximum. Using fuzzy sets we can represent world knowledge affected by uncertainty in digital computers, as fuzzy linguistic variables, perfectly compatible to human reasoning. Further on, fuzzy logic is able to produce inferences using fuzzy variables and specific yet very simple operations: *min-max*, *prod-sum*, etc. The specific software items that are producing logic inferences by control rules, based on previous human expert knowledge, are *the expert systems*. The postmodernist version of the expert systems are the *fuzzy expert systems*.

In science and technology uncertainty may be caused by our poor knowledge or incorrect information on the system we are dealing with. This is happening when we are not disposing of an appropriate mathematical model of the system, by different reasons: too much complexity, inappropriate sensors, insufficient experimental data, etc. In these circumstances fuzzy logic is producing feasible solutions. Besides the uncertainty caused by our qualitative reasoning and our lack of knowledge, the result of our senses - our perceptions - are uncertain too [6]. Generally speaking, uncertainty is an fundamental attribute of life. That is why fuzzy logic may be successfully applied whenever applications address human beings, or any other biological system. This is the case of air conditioning systems, greenhouses and other related applications. For instance the flexibility offered by the very nature of the fuzzy expert systems and the vague perception of the "comfortable temperature" concept can be converted into energy savings, by means of few specific very simple control rules. Here is an example of such a rule:

IF **temperature** is ***moderate*** **low** AND **change of temperature** is **positive** THAN **save energy**

As a conclusion, we think that what Constantin Virgil Negoita was writing about the echoes of fuzzy concept in Eastern Europe, was crisply true [7]:

"In Eastern Europe, everybody liked the idea of a fuzzy set. Probably because it was coming from California, promising liberties."

References

[1] *** Wikipedia The Free Encyclopedia, http://en.wikipedia.org/wiki/Postmodernism.

[2] C. V. Negoita, Postmodernism, Cybernetics and Fuzzy Set Theory, *Proc. of IPMU'02*, July 1-5, Annecy, France, pp. 681-686.

[3] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning I, *Information Sciences* no. 8 (1975), pp. 199-257.

[4] T. Samad, Postmodernism and Intelligent Control, Proc. of the 12th IEEE International *Symposium on Intelligent Control*, 16-18 July, 1997, Istambul, Turkey, pp. 37-42.

[5] D. Simon, Truth, American Culture and Fuzzy Logic, http://academic.csuohio.edu/simond/pubs/317101.pdf

[6] L. A. Zadeh, Toward a Logic of Perceptions Based on Fuzzy Logic, in: V. Novak, I. Perfilieva (Eds.), *Discovering the World With Fuzzy Logic*. Springer-Verlag, Heidelberg (Studies in Fuzziness and Soft Computing, Vol. 57), 2000, pp. 4-28 .

[7] C. V. Negoita, Fuzzy Sets, *New Falcon Publications*, Tempe, Arizona, 2000.

[8] M. M. Balas, V. E. Balas, T. L. Dragomir, On the Switching Controllers' Issue and Smooth Controller Switching by Phase Trajectory Qualitative Analysis, *The 6th International Conf. on Recent Advances in Soft Computing (Abstracts + CD),* Canterbury, 10-12 July, 2006, pp. 63-68.

[9] M. Balas, J. Duplaix, Switching Control by Phase Trajectory. *International Symposium Aurel Vlaicu Univ*., Arad, Nov., 2006, pp. 518-525.

[10] A. Toffler, The Third Wave, *Bantam Books*, 1980.

Power Sets and Implication Operators Revisited: A Retrospective Look at the Foundational and Conceptual Issues in Bandler and Kohout's Paper After 29 Years

Ladislav J. Kohout
Dept of Computer Science, Florida State University,
Tallahassee, Florida 32306-4530, U.S.A.
kohout@cs.fsu.edu

Abstract

In our case study we look at one of the early papers that interrelates the concept of fuzzy set inclusion, power set and many-valued implication operators, namely the paper of Bandler and Kohout [3]. This is followed by discussion of the subsequent related work by the fuzzy community.

Keywords: Fuzzy set inclusion, power set, many-valued implication operators. Fuzziness, crispness, fuzzy disjointness, fuzzy equality.

1 Introduction

It is an interesting story to look at the development of new concepts in fuzzy set theory. One looks at the motivation, the first formulation and the subsequent development.

We shall take as our case study the one of the early papers that look at the relationship of the concept of fuzzy set inclusion, power set and many-valued implication operators. It is the paper of Bandler and Kohout [3] that was presented at an international workshop organised by Professor Mamdani in 1978. The participants of the workshop were people from different disciplines, pure mathematicians, scientists from various fields including brain modelling, and understandably, with strong representation of people from the fuzzy control community. The fuzzy control community had that time strong interest in investigating different types of implications. That established Bandler and Kohout's paper as a repository of useful information about various implication operators and about the bootstrap of their properties into fuzzy sets. The extended version was submitted to Fuzzy Sets and Systems, but the manuscript was considered to be too large, so the Editor-in-Chief Professor Zimmerman recommended to be limited to discussion of different kinds of fuzzy set inclusions and of their link to many-valued logic operators. This paper [5] has become well known in the fuzzy community and was reprinted in a collection edited by Dubois, Prade and Yager. Other parts of the 1978 paper were extended and published as separate papers [7],[6].

We shall examine the historical trace of the development of concept of fuzzy inclusion in the following ways:

1. What part of the original approach has been retained;
2. how was it used in the further development of the concepts involved;
3. what aspects of the paper were understood well, which have been neglected, and what was misunderstood or misinterpreted.
4. We shall also look at the links to other concepts that were only in the original [3] but unfortunately did not appear in the reduced version [5] published in Fuzzy Sets and Systems.

2 Paper of 1978: 'Fuzzy Relational Products and Fuzzy Implication Operators"

The first paper was entitled *Fuzzy Relational Products and Fuzzy Implication Operators*. The list of contents of this paper listed the following topics:

1. Various Products of Crisp Relations.
2. Towards a Theory of Fuzzy Power-Sets.
3. Possibilistic Notation.
4. Comparative Semantics of Fuzzy Implication Operators.
5. Height, Plinth and Crispness of Fuzzy Sets.

6. Fuzzy Set-Inclusions arid Equalities.
7. Disjointness of Fuzzy Sets.
8. A Fuzzy Set and its Complement.
9. Choice of System and Further Aspects.

The main motivation for fuzzification of Zadeh's set inclusion predicate the membership function of which is given by the formula

$$\mu_{A\subseteq B}(x) = \mu_A(x) \leq \mu_B(x)$$

came from the need to fuzzify the crisp BK-products of relations [2]. That is clearly stated in the abstract of the 1978 paper [3]:

> *Besides the usual circlet product of crisp relations, there are three others which are natural and of interest and of use. Their fuzzification requires the choice of a fuzzy implication operator, and will vary with the choice made (Section 1). The reason why this is so leads the problem back to a fundamental and hitherto neglected aspect of fuzzy set theory: the appropriate definition of a fuzzy power-set; thus the motivation for choosing a suitable* **internal** *implication operator is much deepened, and by the use of a possibilistic notation is also somewhat broadened (Sections 2 and 3).* (From the abstract of [3]).

Because the paper links various concepts, it has became the seminal ground for other work of Bandler and Kohout. In particular, the paper [3] introduced the fuzzy non-associative products $(\triangleleft, \triangleright, \square)$ also called BK-products in the literature. That was a successful fuzzification of crisp BK-products introduced by Bandler and Kohout earlier [2]. That paved way to development of Enriched theory of fuzzy relations (ETFR) which successfully extended the crisp enriched theory of relations of [2] into the fuzzy realm.

2.1 Power Sets, Inclusion Predicate and Implication Operators

We shall briefly survey the key concepts of [3] section by section first, and then look at subsequent developments.

2.1.1 Section 2: Towards a Theory of Fuzzy Power-Sets

The situation where sets B and A are both crisp subsets of some universe U is considered first. The standard definition of the subset relation between them is

$$A \subseteq B \quad \text{means } (b \in A \to b \in B).$$

This is the connection between $\subseteq$ and the implication operator $\to$. Now, the subset relation itself is expressible in terms of the *belonging relation* and the *power-set* $\mathcal{P}(B)$ of B:

$$A \subseteq B \quad \text{means} \quad A \in \mathcal{P}(B).$$

$$\text{Thus} \qquad A \in \mathcal{P}(B) \quad \text{means } (b \in A \to b \in B).$$

This formulation is subject to immediate fuzzification as follows:

Definition 1 ([3], Def. 2.1) *Given a fuzzy implication operator $\to$, and a fuzzy subset B of a crisp universe U, the* ***fuzzy power-set*** *$\mathcal{P}(B)$ of B is given by the membership function with*

$$\mu_{\mathcal{P}B}A = \bigwedge_{x\in U} (\mu_A x \to \mu_B x).$$

This is well defined in terms of each suitable $\to$ operator, for every argument $A \in (F)U$.

Hence, the degree to which A is a subset of B is

$$\pi(A \subseteq B) = (\mu_A x \to \mu_B x).$$

The symbol π indicates that, in fact, that the degree assigned to the statement $(A \subseteq B)$ is *degree of possibility.*

Note that, where I is the unit real closed interval, the fuzzy set B is an element of I^U while its power-set B is an element of I^{U^U} (Otherwise put, $B \in \mathcal{F}(U)$, while $\mathcal{P}(B) \in \mathcal{F}(\mathcal{F}(U)$.)

The Mean Inclusion Bandler and Kohout also introduced the mean inclusion in [3] (Prop. 3.2) replacing *inf* by the *mean* value:

$$\pi_m(A \subseteq B) = \frac{\sum_{x\in U}(\mu_A \to \mu_B)}{card(supp\ A \cup supp\ B)}$$

Properties of these have been investigated by Willmott later.

2.1.2 Section 3: Possibilistic Notation

Following Zadeh (1971) in using π for "possibility" in comparison to p for "probability", Bandler and Kohout extend the analogy-or-contrast by enclosing statements in brackets after π to indicate their degree of possibility. On the interpretation of π they say the following:

> *One (but not the only) interpretation of this is, "the degree to which the bracketed statement is true." In particular, the previous section shows that we will wish to have $\pi(A \subseteq B) = (\mu_A x \to \mu_B x)$. "the degree to which A is a subset of B."*

2.1.3 Section 4: Comparative Semantics of Fuzzy Implication Operators

Implication operators play crucial role in linking sets with their power sets as well as with the inclusion predicate. In order to investigate the properties of fuzzy set operations we need to start with examining the properties of logic formulas on which specific set theories are based. The properties of logic connectives are then reflected in the properties of fuzzy sets and set operations as shown in sections 6 – 9 of [3]. The criteria for evaluation outlined in [3]are as follows:

Does an implication operator used in a formula yield

1. a strong, or moderate tautology for $a \to a$?
2. the flat contradiction, or a moderate contradiction for $a \to \neg a$?
3. is the implication operator contrapositive?
4. is the implication operator continuous?

Such questions are, however, meaningful and unambiguous if and only if they are asked in an appropriate context. Bandler and Kohout point out that two entirely different contexts are often not sufficiently distinguished [3],[5].

> *Logic has long been beset with the often-muddied distinction between*
>
> 1. *inferences made in a meta-language from statements in an object-language, on the one hand,*
> 2. *and on the other, the formation in the object-language itself of an implicative combination of its c statements.*
>
> *Both the need for this distinction and the difficulty of keeping to it become more acute in the fuzzy environment.*

They continue [3],[5]

> *Our present need is for a "suitable" generalisation of the second of the distinguenda, the internal implication operator in the object language*[1].

In order to detach this notion from the first one, that of (meta-)reasoning with fuzzy premises, they use the unemotive term favoured by Curry: *PLY operator*; the arrow itself is then read a "ply".

The problem is posed very explicitely [3],[5]:

1. We are working in a Multi-Valued System V, which for present purposes is all or some of the real interval I = [0, 1]. The rationals there are more than ample for their purposes (so: cardinality at most $\aleph_0$).
2. Whatever V is, it is furnished with the uncontroversial operators $\wedge$ and $\vee$, and with the accepted negation $\neg$, with $\neg a = 1 - a$.
3. One seeks within this system a ply operator $\to$, that is, a mapping from $V \times V$ to V, suitable for the concepts of the previous sections, which is to say chiefly for defining the degree to which one fuzzy set is to be said to be a subset of another.
4. The fuzzy sets themselves are mappings from some crisp universe U into our V, that is, the membership degrees of elements are numbers in V.
5. The ply operator will take two such degrees and make another out of them. The natural anticipation, is that the fuzziness will not thereby be diminished in this process.

For further investigations of specific power set theories and inclusion predicates, Bandler and Kohout [3],[5] chose five representative ply operators 1–6.

1. **S**$^{\#}$ Rescher [19] (p. 344).
   $a \to_1 b = 1$ if $a \neq 1$ or $b = 1$, 0 otherwise.
2. **S** The "standard sequence" of Rescher [19] (pp. 46–52, 343–344), Gaines [10] formula (40).
   $a \to_b = 1$ if $a \leq b, 0$ otherwise.
3. **S*** Gödel. $a \to_3 b = 1$ if $a \leq b, b$ otherwise.
4. **G43** Goguen–Gaines. Recommended by Gaines, formula (43), for further investigation.
   $a \to_4 b = min(1, b/a)$.
5. **Ł** Łukasiewicz. $a \to_5 b = min(1, 1 - a + b)$
6. **KD** Kleene–Dienes. $a \to_6 b = max(1 - a, b)$
7. **EZ** Zadeh [27]. $a \to_7 b = max(min(a, b), (1 - a))$
8. **Wm** Willmott [24], [25]
   min(max(1-a, b), min(max(a, 1-b), max(b, 1-a))).

[1] When [3] was written, most of the considerations of fuzzy ply operators in the literature had been from quite a different point of view: an operator suitable for the first of the distinguenda had been sought, a means of meta reasoning from fuzzy data.

2.1.4 Section 5: Interrelating Height, Plinth, Crispness and Fuzziness of Fuzzy Sets

Connected with semantics of PLY is the natural notion of natural crispness and fuzziness of an MVL proposition and of a fuzzy set. This was utilised for assessing PLY by their degrees of crispness and investigating how this bootstraps onto the various constructs made of fuzzy sets by set operations.

In [3] Bandler and Kohout introduced the *crispness* of a proposition $a \in V$ as $\kappa a = a \vee (1-a)$. The *fuzziness* $\phi a = 1 - \kappa a$ as its *dual*. The crispness of a fuzzy set is then

$$\kappa B = \bigwedge_U \kappa(\mu_B x).$$

The *mean* crispness of a fuzzy set

$$\kappa_m B = \frac{\sum_U \kappa(\mu_B x)}{card\ suppB}.$$

Later these concepts played a role in assessing the width of intervals in interval fuzzy logic, thus evaluating quality of inference as a function of data on which inference is performed.

2.2 Diversification As a Result Of Fuzzification: Split of a crisp concept into several fuzzy concepts

The paper [3] clearly demonstrates that from a mathematical viewpoint the important feature of fuzzy set theory is the replacement of the two valued logic by a multiple-valued logic. Since every mathematical notion can be written as a formula in a formal language, we have only to internalise, i.e. to interpret these expressions by the given multiple-valued logic. For that reason, it was important to "internalise", i.e. to form in the object language itself an implicative combination of its statements as pointed out above in section 2.1.3.

One important aspect of fuzzification that 1978 paper and 1980 paper demonstrated was the fact that two or more equivalent crisp definitions are not any more equivalent for their fuzzy counterparts. For example the Definition 5.1 [5] provides 2 formulas for disjointness of two sets that are different in the fuzzy case, despite the fact that they are equivalent in the crisp case.

2.2.1 Section 6. Fuzzy Set-Inclusions and Equalities

In addition to fuzzy set inclusions and equalities, this section also looks at the degree to which a fuzzy set is empty.

2.2.2 Section 7. Disjointness of Fuzzy Sets

In ordinary set theory

$$A \cap B = \emptyset \text{ iff } A \subseteq A^c \text{ iff } A \cap B^c = A.$$

The first two characterisations were investigated by Bandler and Kohout [3][5] while the last characterisation leads to a third possible definition of the degree of disjointness between sets which gives investigated by Willmott.

2.2.3 Section 8. A Fuzzy Set and its Complement

The three distinct concepts of disjointness are also reflected in the issue to what extend a set is disjoint from its complement.

2.2.4 Section 9. Choice of System and Further Aspects

The fuzzier implication operators exhibit a certain property of invariance that has been called by Bandler and Kohout "the conservation of crispness". This is useful in deciding which system to adopt for a particular purpose.

3 Response to the Paper Within the Fuzzy Community

3.1 Connectives

The paper deals with a family of implicational fragments of logics, the properties of which are bootstrapped into the properties of sets. While Willmott extends this by two more PLY operators, Weber looks at link of implications to other connectives.

The six operators of Bandler and Kohout are ordered by them 'in decreasing order of rigidity' or in increasing order of fuzziness, i.e. the later ones give decreasingly many crisp, or increasingly many fuzzy results in the case of non-crisp or fuzzy antecedents.

3.1.1 Willmott [24]

The investigation of Bandler and Kohout in [3] is repeated by Willmott [24] for two more implication operators (EZ, Wm) which follow the above six in this ordering (see Sec. 2.1.3 above, $\leftarrow_7, \leftarrow_8$.). Both EZ and Wm are fuzzier than any considered in [3]. The first was suggested by Zadeh [27] previously. The second is new and probably represents the extreme in fuzziness for a usable operator of this kind, according to Willmott "realising natural anticipation that the fuzziness (the value of an implication compared to that of its components) will not be diminished". In

terms of fuzzy sets, while using this operator, the degree of possibility of any relation between two fuzzy sets cannot be larger than the crispness of the less crisp of the two. The operator retains virtually all of the favourable features of the sixth operator (i. e. Kleene-Dienes) investigated by Bandler and Kohout in Sec. 4 of their paper.

Willmott states [24],[25] *This note should be considered as an addendum to the paper by Bandler and Kohout. It assumes all their notation, definitions and results and uses the same section and item numbering, so that some section numbers are here missing.*

3.1.2 Weber [23]

In function of connectives and negation three types of fuzzy implication operators are introduced, which include almost all known implications, and that of type I using AND only; of type II using OR/NON; of type lII using AND/NON. In the non-strict Archimedean cases the formulas become particularly lucid (Section 5). Finally the different types are compared with respect to some logical properties (Section 6).

The rest of Weber's paper was motivated by Bandler and Kohout [5]. In Section 5 Weber presents the construction of three types of implication operators, which contain the known ones.

Section 6.1 compares the three types of implication operators concerning contrapositive symmetry and contradiction. Remarks on natural crispness and fuzziness that was introduced by Bandler and Kohout in [3],[5] conclude the paper of Weber.

3.2 Power Set

Although [3],[5] introduce power sets, the importance of it has been overlooked by most of the papers that quote Bandler and Kohout. The reason for this is clearly indicated in comments of Höhle and Stout [14]:

> *For fuzzy mathematics we would like to have a foundation for higher order structures as well as for the propositional logic of fuzzy sets. To develop such a foundation we need to ask to what extent it makes sense to talk about a fuzzy power object.*
>
> *This can be internal (in which case an individual could have fuzzy membership in such a power set) or external as a construction in classical mathematics (the usual practice in current fuzzy topology). Indeed we claim that the first fifteen years of fuzzy set theory was dominated by the fuzzy power-set problem.*
>
> *In L.A. Zadeh's pioneering paper of 1965 it is obvious that he defines intersection, union, and complement of fuzzy subsets, but he hesitates to specify the fuzzy power set of a given fuzzy set.*

Indeed, as we remarked above, Zadeh's theory is a theory of fuzzy subsets of a crisp set, not a theory of fuzzy sets.

Bandler and Kohout clearly state that they are *looking for an internal implication operator within an object language."* This yields an individual that has fuzzy membership in a power set.

Stout [21] says about attempts to handling the power set problem:

> *In fuzzy set theory the approach has been more external, at least for second order theories, in part because there is no single fully satisfactory fuzzy power set operator. For example, in fuzzy topology one approach which has been developed at length (Wong uses crisp sets together with a topology which is a crisp set of fuzzy subsets). This uses only the properties of the sub-object lattice, an external propositional level approach. Several attempts have been made by Pultr [18], Bandler and Kohout [5], Gottwald [12],[13] to provide a suitable theory of* **fuzzy power sets.**

Bandler and Kohout [3] approach the problem of power set "top down", and algebraically. Note that in [3],[5], where I is the unit real closed interval, the fuzzy set B is an element of I^U while its power-set B is an element of I^{U^U} (Otherwise put, $B \in \mathcal{F}(U)$, while $\mathcal{P}(B) \in \mathcal{F}(\mathcal{F}(U)$.)

The axiomatic approach within a logic on the other hand was provided by other authors. There have been two set-theoretic approaches to the foundations of fuzzy sets with a power set: one by Gottwald [13], and Klaua [15],[16], based on Lukasiewicz connectives. For summaries of this work see also [14], [11]. The work of Takeuti and Titani [22] is based on intuitionistic connectives. They abandon the Lukasiewicz connectives because of problems with extensionality resulting from the fact that $(p * (p \leftarrow q)) \leftarrow q)$ need not be valid.

Gottwald parallels the construction of Boolean-valued models to get a hierarchical system of fuzzy sets with membership values in a ruler by giving an inductive definition. There is a sense in which each fuzzy set x in his hierarchy has a natural ordinal rank given by the smallest a such that $x \in R(\alpha)$). Gottwald calls the elements with rank > 0 fuzzy sets to distinguish them from the ur-elements with rank 0. The empty

fuzzy set has rank 1, as do other fuzzy subsets in the sense of Zadeh of the set of ur-elements.

The question of ur-elements needs to be revisited, as these may be important for applications. Unfortunately, almost no attention has been given to this important aspect of Gottwald's paper.

The power set problem has been resolved only comparatively recently within the setting of monoidal and other kinds of categories (cf. work of Höhle, Stout, Rodabaugh and others). Approach using the apparatus of (mathematical) categories is useful from the foundational point of view.

The algebraic approach of Bandler and Kohout that uses the many-valued logic connectives directly, is more suitable for development of calculi of fuzzy relations, interval fuzzy logics, and knowledge elicitation. From categorial point of view it is related to *esomathematical* use of category theory pioneered by Bandler [1].

3.3 Fuzzy Set Inclusion

There is rather large literature on this aspect. This unfortunately most papers in this category develop idea in isolation from the power set concept. The papers divide into two groups:

1. papers taking the inclusion predicate just as an index of subsetness, and
2. papers that provide axioms for various "desirable" properties of inclusion predicate.

In the first group are papers by Young [26], Kosko, Bustince [9], Bodenhofer, [8] and others. In the second group (axioms of desirable properties) are papers by Singha and Dougherty [20], Pappis et al, important paper by Kitainik, etc. The extended version of this paper will provide a detailed survey with bibliography which for lack of space cannot be provided here. There are several hundred of quotations of [5], mostly related to the viewing set inclusion as a measure, a subsetness indicator; or quoting Bandler and Kohout's work as a useful repertory of properties of implication operators[2]. Only scant attention is paid to other, equally important aspects of the papers that have been surveyed in greater detail in the previous sections.

[2]None of the authors quoted in this section seem to realize that we have also provided the definition of the mean subsethood [3] and used it extensively in applications since 1979 [4],[7]. Even for crisp sets the mean subsetness yields fuzzy values [7]. Willmott's interest in mean inclusion was triggered by [3],[4] while visited us at Essex. His visit was supported by an SERC grant that was obtained by Bandler for this visit.

4 Summary of Responses

The response in the literature and the influence of the paper on the subsequent work can be summarised as follows.

The paper of Bandler and Kohout [3] presented new concepts and stated also their mutual relationships. The results of investigation of properties of various implication operators $\rightarrow$, of various inclusion predicates $\subseteq$, and of various constructs made from fuzzy sets by fuzzy operators have been recognised and quoted. On the other hand, the important links between these have scarcely been noticed, and the relation of these to the concept of the power set have been almost completely overlooked.

Bandler and Kohout's ideas that were first outlined in [3] further branched into fuzzy relational calculi exploring BK-products of relations – here the first bridging papers are [7],[6]. The paper [7] also contained the Checklist Paradigm that has provided the semantics and the tools for interval fuzzy logics [17].

Unfortunately, because in the truncated version of [3] published in Fuzzy Sets and Systems did not contain the section 1 of [3], the fuzzy community views these three branches that stem from unified foundational study presented in [3] as completely unrelated.

The full report on which this brief EUSFLAT 07 paper is based, lists and more fully describes these developments that are related despite of the received view in the fuzzy community that that represent totally different branches of fuzzy set theory. Because the fuzzy field is rapidly dividing into many specialities and fragmenting view examining these old links may help to make new connection that counteract this undesirable fragmentation. It is clear that we need more and deeper foundational studies.

5 The Need for Foundational Studies

Contemporary mathematical logic is conveniently classified into the parts listed below, which extend into the many-valued domain of fuzzy structures by means of judicious fuzzification. It can be seen that Zadeh and his disciples attempted to fuzzify with success some of these, now classical parts of mathematical logic. Some selected references of such attempts are listed together with the classification below.

- Recursive functions.
- Set theory.
- Arithmetics.

- Quantification & identity.
- Propositional logic.

Although the above hierarchy covers what is known as mathematical logic – the logic intimately linked with the foundations of mathematics and computation, other approaches to logic stem from the linguistic philosophy and the linguistic proper. So, in any foundational studies, attention has to be paid also to these.

Höhle and Stout ask a pertinent question in the context of foundational studies, and offer an answer [14]

> What should the study of foundations of fuzzy sets offer? Certainly it should place fuzzy sets in a longer and broader tradition of many- valued mathematics ... but it must also speak to the needs of the practitioner of applied fuzzy set theory. A foundation for fuzzy set theory should provide a rigorous base for the actual practice of those applying the theory. ... people working with fuzzy sets want to use them for practical purposes ... These practitioners need a fuzzy set theory which is robust ... not particularly sensitive to the details of the model and connectives used but flexible enough so that the model can be tuned to provide high levels of performance. Thus a foundation for fuzzy sets needs to provide for a variety of connectives while clarifying the bounds on choices available.
>
> The second property that foundation should have is elegance. ... We can also ask if a foundation can take into account the 'linguistic variables' and experimental, computational approach

The suggestions are a good start, but in my opinion, one has to go even further. One has to build on algebraic strength of many-valued logic also learning from its failure to tap the conceptual and formal resources of contemporary philosophical logic.

The foundations of fuzzy sets, logics and systems contain some general systemic concepts that run across the boundary between theory and methodology. Although the initial motivation came from Systems Science through the important work of Zadeh that predated his first paper on fuzzy sets in 1965, the field has become rather fragmented in the last decade, losing to a great extent its initial cross disciplinary character. There is also a wide gap between mathematical and philosophical formal logic. Mathematical theory of General Systems has some features that may help to bridge this gap by mediating communication between the two disparate logic disciplines. Also the notions of *dynamics, stability, approximation, optimisation* etc. may provide a fertile ground for formalisation employing the notion of many-valuedness; in particular in the form of many-valued logic based algebraic theories of relations.

So, in a foundational analysis we have to distinguish sharply not only

1. mathematical questions,
2. logical questions,
3. ontological, epistemological and metaphysical questions,

but also look at their interrelationship, with particular emphasis on many-valued systems. For example, there are some interesting links between the mathematical and logical features of fuzzy structures of any kind and the ontological and epistemological questions of the foundational concepts. In order to bring these out explicitly, we need to employ an adequate method of conceptual analysis. In (1) we deal with the structure, in (2) we add to the structure the logical form. In (3) we deal with the problem of ontology, epistemology of the primitive concepts and perhaps, some *minimal* metaphysics of the systems involved; and also with with the questions of selection and justification of the appropriate meaning of the concepts employed. We have also to add the problematics of methods of enquiry and problem solving. This provides us with a conceptual framework, on the backcloth of which we should judge the issues dealt with comparative studies of various approaches in the field of fuzzy sets and systems.

References

[1] W. Bandler. Some esomathematical uses of category theory. In G. Klir, editor, *Applied General Systems Research: Recent Developments and Trends*, NATO Conference Series II, Vol. 5, pages 243–255. Plenum Press, New York, 1978.

[2] W. Bandler and L. Kohout. *Mathematical Relations, their Products and Generalized Morphisms.* Tech. report, Man-Machine Systems Laboratory, EES-MMS-REL 77-3, Dept. of Electrical Eng., University of Essex, Colchester, Essex, U.K., 1977.

[3] W. Bandler and L. Kohout. Fuzzy relational products and fuzzy implication operators. In *International Workshop on Fuzzy Reasoning Theory and Applications*, London, September 1978. Queen Mary College, University of London. Also available as FRP-1 report.

[4] W. Bandler and L. Kohout. The use of new relational products in clinical modelling. In

B. Gaines, editor, *General Systems Research: A Science, a Methodology, a Technology (Proc. 1979 North American Meeting of the Society for General Systems Research).*, pages 240–246, Louisville KY, January 1979. Society for General Systems Research.

[5] W. Bandler and L. Kohout. Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, 4:13–30, 1980. Reprinted in: *Readings in Fuzzy Sets for Intelligent Systems*, D. Dubois, H. Prade and R. Yager (eds.), Morgan Kaufmann Publishers, San Mateo, Calif., 1993, pages 88-96.

[6] W. Bandler and L. Kohout. Fuzzy relational products as a tool for analysis and synthesis of the behaviour of complex natural and artificial systems. In P. Wang and S. Chang, editors, *Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems*, pages 341–367. Plenum Press, New York and London, 1980.

[7] W. Bandler and L. Kohout. Semantics of implication operators and fuzzy relational products. *Internat. Journal of Man-Machine Studies*, 12:89–116, 1980. Reprinted in Mamdani, E.H. and Gaines, B.R. eds. *Fuzzy Reasoning and its Applications*. Academic Press, London, 1981, pages 219-246.

[8] U. Bodenhofer. Representations and constructions of similarity-based fuzzy orderings. *Fuzzy Sets and Systems*, 137:113–136, 2003.

[9] H. Bustince, V. Mohedano, E. Barrenechea, and M. Pagola. Definition and construction of fuzzy DI-subsethood measures. *Information Sciences*, 176:3190–3231, 2006.

[10] B. Gaines. Foundatons of fuzzy reasoning. *Internat. J. of Man-Machine Studies*, 8:623–668, 1976.

[11] B. Gaines and L. Kohout. The fuzzy decade: a bibliography of fuzzy systems and closely related topics. *Internat. Journal of Man-Machine Studies*, 9:1–68, 1977. Reprinted in: Gupta, M.M., Saridis, G.N. and Gaines, B.R., editors, Fuzzy Automata and Decision Processes, Elsevier North-Holland, New York and Amsterdam, 1988, pages 403-490.
(A critical survey with bibliography).

[12] S. Gottwald. A cummulative system of fuzzy sets. In *Lecture Notes in Mathematics 537*, pages 109–120. Springer Verlag, Berlin, 1976. (Collection Set Theory and Hierarchy Theory: A Memorial Tribute to Andrzej Mostovski).

[13] S. Gottwald. Set theory for fuzzy sets of higer level. *Fuzzy Sets and Systems*, 2:125–151, 1979.

[14] U. Höhle and L. Stout. Foundations of fuzzy sets. *Fuzzy Sets and Systems*, 40(2):257–296, 1991.

[15] D. Klaua. Über einen Ansatz zur mehrwertigen Mengenlehre. *Monatsb. Deutsch. Akad. Wiss. (Berlin)*, 7:859–867, 1965.

[16] D. Klaua. Einbettung der klassischen Mengenlehre in die mehrwertige. *Monatsb. Deutsch. Akad. Wiss. (Berlin)*, 9:258–272, 1967.

[17] L. J. Kohout and W. Bandler. Interval-valued systems for approximate reasoning based on the checklist paradigm. In P. Wang, Paul, editor, *Advances in Fuzzy Theory and Technology, vol. 1*, pages 167–193. Bookwrights Press, Durham, N.C., 1993.

[18] A. Pultr. Fuzzy mappings and fuzzy sets. *Comment. Math. Univ. Carolinae*, 17(3):441–459, 1976.

[19] N. Rescher. *Many-Valued Logic.* McGraw-Hill, New York, 1969.

[20] P. Sinha and E. Dougherty. Fuzzification of set inclusion theory and applications. *Fuzzy Sets and Systems*, 55(1993):15–42.

[21] L. Stout. Topoi and categories of fuzzy sets. *Fuzzy Sets and Systems*, 12:169–184, 1984.

[22] G. Takeuti and S. Titani. Intuitionistic fuzzy logic and intuitionistic fuzzy set theory. *Journal of Symbolic Logic*, 49(3):851–866, 1984.

[23] S. Weber. A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. *Fuzzy Sets and Systems*, 11:115–134, 1983.

[24] R. Willmott. Two fuzzier implication operators in the theory of fuzzy power sets. In *Fuzzy Research Project*, University of Essex, Colchester, U.K., December 1978. Dept. of Mathematics. FRP-2 report.

[25] R. Willmott. Two fuzzier implication operators in the theory of fuzzy power sets. *Fuzzy Sets and Systems*, 4:31–36, 1980.

[26] R. Young. Fuzzy subsethood. *Fuzzy Sets and Systems*, 77:371–384, 1996.

[27] L. Zadeh. An outline of a new approach to the analysis of complex systems and decision processes. In J. Cochrane and M. Zelený, editors, *Multiple Criteria Decision Making*, pages 686–725. University of South Carolina Press, Columbia, S.C., USA, 1973.

Exploring Dialogue Games as Foundation of Fuzzy Logic

Christian G. Fermüller
Theory and Logic Group 185.2
Technische Universität Wien
chrisf@logic.at

Abstract

A dialogue game based approach to the problem of providing a deeper semantic foundation for t-norm based fuzzy logics is explored. In particular, various versions, extensions and alternatives to Robin Giles's dialogue and betting game for Łukasiewicz logic are re-visited and put in the context of other foundational research in logic. It emerges that dialogue games cover a wide range of topics relevant to approximate reasoning.

Keywords: foundations, dialogue games, vagueness, analytic reasoning

1 Introduction

The adequate formalization of correct reasoning with vague notions and propositions is an important challenge in logic, computer science, as well as in philosophy. Many experts agree that modeling vagueness triggers reference to *degrees of truth* (but should strictly be distinguished from *degrees of belief*, and therefore requires methods different from probability theory and from modal logic). *Fuzzy Logics*, taken here in Zadeh's 'narrow sense' [18], are based on the extension of the two classical truth values by infinitely many intermediary degrees of truth. Formal deduction systems for specific fuzzy logics abound; however these systems are hardly ever explicitly related to models of correct reasoning with vague information. In other words: the challenge to derive inference systems for fuzzy logics from first principles about approximate reasoning is as yet largely unmet. The reference to general models of reasoning and to theories of vagueness—a prolific discourse in contemporary analytic philosophy—is only implicit, if not simply missing, in most presentations of inference systems for fuzzy logics. Some notable exceptions are: Ruspini's similarity semantics [25]; voting semantics [16]; 're-randomizing semantics' [15]; measurement-theoretic justifications [4]; the Ulam-Rényi game based interpretation of D. Mundici [20]; and 'acceptability semantics' of J. Paris [22]. As we have argued elsewhere [8], these formal semantics for various t-norm based logics (in particular Łukasiewicz logic) should be placed in the wider discourse on adequate *theories of vagueness*, a prolific subfield of analytic philosophy.

Here, we focus on a specific approach to derive logics from fundamental reasoning principles that was initiated by Robin Giles already in the 1970s [12]. This concept combines Paul Lorenzen's attempt to provide a dialogical foundation of logic in general (explained, e.g., in [17] and [2]) with a 'risk based' evaluation of atomic propositions that is specific to the context of vagueness understood as a phenomenon implying 'dispersion'. By the latter notion Giles refers to the fact that binary (yes/no)-experiments set up to test the acceptability of vague atomic assertions may show different outcomes when repeated. As we will see in Section 8, this concept allows to relate two seemingly very different theories of vagueness. Namely the familiar degree based, truth functional, approach of t-norm based fuzzy logic, on the one hand side, and 'supervaluationism', introduced by Kit Fine [11] and currently very popular among philosophers of vagueness, on the other hand side.

We will briefly review Giles's characterization of Łukasiewicz logic and provide an overview over more recent results covering a wider range of logics. As a sort of conclusion, we will indicate connections of the dialogical paradigm to other important foundational research programs in contemporary logic.

For brevity we restrict our attention to propositional logics, here. We assume the reader to be familiar with the basic concepts of t-norm based fuzzy logics as presented, e.g., in [14]. On the other hand we aim at a self-contained presentation as far as the (presumably less well known) concept of dialogue games as formal foundation of logical reasoning is concerned.

2 Giles's game for Łukasiewicz logic

Giles's analysis [12] of approximate reasoning originally referred to the phenomenon of 'dispersion' in the context of physical theories. Later Giles [13] explicitly applied the same concept to the problem of providing 'tangible meanings' to (logically complex) fuzzy propositions. For this purpose he introduces a game that consists of two independent components:

(1) Betting for positive results of experiments.

Two players—say: *me* and *you*—agree to pay 1€ to the opponent player for every false statement they assert. By $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ we denote an *elementary state* of the game, where I assert each of the q_i in the multiset $\{q_1, \ldots, q_n\}$ of atomic statements (represented by propositional variables), and you, likewise, assert each atomic statement $p_i \in \{p_1, \ldots, p_m\}$.

Each propositional variable q refers to an experiment E_q with binary (yes/no) result. The statement q can be read as 'E_q yields a positive result'. Things get interesting as the experiments may show dispersion; i.e., the same experiment may yield different results when repeated. However, the results are not completely arbitrary: for every run of the game, a fixed *risk value* $\langle q \rangle^r \in [0,1]$ is associated with q, denoting the probability that E_q yields a negative result. For the special atomic formula $\bot$ (*falsum*) we define $\langle \bot \rangle^r = 1$. The risk associated with a multiset $\{p_1, \ldots, p_m\}$ of atomic formulas is defined as $\langle p_1, \ldots, p_m \rangle^r = \sum_{i=1}^{m} \langle p_i \rangle^r$. The risk $\langle\rangle^r$ associated with the empty multiset is defined as 0. The risk associated with an elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ is calculated from my point of view. Therefore the condition $\langle p_1, \ldots, p_m \rangle^r \geq \langle q_1, \ldots, q_n \rangle^r$ expresses that I do not expect any loss (but possibly some gain) when betting on the truth of atomic statements, as explained above.

(2) A dialogue game for the reduction of compound formulas.

Giles follows ideas of Paul Lorenzen and his school that date back already to the 1950s (see, e.g., [17]) and constrains the meaning of logical connectives by reference to rules of a dialogue game that proceeds by systematically reducing arguments about compound formulas to arguments about their subformulas.

We at first assume that formulas are built up from propositional variables, the falsity constant $\bot$, and the connective $\to$ only.[1] The central dialogue rule can then be stated as follows:

($R_\to$) If I assert $A \to B$ then, whenever you choose to attack this statement by asserting A, I have to assert also B. (And vice versa, i.e., for the roles of me and you switched.)

This rule reflects the idea that the meaning of implication is specified by the principle that an assertion of 'if A, then B' ($A \to B$) obliges one to assert B, if A is granted.

In contrast to dialogue games for intuitionistic logic [17, 7], no special regulation on the succession of moves in a dialogue is required here. However, we assume that each assertion is attacked at most once: this is reflected by the removal of $A \to B$ from the multiset of all formulas asserted by a player during a run of the game, as soon as the other player has either attacked by asserting A, or has indicated that she will not attack $A \to B$ at all. Note that every run of the dialogue game ends in an elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$. Given an assignment $\langle\cdot\rangle^r$ of risk values to all p_i and q_i we say that I *win* the corresponding run of the game if I do not expect any loss, i.e., if $\langle p_1, \ldots, p_m \rangle^r \geq \langle q_1, \ldots, q_n \rangle^r$.

As an almost trivial example consider the game where I initially assert $p \to q$ for some atomic formulas p and q; i.e., the initial state is $[\|p \to q]$. In response, you can either assert p in order to force me to assert q, or explicitly refuse to attack $p \to q$. In the first case, the game ends in the elementary state $[p\|q]$; in the second case it ends in state $[\|]$. If an assignment $\langle\cdot\rangle^r$ of risk values gives $\langle p \rangle^r \geq \langle q \rangle^r$, then I win, whatever move you choose to make. In other words: I have a winning strategy for $p \to q$ in all assignments of risk values where $\langle p \rangle^r \geq \langle q \rangle^r$.

Recall that a *valuation* v for Łukasiewicz logic Ł is a function assigning values $\in [0,1]$ to the propositional variables and 0 to $\bot$, extended to compound formulas using the truth function $x \Rightarrow_Ł y = \inf\{1, 1 - x + y\}$.

Theorem. (R. Giles [12])
Every assignment $\langle\cdot\rangle^r$ of risk values to atomic formulas occurring in a formula F induces a valuation $v_{\langle\cdot\rangle^r}$ for Ł such that $v_{\langle\cdot\rangle^r}(F) = 1$ iff I have a winning strategy for F in the game presented above.

Corollary.
F is valid in Ł iff, for all assignments of risk values to atomic formulas occurring in F, I have a winning strategy for F.

3 Other connectives

Although all other connectives can be defined in Łukasiewicz logic from $\to$ and $\bot$ alone, it will be helpful to illustrate the idea that the meaning of *all* rele-

[1] Remember that in Ł all other connectives can be defined from $\to$ and $\bot$ alone. (See, e.g., [14].)

vant connectives can be specified directly by intuitively plausible dialogue rules. Interestingly, for conjunction *two* different rules seem to be plausible candidates at a first glance:

($R_\wedge$) If I assert $A_1 \wedge A_2$ then I have to assert also A_i for any $i \in \{1, 2\}$ that you may choose.

($R_{\wedge'}$) If I assert $A_1 \wedge' A_2$ then I have to assert also A_1 as well as A_2.

Of course, both rules turn into rules referring to *your* claims of a conjunctive formula by simply switching the roles of the players ('I' and 'you').

Rule ($R_\wedge$) is dual to the following natural candidate for a disjunction rule:

($R_\vee$) If I assert $A_1 \vee A_2$ then I have to assert also A_i for some $i \in \{1, 2\}$ that I myself may choose.

Moreover, it is clear how ($R_\wedge$) generalizes to a rule for universal quantification.

It follows already from results in [12] that rules ($R_\wedge$) and ($R_\vee$) are adequate for 'weak' conjunction and disjunction in Ł, respectively. $\wedge$ and $\vee$ are also called 'lattice connectives' in the context of fuzzy logics, since their truth functions are given by $v(A \wedge B) = \inf\{v(A), v(B)\}$ and $v(A \vee B) = \sup\{v(A), v(B)\}$.

The question arises, whether one can use the remaining rule ($R_{\wedge'}$) to characterize strong conjunction ($\&$) which corresponds to the t-norm $x *_Ł y = \sup\{0, x + y - 1\}$.

However, rule ($R_{\wedge'}$) is inadequate in the context of our betting scheme for random evaluation in a precisification space. The reason for this is that we have to ensure that for each (not necessarily atomic) assertion that we make, we risk a *maximal* loss of 1€ only. It is easy to see that rules ($R_\rightarrow$), ($R_\wedge$), and ($R_\vee$) comply with this 'principle of limited liability'. However, if I assert $p \wedge' q$ and we proceed according to ($R_{\wedge'}$), then I end up with a loss of 2€, in case both experiments E_p and E_q fail. There is a simply way to redress this situation to obtain a rule that is adequate for ($\&$): Allow any player who asserts $A_1 \,\&\, A_2$ to hedge her possible loss by asserting $\bot$ instead of A_1 and A_2, if whished. Asserting $\bot$, of course, corresponds to the obligation to pay 1€ (but not more) in the resulting final state. We obtain the following rule for strong conjunction:

($R_\&$) If I assert $A_1 \,\&\, A_2$ then I either have to assert A_1 as well as A_2, or else I have to assert $\bot$.

In a similar way, also dialogue rules for negation, 'strong' disjunction, and equivalence can be formulated directly, instead of just derived from ($R_\rightarrow$).

4 Beyond Łukasiewicz logic

There is an interesting ambiguity in the phrase 'betting for positive results of (a multiset of) experiments' that describes the evaluation of elementary states of the dialogue game. As explained above, Giles identifies the combined risk for such a bet with the *sum* of risks associated with the single experiments. However, other ways of interpreting the combined risk are worth exploring. In [6] we have considered a second version of the game, where an elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ corresponds to my single bet that *all* experiments associated with the q_i, where $1 \leq i \leq n$, show a positive result, against your single bet that *all* experiments associated with the p_i ($1 \leq i \leq m$) show a positive result. A third form of the game arises (again, see [6]) if one decides to perform only *one* experiment for each of the two players, where the relevant experiment is chosen by the opponent.

It turns out that these three betting schemes constitute three versions of Giles's game that are adequate for the three fundamental logics Ł (Łukasiewicz logic), P (Product logic), and G (Gödel logic), respectively. To understand this result it is convenient to invert risk values into probabilities of *positive* results (yes-answers) of the associated experiments. More formally, the *value* of an atomic formula q is defined as $\langle q \rangle = 1 - \langle q \rangle^r$; in particular, $\langle \bot \rangle = 0$.

My expected gain in the elementary state $[p_1, \ldots, p_m \| q_1, \ldots, q_n]$ in Giles's game for Ł is the sum of money that I expect you to have to pay me minus the sum that I expect to have to pay you. This amounts to $\sum_{i=1}^m (1 - \langle p_i \rangle) - \sum_{i=1}^n (1 - \langle q_i \rangle)$ €. Therefore, my expected gain is greater or equal to zero iff $1 + \sum_{i=1}^m (\langle p_i \rangle - 1) \leq 1 + \sum_{i=1}^n (\langle q_i \rangle - 1)$ holds. The latter condition is called winning condition $W_\sum$. (Note that 'winning' here refers to *expected* gain: although, in this sense, I 'win' in state $[p \| p]$, I still loose 1€ in those concrete runs of the game, where the instance of the experiment E_p referring to *my* assertion of p results in 'no', but where the instance of E_p referring to *your* assertion of p end positively (answer 'yes').

In the second version of the game, you have to pay me 1€ unless all experiments associated with the p_i test positively, and I have to pay you 1€ unless all experiments associated with the q_i test positively. My expected gain is therefore $1 - \prod_{i=1}^m \langle p_i \rangle - (1 - \prod_{i=1}^n \langle q_i \rangle)$ €; the corresponding winning condition $W_\prod$ is $\prod_{i=1}^m \langle p_i \rangle \leq \prod_{i=1}^n \langle q_i \rangle$.

To maximize the expected gain in the third version of the game I will choose a $p_i \in \{p_1, \ldots, p_m\}$ where the probability of a positive result of the associated experiment is least; and you will do the same for the

q_i's that I have asserted. Therefore, my expected gain is $(1 - \min_{1\leq i\leq m}\langle p_i\rangle) - (1 - \min_{1\leq i\leq n}\langle q_i\rangle)$ €. Hence the corresponding winning condition $W_{\min}$ is $\min_{1\leq i\leq m}\langle p_i\rangle \leq \min_{1\leq i\leq n}\langle q_i\rangle$.

We thus arrive at the following definitions for the value of a multiset $\{p_1, \ldots, p_n\}$ of atomic formulas, according to the three versions of the game:

$\langle p_1, \ldots, p_n\rangle_{\text{Ł}} = 1 + \sum_{i=1}^{n}(\langle p_i\rangle - 1)$
$\langle p_1, \ldots, p_n\rangle_{\text{P}} = \prod_{i=1}^{n}\langle p_i\rangle$
$\langle p_1, \ldots, p_n\rangle_{\text{G}} = \min_{1\leq i\leq n}\langle p_i\rangle$.

For the empty multiset we define $\langle\rangle_{\text{Ł}} = \langle\rangle_{\text{P}} = \langle\rangle_{\text{G}} = 1$.

In contrast to Ł, the dialogue game rule (R) does not suffice to characterize P and G. To see this, consider the state $[p \to \bot \| q]$. According to rule (R) I may assert p in order to force you to assert $\bot$. Since $\langle\bot\rangle = 0$, the resulting elementary state $[\bot \| p, q]$ fulfills the winning conditions $\langle\bot\rangle \leq \langle p\rangle \cdot \langle q\rangle$ and $\langle\bot\rangle \leq \min\{\langle p\rangle, \langle q\rangle\}$, that correspond to P and G, respectively. However, this is at variance with the fact that for assignments where $\langle p\rangle = 0$ and $\langle q\rangle < 1$ you have asserted a statement $(p \to \bot)$ that is definitely true ($v(p \to \bot) = 1$), whereas my statement q is not definitely true ($v(q) < 1$).[2]

There are different ways to address the indicated problem. They all seem to imply a break of the symmetry between the roles of the two players (me and you). We have to distinguish between elementary states in which my expected gain is non-negative and those in which my expected is strictly positive. Accordingly, we introduce a (binary) signal or *flag* ¶ into the game which, when raised, announces that I will be declared the winner of the current run of the game, only if the evaluation of the final elementary state yields a *strictly positive* (and not just non-negative) expected gain for me. This allows us to come up with a version of the dialogue rules for implication that can be shown to lead to adequate games for all three logics consider here (Ł, P, G):

($R^{I*}_{\to}$) If I assert $A \to B$ then, whenever you choose to attack this statement by asserting A, I have the following choice: either I assert B in reply or I challenge your attack on $A \to B$ by replacing the current game with a new one in which you assert A and I assert B.

In formulating an adequate rule for my attacks on your assertions of an implicative formulas we have to use the flag signalling the strict case of the winning condition:

($R^{Y*}_{\to}$) If you assert $A \to B$ then, whenever I choose to attack this statement by asserting A, you have the following choice: either you assert B in reply or you challenge my attack on $A \to B$ by replacing the current game with a new one in which the flag ¶ is raised and I assert A while you assert B.

In contrast to Ł, in G and P the other connectives cannot be defined from $\to$ and $\bot$ alone. However, the rules presented in Section 3 turn out to be adequate for G and P, too. In the case of Gödel logic (G), the two versions of conjunction ('strong' and 'weak') coincide. This fact, that is well known from the algebraic view of t-norm based logic (see, e.g., [14]) can also be obtained by comparing optimal strategies involving the rules ($R_\wedge$) and ($R_\&$), respectively.

5 Truth comparison games

In [9] yet another dialogue game based approach to reasoning in Gödel logic G has been described. It relies on the fact that G is the only t-norm based logic, where the validity of formulas depends only on the *relative order* of the values of the involved propositional variables. This observation arguably is of philosophical interest in the context of scepticism concerning the meaning of particular real numbers $\in [0, 1]$ understood as 'truth values'. To emphasize that only the *comparison* of degrees of truth, using the standard order relations $<$ and $\leq$, is needed in evaluating formals in G, one may refer to a dialogue game which is founded on the idea that any logical connective $\circ$ of G can be characterized via an adequate response by a player **X** to player **Y**'s attack on **X**'s claim that a statement of form $(A \circ B) \lhd C$ or $C \lhd (A \circ B)$ holds, where $\lhd$ is either $<$ or $\leq$.

We need the following definitions. An assertion $F \lhd G$ is *atomic* if F and G are either propositional ; otherwise it is a *compound assertion*. Atomic assertions of form $p < p$, $p < \bot$, $\top < p$ or $\top \leq \bot$ are called *elementary contradictions*. An *elementary order claim* is a set of two assertions of form $\{E \lhd_1 F, F \lhd_2 G\}$, where E, F, and G are atoms, and $\lhd_1, \lhd_2 \in \{<, \leq\}$.

Following traditional terminology, introduced by Paul Lorenzen, we call the player that initially claims the validity of a chosen formula the *Proponent* **P**, and the player that tries to refute this claim the *Opponent* **O**. The dialogue game proceeds in rounds as follows:

1. A dialogue starts with **P**'s claim that a formula F is valid. **O** answers to this move by contradicting this claim with the assertion $F < \top$.
2. Each following round consists in two steps:
   (i) **P** either attacks a compound assertion or an elementary order claim, contained in the set

[2] The problem does not arise in logic Ł, since there the expected gain for state $[\bot \| p, q]$ is $\langle p, q\rangle_{\text{Ł}} - \langle\bot\rangle_{\text{Ł}} = 1 - (\langle p\rangle - 1) - (\langle q\rangle - 1) - (1 - 1) = \langle p\rangle + \langle q\rangle - 1$ and therefore, indeed, negative, as expected, if $\langle p\rangle = 0$ and $\langle q\rangle < 1$.

of assertions that have been made by **O** up to this state of the dialogue, but that have not yet been attacked by **P**.

(ii) **O** answers to the attack by adding a set of assertions according to the rules of Table 1 (for compound assertions) and Table 2 (for elementary order claims).

3. The dialogue ends with **P** as winner if **O** has asserted an elementary contradiction. Otherwise, **O** wins if there is no further possible attack for **P**.

Table 1: Rules for connectives

P attacks:	**O** asserts as answer:
$A\&B \lhd C$	$\{A \lhd C\}$ or $\{B \lhd C\}$
$C \lhd A\&B$	$\{C \lhd A,\ C \lhd B\}$
$A \vee B \lhd C$	$\{A \lhd C,\ B \lhd C\}$
$C \lhd A \vee B$	$\{C \lhd A\}$ or $\{C \lhd B\}$
$A \to B < C$	$\{B < A,\ B < C\}$
$C < A \to B$	$\{C < B\}$ or $\{A \leq B,\ C < \top\}$
$A \to B \leq C$	$\{\top \leq C\}$ or $\{B < A, B \leq C\}$
$C \leq A \to B$	$\{A \leq B\}$ or $\{C \leq B\}$

In the first four lines, $\lhd$ denotes either $<$ or $\leq$, consistently throughout each line. Assertions, which involve a choice of **O** in the answer (indicated by 'or') are called *or-type* assertions. All other assertions are of *and-type*.

Table 2: Rules for elementary order claims

P attacks:	**O** asserts as answer:
$\{A \leq B,\ B \leq C\}$	$\{A \leq C\}$
$\{A < B,\ B \lhd C\}$	$\{A < C\}$
$\{A \lhd B,\ B < C\}$	$\{A < C\}$

where $\lhd$ is either $<$ or $\leq$.

Remark. Instead of considering the rules of Table 1 and 2 as derived from the truth functions for G, one may argue that the dialogue rules are derived from fundamental principles about reasoning in a truth functional, order based fuzzy logic.

Consider the example of conjunction. We contend that anyone who claims '$A\&B$ is at least as true as C' (for any concrete statements A, B, and C) has to be prepared to defend the claim that 'A is at least as true as C' and the claim that 'B is at least as true as C'. On the other hand, the claim that 'C is at least as true as $A\&B$', arguably, should be supported either by 'C is at least as true as A' or by 'C is at least as true as B'. (Likewise, if we replace 'at least as true' by 'truer than'.) One may then go on to argue that this reading of the rules for & in Table 1 completely determines correct reasoning about assertions of this form. Form this assumption, one can *derive* that $v(A\&B) = \min(v(A), v(B))$ is the only adequate definition for the semantics of conjunction in this setting.

The case for disjunction is very similar. Implication, as usual, is more controversial. However, it is easy to see that there are hardly any reasonable alternatives to our rules, if the truth of any assertion involving a formula $A \to B$ should only depend on the relative degree of truth of A and B (but should not depend on the result of an arithmetical operations that had to be performed on the values of A and B, respectively).

Formally we may summarize this analysis of Gödel logic as follows:

Theorem. [9]
A formula F is valid in G iff there exists a winning strategy for **P** *on F in the presented comparison game.*

6 Pavelka style reasoning

An important paradigm for approximate reasoning has been explored in a series of papers by J. Pavelka [23]. It is sometimes also advocated as 'fuzzy logic with evaluated syntax' (see, e.g., [21]). In this approach one makes the reference to degrees of truth explicit by considering *graded formulas* $r : F$ as basic objects of inference, where r is a rational number $\in [0, 1]$ and F is an Ł-formula, with the intended interpretation that F is evaluated to a value $\geq r$. The resulting logic is called *rational Pavelka logic* RPL in [14].

Inference systems for RPL can be obtained by using the following graded version of modus ponens as rule of derivation:

$$\frac{r : A \quad s : A \to B}{r *_{Ł} s : B}$$

Completeness and soundness of such systems can be stated as the coincidence of the *truth degree* $\|F\|_T$ of F over some theory (set of graded formulas) T with the *provability degree* $|F|_T$ of F over T. Here $\|F\|_T$ is defined as $\inf_{v \in I_T} v(F)$, where I_T is the set of all Ł-valuations satisfying T; and $|F|_T$ is defined as $\sup\{r \mid T \vdash r : F\}$, where $\vdash$ denotes the indicated derivability relation. (See, e.g., [14] for details.)

It is easy to see that Giles's dialogue game for Ł can be adapted to RPL, since a graded formula $r : F$ can be expressed as $\bar{r} \to F$ in Ł if Ł is extended by truth constants $\bar{r}$ for all rationals $\in [0, 1]$, interpreted by stipulating $v(\bar{r}) = r$. The only change in Giles's original dialogue and betting scenario (explained in Section 2) is the additional reference to special elementary experiments $E_{\bar{r}}$ with fixed success probabilities r. Such experiments can easily be defined for all rational p by referring to a certain number of fair coin tosses and an adequate definition of a 'positive result'. According to the dialogue rule $(R_{\to})$ of Section 2 an attack on the graded statement $r : F$ $(=\bar{r} \to F)$ indicates the willingness of the attacking player to bet on a positive result of $E_{\bar{r}}$ in exchange for an assertion of F by the

other player. Clearly, one can simplify the overall pay-off scheme by stipulating that an attack by player **X** on a graded formula $r : F$ consists in paying $(1 - r)$€ to the opponent player **Y** and thereby forcing **Y** to continue the game with an assertion of F.

Since Ł is the only fuzzy logic, where also the residuum $\Rightarrow_Ł$ of the underlying t-norm is a continuous function, one cannot readily transfer the concept of provability degrees that match truth degrees to other logics. Nevertheless, it makes sense to enrich the syntax of Gödel logic G and Product logic P not only by rational truth constants, but also by a binary connective ':' with the corresponding truth function $\tilde{:}$ given by

$$x \,\tilde{:}\, y = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{otherwise} \end{cases}$$

Whereas Hilbert-style axiomatizations of such enriched logics, which contain 'evaluated syntax', seem elusive (beyond RPL), it is rather straightforward to define dialogue game rules and pay-off schemes that capture the intended meaning of the thus extended versions of G and P. Moreover, in the case of G, it is also possible to extend the truth comparison game sketched in Section 5 to include evaluated syntax. (Details are left to future work.)

7 Connections to supervaluation

Supervaluation is a widely discussed concept in philosophical logic. Kit Fine has pioneered its application to formal languages that accommodate vague propositions in [11], a paper that remains an important reference point for philosophers of language and logic. The main idea is to evaluate propositions not simply with respect to classical interpretations—i.e., assignments of the truth values 0 ('false') and 1 ('true') to atomic statements—but rather with respect to a whole *space* Π of (possibly) partial interpretations. For every partial interpretation I in Π, Π is required to contain also a classical interpretation I' that extends I. I' is called an *admissible (complete) precisification* of I. A proposition is called *supertrue* in Π if it evaluates to 1 in all admissible precisifications, i.e., in all classical interpretations contained in Π.

Supervaluation and fuzzy logics can be viewed as capturing contrasting, but individually coherent intuitions about the role of logical connectives in vague statements. Consider a sentence like

(*) "The sky is blue and is not blue".

When formalized as $b \,\&\, \neg b$, (*) is superfalse in all precisification spaces, since either b or $\neg b$ is evaluated to 0 in each precisification. This fits Kit Fine's motivation in [11] to capture 'penumbral connections' that prevent any mono-colored object from having two colors at the same time. According to Fine's intuition the statement "The sky is blue" absolutely contradicts the statement "The sky is not blue", even if neither statement is definitely true or definitely false. Consequently (*) is judged as definitely false, although admittedly composed of vague sub-statements. On the other hand, by asserting (*) one may intend to convey the information that both component statements are true *only to some degree*, different from 1 but also from 0. Under this reading and certain 'natural' choices of truth functions for & and ¬ the statement $b \,\&\, \neg b$ is *not* definitely false, but receives some intermediary truth value.

In [10], we have worked out a dialogue game based attempt to reconcile supervaluation and t-norm based ('fuzzy') evaluation within a common formal framework. To this aim we interpret 'supertruth' as a modal operator and define a logic SŁ that extends both, Łukasiewicz logic Ł, as well as the classical modal logic S5.

Formulas of SŁ are built up from the propositional variables $p \in V = \{p_1, p_2, \ldots\}$ and the constant $\bot$ using the connectives & and $\rightarrow$. The additional connectives $\neg$, $\wedge$, and $\vee$ are defined as explained above. In accordance with our earlier (informal) semantic considerations, a precisification space is formalized as a triple $\langle W, e, \mu\rangle$, where $W = \{\pi_1, \pi_2, \ldots\}$ is a non-empty (countable) set, whose elements π_i are called *precisification points*, e is a mapping $W \times V \mapsto \{0, 1\}$, and μ is a probability measure on the σ-algebra formed by all subsets of W. Given a precisification space $\Pi = \langle W, e, \mu\rangle$ a *local truth value* $\|A\|_\pi$ is defined for every formula A and every precisification point $\pi \in W$ inductively by

$$\begin{aligned}
\|p\|_\pi &= e(\pi, p), \text{ for } p \in V \\
\|\bot\|_\pi &= 0 \\
\|A \,\&\, B\|_\pi &= \begin{cases} 1 & \text{if } \|A\|_\pi = 1 \text{ and } \|B\|_\pi = 1 \\ 0 & \text{otherwise} \end{cases} \\
\|A \rightarrow B\|_\pi &= \begin{cases} 1 & \text{if } \|A\|_\pi = 1 \text{ and } \|B\|_\pi = 0 \\ 0 & \text{otherwise} \end{cases} \\
\|\mathsf{S}A\|_\pi &= \begin{cases} 1 & \text{if } \forall \sigma \in W : \|A\|_\sigma = 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Local truth values are classical and do not depend on the underlying t-norm $*_Ł$. In contrast, the *global truth value* $\|A\|_\Pi$ of a formula A is defined by

$$\begin{aligned}
\|p\|_\Pi &= \mu(\{\pi \in W | e(\pi, p) = 1\}), \text{ for } p \in V \\
\|\bot\|_\Pi &= 0 \\
\|A \,\&\, B\|_\Pi &= \|A\|_\Pi *_Ł \|B\|_\Pi \\
\|A \rightarrow B\|_\Pi &= \|A\|_\Pi \Rightarrow_Ł \|B\|_\Pi \\
\|\mathsf{S}A\|_\Pi &= \|\mathsf{S}A\|_\pi \text{ for any } \pi \in W
\end{aligned}$$

Note that $\|\mathsf{S}A\|_\pi$ is the same value (either 0 or 1) for all $\pi \in W$. In other words: 'local' supertruth is in fact already global; which justifies the above clause for $\|\mathsf{S}A\|_\Pi$. Also observe that we could have used the global conditions, referring to $*_Ł$ and $\Rightarrow_Ł$, also to define $\|A \,\&\, B\|_\pi$ and $\|A \to B\|_\pi$, since the t-norm based truth functions coincide with the (local) classical ones, when restricted to $\{0, 1\}$. (However that presentation might have obscured their intended meaning.)

Most importantly for our current purpose, it has been demonstrated in [10] that the evaluation of formulas of SŁ can be characterized by a dialogue game extending Giles's game for Ł, where 'dispersive elementary experiments' (see Section 2) are replaced by 'indeterministic evaluations' over precisification spaces. The dialogue rule for the supertruth modality involves a relativization to specific precisification points:

(R_S) If I assert $\mathsf{S}A$ then I also have to assert that A holds at any precisification point π that you may choose. (And *vice versa*, i.e., for the roles of me and you switched.)

The resulting game is adequate for SŁ:

Theorem. [10]
A formula F is valid in SŁ *iff for every precisification space* Π *I have a winning strategy for the game starting with my assertion of F.*

8 Dialogue games in a wider context

Having sketched the rather varied landscape of dialogue game based approaches to the foundations of fuzzy logic—following Giles's pioneering work in the 1970s—we finally want to hint briefly at some connections with other foundational enterprises in logic. We think that these connections indicate potential benefits that the dialogue game approach might enjoy relative to alternative semantic frameworks mentioned in the introduction (Section 1).

Connections to Lorenzen style constructivism. It is certainly true that reasoning with vague notions and propositions poses challenges to philosophical logic that are different from well known concerns about, e.g., constructive meaning, adequate characterization of entailment ('relevance'), or intentional logics. However, one should not dismiss the possibility that traditional approaches to foundational problems in logic may benefitly be employed to enhance the understanding of fuzzy logics, too. Lorenzen's dialogical paradigm is a widely discussed, flexible tool in such foundational investigations. (See, e.g., [2, 24, 17].) Its philosophical underpinnings can assist in the difficult task to derive mathematical structures that are used in fuzzy logics from more fundamental assumptions about correct reasoning. In this context, the fact that Lorenzen and his collaborators have (somewhat narrowly) focussed on intuitionistic logic, may help to uncover deep connections between constructive reasoning and reasoning under vagueness.

Connections to 'game logics' and 'logic games'. In recent years the logical analysis of games as well as game theoretic approaches to logic emerge as prolific foundational research areas that entail interest in topics like dynamics and interaction of reasoning agents, analysis of strategies and different forms of knowledge. (See, e.g., [3] or `www.illc.uva.nl/lgc/` for further references.) It is clear that dialogue games, like the ones described in this paper, nicely fit in this framework. Foundational research in fuzzy logic, along the lines indicated here, will surely profit from new results about games in logic and logic in games. Moreover, it is not unreasonable to hope that, *vice versa*, fuzzy logic has to offer interesting new perspectives on agent knowledge and interaction that will be taken up by 'game logics' in future research.

Connections to proof search and analytic calculi. The renewed interest in Giles's game for Ł, indicated in our brief survey, above, was in fact triggered by the discovery of relations between corresponding winning strategies, on the one hand side, and cut-free derivations in a so-called hypersequent system for Ł[19], on the other hand side. The logical rules of the uniform analytic proof system for Ł, G, and P introduced in [6], correspond directly the dialogue rules of the modified dialogue game described in Section 4, above. This entails a close correspondence between the systematic construction of winning strategies for the dialogue game and proof search strategies in the uniform analytic calculus. Moreover, this correspondence can be viewed as a general principle for the interpretation of logical rules in analytic (i.e., cut-free) hypersequent or sequent calculi in terms of the options of the two players in a dialogue game. It remains to be seen whether, as a consequence, the dialogue game approach can be used (beyond foundational concerns) also to model and plan efficient proof search.

Connections to substructural logics. Games that have been inspired by Lorenzen's original dialogue game for intuitionistic logic are widely used in the analysis of (fragments of) linear logic and related formalism (see, e.g., [5, 1]). This research field, often simply called 'game semantics', highlights applications of rather abstract forms of dialogue games, where logical connectives are viewed as certain operators on formal games. While the emphasis in dialogical approaches to fuzzy logics, arguably, is closer to philosophical concerns about providing 'tangible meaning' (to use a phrase

of Robin Giles), it is nevertheless evident that there are common interests in the search for alternative semantics of linear logic and t-norm based fuzzy logics, respectively. To name just one corresponding problem: How can the feature of 'resource consciousness' of logics be adequately characterized at the level of analytic reasoning? Dialogue semantics clearly aims at a direct model of this and related features of information processing, thus stressing the well known fact that t-norm based fuzzy logics can be viewed as a particular type of substructural logics.

Let us finally point out that this short survey on dialogue games for fuzzy logics is far from complete. Among related topics, pursued elsewhere, we just mention evaluation games, parallel dialogue games for intermediate logics (including G) and connections to Mundici's analysis of the Ulam-Rényi game. However, already the results described here allow us to conclude that the dialogical approach, although originally developed in a quite different philosophical context, bears fruits also in the realm fuzzy logic.

Acknowledgement

This work has been partly supported by FWF, project MAT P18563-N12.

References

[1] S. Abramsky, G. McCusker. Game semantics. In H. Schwichtenberg and U. Berger, editors, *Computational Logic*, 1–56. Springer-Verlag, 1999.

[2] E.M. Barth, E.C.W. Krabbe. From Axiom to Dialogue. De Gruyter, 1982.

[3] J. van Benthem. Logic in Game (Book). In preparation,

[4] T. Bilgic, I.B. Türksen. Measurement-theoretic justification of connectives in fuzzy set theory. *Fuzzy Sets and Systems*, 76(3):289–308, 1995.

[5] A. Blass. A game semantics for linear logic. *Annals of Pure and Applied Logic*, 56:183–220, 1992.

[6] A. Ciabattoni, C.G. Fermüller, G. Metcalfe. Uniform rules and dialogue games for fuzzy logics. LPAR 2004, Springer LNCS 3452, 496–510, 2005.

[7] W. Felscher. Dialogues, strategies, and intuitionistic provability. *Annals of Pure and Applied Logic*, 28:217–254, 1985.

[8] C.G. Fermüller. Theories of vagueness versus fuzzy logic: Can logicians learn from philosophers. *Neural Network World*, 13(5):455–466, 2003.

[9] C.G. Fermüller, N. Preining. A Dialogue Game for Intuitionistic Fuzzy Logic Based on Comparisons of Degrees of Truth. Proceedings of *InTech'03*, (no page numbers), 2003.

[10] C.G. Fermüller, R. Kosik. Combining Supervaluation and Degree Based Reasoning Under Vagueness. LPAR 2006, Springer LNCS 4246, 212-226

[11] K. Fine. Vagueness, truth and logic. *Synthése*, 30:265–300, 1975.

[12] R. Giles. A non-classical logic for physics. *Studia Logica*, 4(33):399–417, 1974.

[13] R. Giles. Łukasiewicz logic and fuzzy set theory. *International Journal of Man-Machine Studies*, 8(3):313–327, 1976.

[14] Petr Hájek. Metamathematics of Fuzzy Logic. Kluwer, 1998.

[15] E. Hisdal. Are grades of membership probabilities? *Fuzzy Sets and Systems*, 25:325–348, 1988.

[16] J. Lawry. A voting mechanism for fuzzy logic. *The International Journal of Approximate Reasoning*, 19:315–333, 1998.

[17] P. Lorenzen. Logik und Agon. In *Atti Congr. Internaz. di Filosofia*, pages 187–194. Sansoni, 1960.

[18] R. J. Marks, editor. *Fuzzy logic technology and applications*. IEEE technology updates. IEEE Press, New York, 1994.

[19] G. Metcalfe, N. Olivetti, and D. Gabbay. Analytic sequent calculi for abelian and Łukasiewicz logics. TABLEAUX 2002, Springer LNCS 2381, 191–205.

[20] D. Mundici. Ulam's game, Łukasiewicz logic and C*-algebras. *Fundamenta Informaticae*, 18:151–161, 1993.

[21] V. Novák, I. Perfilieva, and J. Mockor. *Mathematical Principles of Fuzzy Logic*. Kluwer, Boston, MA, 1999.

[22] J.B. Paris. Semantics for fuzzy logic. *Soft Computing*, 1:143–147, 1997.

[23] J. Pavelka. On fuzzy logic I, II, III. *Z. Math. Logik Grundl. Math.* 25:45–52; 119–134; 447–464.

[24] S. Rahman, H. Rückert (editors). Special issue on 'Dialogic Logic'. *Synthese* 127:1-2, April 2001.

[25] E.H. Ruspini. On the semantics of fuzzy logic. *International Journal of Approximate Reasoning*, 5:45–88, 1991.

Representing Groups with Imprecise Opinions

Hannu Nurmi*
Academy of Finland and
Department of Political Science
University of Turku
FI-20014 TURKU
Finland
e-mail:hnurmi@utu.fi

Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
Ul. Newełska 6
01-447 Warsaw
Poland
e-mail: kacprzyk@ibspan.waw.pl

Abstract

This paper deals with the composition of multi-member decision making bodies (committees) on the basis of fuzzy individual preference relations over candidates. The methods discussed are basically extensions of the principles of optimal representation as spelled out by Chamberlin and Courant. Since these principles are defined for non-fuzzy individual preference rankings, we focus primarily on the essential ambiguity in determining the degree of representation associated with each possible committee when the individual preferences are fuzzy. Once this ambiguity is solved, the composition of an optimal committee is relatively straightforward.

1 Introduction

The setting studied by social choice theory is typically one where we are given a set N of voters, set K of alternatives, a profile R of voter opinions over alternatives and our task is to design a method for electing a subset of K, i.e. a set of best alternatives or winner so that the voter opinions are reflected in the choice in a plausible (just, democratic, rational) way. Typically, it is assumed that the voter opinions can be represented as complete and transitive binary relations over K. The plausibility of the choice procedure is determined by the conditions one imposes on the method. The social choice theory is well known for the plethora of negative results stating the incompatibility of several intuitively desirable properties of choice methods. The best-known of these results is undoubtedly Arrow's (1961) theorem which states the incompatibility of universal domain, Pareto optimality, independence of irrelevant alternatives and non-dictatorship insofar as one is focusing on methods that result in complete and transitive relations over the alternatives. In other words, the result says that any rule that ends up with a ranking over alternatives must violate one or more of the just mentioned properties.

Arrow's result has given rise to a vast literature which we shall not go into here. For the purposes of the present paper it is important to point out, however, that Arrow's choice rule concept – a method that results in a (collective) ranking – was eventually replaced by another concept, viz. that of a social choice correspondence which associates with each preference profile and an alternative set a subset of alternatives. In other words, the standard literature settled for a somewhat less demanding output of the choice rule.

While the choice of a winning alternative or – should there be a tie between several alternatives – alternatives has received a lot of attention in the literature, the composition of multi-member bodies has been the focus of relatively few scholarly works. Yet, multi-member bodies exhibit new types of challenges for research that cannot be answered by simply iterating choices of a method that results in one winner. One of these challenges is the notion of representation.

2 Optimal representation

According to a widely held view a voting body is representative to the extent its composition is a "mirror image" of the composition of its constituency. This is, of course, a very crude way of describing the degree of representation. A somewhat more precise way is to require that the distribution of opinions in the representative body is the same as in the electorate. Chamberlin and Courant impose the following requirements on representation [1]. A committee member represents a voter to the extent that

1. the committee member "makes present" the

*This is a considerably abridged and modified version of [4]

voter's opinions in the deliberations that take place within the committee,

2. the committee member is similarly responsive to various kinds of arguments presented in those deliberations as the voter, and

3. the committee member votes in the same way as the voter should the latter be present in the committee.

This is still too vague to provide a direct guideline for designing representative committees. To make the problem tractable we make resort to the standard assumptions of social choice theory mentioned above. More specifically, we assume that the voters are endowed with complete and transitive preference relations over candidates. In other words, rather than determining the degree of representation *a posteriori*, i.e. after the committee has been in office for some time, we assume that the voters know *a priori* which candidates best represent them.

We, thus, assume that the voters have complete and transitive binary preference relations over candidates and that we are to choose a maximally representative committee with k members on the basis of the preference information provided by the voters. Chamberlin and Courant suggest the following procedure. We first generate all k-member committees that can be formed out of the candidate set K. For each possible committee, we, then, compute the number of individuals whose most preferred candidate is present in the committee. Denote this number by n_1. It can obviously be any number between 0 and n, the total number of voters. Then count the number of voters whose first or second preference candidate is present in the committee and denote this by n_2. Continue in this manner until all ranks $1, \ldots, k$ have been considered. We denote the set of all k-member committees by C^k with elements $c_1, c_2, \ldots, c_s$. The value $C(c_i) = \sum_j^k n_j$, for each $i = 1, \ldots, s$ is the indicator of the representativeness of a committee: the higher the value, the better represented are the voters. Clearly, $k \times n$ is the maximum attainable value and is associated with a committee where each voter's first ranked candidate is present. Similarly, 0 is the minimum value of $C(c_i)$. This "worst possible" committee has the distinction that no voter ranks any committee member higher than $k+1$th in his/her ranking.

It turns out that maximizing $C(c_i)$ over all possible k-member committees amounts to maximizing the sum of the Borda scores of committee members. The most representative one-member committee is the one consisting of the candidate with the maximum Borda score. A maximally representative k-member committee, on the other hand, is determined by a modified Borda count as follows. Let us define each voter's representative in a committee as the committee member getting the largest number of Borda points from that voter, i.e. the member ranked highest in the voter's ranking over candidates. Thus, each voter has a representative in each committee. Now, let $B(c_i)$ denote the sum over voters of the Borda points given to their representative in the committee c_i. The most representative committee is then defined as $c = arg\ max_i B(c_i)$, i.e. the committee where the sum of the Borda points given by each voter to his/her representative is maximal. This is, indeed, a modified Borda count since each voter gives only one score, viz. that of his/her representative.

3 Maximizing representation under fuzziness

Consider now the concept of representation in the context of fuzzy individual preference relations. Voter i's preference relation over candidates can be presented as:

$$\begin{array}{cccc} - & r^i_{12} & \ldots & r^i_{1k} \\ r^i_{21} & - & \ldots & r^i_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ r^i_{k1} & r^i_{k2} & \ldots & - \end{array}$$

Consider now voter i and a committee c_t consisting of k candidates as required. We are now primarily interested in finding the members of c_t that best represent i. Denote the set of these representatives by $B(i, c_t)$. Several plausible ways of finding the best representatives can be envisioned:

1. $B^i_{\text{sum}}(c_t) = \{j \in c_t | \sum_l r_{jl} \geq \sum_l r_{ql}, \forall q \in c_t\}$,
2. $B^i_{\text{min}}(c_t) = \{j \in c_t | \min_l r_{jl} \geq \min_l r_{ql}, \forall l \in K, \forall q \in c_t\}$,
3. $B^i_h(c_t) = \{j \in c_t | h(j) \geq h(q), \forall q \in c_t\}$ where $h(j) = p\ (\max_l r_{jl}) + (1-p)(\min_l r_{jl})$,
4. $B^i_{\text{cop}}(c_t) = \{j \in c_t | cop(j) \geq cop(q), \forall q \in c_t\}$ where $cop(j) = |\{l \in c_t | r_{jl} > r_{lj}, \forall l \in K\}|$

The first one determines the best representatives on the basis of the sums of the preference degrees obtained by candidates in all pairwise comparisons. This method is very much in the spirit of the Borda count. The second method looks at the minimum preference degree of each candidate when compared with all others and picks the candidate with the largest minimum. It is a variant of the min-max method in social choice theory. The third method is a version of Hurwicz's

rule which maximizes the weighted sum of the smallest and largest preference degrees [2]. The fourth method is motivated by Copeland's rule in social choice theory. The Copeland winner is the candidate that defeats more candidates than any other candidate. In the setting of fuzzy preference relation $cop(j)$ is the number of candidates in c_s that are less preferred to j than j is preferred to them. In reciprocal preference matrices, $cop(j)$ is simply the number of entries larger than 0.5 on the j'th row.

Each of these methods singles out the best representatives of every voter in any given committee. Since each of the methods is based on a score, we can define a ranking of candidates in accordance with those scores. From the point of view of representation more important is, however, the ranking over committees ensuing from these methods. The most straightforward way to accomplish this is to define the score of committee c_t as follows:

$$S_t = \sum_{i \in N} \sum_{a \in c_t} \sum_{j \in K} r^i_{aj}.$$

Thus, the score of a committee is the sum of values given by voters to each of its members. The values, in turn, are the sums of preference degrees in all pairwise comparisons. This method is a variation of the Borda count. The most representative committee RC^B would then be:

$$RC^B = \{c_i \in C^k | S_i \geq S_j, \forall c_j \in C^k\}.$$

Although the Chamberlin-Courant approach is very close to the Borda count as well, the above method is not its most plausible fuzzy counterpart. Rather than summing the preference degrees over alternatives and voters, the Chamberlin-Courant approach sums the Borda scores of each voter's representative in any given committee. First we define

$$r^i_j = \sum_{q \in K} r^i_{jq}.$$

Then, for each committee c_t we define:

$$V_{it} = max_{j \in c_t} r^i_j.$$

This can be viewed as the value of the committee c_t to voter i as reflected by the value i assigns to his/her representative in c_t.

Now, the most representative committee in the sense of Chamberlin-Courant is:

$$RC^{CC}_{\text{sum}} = \{c_j \in C^k \mid \sum_i V_{ij} \geq \sum_i V_{iq}, \forall c_q \in C^k, i \in N, j \in K\}.$$

The RC^{CC}_{Sum} committee thus defined is based on the summation of preference degrees in individual preference matrices. In analogous manner one can define the most representative committee in the min-max sense. Let $\underline{r^i_j} = min_{q \in K} r_{jq}$. Now define, for each committee c_t and each voter i:

$$V'_{it} = max_{j \in c_t} \underline{r^i_j}.$$

Then the most representative committee in the min-max sense is:

$$RC'^{CC}_{\text{min}} = \{c_j \in C^k \mid \sum_i V'ij \geq \sum_i V'_{iq}, \forall c_q \in C^k\}.$$

The RC'^{CC}_{min} differs from the previous committee in using the min-max calculus to determine each voter's representative. In a way, RC'^{CC}_{min} mixes two kinds of maximands: the "utilitarian" and "Rawlsian". The former maximizes the average utility, while the latter maximizes the utility of the worst-off individual [5].

A purely Rawlsian committee can also be envisioned. This is obtained as follows:

$$RC^R = \{c_j \in C^k \mid min_i V'_{ij} \geq min_i V'_{iq}, \forall c_q \in C^k\}.$$

In similar vein, one can define Hurwicz and Copeland committees, RC^H and RC^{Co}, respectively. For a fixed value of $p^i \in [0,1]$, let $r^{iH}_j = p^i(\max_q r_{jq}) + (1 - p^i)(\min_q r_{jq})$ and $V^H_{it} = max_{j \in c_t} r^{iH}_j$. The set of most representative Hurwicz-type committees is, then:

$$RC^H = \{c_j \in C^k \mid \sum_i V^H_{ij} \geq V^H_{iq}, \forall c_q \in C^k\}.$$

Note that the value p^i is voter specific measure of his/her "optimism", i.e. the weight assigned to $\max_j r^i_{ij}$, i.e. the degree of preference assigned to each candidate in the comparison of its weakest competitor. Intuitively speaking the exclusive emphasis on strongest and weakest pairwise comparisons is somewhat questionable in voting contexts.

To define, the Copeland-type committee, let RC^{Co}, in turn, is based on the voters' value function $r^{iCo}_j = | \{q \in K \mid r_{jq} > r_{qj}\} |$ and the value function $V^{iCo}_{it} = max_{j \in c_t} r^{iCo}_j$. Now,

$$RC^{Co} = \{c_j \in C^k \mid \sum_i V_{ji}^{iCo} \geq \sum_i V_{qi}^{iCO}, \quad \forall c_q \in C^k\}.$$

Of these four types of committees, the Rawlsian and Copeland types utilize the least amount of the voter preference information. The former looks at the minimal level preference of each candidate when compared with all others. The latter uses only the order information of preference degrees. Of course, if the aim is to economize on information usage, the very idea of resorting to fuzzy preference degrees loses much of its appeal.

4 Concluding remarks

The computational complexity issues notwithstanding the design of representation maximizing committees turns on the fundamental question: given a complete and transitive preference relation over candidates, how to determine the "winner"? This question has dominated much of the applied social choice literature, especially after Donald Saari presented a strong case for the Borda count which for a long time was considered inferior to Condorcet extension methods, e.g. Copeland's or Nanson's rule [6, 7, 3]. When the individual preferences are fuzzy we have several alternative ways of defining the winner, i.e. the best candidate in the voter's view. Some of these ways extend Condorcet's notion of winning into fuzzy environments, while others are more in accordance with Borda's views. Given the completeness of the individual preference relation, we can not only define the winner but also the ranking of the candidates. These rankings can then be used in defining the maximally representative committee.

References

[1] J. Chamberlin and P. Courant, Representative deliberations and representative decisions: Proportional representation and the Borda rule, The American Political Science Review 77 (1983) 718–733.

[2] J. Milnor, Games against nature, in: R. Thrall, C. Coombs and R. Davis (eds.) *Decision Processes*. Wiley, New York, 1954.

[3] H. Nurmi, Comparing Voting Systems, Dordrecht, D. Reidel 1987.

[4] H. Nurmi and J. Kacprzyk, Political representation: a perspective from fuzzy systems theory, New Mathematics and Natural Computation, forthcoming 2007.

[5] J. Rawls, A Theory of Justice, Oxford University Press, Oxford, 1971.

[6] D. Saari, Basic Geometry of Voting, Berlin-Heidelberg-New York, Springer Verlag, 1995.

[7] D. Saari, Which is better: The Condorcet or Borda winner?, Social Choice and Welfare 26 (2006) 107–129.

Fuzziness - Representation of Dynamic Changes ?

Witold Kosiński
Polish-Japanese Institute
of Information Technology
Research Center
ul. Koszykowa 86, 02-008 Warsaw, PL
wkos@pjwstk.edu.pl

Piotr Prokopowicz
Kazimierz Wielki University in Bydgoszcz
Institute of Environmental Mechanics
and Applied Computer Science
ul. Chodkiewicza 30, 85-064 Bydgoszcz, PL
piotrekp@ukw.edu.pl

Abstract

The paper brings a discussion about the source of the inaccuracy in observations of objects and demonstrates that the essential reason of the lack of precision is changeability, and the more changeability, i.e. more dynamics, can be experienced the more inaccurate, more fuzzy judges can be. The space of ordered fuzzy numbers (OFN), the new model of fuzzy numbers that make possible to deal with fuzzy inputs quantitatively, exactly in the same way as with real numbers, is shortly presented. The new model possesses a set of properties which are in accordance with the influence of changeability on the increase of the inaccuracy in observations of the environment. The use of OFN is getting rid of the main problem in a classical fuzzy numbers - an unbounded increase in inaccuracies with next calculations. Moreover, new interpretation can be treated as an extend of classic proposals so there is no need to abandon existing ideas to deal with the new model of fuzzy numbers..

Keywords: Source of uncertainty, Fuzzy number, Ordered fuzzy number, Interpretation of fuzzy numbers, Algebraic operations.

1 Introduction

Fuzzy concept have been introduced in order to model such vague terms as observed values of some physical or economical terms, like pressure values or stock market rates, that can be inaccurate, can be noisy or can be difficult to measure with an appropriate precision because of technical reasons. In our daily life there are many cases that observations of objects in a population are fuzzy.

Discussion about the source of that inaccuracy is an aim of this publication. Authors want here to demonstrate that the essential reason of lack of precision in world's observing is changeability and the more changeability can be experienced the more inaccurate (more fuzzy) judges can be. Authors are introducing the new model of fuzzy numbers [9],[10],[11] defined by themselves together with Dominik Ślęzaka. The new model possesses a set of properties which are in accordance with the influence of the changeability on the increase of the inaccuracy in observations of the environment. Interesting thing is that the new interpretations supplied by the new model can be treated as an extend of classic proposals so we do not need abandon existing ideas to deal with new ones. Beside a little bit of different interpretation, the new model of fuzzy numbers has a lot of useful mathematical properties, in the particular we are getting rid of the main problem in a classical fuzzy numbers - the unbounded increase of inaccuracies with next calculations. Moreover, thanks to the new attempt we can define new methods - based on the arithmetic of ordered fuzzy num-

bers - of processing information in processes dealing with fuzzy control [14],[15].

2 Changes as source of uncertainty

We can ask a question: *which kind of person is an expert?* A possible answer seems obvious - he/she is a specialist in solving some kind of problems which can be described by a set of parameters. Those parameters should be at least in a number of few variables, in other way he/she could solve only one unique problem and it could be difficult say about him/her - the expert. So we can say: *more solvable problems with more variables and with wider ranges of values the person can describe, the better expert he/she is.* In fact if the one is a high class expert then he/she probably does not call a changeable his/her common situations, however another non-expert will see many changes around on the expert place. Point is the *changes* in this article should be treated relatively, not only in straight meaning of word changes. Now we can analyze some examples.

Let us imagine a situation, in which Mr. D. - an expert in assessing the distance - came on picnic out of the city. Let us establish, that while resting on the grass he has a good view on the nearby valley, in which a supermarket was built and many people are arriving for shopping. There is a crossroad with a quite

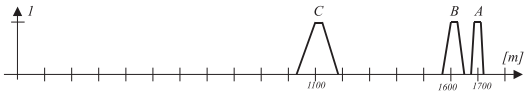


Figure 1: Assessments of Mr.D.

busy way at the end of the valley, and the majority of customers must stop there before living the valley. Observing cars which are starting from the parking lot Mr. D. can very accurate (the more accurate, the better expert he is) assess how long the road distance they must pass before reaching the crossroad. Now let us suppose a fuzzy number A (Fig.1) represents his assessment.

However, Mr. D.'s assessment of the distance from the place a given car stars to the crossroad becomes less precise when the car is in motion. The cause is the dynamics of the observed car. Faster the car drives, the less certain assessment is. Now let us allow fuzzy numbers B and C to represent the opinion about the distance in the tenth and twentieth seconds of observation of the moving car. It is of course pre-arranged script of assessments, however, intuitively the majority of people will confirm the fact that "fuzziness" of consecutive numbers should increase, at least till the moment of reaching the monotonous speed of the observed phenomenon.

Let us elaborate the example. Let us suppose the Mr. D. is great enjoying the picnic in the company of his friends and Mr. V. - an expert in assessing a velocity of moving objects. Mr. V. is observing the valley and he is able to describe with a high precision the speed of monotonously moving lorry, and this represents a fuzzy number S. However, the certainty of his assessment is less when he is trying to establish a velocity of a motorbike which is overtaking the lorry; in this case he gives a fuzzy number T. Moreover, if the motorbike all the time is speeding up and then slowing down overtaking next vehicles on the road, the precision of the assessment of Mr. V. is smaller and smaller. This represents a fuzzy number U.

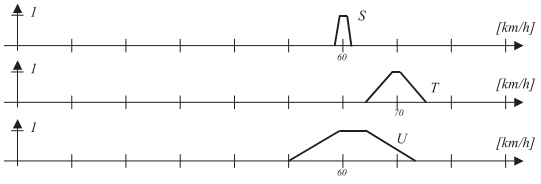


Figure 2: Assessments of Mr.V.

Alike as in the case of Mr. D. in the moment when well identified situations (i.e. monotonous speed of the object) begin to change, the uncertainty of assessments of Mr. V. is growing.

One can look for different examples showing the more changeable situations in which the uncertainty (as well as fuzziness) of assessments is growing. They could concern very different situations e.g. the teacher

does not have a problem with assessing the pupil if his progress for the entire semester is monotonously growing, however, when the pupil once writes a very good work, another time a very crummy one, so in the course of the semester, the justice assessment is difficult and doubts can easily appear. Another example refers to prices of shares on stock exchange. When changes are very dynamic even the best experts will have some difficulties in assessing and a large portion of the uncertainty of their predictions will appear.

Perhaps one should not regard dynamics of changes in observed parameters as the only source of uncertainties, however, we can see that it obviously influences the precision of experts' assessments. One can give some reasons for linking uncertainty, and inaccuracy with dynamics of changes. Certainly one of them is rather the imprecise term: *now*. Since it is very hard for people to determine the exact moment of carrying the assessment out. Very notion *now* is a very inaccurate term. Sometimes it is indicating the given second, other time an hour and yet another time can mean even years (especially at economic assessments). Every *change* has a specific property which is a direction. In next part of this publication a new model of fuzzy numbers will be introduced - the *ordered fuzzy numbers*. They form a good tool to represent the imprecision understood exactly as a result of changes observed in values of parameters.

3 Critiques of convex fuzzy numbers

As long as one works with fuzzy numbers that possess continuous membership functions the two procedures: the extension principle and the α-cut and interval arithmetic method give the same results (cf. [1]) as far as their arithmetic. However, approximations of fuzzy functions and operations are needed if one wants to follow the extension principle and stay within (L, R)–numbers. It leads to some drawbacks as well as to unexpected and uncontrollable results of repeatedly applied operations [16].

Classical fuzzy numbers are very special fuzzy sets defined on the universe of all real numbers. If for a fuzzy set A defined on reals $\mathbf{R}$, we call

- the core of A as the (classical) set of those $x \in \mathbf{R}$ for which its membership function $\mu_A(x) = 1$, and
- the α-cut of A as a (classical) set $A[\alpha] = \{x \in \mathbf{R} : \mu_A(x) \geq \alpha\}$, for each $\alpha \in [0, 1]$, and
- the support of A as the (classical) set $\text{supp } A = \{x \in \mathbf{R} : \ \mu_A(x) > 0\}$,

then we are ready to define the so-called convex fuzzy numbers as those fuzzy sets A's on $\mathbf{R}$ that satisfy three conditions (compare [1],[2],[3],[13],[16]): a) the core of a fuzzy number A is nonempty, b) α-cuts of A are closed, bounded intervals, and c) $\text{supp}A$ is bounded. Since no assumption about continuity of the membership function μ_A of the fuzzy number has been made all crisp numbers are fuzzy numbers, as well.

The results of multiply operations on the convex fuzzy numbers are leading to the large grow of the fuzziness, and depend on the order of operations since the distributive law, which involves the interaction of addition and multiplication, does not hold there.

In this paper we will repeat our main arguments presented in the series of papers [7],[8],[9],[10],[11],[14],[15], that lead to a generalization of the classical concept of fuzzy numbers and then to new definition of ordered fuzzy numbers and their algebra which brings an evolutionary algorithm making possible its determination.

4 Inverse representation of membership functions

Our main observation made in [8] was: a kind of quasi-invertibility of membership functions is crucial and one has to define arithmetic operations on their inverse parts to be in agreement with operations on the crisp real numbers. Consequently, assuming this, the invertibility of membership functions of convex

fuzzy number A makes it possible to define two functions a_1, a_2 on $[0,1]$ that give lower and upper bounds of each α-cut of the membership function μ_A of the number A

$$A[\alpha] := \{x \in \mathbf{R} : \mu_A(x) \geq \alpha\} = [a_1(\alpha), a_2(\alpha)], \tag{1}$$

where boundary points are given for each $\alpha \in [0,1]$ by

$$a_1(\alpha) = \mu_A|^{-1}_{incr}(\alpha) \text{ and } a_2(\alpha) = \mu_A|^{-1}_{decr}(\alpha) \ . \tag{2}$$

In (2) the symbol $\mu_A|^{-1}_{incr}$ denotes the inverse function of the increasing part of the membership function $\mu_A|_{incr}$, the other symbol refers to the decreasing part $\mu_A|_{decr}$ of μ. Then we can see that the membership function μ_A of A is completely defined [1] by two functions $a_1 : [0,1] \to \mathbf{R}$ and $a_2 : [0,1] \to \mathbf{R}$. In terms of them arithmetic operations on the set of fuzzy numbers are defined [1],[2],[13],. For example: if A and B are two (convex) fuzzy numbers with the corresponding functions a_1, a_2 and b_1, b_2 for A and B, respectively, then in terms of their α-cuts the result $C = A + B$ of addition is defined as follows:

$$C[\alpha] = A[\alpha] + B[\alpha], \ \alpha \in [0,1] \ , \tag{3}$$

$$C[\alpha] = [a_1(\alpha) + b_1(\alpha), \ a_2(\alpha) + b_2(\alpha)].$$

For subtraction, however, according to the interval arithmetic [5] the difference $D = A - B$ is defined

$$D[\alpha] = [a_1(\alpha) - b_2(\alpha), \ a_2(\alpha) - b_1(\alpha)], \ \alpha \in [0,1] \ . \tag{4}$$

Notice, that in subtraction of the same fuzzy number A, i.e. for $C = A - A$, we get $C[\alpha] = [a_1(\alpha) - a_2(\alpha), a_2(\alpha) - a_1(\alpha)]$ which represents non-crisp, fuzzy zero, unless $a_1(\alpha) = a_2(\alpha)$ for each α.

However, when the classical denotation for independent and dependent variables of the membership functions, namely x and y is used, and we look once more at (1)-(2), and if we put $y = \alpha$ and use x for the denotation of values of the functions a_1 and a_2, then we will get for two "wings" of the graph of A possible representations:

$$x = a_1(y) \text{ and } x = a_2(y) \ , \ y \in [0,1] \ . \tag{5}$$

In what follows we will use the approach (5) in the representation of so-called ordered fuzzy numbers identified with pairs of continuous functions of the interval $[0,1]$.

5 Ordered fuzzy numbers

In the series of papers [7],[6],[9],[10],[11],[12], [14],[15] we have introduced and then developed main concepts of the space of ordered fuzzy numbers. In our approach the concept of membership functions has been weakened by requiring a mere *membership relation*. Following our observations made in section 4 a fuzzy number A will be identified with the pair of functions a_1 and a_2 (cf. (1) - (2)) defined on the interval $[0,1]$, i.e.

Definition 1. *By an ordered fuzzy number A we mean an ordered pair of two continuous functions*

$$A = (x_{up}, x_{down})$$

called the up-branch and the down-branch, respectively, both defined on the closed interval $[0,1]$ with values in $\mathbf{R}$.

Notice, however, that in our definition we do not require that two continuous functions are inverse functions of some membership function. Moreover, in general a membership function corresponding to A may not exist.

The continuity of both parts implies their images are bounded intervals, say UP and $DOWN$, respectively (Fig. 2a)). If we used the symbols $UP = [l_A, 1^-_A]$ and $DOWN = [1^+_A, p_A]$ to mark boundaries and add the third interval $CONST = [1^-_A, 1^+_A]$, then we can see that are in fact three subintervals appearing in splitting the support of each convex fuzzy number, discussed above. Notice that in general neither $l_A \leq 1^-_A$ nor $1^+_A \leq p_A$ must hold (i.e. $x_{up}(1)$ does not need to be less than $x_{down}(1)$). In this way we can reach improper intervals, which have been already discussed in the framework of the extended interval arithmetic by Kaucher in [4] and called by

[1] The boundary points of the core of A, i.e. the set on which the membership function attains value one, are defined by two values $a_1(1)$ and $a_2(1)$.

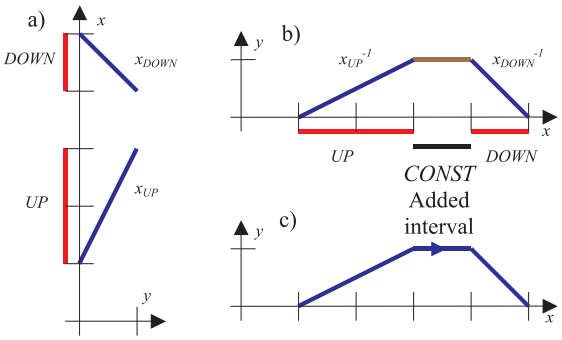


Figure 3: a) Ordered fuzzy number, b) Ordered fuzzy number with membership function, c) Arrow denotes the order of inverted functions and the orientation.

him directed intervals, i.e. such $[n, m]$ where n may be greater than m.

In general, the functions x_{up}, x_{down} need not to be invertible, if we assume, however, that they are monotonous: x_{up} is increasing, and x_{down} is decreasing, and such that $x_{up} \leq x_{down}$ (pointwise), we may define the membership function $\mu(x) = x_{up}^{-1}(x)$, if $x \in [x_{up}(0), x_{up}(1)] = [l_A, 1_A^-]$, and $\mu(x) = x_{down}^{-1}(x)$, if $x \in [x_{down}(1), x_{down}(0)] = [1_A^+, p_A]$ and $\mu(x) = 1$ when $x \in [1_A^-, 1_A^+]$.

In this way we have obtained the membership function $\mu(x), x \in \mathbf{R}$. When the functions x_{up} and/or x_{down} are not invertible or the second condition is not satisfied then the membership curve (or relation) can be defined, composed of the graphs of x_{up} and x_{down} and the line $y = 1$ over the core $\{x \in [x_{up}(1), x_{down}(1)]\}$.

It is worthwhile to point out that a class of ordered fuzzy numbers (OFNs) represents the whole class of convex fuzzy numbers ([1],[2],[3],[13],[16]) with continuous membership functions. In Fig. 3 c) to the ordered pair of two continuous functions (here just two affine functions) x_{up} and gx_{down} corresponds a membership function of a convex fuzzy number with an extra arrow which denotes the orientation of the closed curve formed below. This arrow shows that we are dealing with the ordered pair of functions.

A pair of continuous functions (x_{down}, x_{up}) determines different ordered fuzzy number than the pair (x_{up}, x_{down}). Graphically the plotes of (x_{up}, x_{down}) and (x_{down}, x_{up}) do not differ, however, the corresponding curves determine two different ordered fuzzy numbers: they differ by the *orientation* which we have denoted in Fig.3c by an arrow.

The original definition of OFNs from [9] has been recently generalized in [12].

Now, in the most natural way, the operation of addition between two pairs of such functions has been defined as the pairwise addition of their elements. This is exactly the same as the operation defined in Sec. 4 on α-cuts of A and B, cf. (3). As long as we are adding ordered fuzzy numbers which possess their classical counterparts in the form of trapezoidal type membership functions, and moreover, are of the same orientation, the results of addition are in agrement with the α-cut and interval arithmetic. However, this does not hold, in general, if the numbers have opposite orientations, for the result of addition may lead to improper intervals as far as some α-cuts are concerned. In this way we are close to the Kaucher arithmetic [4] with improper intervals.

Definition 2. *Let $A = (f_A, g_A), B = (f_B, g_B)$ and $C = (f_C, g_C)$ are mathematical objects called ordered fuzzy numbers. The sum $C = A + B$, subtraction $C = A - B$, product $C = A \cdot B$, and division $C = A \div B$ are defined by formula*

$$f_C(y) = f_A(y) \star f_B(y)\ , g_C(y) = g_A(y) \star g_B(y) \tag{6}$$

where "$\star$" works for "$+$", "$-$", "$\cdot$", and "$\div$", respectively, and where $A \div B$ is defined, if the functions $|f_B|$ and $|g_B|$ are bigger than zero.

As it was already noticed in the previous section the subtraction of B is the same as addition of the opposite of B, i.e. the number $(-1) \cdot B$.

6 Ordered fuzzy numbers around us

Model of ordered fuzzy numbers provides some interesting properties ([7],[9],[14],[15]), which open new areas for calculating and pro-

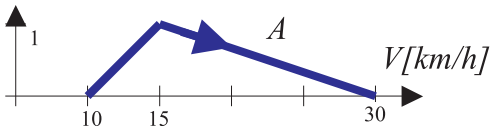


Figure 4: An example of the OFNs describing "slow in speed-up process".

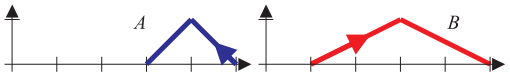


Figure 5: An example of the OFNs that describes income in two units of a financial company.

cessing vague information. Very important for every idea is how it refers to the real life. The interpretation of OFNs together with their orientation will be presented here.

The common use of fuzzy numbers is the presentation and the operation on imprecise data. In general, that is also a source of the idea of all fuzzy sets. Interpretation of the ordered fuzzy numbers is compatible with the general idea of the fuzzy sets. However, there exists a new property - the orientation. By using OFNs we can describe any imprecise value in the real-life processes. The parts up-branch and down-branch of OFN can be related to an opinion of an expert about dynamic changes of the analyzed value. The up-branch describes the behaviour of the value before the very moment when the opinion was made, and the down-branch describes value in afterwards. In that way we expand existing interpretation of fuzzy numbers. We can still use OFNs in the way as usual when we ignore the orientation, but we can also use the orientation to put more complex information about the evaluation made by OFNs.

Let us look at the example in which we have an imprecise opinion "slow" about the speed of a vehicle as OFN A (see Fig. 4). We can ignore the orientation and use this OFN as fuzzy data by saying speed 15 is surely slow and speeds 13 and 20 are slow in degree little more than 50%.We can also take into consideration the orientation of OFN and can say: it is "slow in the speed-up process".

Let us look for another example from the economy and consider a financial company, which has two units A and B. Expert made opinion about the income of both units. For A he said: "income is stated on level 4 millions and this is a downward trend". For B he said: "income is stated on level 3 millions and this is a upward trend". He described incomings of both units by two OFNs (Fig. 5) A and B. By using OFNs the expert can describe not only the value and the trend but also the escalation of that trend.

We have two OFNs where "wide" of branches (up and down) are different. Number B is more "wide" than A. What does it mean? We can find answer if we make more deep (but simply) analysis. If the expert has made up-branch of A from 5 to 4 millions then he considers possible range of changes as 1 million. Up-branch of B was made from 1 to 3 millions so he considers range of changes as 2 millions. To sum up, we understand the number B as an information about a process which us more dynamic than A. Another thing is the direction that shows that A is the decreasing process and B is the increasing one.

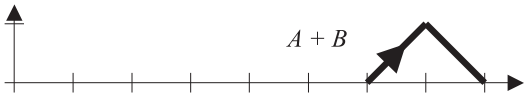


Figure 6: The total income of the company as the sum of A and B.

In real life we could expect total income of analyzed company about 7 millions. Additional, if the increasing process of B was more dynamic than decreasing of A then we expect in total also increasing process, however less dynamic than for B. If we use OFN model and add numbers A and B according to (6) then we get expected results (Fig. 6).

7 Conclusions

The ordered fuzzy numbers form a tool for describing and processing vague information. They expand existing ideas. Their "good" algebra opens new areas for calculations. Beside that, new property (orientation) and its

interpretation presented in this paper can open new areas for using fuzzy numbers. Important fact (in author's opinion) is that thanks to OFNs we can join without complication classical field of fuzzy numbers with new ideas. We can use the OFNs instead the convex fuzzy numbers and if we need to use extended properties we can use them easily. One of directions of the future work with the OFNs are rules in the inference system for a fuzzy controller with new rules. The OFN can contain much more information than the classical fuzzy number - so why do not use it?

References

[1] Buckley James J. and Eslami E., *An Introduction to Fuzzy Logic and Fuzzy Sets*, Physica-Verlag, Springer-Verlag Company, Heidelberg 2005.

[2] Chen Guanrong, Pham Trung Tat, *Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems*, CRS Press, Boca Raton, London, New York, Washington, D.C., 2001.

[3] Drewniak J., Fuzzy numbers (in Polish), in: *Fuzzy Sets and their Applications* (in Polish), J. Chojcan, J. Łęski (Eds.), WPŚ, Gliwice, Poland, 2001, pp. 103–129.

[4] Kaucher E. , Interval analysis in the extended interval space IR, *Computing, Suppl.* **2**,(1980) 33-49.

[5] Kaufman A. and Gupta M. M. , *Introduction to Fuzzy Arithmetic*, Van Nostrand Reinhold, New York 1991.

[6] Koleśnik R., Prokopowicz P., Kosiński W., Fuzzy calculator - useful tool for programming with fuzzy algebra, in: *ICAISC 2004, 7th Int. Conference, Zakopane, Poland, June 2004*, L. Rutkowski, Jörg Siekmann, Ryszard Tadeusiewicz, Lofti A. Zadeh (Eds.) LNAI, vol. 3070, Springer-Verlag, Berlin, Heidelberg, 2004, pp. 32–325.

[7] Kosiński W., On defuzzyfication of ordered fuzzy numbers, in: *ICAISC 2004, 7th Int. Conference, Zakopane, Poland, June 2004*, L. Rutkowski, Jörg Siekmann, Ryszard Tadeusiewicz, Lofti A. Zadeh (Eds.) LNAI, vol. 3070, Springer-Verlag, Berlin, Heidelberg, 2004, pp. 326–331.

[8] Kosiński W., Piechór K., Prokopowicz P., Tyburek K.: On algorithmic approach to operations on fuzzy numbers, in: *Methods of Artificial Intelligence in Mechanics and Mechanical Engineering*, T. Burczyński, W. Cholewa (Eds.), PACM, Gliwice, Poland, 2001, pp. 95–98.

[9] Kosiński W., P. Prokopowicz P., Ślęzak D., Ordered fuzzy numbers, *Bulletin of the Polish Academy of Sciences, Ser. Sci. Math.*, 51 (3), 2003, 327-339.

[10] Kosiński W., P. Prokopowicz P., Ślęzak D., On algebraic operations on fuzzy numbers, in: *Intelligent Information Processing and Web Mining*, Proc. of the International IIS: IIPWM,03 Conference held in Zakopane, Poland, June 2-5,2003, M. Klopotek, S.T. Wierzchoń, K. Trojanowski(Eds.), Physica-Verlag, Heidelberg, 2003, pp. 353-362.

[11] Kosiński W., Prokopowicz P., Ślęzak D., On algebraic operations on fuzzy reals, in: *Advances in Soft Computing*, Proc. of the Sixth Int. Conference on Neutral Networks and Soft Computing, Zakopane, Poland June 11-15, 2002, L. Rutkowski , J. Kacprzyk (Eds.), Physica-Verlag, Heidelberg, 2003, pp.54-61.

[12] Kosiński W. , On fuzzy number calculus, *Int. J. Appl. Math. Comput. Sci.*, **16** (1), (2006) 51–57.

[13] Nguyen H.T. , A note on the extension principle for fuzzy sets, *J. Math. Anal. Appl.* **64**,(1978) 369-380 .

[14] Prokopowicz P., *Algorithmization of Operations on Fuzzy Numbers and its Applications* (In Polish: Algorytmizacja działań na liczbach rozmytych i jej zastosowania), Ph. D. Thesis, IPPT PAN, kwiecień 2005.

[15] Prokopowicz P., Using ordered fuzzy numbers arithmetic, in: *Fuzzy Control in Artificial Intelligence and Soft Computing*, Proc. of the 8th International Conference on Artificial Intelligence and Soft Computing Zakopane, Polska, 2004, June 25-29, 2006, Cader A, Rutkowski L., Tadeusiewicz R., Zurada J. (Eds.), Academic Publishing House EXIT, Warsaw, 2006, pp.156-162.

[16] Wagenknecht M., On the approximate treatment of fuzzy arithmetics by inclusion, linear regression and information content estimation, in: *Fuzzy Sets and their Applications* (In Polish), J. Chojcan, J. Łęski (Eds.), Wydawnictwo Politechniki Śląskiej, Gliwice, 2001, pp. 291-310.

Session 9

The Two Decades of Fuzzy Research at the IPM (Zittau) – Past and Recent Developments –

M. Wagenknecht and R. Hampel

Soft Computing at the Zittau IPM – an overview

Wolfgang Kästner **Rainer Hampel**

University of Applied Sciences Zittau/Görlitz,
Institute of Process Technology, Process Automation and Measuring Technology (IPM)
P.O. BOX 1455
D – 02763 Zittau; Germany
Tel.: +49 (0) 3583 61 1282, Fax: +49 (0) 3583 61 1288

{w.kaestner, r.hampel}@hs-zigr.de

Abstract

The paper gives an overview about the activities in research and development at the University of Applied Sciences Zittau/Görlitz and the IPM-Institute in the field of Soft Computing. The main application fields as well as the main research projects are explained. The motivation is to invite scientists from East and West for further successful collaboration.

Keywords: Soft Computing, Artificial Neural Networks, Fuzzy Modelling, Fuzzy Control.

1 Preamble

Traditionally, research and development (R&D) at the Institute of Process Technology, Process Automation and Measuring Technology (IPM) at the University of Applied Sciences Zittau/Görlitz are focussed on power engineering and power plant technology. Studies are realised not only for conventional power plants also for nuclear power plants (NPP). The R&D projects are financed by industry (IP) as well as by the government (GP). The topics in the nuclear field cover the reactor safety research in Germany and serve the preservation of competence. The results are also applicable to other industrial applications (quality management).

2 Main application fields and projects

The main application fields are:

- Modelling,
- Monitoring,
- Diagnosis,
- Control.

To reach the above mentioned aims soft computing methods are applied or combined with classical algorithms because complexity of the investigated processes (non-linearity, dynamics, ...) makes conventional methods not always efficient.

For about 20 years the development and application of soft computing methods especially Fuzzy Systems and Artificial Neural Networks (ANN) have been one of the main topics in R&D at the University of Applied Sciences and the IPM.

Selected projects:

- Fuzzy-Control of mixture level using conductivity sensors (1992)
- Application of conventional methods and Fuzzy-Logic to improve the hydrostatic water level measurement at pressure vessels (GP: BMFT 1500855/7, 1994; PhD thesis Worlitz, 1993)

- Implementation of Fuzzy Controllers within industrial control and instrumentation systems (IP: Mauell, 1995)

- Combination of conventional observer structures with Fuzzy Algorithms to improve the water level monitoring in pressure vessels (GP: BMFT 1500855/7, 1994; PhD thesis Kästner, 1996)

- Development and realisation of Fuzzy Control conceptions for steam turbine control (IP: Mauell, 1996)

- Fuzzy-Control for optimal operation of a neutralisation facility in a NPP (IP: NPP Brunsbüttel, 1998)

- Combination of conventional and Fuzzy Control to optimise (low-emission) the fire control of conventional steam generator (GP: BMBF 1703898, 2000)

- Control and diagnosis of Magnetic bearings using Fuzzy-Control (GP: BMBF, SMWA, 2002, 2005)

- Application of cluster algorithms for the diagnosis of hydrostatic water level measuring systems (GP: BMWA 1501204, 2003)

- Development of static and dynamic Fuzzy Models to detect the mixture level within pressure vessels during negative pressure gradients (GP: BMWA 1501015, 1999; PhD thesis Traichel, 2005)

- Application of ANN for the modelling of coal-fired steam generators (IP: Vattenfall, 2005)

- Application of ANN and Fuzzy Systems of TSK-type for the modelling of the relationship between water level within pressure vessel and gamma radiation for Boiling Water Reactors (GP: BMWA 1501248, 2007)

- Application of ANN and Fuzzy Systems of TSK-type for the modelling of the relationship between differential pressure and agglomeration of isolation material at strainers (GP: BMWA 1501270, 1501307, 2007)

- Application of Fuzzy Systems of TSK-type for particle classification (GP: BMWA 1501270, 2007)

Besides industrial applications, methodical investigations have been carried out:

- Implementation of a Fuzzy Shell within the simulation tool DynStar (1992)

- Contributions to structure analysis and optimisation of Fuzzy Controllers demonstrated at a turbine control system (PhD thesis Chaker, 1996)

- Comprehension of fuzziness within Probabilistic Safety Analysis (GP: BMWA 1501249, 2002)

- Studies on fuzzy equation systems (Fuzzy Arithmetic) (GP: DFG, 2002)

- Weight analysis of Multilayer Perceptron (MLP) and coefficient analysis of Fuzzy Systems of TSK-type to generate characteristic values for the valuation of model quality (GP: BMWA 1501248, 2004)

- Studies about compact and hierarchical structures of MLP (IP: Vattenfall, 2005)

3 Summary

The close connection between methodical studies (development and methods) and experimental research (design and validation) have been proven to be highly efficient.

Albeit, further methodical research is necessary to improve the acceptance of Soft Computing methods especially in connection with safety-related applications.

This will be a research field in future particularly with regard to the application of digital control and instrumentation systems.

We think the annual Zittau Fuzzy Colloquium offers an excellent panel for discussion about perspectives of Soft Computing particularly concerning R&D in safety-related problems.

Acknowledgement

The grant by the German Government is gratefully acknowledged.

References

[1] F. Worlitz, Anwendung klassischer Verfahren und Fuzzy Logik zur Verbesserung der hydrostatischen Höhenstandsmessung, Technical University Zittau, PhD thesis, 1992.

[2] R. Hampel, N. Chaker, Minimizing the Number of Variable Parameters for Optimizing the Fuzzy Controller, Fuzzy Sets and Systems, 1997.

[3 R. Hampel, M. Wagenknecht, N. Chaker (Eds.), Fuzzy Control: Theory and Practice, Physica-Verlag, Heidelberg; New York, 2000.

[4] R. Hampel, W. Kästner, A. Fenske, B. Vandreier, S. Schefter, Analysis of Selected Structures for Model-Based Measuring Methods using Fuzzy Logic in Da Ruan (Editor), *Fuzzy Systems and Soft Computing in Nuclear Engineering*, Physica Verlag, 2000, pp. 1 – 38.

[5] A. Traichel, W. Kästner, R. Hampel, Implementation of Fuzzy Set Theory in the Field of Energy Processes Applied in Nuclear Power Technology, University of Applied Sciences Zittau/Goerlitz, Institute of Process Technology, Process Automation, and Measuring Technology Zittau, BISCSE 05, Berkeley, USA, 2.-5.11.2005

Experiences in Soft Computing and dynamical simulation

Wolfgang Kästner **Rainer Hampel**

University of Applied Sciences Zittau/Görlitz,
Institute of Process Technology, Process Automation and Measuring Technology (IPM)
P.O. BOX 1455
D – 02763 Zittau; Germany
Tel.: +49 (0) 3583 61 1282, Fax: +49 (0) 3583 61 1288

{w.kaestner, r.hampel}@hs-zigr.de

Abstract

The paper deals with the modelling of dynamical processes based on Soft Computing methods. The structure of a Dynamic Fuzzy Model is illustrated on a linear dynamic system of first order. The influence of impreciseness of the characteristic map is demonstrated on an oscillating non-linear model. At last the results of simulation are valuated.

Keywords: Soft Computing, Fuzzy Modelling, Dynamic Systems.

1 Preamble

Soft Computing algorithms like Fuzzy Systems (Mamdani, TSK) and Artificial Neural Networks (ANN) are efficient methods for the modelling of relationships between input and output variables of complex non-linear processes. The modelling of dynamic processes based on Soft Computing algorithms can be realised by a combination of Soft Computing Model (SCM) and functional blocks which consider the dynamics of the process. Structures consist of a static Soft Computing Model and an external integrator whose output variable is fed back as input of the SCM have to been proven. Figure 1 illustrates the structure of the Dynamic Soft Computing Model (DSCM).

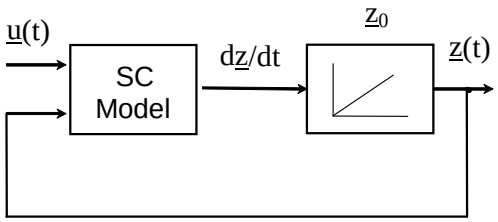


Figure 1: Structure of Dynamic Soft Computing Model.

In the following sections selected aspects concerning the modelling by Dynamic Soft Computing Models are analysed. The study bases on a presetting of an analytical dynamic model and its reproduction by a Dynamic Soft Computing Model. The valuation of the DSCM quality is realised in comparison to the analytical model (ideal result). For the purpose of visualisation the analysis is restricted to low-dimensional systems.

2 Linear Dynamic Model

Based on a linear differential equation of first order (1) a DSCM is designed. As Soft Computing Model a Mamdani Fuzzy System is applied (Dynamic Fuzzy Model).

$$\dot{z}(t) = a \cdot z(t) + b \cdot u(t) \qquad (1)$$

Inputs of the Fuzzy Model are the input variable u(t) and the state variable z(t), the output is

given by the time derivation of the state variable dz/dt.

The state variable z(t) and the input variable u(t) are represented by 3 Fuzzy Sets on the interval [-1, 1], the time derivation dz/dt is represented by 5 Fuzzy Sets on the interval [-1, 1]. Linear membership functions, the fuzzy operators S2 and T2 and the Singleton defuzzification method are used.

The linear characteristic map of the Fuzzy Model is shown in Figure 2. The slope of the map is characterized by parameters a and b.

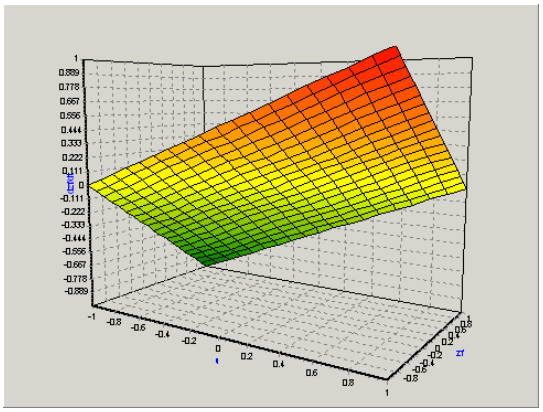


Figure 2: Characteristic map of Fuzzy Model (stable system).

Figure 3 presents the time response of the state variable z(t) (analytical model) and $z_f(t)$ of the Dynamic Fuzzy Model (DFM) for a sinusodial input variable u(t). The error between the state variable of the analytical model and the DFM e_z is negligible.

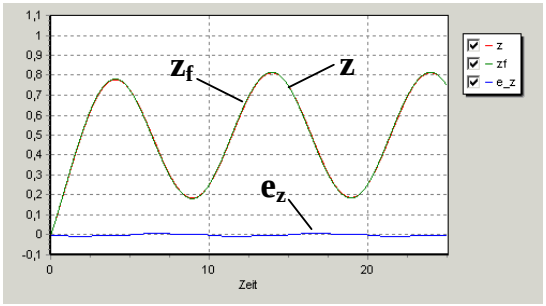


Figure 3: Time response of z(t), $z_f(t)$ and error e_z for a sinusodial input u(t) (stable system).

Figure 4 illustrates the embedding of the trajectory dz/dt = f (z, u) in the linear characteristic map.

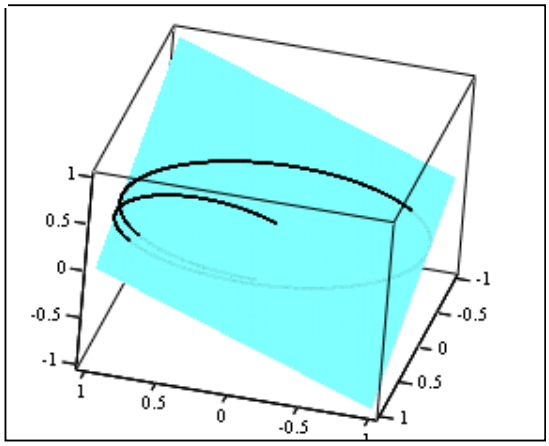


Figure 4: Embedding of trajectory within the characteristic map.

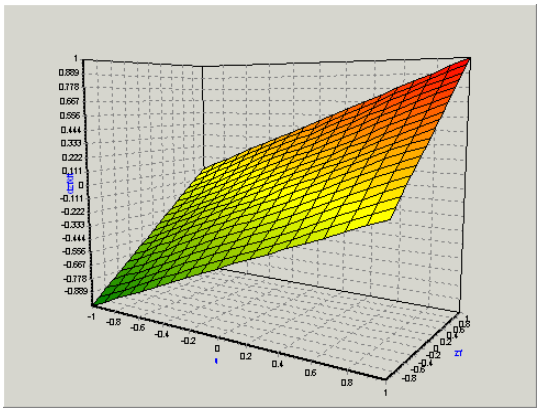


Figure 5: Characteristic map of Fuzzy Model (unstable system).

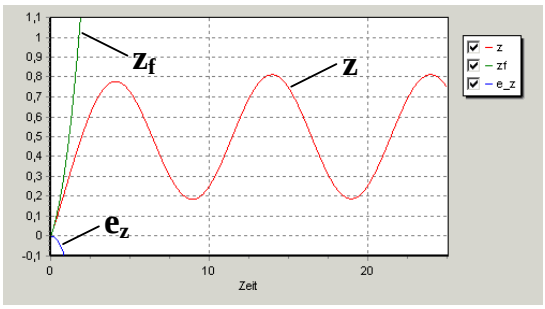


Figure 6: Time response of z(t), $z_f(t)$ and error e_z for a sinusodial input u(t) (unstable system).

The assignment of the rule matrix has an influence on the orientation of the characteristic map and is an essential criterion for the stability of the model. Figure 5 shows the mirrored map resulting from a modified assignment of the rule matrix. The time response of the state variable $z_f(t)$ of DFM is characterised by instability (Figure 6).

3 Non-linear Dynamic Model

Based on a non-linear dynamic model a Dynamic Fuzzy Model is designed. It has to be shown that seeming negligible impreciseness of the fuzzy characteristic map leads to a significant difference between the ideal model (analytical model) and the reproduction based on DFM. This can be observed particularly for models characterised by oscillating time response of state variables.

To demonstrate this effect a model for the description of a Predator-Prey-relationship is used. The model is defined as an autonomous system with the state variables X (prey) and Y (predator). The model structure (2) as well as the parameterisation (3) is designed with the aim to create a system at the stability limit. The time response of the state variables is characterised by an oscillation with constant amplitude.

$$\dot{X}(t) = a \cdot X(t) - b \cdot X(t) \cdot Y(t) - e \cdot X(t)^2$$

$$\dot{Y}(t) = c \cdot X(t) \cdot Y(t) - d \cdot Y(t) \qquad (2)$$

with

$$a = b = = c = d = 1 \qquad e = 0$$

$$X_0 = Y_0 = 1.5 \qquad (3)$$

The state variables X and Y are represented by 5 Fuzzy Sets on the interval [0, 2]. The time derivations of state variables dX/dt and dY/dt are represented by 13 Fuzzy Sets on the interval [-2, 2]. Linear membership functions, the fuzzy operators S2 and T2 and the Singleton defuzzi-fication method are used.

The characteristic maps dX/dt = f (X, Y) and dY/dt = f (X, Y) are shown in Figure 7 and 8.

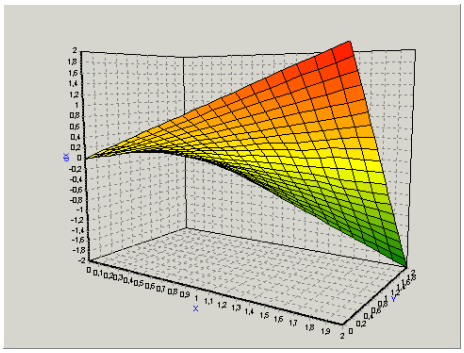


Figure 7: Characteritic map of Fuzzy Model dX/dt = f (X, Y).

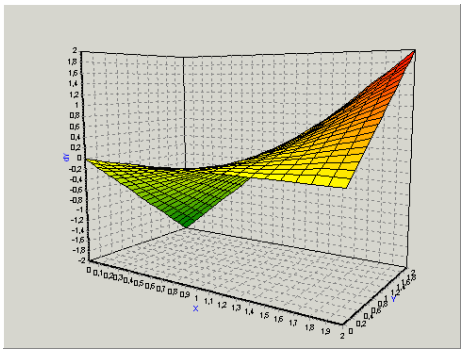


Figure 8: Characteristic map of Fuzzy Model dY/dt = f (X, Y).

Figure 9 presents the time response of the state variable X (prey) for the analytical model X(t) and the DFM $X_f(t)$.

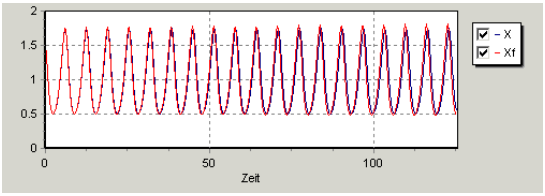


Figure 9: Time response of state variables X(t) and $X_f(t)$.

It can be recognised that the difference between the state variable X(t) (analytical model) and $X_f(t)$ (DFM) increases with time. This effect can be also identified for the state variable Y (predator). There is a difference between the amplitudes observable as well as a phase shift.

Figure 10 illustrates the appropriate state errors e_X and e_Y between the state variables of the analytical model and the DFM.

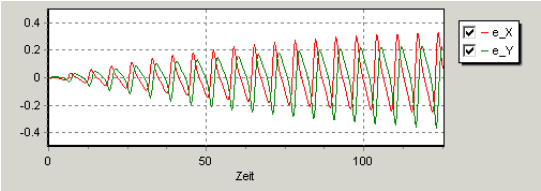


Figure 10: Time response of the state errors e_X und e_Y.

An analysis of sensibility leads to the result that a variation of the:

- number of Fuzzy Sets
- shape of membership functions
- kind of Fuzzy Operators
- kind of Defuzzification Method

has no influence on the result.

As a reason for the increasing error small differences between the characteristic maps of analytical model and DFM is detected.

Figure 11 exemplarily shows the characteristic map of the error $e_{dX} = f(X, Y)$ between the analytical model and Dynamic Fuzzy Model for the relationship $dX/dt = f(X, Y)$. The error ranges between [-0.02, 0.02], i. e. ±1% of the time derivation range dX/dt.

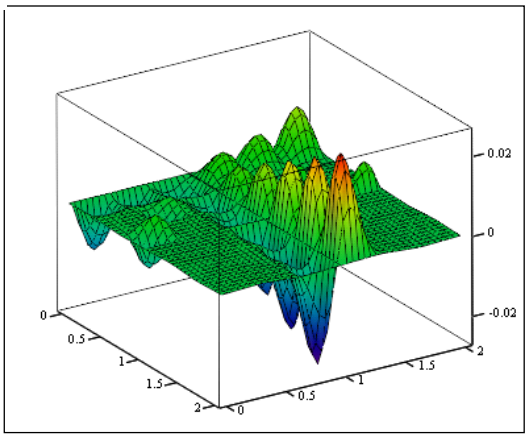


Figure 11: Characteristic map of error e_{dX} between analytical model and DFM for dX/dt.

Normally, the model quality of the DFM can be evaluated as “good” or “very good”.

It turned out that the oscillating time response of the state variables is significantly influenced by the map error. The trajectory of state variable passes through the map regions which are inaccurate. As a result the difference between the ideal model (analytical model) and the DFM increases more and more. Figure 12 illustrates the embedding of the error trajectory in the error map $e_{dX} = f(X, Y)$ (situation after 50 s simulation time).

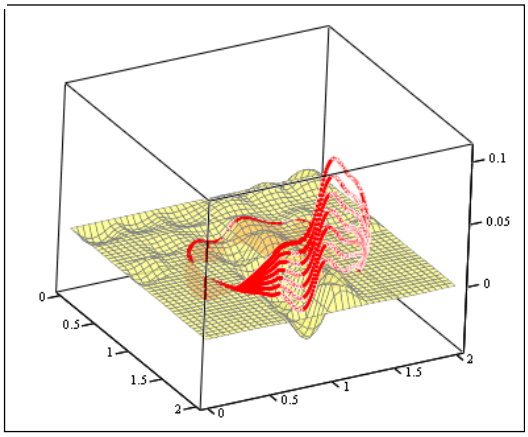


Figure 12: Error trajectory embedded in the error map for dX/dt after 50 s simulation time.

It can be recognised that the cyclic passing through the inaccurate map regions leads to a increasing error of the state variables.
Long-term simulation (simulation time 1000 s) shows that the error of the state variables e_X and e_Y is characterised by periodic variations. The errors e_X and e_Y reach the magnitude of the wanted signal (Figure 13).

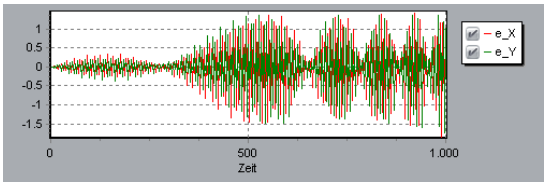


Figure 13: State errors e_X and e_Y between analytical model and DFM for long-term simulation.

Further studies show that this effect is independent from the applied Soft Computing method (Fuzzy system of Mamdani-type, Fuzzy system of TSK-type, Multilayer Perceptron).

4 Summary

Admittedly, the investigated case of a dynamical model at the stability limit represents an extreme example. However the described effects can also be observed for oscillating systems converging to a point of equilibrium.

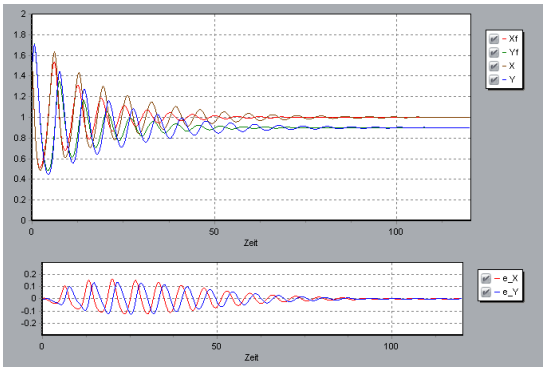


Figure 14: Time response of state variables X, Y (analytic model) and X_f, Y_f (DFM) as well as state errors e_X and e_Y .

Figure 14 shows the time response of state variables X, Y (analytic model) and X_f, Y_f (DFM) as well as state errors e_X and e_Y for a design of the Predator-Prey-Model in form of a fixed-point attractor (e = 0.1).

The simulation confirms that in the case of oscillating state variables small deviations within the characteristic map leads to significant differences between DFM reproduction and ideal model (analytical model). This has to be taken into consideration for the modelling of dynamical processe based on Dynamic Soft Computing Models.

Acknowledgement

The grant by the German Government is gratefully acknowledged.

References

[1] R. Hampel, A. Fenske, W. Kästner, Model-based and knowledge-based measuring methods for the observation of non-linear processes. in Proceedings 5th Zittau Fuzzy Colloquium, Sept. 4-5, 1997, pp. 118-127, Zittau 1997.

[2] A. Traichel, W. Kästner, R. Hampel, Fuzzy Modelling of dynamic non-linear processes – applied for water level measurement. in Proceedings 7th Zittau Fuzzy Colloquium, Sept. 8-10, 1999, pp. 119-134, Zittau 1999.

[3] R. Hampel, A. Traichel, W. Kästner, N. Chaker, System Control and Diagnosis using Soft Computing Methods. Presentation at the conference “Control 2003”, 22.10.-24.10.2003, Moscow, MPEI.

[4] W. Kästner, Dynamisches Fuzzy Modell (Nachbildung linearer dynamischer Strukturen 1. und 2. Ordnung auf der Grundlage von Fuzzy Modellen). Tutorial, University of Applied Sciences Zittau/Görlitz, IPM, 2007 (in German).

[5] W. Kästner, Nachbildung des klassischen RB-Modells (Grenzzyklus) mittels eines Dynamischen Fuzzy Modells (DFM) – Sensibilitätsanalyse. Arbeitsbericht, University of Applied Sciences Zittau/Görlitz, IPM, 2007 (in German).

Modelling the differential pressure at sieves with artificial neural networks (multilayer perceptron) – a contribution to reactor safety research

Alexander Kratzsch **Wolfgang Kästner** **Rainer Hampel**

University of Applied Sciences Zittau/Goerlitz,
Institute of Process Technology, Process Automation and Measuring Technology (IPM)
P.O. BOX 1455
D – 02763 Zittau; Germany
Tel.: +49 (0) 3583 61 1282, Fax: +49 (0) 3583 61 1288

{akratzsch, w.kaestner, r.hampel}@hs-zigr.de

Abstract

In this contribution we describe the modelling of the differential pressure behavior of isolation materials at a sieve by artificial neural networks (ANN). The subject arranges itself in the area of the reactor safety research. Compared with [3] the number of the inputs for the connection which can be modelled was increased. Thereby the number of necessary connections is reduced and the model quality is improved.

Keywords: artificial neural network, reactor safety research.

1 Preamble

One of the main features in reactor safety research is the safe heat dissipation from the reactor core and the reactor containment of light-water reactors. In the case of loss of coolant accident the possibility of the entry of isolation material into the reactor containment and the building sump of the reactor containment and into the associated systems to the residual heat exhaust is a serious problem. That can lead to a handicap of the system functions. To ensure the residual heat exhaust if necessary the emergency cooling systems are put in operation, which transport the water from the sump to the condensation chamber and directly to the reactor pressure vessel. A high allocation of the sieves with fractionated isolation material, in the sump, can lead to the blockage of the sieves, inadmissibly increased pressure build-up at the sieves and to malfunctioning pumps.

2 Modelling the differential pressure at sieves with ANN

2.1 Basics

Our goal is to determine the differential pressure Δp in dependency on the mass allocation MB and the flow rate v at the sieves. For this in Germany a lot of experimental investigations were performed. For the modelling the following nonlinear relationship is indicated.

$$\Delta p = f\left(MB, v, T, L\right), \tag{1}$$

i. e. the differential pressure Δp depends on the mass allocation MB, the flow rate v, the coolant temperature T and a characteristic length L.

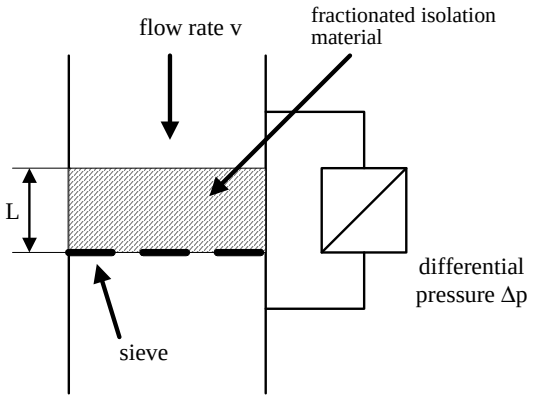


Figure 1: Schematic representation of the filter bed

For the characteristic length L the filter bed depth is used (Figure 1)

$$\varepsilon = 1 - \frac{m}{\rho \cdot A \cdot L}, \tag{2}$$

finally leading to

$$\Delta p = f\left(MB, v, T, \varepsilon\right), \tag{3}$$

where L is substituted by the porosity ε of the filter bed.

2.2 The databases

To build an artificial neural network [1] from the available experimental data a database is created, which is used as basis for the training of the ANN. Figure 2 shows the typical process of a differential pressure experiment with stepwise increase of the flow rate v.

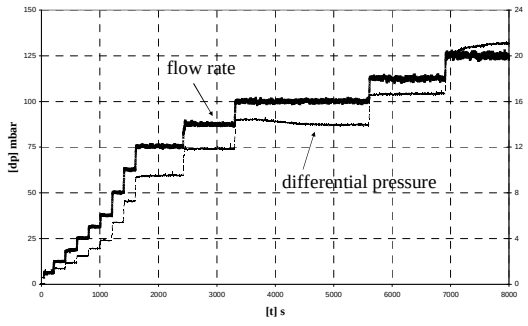


Figure 2: Differential pressure process as a function of the flow rate with stepwise increase of the flowing speed

From „the pressure – time“ process and from „flow rate – time“ process at stationary values of the flow rate v the corresponding stationary values for the differential pressure Δp are determined and gathered in a database $D^{\langle G \rangle}$. The isolation materials type and geometry of the sieve were constant during the test series. The variation of the mass allocation MB, flow rate v and the coolant temperature T took place as below indicated:

$$MB = \{2,4,6,8,10,12\}\frac{kg}{m^2},$$

$$v = \{0,1,2,3,4,5,6,8,10,12,14,16,18,20\}\frac{cm}{s},$$

$$T = \{25,45,70\}°C.$$

2.3 The modelling procedure

The existing database $D^{\langle G \rangle}$ is split into two databases. The training database $D^{\langle 1 \rangle}$ is used for the training of the artificial neural network. With test database $D^{\langle 2 \rangle}$ the trained network is tested. Table 1 shows the structure of the databases.

Table 1: Structure of the databases

	$D^{\langle G \rangle}$	$D^{\langle 1 \rangle}$	$D^{\langle 2 \rangle}$
Number of samples	233	143	90
v [cm/s]	0,1,2,3, 4,5,6,7, 8,10,12, 14,16,18 ,20	1,2,5,8, 12,16,18, 20	3,4,6,10,14,
MB [kg/m²]	2, 4, 6, 8, 10, 12		
T [°C]	35, 45, 70		

As transfer function the following piecewise Parabola – function [2] is used (equation (4), Figure 3).

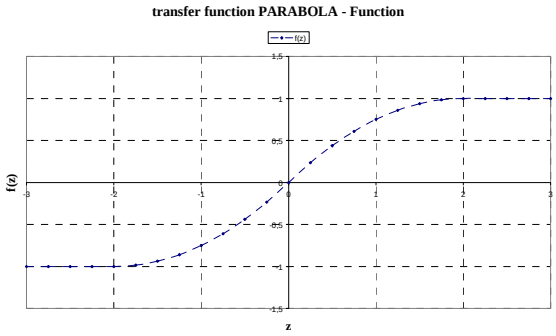


Figure 3: Parabola – Function [2]

$$f(z)=\begin{cases}-1 \text{for } z<2;\\ \frac{1}{4}\cdot(z+2)^2-1 \text{ for } -2\le z\le 0;\\ 1-\frac{1}{4}\cdot(z-2)^2 \text{ for } 0\le z\le 2;\\ 1 \text{ for } z>2.\end{cases} \quad (4)$$

At the beginning an oversized net is designed. First training runs are carried out with activated Pruning. Subsequently, a neural net is available, which contains the necessary number of connections, to successfully reproduce the interrelationships. The final artificial neural network possesses the following architecture.

Table 2: Network architecture

Number of neurons in the input layer	4
Number of neurons in the first hidden layer	9
Number of neurons in the second hidden layer	3
Number of neurons in the output layer	1

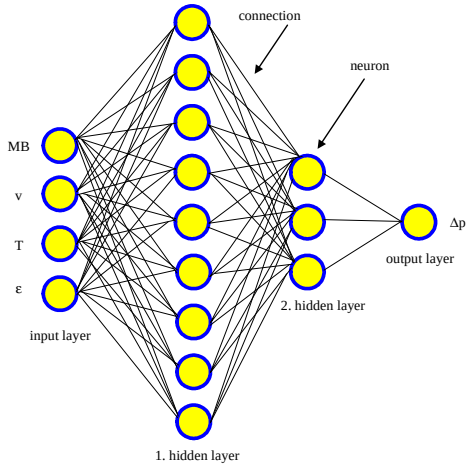


Figure 4: Schematic representation of the artificial neural network

2.4 Results

For the successful reproduction of the 143 training data samples 66 connections are necessary. The trained artificial neural network reproduces the training database $D^{\langle 1\rangle}$ with a relative maximum error of 2.94 % and a mean relative error of 0.51 %. Figure 5 shows the result for the training data record in dependence on the data samples.

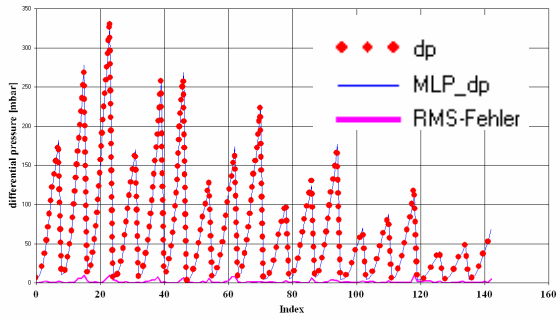


Figure 5: Training result

The test data basis $D^{\langle 2\rangle}$ is reproduces with a relative maximum error of 1.37 % and a mean relative error of 0.33 %. Figure 6 show the result for the test data record in dependence of the data samples.

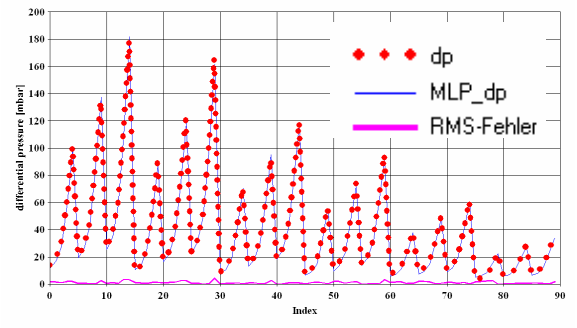


Figure 6: Test result

By adjusting the ANN weights we obtain a satisfactory model for the reproduction of the connection between mass allocation *MB*, flow rate *v*, temperature *T* and porosity ε.

3 Summary

The developed NN is able to model the input – output connection with good accuracy. The model is not limited to the development environment. It can be applied in different simulation programs.

Acknowledgement

The grant by the Vattenfall Europe Nuclear Energy is gratefully acknowledged.

References

[1] A. Zell, Simulation Neuronaler Netze (in German), R. Oldenbourg Verlag, 1997.

[2] DataEngine, Überblick und Benutzerdokumentation (in German), MIT GmbH, 2001.

[3] W. Kästner, T. Förster, A. Kratzsch, R. Hampel, Modelling of the differential pressure behavior of MD2 with multilayer perceptrons and Takagi Sugeno Models. 13th Zittau Fuzzy Colloquium, 2006.

On the reliability of multistate systems with imprecise probabilities

M. Wagenknecht, U. Gocht
University of Applied Sciences Zittau/Goerlitz, IPM
Theodor-Koerner-Allee 16, 02763 Zittau, Germany
{m.wagenknecht,u.gocht}@hs-zigr.de

Abstract

We consider the computation of multistate systems reliabilities in the presence of random set estimations for the elements' working abilities. It turns out that the Dempster-Shafer (DS) approach is a suitable mathematical tool. For the case that the interdependence of the elements is unknown, bounds for the system's performance belief and plausibility functions are given as well.

Keywords: multistate systems, reliability, DS approach.

1 Introduction

Consider a system $\mathbf{\Sigma}$ with n components $E_1,...,E_n$ (e.g., parallel, serial, etc...). The performance of each component is described by $x_i \in L_i$ for $i = 1,...,n$ with L_i being a complete lattice. Moreover, let $\boldsymbol{x} = (x_1,...,x_n)$. These are the basics for a rather general mathematical model of multistate systems where performance often means "working ability".

In applications, the L_i are usually finite sets (e.g. nonnegative integers) or real numbers from [0,1]. The system's performance is computed via the structure function $\Phi(\boldsymbol{x})$ (see Def. 2.1). Concerning the elements performance, it is assumed that $p_i(x_i)$, i.e. the probability (density) for x_i taking values from L_i is known. (Thus the performance of E_i can be interpreted as a random variable on the states of E_i with range L_i.) This, however, may be unrealistic, because the available information for E_i often concerns regions of performances rather than single values.

Take for example L_i = [0,1]. Then the performance of E_i might be characterised by the statement "the probability of high performance is medium", "mean performance is likely", "low performance is not very probable". These linguistic statements are vague and one could try to grasp notions like "high", "medium", etc. by fuzzy sets on L_i (for the performance) and on [0,1] (for the probabilities). For the sake of lucidity we will, however, assume the performance regions to be crisp subsets of L_i and the probabilities to be crisp numbers. Thus we are led to classical the Dempster-Shafer Theory (DST).

Another problem concerns the correlation of the elements with respect to their performance. The assumption often made is that the elements behave independently, what is not always the case. Here, estimations for dependent elements are necessary.

2 Mathematical Prerequisites

Suppose to be given a system $\mathbf{\Sigma}$ with the above properties. Then the Cartesian product $P = L_1 \times ... \times L_n$ is a complete lattice as well, and we obviously have $\boldsymbol{x} \in P$. Further, let L be another complete lattice. We suppose all lattices to be bounded, i.e. for any of them there exist largest and smallest elements which we uniformly denote by 0 and 1. For the different partial orders within the lattices we always use "≤". The following definitions are well-known [3].

Definition 2.1. Let $\Phi: P \to L$ be an isotonic (non-decreasing) function (with respect to the partial order in P) with $\Phi(0,...,0)=0$, $\Phi(1,...,1)=1$. We call Φ the structure function of **Σ**.

Definition 2.2. Let Ω be a sample space, P be a probability measure defined on a suitable σ-algebra over Ω (e.g. the set of all subsets of Ω). Further, let be given a system of sets $\mathfrak{D}$ (σ-algebra) and a set-valued function (random set) $X: \Omega \to \mathfrak{D}$. Then we define for any set $A \in \mathfrak{D}$ the function $m_X: \mathfrak{D} \to [0,1]$ by

$$m_X(A) = P(\omega : X(\omega) = A) \qquad (1)$$

where problems of measurability are left outside for simplicity. The lower index "X" will be omitted if misinterpretation is impossible. The system

$$\{A_1,...,A_N\}$$

with $A_i \in \mathfrak{D}$ is called focal (w. r. to X) if all A_i are nonempty, the mass assignments $m(A_i)$ are positive for all i and the normalisation condition

$$\sum_i m(A_i) = 1$$

is fulfilled. Hence, the random set X can be given by

$$\{(A_1; m(A_1)),...,(A_N; m(A_N))\}.$$

Now we define the functions *bel*, *pl* (belief, plausibility) : $\mathfrak{D} \to [0,1]$ by

$$bel(A) = \sum_{A_i \subseteq A} m(A_i), \quad pl(A) = \sum_{A_i \cap A \neq \varnothing} m(A_i). \quad (2)$$

Obviously, $bel(A) \le pl(A)$. We emphasise that the elements of $\mathfrak{D}$ may intersect. This is typical for situations with incomplete information. Presentations (1) and (2) are generalisations of the classical random variable which is recovered for atomic A_i (i.e., they are pairwise disjoint and $A_i \cap A \neq \varnothing$ implies $A_i \subseteq A$).

Next we need the following generalisation of DST to functions of random variables.

Definition 2.3. Suppose to be given M random sets X_i with ranges $rg(X_i) \in \mathfrak{D}_i$ characterised by focal elements $\{A^i_{k_i}\}$ and corresponding mass assignments $\{m^i_{k_i}\}$; $i = 1,...,M$. Here, $m^i_{k_i} = m(A^i_{k_i})$. Further, let be given a function

$$f: \mathop{\mathbf{X}}_{i=1}^{M} rg(X_i) \to \mathfrak{D},$$

where $\mathfrak{D}$ is a suitable σ-algebra and **X** means the Cartesian product. Then we get the induced random set $Y = f(X_1,...,X_M)$ with focal elements $B_{k_1...k_M} = f(A^1_{k_1},...,A^M_{k_M})$ and given mass assignments

$$m_{k_1...k_M} = m(B_{k_1...k_M}) = P(X_1 = A^1_{k_1},...,X_M = A^M_{k_M}).$$

Notice that the entity $\{m_{k_1...k_M}\}$ is not necessarily normalised, because some of the $B_{k_1...k_M}$ may happen to be empty thus being excluded from further consideration. Hence, a normalisation should be performed in those cases and we may assume the above entity to be normal.

Now, for any $B \in \mathfrak{D}$ we get in analogy to (2)

$$bel(B) = \sum_{\substack{k_1,...,k_M \\ B_{k_1,...,k_M} \subseteq B}} m_{k_1,...,k_M},$$

$$pl(B) = \sum_{\substack{k_1,...,k_M \\ B_{k_1,...,k_M} \cap B \neq \varnothing}} m_{k_1,...,k_M}. \qquad (3)$$

The assumption that the $m_{k_1...k_M}$ are known is rather restricting and may be unrealistic (as in statistics). If the random sets X_i are independent then one can set

$$m_{k_1...k_M} = m^1_{k_1} \cdot ... \cdot m^M_{k_M}.$$

The case that information on X_i originates from several experts leads to Dempster's rule of combination and is considered, e.g. in [5].

In the case that the correlation between $X_1,...,X_M$ is unknown one can derive estimations as solutions of the following optimisation tasks (omitting non-negativity conditions)

$$\sum_{\substack{k_1,\dots,k_M \\ B_{k_1\dots k_M}\subseteq B}} m_{k_1,\dots,k_M} \xrightarrow[\{m_{k_1\dots k_M}\}]{} \min$$

$$\sum_{\substack{k_1,\dots,k_M \\ B_{k_1\dots k_M}\cap B\neq\varnothing}} m_{k_1,\dots,k_M} \xrightarrow[\{m_{k_1\dots k_M}\}]{} \max \qquad (4)$$

$$\sum_{k_1,\dots,k_M}{}' m_{k_1,\dots,k_M} = m^i_{k_i};\ i=1,\dots,M$$

(here, prime means that the *i*th summand is omitted).

Denoting the extremal values of (4) by $\underline{bel}(B)$ and $\overline{pl}(B)$ one gets the obvious inclusion

$$\underline{bel}(B) \le bel(B) \le pl(B) \le \overline{pl}(B). \qquad (5)$$

Remark 2.1. Solving (3) and (4) becomes rather time-consuming for higher dimensions. To keep efforts minimal, one should take sets *B* which are of special interest for the random set *Y*. In practise, often $\mathfrak{D}_i$ and $\mathfrak{D}$ are set systems on the real axis. This may lead to interval computation for (3) and (4). For *B* one can take the set $\mathfrak{I}(z) = \{x\in\mathbb{R} : x \le z\}$ thus obtaining the plausibility and belief distribution functions $\overline{F}$, $\underline{F}$ from

$$\overline{F}(z) = pl(\mathfrak{I}(z)),\quad \underline{F}(z) = bel(\mathfrak{I}(z)). \qquad (6)$$

Example 2.1. Consider two independent random sets $X_1 = \{([0,0.4];0.2),([0.3,0.8];0.67),([0.7,1];0.13)\}$ and $X_2 = \{([0,0.6];0.67),([0.8,1];0.33)\}$ characterising the working ability of the two elements in a serial system.

Hence, we take function *f* as *min* (acting on intervals by bounds). After simple computations we get

$Y = \{([0,0.4];0.2),([0.3,0.6];0.44),([0.3,0.8];0.22),([0.4,0.6];0.09),$
$([0.7,1];0.05)\}$.

Assume we want to know *bel* and *pl* for an "acceptable" work ability of the system characterised by the interval $B = [0.65,1]$. From (3) we get

$bel(B) = 0.05$,

$pl(B) = 0.22+0.05 = 0.27$,

what is not very high, because both systems mainly work at medium level.

Therefore, the question for "medium" working ability given by $B = [0.3,0.6]$ will be answered by $bel(B) = 0.44+0.09 = 0.53$,

$pl(B) = 0.44+0.22+0.09 = 0.75$.

3 Application to System Reliability

In principle, the above apparatus easily applies to reliability determination of multistage systems. The information on the elements performance is given by the random sets X_i with focal elements

$$A^i_{k_i} \subseteq \mathfrak{D}_{L_i}$$

(the latter being a suitable extension of L_i).

The role of the function *f* is now played by the structure function Φ that maps (in analogy to *f*) into $\mathfrak{D}_L$, the latter being the corresponding extension of *L*. Often, the system is a connection of parallel-serial sub-systems what may ease the computation of Φ (e.g. by paths or cuts). A popular choice for L_i and *L* is the unit interval [0,1]. Usually, one aims at computing the probability for a certain minimal level α of the system's performance, it is $\Phi(\boldsymbol{x}) \ge \alpha$. This leads to

$$B_{k_1\dots k_M} = \Phi\left(A^1_{k_1},\dots,A^n_{k_M}\right)$$

whereby the focal elements of X_i may be taken as intervals in the continuous case, i.e.

$$A^i_{k_i} = \left[\underline{a}^i_{k_i}, \overline{a}^i_{k_i}\right].$$

For *B* we take [α,1]. Due to the isotonicity of Φ we get for (3)

$$bel(\alpha) = \sum_{\substack{k_1,\dots,k_n \\ \Phi\left(\underline{a}^1_{k_1},\dots,\underline{a}^n_{k_n}\right)\ge\alpha}} m_{k_1,\dots,k_M},$$

$$pl(\alpha) = \sum_{\substack{k_1,\dots,k_n \\ \Phi\left(\overline{a}^1_{k_1},\dots,\overline{a}^n_{k_n}\right)\ge\alpha}} m_{k_1,\dots,k_M} \qquad (7)$$

where we used $bel(\alpha)$, $pl(\alpha)$ for $bel(B)$, $pl(B)$.

Though (7) is computationally easier to handle than the general task (3), it may be of advantage to decompose the system $\mathbf{\Sigma}$ into smaller parts what is typical for parallel-serial systems. The most elementary subsystems are those consisting of two elements. As a result we obtain random sets describing the behaviour of the subsystems and which can be combined to get the final estimation with respect to (7) or (4).

4 Conclusion

In the present paper we considered possibilities to compute reliabilities of multistate systems in the presence of random set estimations for the elements' working ability (performance). It turned out that the Dempster-Shafer approach is a suitable mathematical tool. For the case that the interdependence of the elements is unknown, bounds for the system's performance belief and plausibility functions are given as well.
From a practical point of view it may be useful to consider fuzzy focal elements and/or fuzzy sets B wich will be a topic for future research. We also refer to [1,2,6] where generalised implication operators are used to characterise the degree of inclusion of fuzzy sets.

References

[1] Th. Denoeux, Modeling vague beliefs using fuzzy-valued belief structures. FSS 116 (2000) 167-199.

[2] I. Kramosil, Belief functions generated by fuzzy and randomized compatibility relations. FSS 135 (2003) 341-366.

[3] A. Lisnianski and G. Levitin, Multi-State System Reliability. World Scientific, N.J., 2003.

[4] H. T. Nguyen and E. A. Walker, A First Course in Fuzzy Logic. Chapman & Hall, Boca Raton , 1999.

[5] G. Shafer, A Mathematical Theory of Evidence. Princeton University Press, Princeton, 1976.

[6] J. Yen, Computing generalized belief functions for continuous fuzzy sets. Int. Journal Appr. Reasoning 6 (1992) 1-31.

Author index

Achs, Á., 221/II
Adlassnig, K. P., 361/I
Agrafonova, T., 377/I
Aguiló, I., 83/I
Alcalde, C., 157/II
Alvoni, E., 185/I
Asmuss, S., 175/II
Averkin, A. N., 377/I, 387/I, 391/I, 395/I

Baczyński, M., 105/I, 111/I
Balas, M. M., 425/I
Balas, V. E., 425/I
Ballester, M. A., 479/II
Ballini, R., 231/II
Baroni, P., 281/I
Barzamini, R., 137/II, 237/II
Běhounek, L., 145/I, 153/I
Belenki, A. G., 387/I
Beliakov, G., 61/I, 67/I
Belluce, L. P., 195/II
Bobillo, F., 213/II
Bodenhofer, U., 153/I, 321/I
Bosc, P., 343/II
Botzheim, J., 361/I
Bou, F., 177/I
Boukezzoula, R., 279/II
Bradley, J., 417/I
Brinker, K., 353/I
Bronevich, A. G., 297/I
Burusco, A., 157/II

Calvo, T., 61/I
Casasnovas, J., 163/II
Chitsaz, M., 237/II
Chojnacki, E., 367/II
Cintula, P., 137/I, 153/I
Cornelis, C., 13/II
Costa, P., 231/II

Čomić, L., 325/II

De Baets, B., 47/I, 95/II
Deschrijver, G., 53/I, 13/II
Destercke, S., 367/II
Detyniecki, M., 345/I
Di Nola, A., 195/II, 207/II
Divari, M., 169/II
Drobics, M., 361/I
Drygaś, P., 185/II
Dubois, D., 367/II
Dudziak, U., 359/II
Durante, F., 185/I, 191/I
Dvořák, A., 103/II
Dyczkowski, K., 129/II

Esteva, F., 177/I, 21/II

Fermüller, Ch. G., 17/I, 437/I
Fiot, C., 329/I
Flaminio, T., 201/II
Foschi, R., 197/I
Foulloy, L., 279/II
Fuentes-González, R., 157/II

Galichet, S., 279/II
García-Lapresta, J. L., 479/II
Garmendia, L., 351/II

Gegov, A., 111/II
Georgescu, I., 437/II, 445/II
Gerla, B., 195/II
Ghiselli Ricci, R., 91/I
Gil-Lafuente, A. M., 463/II
Gispert, J., 19/I
Gobalakrishnan, N., 111/II
Gocht, U., 473/I, 425/II
Godo, L., 177/I, 21/II
Gomide, F., 231/II
Gouveia, F., 243/II
Grabisch, M., 257/I, 289/I, 313/I
Grbić, T., 221/I
Grigolia, R., 207/II
Gusev, A., 395/I

Hájek, P., 21/I
Habiballa, H., 27/II
Hadjali, A., 343/II
Hampel, R., 459/I, 463/I, 469/I
Hashemi Golpayegani, M., 63/II, 143/II, 149/II
Havranová, Z., 331/II, 337/II
Hell, M., 231/II
Hliněná, D., 431/II
Holčapek, M., 409/II
Horčík, R., 137/I
Hüllermeier, E., 353/I
Hürlimann, W., 205/I
Huňka, F., 273/II

Igrejas, G., 243/II
Imai, H., 311/II
Iyer, V., 71/II

Jaafari, A., 63/II, 143/II, 149/II
Jacas, J., 127/I
Jaworski, P., 213/I
Jayaram, B., 105/I, 111/I
Jenei, S., 23/I
Jouzdani, J., 237/II

Kacprzyk, J., 445/I
Kalganova, I., 381/I
Kalina, M., 331/II, 337/II
Kästner, W., 459/I, 463/I, 469/I
Kawaguchi, M. F., 45/II
Kerre, E. E., 13/II, 49/II
Khosravi, A., 137/II, 237/II
Klawonn, F., 321/I
Klement, E. P., 253/I
Kohout, L. J., 161/I, 429/I
Kojadinovic, I., 305/I
Kolesárová, A., 75/I
Kondo, M., 45/II
Kosiński, W., 449/I
Kosterev, V. V., 391/I
Kráľ, P., 431/II, 471/II
Kramosil, I., 375/II
Kratzsch, A., 469/I
Kroupa, T., 145/I
Křivý, I., 251/II

Labreuche, Ch., 289/I, 295/II, 303/II
Lakov, D. V., 79/II
Laurent, A., 329/I
Lebedeva, O., 403/II
Lendelová, K., 231/I
Lepskiy, A. E., 297/I
Limberg, J., 409/I
Lin, R., 437/II
Lu, J., 137/II
Lucas, C., 149/II

Maes, K. C., 47/I
Marchioni, E., 169/I
Mareš, M., 179/II
Marichal, J. L., 305/I
Marsala, C., 345/I
Mas, M., 97/I
Mayor, G., 75/I
Medina, J., 209/II

Merigó, J. M., 463/II
Mesiar, R., 75/I, 253/I
Michalíková, A., 237/I
Mihailović, B., 265/I
Minoux, M., 25/I
Miranda, P., 257/I
Miyakoshi, M., 45/II
Močkoř, J., 417/II
Molina, C., 337/I
Monserrat, M., 97/I
Montagna, F., 201/II
Motie Nasrabadi, A., 63/II, 143/II, 149/II
Murthy, G. R., 71/II

Nasibov, E., 59/II
Navara, M., 87/II
Nedović, L., 221/I
Noguera, C., 21/II
Nosková, L., 95/II
Novák, V., 103/II
Nurmi, H., 445/I

Ohnishi, S., 311/II
Ojeda-Aciego, M., 209/II
Olivas, J. A., 119/II

Pankowska, A., 317/II
Pap, E., 253/I, 265/I
Papini, P. L., 185/I
Pavlačka, O., 455/II
Pelessoni, R., 281/I
Petrík, M., 87/II
Pivert, O., 343/II
Pourmorteza, A., 63/II
Prieto, M. E., 119/II
Prokopowicz, P., 449/I

Qiu, X., 437/II
Quesada-Molina, J. J., 191/I

Radojević, D. G., 119/I
Ralević, N. M., 325/II
Recasens, J., 127/I, 351/II
Riečan, B., 227/I
Riera, J. V., 163/II
Rodríguez-Muñiz, L. J., 381/II
Rossi, G., 273/I
Ruan, D., 27/I, 49/II
Ruiz-Calviño, J., 209/II
Ruzha, V., 175/II

Sadati, N., 237/II
Salgado, P., 243/II
Salimi Khorshidi, G., 63/II, 143/II, 149/II
Saminger, S., 35/I
Sánchez, D., 337/I
Saralieva, M. R., 79/II
Sarkoci, P., 35/I
Seising, R., 401/I, 409/I
Sempi, C., 185/I
Serrano, J. M., 337/I
Shestakov, M., 395/I
Shi, Y., 49/II
Sokolov, O., 425/II
Solovjovs, S., 395/II
Soto, A., 119/II
Spizzichino, F., 197/I
Stankevich, K. V., 391/I
Stefanini, L., 259/II
Straccia, U., 213/II
Suner, J., 83/I

Šabo, M., 41/I
Škrbić, S., 247/I
Šostak, A., 387/II
Špirková, J., 471/II
Štajner-Papuga, I., 221/I
Štěpnička, M., 95/II

Takači, A., 247/I
Talašová, J., 455/II
Teisseire, M., 329/I

Tikhonov, E., 369/I
Titova, N., 377/I
Torrens, J., 83/I, 97/I
Troiano, L., 67/I
Tsoy, Y. R., 267/II
Tvrdík, J., 251/II

Úbeda-Flores, M., 191/I
Uljane, I., 387/II
Ulutagay, G., 59/II

Van Gasse, B., 13/II
Vetterlein, T., 37/II
Vicig, P., 281/I
Vila, M. A., 337/I
Vivona, D., 169/II

Wagenknecht, M., 473/I, 425/II
Watari, O., 45/II
Wygralak, M., 287/II

Xie, L., 313/I

Yamanoi, T., 311/II
Yankovskaya, A. E., 267/II

Zadeh, L. A., 29/I
Zheng, X., 137/II

Žikić-Došenović, T., 241/I