Pardubice, Czech Republic

September 17–20, 2017

Proceedings of the 20th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty

Edited by

Vilém Novák

Masahiro Inuiguchi

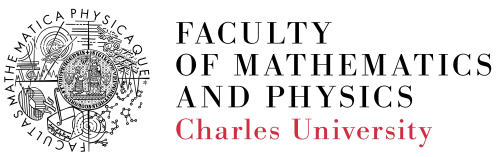
Martin Štěpnička

Organizing institution:

Institute for Research and Applications of Fuzzy Modeling

Co-organizing institutions:

- Faculty of Mathematics and Physics of the Charles University
- Institute of Information Theory and Automation (ÚTIA) of the Czech Academy of Sciences



Title: **Proceedings of the 20th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty**
CZECH-JAPAN SEMINAR 2017, Pardubice, Czech Republic, September 17–20, 2017

Editors: Vilém Novák, Masahiro Inuiguchi, Martin Štěpnička
Publisher: University of Ostrava

ISBN: 978-80-7464-932-5

Committees

Programme Chairs
Vilém Novák (University of Ostrava)
Masahiro Inuiguchi (Osaka University)

Programme Committee:
Ondřej Čepek (Charles University in Prague)
Martin Gavalec (University of Hradec Králové)
Takeshi Itoh (Tohoku University)
Radim Jiroušek (University of Economics, FM Jindrichuv Hradec)
Irina Perfilieva (University of Ostrava)
Jaroslav Ramík (Silesian University in Opava)
Shogo Shiode (Kobe Gakuin University)
Shao Chin Sung (Aoyama Gakuin University)
Jiřina Vejnarová (Czech Academy of Sciences, Prague)
Milan Vlach (Charles University in Prague)

Financial Chair
Martin Štěpnička (University of Ostrava)

Organizing Chair
Petra Števuliáková (University of Ostrava)

Organizing Committee:
Martin Dyba (University of Ostrava)
Petr Hurtík (University of Ostrava)
Vojtěch Molek (University of Ostrava)

Foreword

Dear Colleagues,

with the great pleasure and at the same time with the great surprise I am writing this foreword. Why surprise? Because this is already 20^{th} Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty! Nobody in the beginning expected that we would have succeeded in establishing such a nice and long tradition.

The idea to establish Czech-Japan seminars came from two people — Jaroslav Ramík who in that time was worker of IRAFM (Institute for Research and Applications of Fuzzy Modeling) and Milan Vlach who was professor in JAIST (Japan Advanced Institute for Science and Technology). Other people who were instrumental in establishing this tradition are Masahiro Inuiguchi, Tetsuzo Tanino and Hiroaki Ishii and Junzo Watada of the Osaka University.

The first Czech-Japan seminar took place in March 1999 in JAIST in Hokuriku and the second one in September of the same year in Čeladná — a holiday resort in Beskydy in the Czech Republic. Since then, the seminar took place each year alternately in Japan and in the Czech Republic. During the years, the seminars were organized in various places both Japan as well as in the Czech Republic: in Japan, e.g., in Kyoto, Fukuoka, Nagasaki, Sendai, Osaka; in Czechia, e.g., in Jindřichův Hradec, Valtice, Liblice, Litomyšl, Mariánské Lázně and few other places.

This year, the seminar takes place in the nice town of East Bohemia — Pardubice. The town has a population of more than ninety thousand and is located at the confluence of two rivers, the Labe and Chrudimka, 96 kilometres east of Prague. The town has a long history: the first mention of Pardubice comes from 1295 and the city was founded in 1340. Until 1918, the town was part of the Austrian monarchy. The history of Pardubice is related to its growth as an important industrial town. In 1845, the first train arrived to Pardubice, in 1911, Jan Kašpar made history by flying the first long-haul flight towards Prague. The present factories include the Synthesia chemical factory — manufacturer of the ingloriously known plastic explosive Semtex and oil refinery Paramo.

The Czech-Japan seminars were quite successful both in producing new scientific results and in putting together an international group of researchers. I wish that the jubilee 20^{th}-meeting will be no exception and hope that all of us meet again next year somewhere in Japan in the 21^{st}-meeting.

Ostrava, September 2017

Vilém Novák
University of Ostrava, IRAFM

Table of Contents

Composition in exponential families: Case of multivariate normal distribution

Vladislav BÍNA

Faculty of Management, University of Economics in Prague
Jarošovská 1117/II, 37701, Jindřichův Hradec
Czech Republic
bina@fm.vse.cz

Abstract

The paper presents an application of operator of composition into the field of multivariate continuous distributions, namely employs the properties of an important class of exponential families. Thanks to the fact that the exponential families are closed with respect to the operations of marginalization, conditioning and multiplication it is possible to define the operator of composition in general. After outlining of general framework author presents an application of composition in case of two bivariate normal distributions.

Keywords: Operator of composition, Continuous random variable, Exponential family, Multivariate normal distribution.

1 Introduction

The presented paper presents a contribution to the field of theory of compositional models, namely the issue of composition of multidimensional densities of continuous random variables. As we already presented in paper [4], the approach is a general alternative to the copula approach based on Sklar's famous theorem (see [12]) elaborated later into the form of a vine approach, e.g., [1]. For a basic discussion of the properties, advantages and disadvantages of the compositional model approach and copula based approaches, see [5] and [2].

The paper recalls a general definition of an operator of composition in continuous case (already introduced in [2]) and the results presented in [4] for case of general exponential families. The basic aim is to provide a first step for building of multidimensional compositional models from low-dimensional normal distributions.

2 Preliminaries and basic notions

Throughout the presented paper we consider a finite index set $N = \{1, \dots, n\}$ together with a set of random variables $\{X_i\}_{i \in N}$ with values, or vectors of values, denoted by the corresponding lowercase letters. The domain of variables will be denoted by the corresponding bold uppercase letter $\mathbf{X}_i$. In general, variables with a finite or countable set of possible *states* are called *discrete*; other variables are called *continuous*. In this paper, we will focus on the later case.

The probability density functions of continuous random variables will be denoted by lowercase letters of the Latin alphabet $(f, g, h, \dots)$, e.g., the abbreviated notation $f(x_K)$ denotes a multidimensional density of variables having indices from set $K \subseteq N$. For a probability density function $f(x_K)$ and any set of variable indices $L \subset K$, a marginal probability density $f(x_L)$ of $f(x_K)$ can be computed for each x_L as follows

$$f(x_L) = \int_{\mathbf{X}_{K \setminus L}} f(x_K) dx_{K \setminus L}$$

where obviously the integration runs over the domains of all variables in $K \setminus L$. We will also employ an equivalent way to denote the marginal $f(x_L)$, namely $f^{\downarrow\{L\}}$ which was introduced by Glenn Shafer (see, e.g., [11]).

Having probability density $f(x_K)$ and two disjoint subsets $L, M \subseteq K$ we define the *conditional probability density* of X_L given a value $x_M = \mathbf{x_M}$ for every $x_{L\cup M}$ as

$$f(x_L \,|\, x_M = \mathbf{x_M}) f(x_M = \mathbf{x_M}) = f(x_L, x_M = \mathbf{x_M}).$$

Let us note that for $f(x_M = \mathbf{x_M}) = 0$ the definition is ambiguous, but we do not need to exclude such cases.

Other important properties of composition in continuous case are summarized in [4].

3 Composition in exponential families

As we mentioned above, in the conference paper [2] we presented properties of composition in certain copula class, namely the formulation of the composition for copula densities from the class of strict Archimedean copulas of general dimension. But the resulting expressions are significantly simplified in case of two dimensional copulas, see [3].

The possibility to define the operation of composition for densities of distributions from *exponential families* was studied in [4]. The exponential family is an interesting set of probability distributions that can be expressed in a certain form, e.g., see [7].

Now, let us recall the most important notions and properties introduced in the context of compositional model in [4]. *Density $f(x_L)$ belongs to the exponential family* if it can be expressed in the form

$$f(x_L; \theta_L) = h(x_L) e^{\eta_L(\theta_L)\cdot T(x_L) - A(\eta_L)}$$

where θ_L is a (real) vector of parameters and $h(x_L)$, $T(x_L)$, $\eta_L(\theta_L)$ and $A(\eta_L)$ are vector functions.

The function $\eta_L(\theta_L)$ is a *natural parameter* (or exponential parameter), $T(x_L)$ is a *sufficient statistic*, $A(\eta_L)$ is a log-partition function and $h(x_L)$ is a non-negative base measure. Obviously, the product of $\eta_L(\theta_L)$ and $T(x_L)$ vector functions is a scalar product. Examples of the most important members of the exponential family, such as Gaussian, binomial, multinomial, Gamma and Beta distributions can be found, e.g., in [8].

It can be shown that exponential family is closed with regard to several important operations, particularly product, marginalization and conditioning, see, e.g., Lemmata 6 and 8 in [7].

If both operands belong to the exponential family, the result of operation of composition is defined and can be expressed in the above form and thus also belongs to the exponential family.

Theorem 1. *For two densities $f(x_L)$ and $g(x_M)$ belonging to an exponential family, i.e. such that $f(x_L) = h_L(x_L) e^{\eta_L \cdot T_L(x_L) - A_L(\eta_L)}$ and $g(x_M) = h_M(x_M) e^{\eta_M \cdot T_M(x_M) - A_M(\eta_M)}$ the* composition also belongs to the exponential family*.*

Proof. For disjoint L and M we get the product of both densities, which obviously also belongs to the exponential family.

If the other possibility realizes, i.e. if $L \cap M \neq \emptyset$ then we can express

$$g(x_M) = h_M(x_M) e^{\eta_{L\cap M}\cdot T_{L\cap M}(x_{L\cap M}) + \eta_{M\setminus L}\cdot T_{M\setminus L}(x_{M\setminus L}) - A_M(\eta_{L\cap M}, \eta_{M\setminus L})}.$$

According to [7] the conditional distribution

$$g(x_{M\setminus L} \,|\, x_{L\cap M} = \mathbf{x_{L\cap M}}) = h_{\mathbf{L\cap M}} e^{\eta_{M\setminus L}\cdot T_{M\setminus L}(x_{M\setminus L}) - A_{\mathbf{L\cap M}}(\eta_{M\setminus L})}$$

where $h_{\mathbf{L\cap M}}$ and $A_{\mathbf{L\cap M}}$ are dependent on the values of conditioning variables. It is now apparent that the product of $f(x_L)$ and $g(x_{M\setminus L} \,|\, x_{L\cap M})$ again belongs to the exponential family since it can be written in the corresponding form, i.e.

$$(f \triangleright g)(x_{L\cup M}) = h_L h_{\mathbf{L\cap M}} e^{\eta_L \cdot T_L(x_L) + \eta_{M\setminus L}\cdot T_{M\setminus L}(x_{M\setminus L}) - A_L(\eta_L) - A_{\mathbf{L\cap M}}(\eta_{M\setminus L})}.$$

□

4 Multivariate normal distribution

The non-degenerate multivariate normal distribution has a symmetric and positive definite covariance matrix $\mathbf{\Sigma}$. In such case, the multivariate normal distribution $f(x_L)$ with vector of means μ_L and covariance matrix $\mathbf{\Sigma}_L$ has a density given by formula

$$f(x_L;\mu_L,\mathbf{\Sigma}_L)=\frac{1}{\sqrt{(2\pi)^\ell|\mathbf{\Sigma}_L|}}\exp\left(-\frac{1}{2}(x_L-\mu_L)^{\mathrm{T}}\mathbf{\Sigma}_L^{-1}(x_L-\mu_L)\right)$$

where ℓ is a dimension (length) of x_L vector, symbol $^{\mathrm{T}}$ stands for a vector transpose, $|\mathbf{\Sigma}_L|$ is determinant of covariance matrix and $\mathbf{\Sigma}_L^{-1}$ is an inverse of covariance matrix.

Thus, multivariate density $f(x_L;\mu_L,\mathbf{\Sigma}_L)$ has variables and functions according to definition of exponential family given in the following way

$$\begin{aligned}
x_L &= (x_1,\ldots,x_\ell)^{\mathrm{T}},\\
\eta_L &= \begin{pmatrix} \mathbf{\Sigma}_L^{-1}\mu_L \\ -\frac{1}{2}\mathbf{\Sigma}_L^{-1}\end{pmatrix},\\
T_L(x_L) &= \begin{pmatrix} x_L \\ x_Lx_L^{\mathrm{T}}\end{pmatrix},\\
A_L(\eta_L) &= \frac{1}{2}\mu_L^{\mathrm{T}}\mathbf{\Sigma}_L^{-1}\mu_L+\frac{1}{2}\log|\mathbf{\Sigma}_L|,\\
h_L(x_L) &= (2\pi)^{-\frac{\ell}{2}}.
\end{aligned}$$

4.1 Conditional multivariate density

Let us have a multivariate density $g(x_M;\mu_M,\mathbf{\Sigma}_M)$ and let us divide index set M into two disjoint parts such that $A=L\cap M$ and $B=M\setminus L$. Thus, the m-dimensional vector x_M can be partitioned into two parts of dimensions m_A and m_B ($m_A+m_B=m$) in such a way that

$$x_M=\begin{pmatrix} x_A \\ x_B\end{pmatrix}$$

and similarly

$$\mu_M=\begin{pmatrix} \mu_A \\ \mu_B\end{pmatrix}.$$

The covariance matrix is partitioned into the corresponding blocks in the following way

$$\mathbf{\Sigma}_M=\begin{pmatrix} \mathbf{\Sigma}_{AA} & \mathbf{\Sigma}_{AB} \\ \mathbf{\Sigma}_{BA} & \mathbf{\Sigma}_{BB}\end{pmatrix}$$

having sizes

$$\begin{pmatrix} m_A^2 & m_Am_B \\ m_Am_B & m_B^2\end{pmatrix}.$$

Thus, having the multivariate density $g(x_M)\sim\mathcal{N}(\mu_M,\mathbf{\Sigma}_M)$ the *conditional multivariate density* $g(x_{M\setminus L}\,|\,x_{L\cap M}=\mathbf{a})=g(x_B\,|\,x_A=\mathbf{a})$ is again a multivariate density distribution (see, e.g., [6]) and $g(x_B\,|\,x_A=\mathbf{a})\sim\mathcal{N}(\overline{\mu}_B,\overline{\mathbf{\Sigma}}_B)$ where

$$\overline{\mu}_B=\mu_B+\mathbf{\Sigma}_{BA}\mathbf{\Sigma}_{AA}^{-1}(\mathbf{a}-\mu_A)$$

and

$$\overline{\mathbf{\Sigma}}_B=\mathbf{\Sigma}_{BB}-\mathbf{\Sigma}_{BA}\mathbf{\Sigma}_{AA}^{-1}\mathbf{\Sigma}_{AB}.$$

We can somewhat surprisingly see, that the known value $\mathbf{a}$ influences the mean of conditional density but not its covariance matrix. Let us note that the formula for $\overline{\mathbf{\Sigma}}_B$ is known as the Schur complement of $\mathbf{\Sigma}_{AA}$ in $\mathbf{\Sigma}_M$ and $\mathbf{\Sigma}_{AA}^{-1}$ is a generalized inverse (see again [6]).

4.2 Product of multivariate densities

Similarly, the product of two multivariate normal densities is again multivariate normal distribution (must be then renormalized, see [9]). For two multivariate densities $f(x_L) \sim \mathcal{N}(\mu_L, \mathbf{\Sigma}_L)$ and $g(x_M) \sim \mathcal{N}(\mu_M, \mathbf{\Sigma}_M)$ we get

$$f(x_L)g(x_M) \sim \mathcal{N}(\overline{\mu}, \overline{\mathbf{\Sigma}})$$

where

$$\overline{\mu} = \overline{\mathbf{\Sigma}} \left(\mathbf{\Sigma}_L^{-1} \mu_L + \mathbf{\Sigma}_M^{-1} \mu_M \right)$$

and

$$\overline{\mathbf{\Sigma}} = \left(\mathbf{\Sigma}_L^{-1} + \mathbf{\Sigma}_M^{-1} \right)^{-1}.$$

The normalizing constant is (see [10]) equal to

$$(2\pi)^{-\frac{\ell+m}{2}} |\mathbf{\Sigma}_L + \mathbf{\Sigma}_M|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\mu_L - \mu_M)^{\mathrm{T}} (\mathbf{\Sigma}_L + \mathbf{\Sigma}_M)^{-1} (\mu_L - \mu_M) \right).$$

5 Example of composition

Now, let us set two multivariate densities. First density $f(x_L; \mu_L, \mathbf{\Sigma}_L)$ has variables, parameters and functions given in the following way

$$\begin{aligned}
x_L &= (x_1, x_2)^{\mathrm{T}}, \\
\mu_L &= (1, 0)^{\mathrm{T}}, \\
\mathbf{\Sigma}_L &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \\
\eta_L &= \begin{pmatrix} \mathbf{\Sigma}_L^{-1} \mu_L \\ -\frac{1}{2} \mathbf{\Sigma}_L^{-1} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & -\frac{1}{3} \end{pmatrix}, \\
T_L(x_L) &= \begin{pmatrix} x_L \\ x_L x_L^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \\ x_1^2 & x_1 x_2 \\ x_1 x_2 & x_2^2 \end{pmatrix}, \\
A_L(\eta_L) &= \frac{1}{2} \mu_L^{\mathrm{T}} \mathbf{\Sigma}_L^{-1} \mu_L + \frac{1}{2} \log |\mathbf{\Sigma}_L| = \frac{1}{3} + \frac{1}{2} \log 3, \\
h_L(x_L) &= (2\pi)^{-\frac{\ell}{2}} = (2\pi)^{-1}.
\end{aligned}$$

This corresponds to the two-dimensional density

$$f(x_L) = f(x_1, x_2) = \frac{1}{2\pi\sqrt{3}} \exp\left(\frac{1}{3} \left(2x_1 - x_2 - x_1^2 + x_1 x_2 - x_2^2 - 1 \right) \right).$$

Second density $g(x_M; \mu_M, \mathbf{\Sigma}_M)$ has variables, parameters and functions given by

$$\begin{aligned}
x_M &= (x_2, x_3)^{\mathrm{T}}, \\
\mu_M &= (0, 2)^{\mathrm{T}}, \\
\mathbf{\Sigma}_M &= \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 2 \end{pmatrix}, \\
\eta_M &= \begin{pmatrix} \mathbf{\Sigma}_M^{-1} \mu_M \\ -\frac{1}{2} \mathbf{\Sigma}_M^{-1} \end{pmatrix} = \begin{pmatrix} -\frac{4}{7} & \frac{8}{7} \\ -\frac{4}{7} & \frac{1}{7} \\ \frac{1}{7} & -\frac{2}{7} \end{pmatrix}, \\
T_M(x_M) &= \begin{pmatrix} x_M \\ x_M x_M^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_2 & x_3 \\ x_2^2 & x_2 x_3 \\ x_2 x_3 & x_3^2 \end{pmatrix}, \\
A_M(\eta_M) &= \frac{1}{2} \mu_M^{\mathrm{T}} \mathbf{\Sigma}_M^{-1} \mu_M + \frac{1}{2} \log |\mathbf{\Sigma}_M| = \frac{8}{7} + \frac{1}{2} \log \frac{7}{4}, \\
h_M(x_M) &= (2\pi)^{-\frac{m}{2}} = (2\pi)^{-1}.
\end{aligned}$$

This corresponds to the two-dimensional density

$$g(x_M) = g(x_2, x_3) = \frac{1}{\pi\sqrt{7}} \exp\left(\frac{2}{7}\left(-2x_2 + 4x_3 - 2x_2^2 + x_2x_3 - x_3^2 - 4\right)\right).$$

Now, we need to compute conditional density $g(x_3 \,|\, x_2 = \mathbf{b})$. And since we have A given by variable x_2 and B by variable x_3, according to Subsection 4.1 parameters and functions given by

$$\begin{aligned}
\overline{\mu}_3 &= \mu_B + \mathbf{\Sigma}_{BA}\mathbf{\Sigma}_{AA}^{-1}(\mathbf{a} - \mu_A) = 2 + \frac{b}{2},\\
\overline{\mathbf{\Sigma}}_3 &= \mathbf{\Sigma}_{BB} - \mathbf{\Sigma}_{BA}\mathbf{\Sigma}_{AA}^{-1}\mathbf{\Sigma}_{AB} = \frac{7}{4},\\
\eta_3 &= \begin{pmatrix} \overline{\mathbf{\Sigma}}_3^{-1}\overline{\mu}_3 \\ -\frac{1}{2}\overline{\mathbf{\Sigma}}_3^{-1} \end{pmatrix} = \begin{pmatrix} \frac{8+2b}{7} \\ -\frac{2}{7} \end{pmatrix},\\
T_3(x_3) &= \begin{pmatrix} x_3 \\ x_3^2 \end{pmatrix},\\
A_3(\eta_3) &= \frac{1}{2}\mu_3^{\mathrm{T}}\mathbf{\Sigma}_3^{-1}\mu_3 + \frac{1}{2}\log|\mathbf{\Sigma}_3| = \frac{(b+4)^2}{28} + \frac{1}{2}\log\frac{7}{4},\\
h_M(x_M) &= (2\pi)^{-\frac{1}{2}}.
\end{aligned}$$

This corresponds to the conditional density

$$g(x_3 \,|\, x_2 = \mathbf{b}) = \frac{1}{\sqrt{\pi\frac{7}{2}}} \exp\left(\frac{8+2b}{7}x_3 - \frac{2}{7}x_3^2 - \frac{(b+4)^2}{28}\right).$$

Finally, the composition is computed as a product of bivariate density $f(x_L)$ and conditional density $g(x_3 \,|\, x_2 = \mathbf{b})$ for all possible values of x_2 inserted.

$$\begin{aligned}
x &= (x_1, x_2, x_3)^{\mathrm{T}},\\
\overline{\mu} &= \overline{\mathbf{\Sigma}}\left(\mathbf{\Sigma}_L^{-1}\mu_L + \mathbf{\Sigma}_3^{-1}\mu_3\right),\\
\overline{\mathbf{\Sigma}} &= \left(\mathbf{\Sigma}_L^{-1} + \mathbf{\Sigma}_3^{-1}\right)^{-1}.
\end{aligned}$$

The result corresponds to the three-dimensional density

$$\begin{aligned}
&h(x_1,x_2,x_3) =\\
&\quad = \frac{1}{2\pi\sqrt{42\pi}} \exp\left(\frac{1}{3}\left(2x_1 - x_2 - x_1^2 + x_1x_2 - x_2^2 - 1\right) + \frac{8+2x_2}{7}x_3 - \frac{2}{7}x_3^2 - \frac{(x_2+4)^2}{28}\right).
\end{aligned}$$

6 Conclusion

Throughout the paper author focuses on the continuous random variables and important subclass of distributions, so called exponential families. A general definition of operator of composition and its properties are in a way restricted within the framework of exponential families. But thanks to the fact that exponential families are closed with respect to the operations necessary for definition of composition it is feasible to define the operator of composition.

After discussion of general properties author presents an example of particular operations leading to the composition of pair of bivariate normal densities. This can be easily generalized and it is feasible to perform the operations automatically within suitable computational environment (capable of matrix manipulations).

Acknowledgement
The research was supported by the National Science Foundation of the Czech Republic (GA ČR) under grant no. 15-00215S.

References

[1] K. Aas, C. Czado, A. Frigessi, and H. Bakken, Pair-Copula Constructions of Multiple Dependence. *Insurance: Mathematics and Economics*, **44**(2):182–198, 2009.

[2] V. Bína, An Operator of Composition for the Multivariate Copula Construction. In *AIP Conference Proceedings*, **1636**, MaxEnt'13, pages 5–10, Canberra, Australia, 2014.

[3] V. Bína, Operator of Composition for Continuous Copulas. In: *Proceedings of the 16th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, pages 62–69, Mariánské Lázně, 2013.

[4] V. Bína, Composition operator for densities of continuous variables. *International Journal of Approximate Reasoning*, Available online 25 January 2017, DOI: 10.1016/j.ijar.2017.01.008.

[5] V. Bína and R. Jiroušek, A Short Note on Multivariate Dependence Modeling. *Kybernetika*, **9**(3):420–432, 2013.

[6] M. L. Eaton, *Multivariate Statistics: a Vector Space Approach*, John Wiley and Sons, 1983.

[7] E. L. Lehmann, *Testing statistical hypotheses*, 2nd Edition. Springer Texts in Statistics, Springer, 1986.

[8] F. Nielsen and V. Garcia, *Statistical exponential families: A digest with flash cards*, 2009. CoRR abs/0911.4863. URL http://arxiv.org/abs/0911.4863

[9] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Technical University of Denmark, 2012.

[10] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, the MIT Press, 2006.

[11] G. Shafer, Probabilistic Expert Systems. *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics, Philadelphia, USA, 1996.

[12] A. Sklar, Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**:229–231, 1959.

Causality and Intervention in Business Process Management

Vladislav BÍNA[1] and Radim JIROUŠEK[1,2]

[1] *Faculty of Management, University of Economics, Prague*
Jarošovská 1117/II, 69121 Jindřichův Hradec
Czech Republic
E-mail: bina@fm.vse.cz

[2] *Inst. of Information Theory and Automation AV ČR*
Pod Vodárenskou věží 4, Praha 8
Czech Republic
E-mail: radim@utia.cas.cz

Abstract

Business processes, as well as various other fields of management and decision-making, involves causal relations. The paper presents an algebraic approach to the modeling of causality in systems of stochastic variables. The methodology is based on an operator of a composition that provides the possibility of composing a multidimensional distribution from low-dimensional building blocks taking advantage of the dependence structure of the problem variables. The authors formally define and demonstrate on a hypothetical example a surprisingly elegant unifying approach to conditioning by a single variable and to the evaluation of the effect of an intervention. Both operations are realized by the composition with a degenerated distribution and differ only in the sequence in which the operator of the composition is performed.

Keywords: Compositional models, operator of composition, causality, conditioning, intervention.

1 Introduction

The purpose of this paper is to foster a new method for causality modeling that can be incorporated into standard business process models. Without a doubt, causality and its proper treatment plays an important role in this modeling and may be used to solve a great variety of problems. For example, Feugas, Mosser, and Duchien [3] use it to predict the effect of business process evolution on the quality of service. They show the possibility of coping with one phenomenon that also will be discussed in this paper. Namely, with the fact that the characteristics of the quality of service, such as the response time, are influenced by *hidden factors.*

As Kirkwood argues in his book [9] *human beings are quick problem solvers,* because *quick problem solvers were the ones who survived.* For this, it is necessary to distinguish correctly between causes and effects. However, in many complex situations, it is not a simple task, as business processes are complex systems as a rule, especially when one considers systems with causal loops. To simplify the problem, we consider only the causal relations in this paper that can be well described by Markovian probabilistic models. In fact, we restrict our attention only to situations that are well described by Pearl's causal diagrams [15].

The exact approaches to causality modeling are often applied to the spheres of social sciences, e.g., applications in strategic management by Michael Ryall [16], in psychology by York Hagmayer et al. [4], in educational research and other social sciences [14]. For other examples the reader can also refer to an introduction to causal inference by Morgan and Winship [13] where the applications are aimed at the fields of sociology, political science, and economics.

Though graphical modeling for business process representation has been considered classical from its very beginning (flowcharts in the 20th, PERT in 50th, influence diagrams in 80th, ...), the

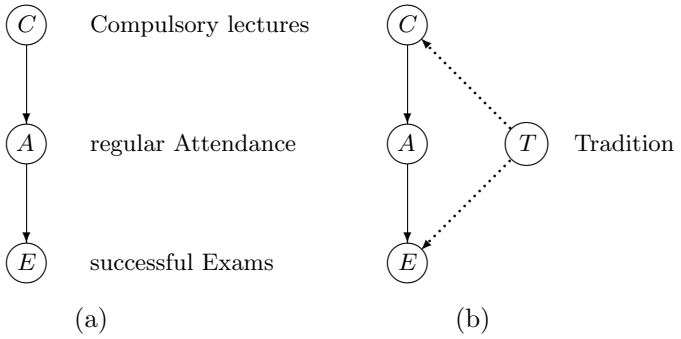


Figure 1: Causal models for the hypothetical example.

application of Pearl's causal diagrams (i.e., employment of directed graphs) to represent causality using arrows pointing from causes to effects, though quite intuitive and generally accepted, may bring some technical complications. Moreover, their advantage is diminishing when the diagrams containing hundreds of nodes are considered. In these situations, one may acknowledge an "algebraic" approach based on compositional models, in which computations of conditioning and interventions are computed in a similar way using the operator of a composition.

The goal of this paper is not to introduce a mathematical theory of compositional models. For this, the reader is referred to [6]. We want to convince the reader that incorporating of causal compositional models into business process models may help them to solve otherwise unsolvable problems. To make the following explanation as simple as possible, we present the results just using a trivial hypothetical example.

1.1 Motivation - Hypothetical Example

Analyzing data from universities, we found out that regular lecture attendance significantly increases the chances of passing exams on the first attempt. In connection with this, we were not surprised that students from universities where attendance at lectures was compulsory were slightly more successful in passing exams than students from universities where attendance was optional. These results led us to construct a simple and "natural" causal model with a causal diagram from Figure 1(a).

After publishing these research results, we succeeded in convincing the public to such an extent that eventually all universities excepted the rules that their lectures henceforth would be compulsory. We were quite satisfied that our research had made a real impact and expected that it would manifest itself in the form of better results at university exams. Unfortunately, the opposite proved to be true; on average, fewer students passed the exam on the first attempt. How could this be possible?

All this could happen because of a wrong causal graph. It was true that the new rules increased the percentage of regular attendance to lectures. It was also true that regular attendants were more successful at exams; yet, the impact of the law was negative because the causal model did not reflect another important cause - the fact that regulations and therefore also the behavior of students is different between traditional universities and new ones. Taking this fact into account, we can construct a new causal diagram (see Figure 1(b)), which describes the situation much better, and which can explain the decrease in success rates at the exams. Unfortunately, when we collected the data we did not recorded the type of university (traditional vs. new one), and therefore we have to treat the respective variable as a "hidden" one, which is why the respective arrows are dotted in Figure 1(b). As it will be shown later, if we had used the second causal model (even if we had to consider the type of university to be hidden - a non-measurable variable) to estimate the impact of an intervention realized by the new regulations, we would have really predicted a slight decline in the number of successful students.

In connection with this failure, one question arises. Was it possible to realize that the causal model from Figure 1(a) did not correspond to the situation described? The answer is positive and will be discussed in more details in Secion 4.

1.2 Causal Models

The purpose why we foster the employment of causal models in business process modeling is evident from the previous hypothetical example (for another example see, e.g., Váchová and Bína [18]). Prior to any serious intervention, the managerial board should take into consideration all possible ways of predicting a potential impact of the intended action.

In [15] (see page 29), Pearl says causal models may help to answer the following three types of queries:

- **predictions/conditioning** - are the students more successful if the lectures are compulsory?
- **interventions** - will the students be more successful if we introduce a regulation making the lectures compulsory?
- **counterfactuals** - would the students be more successful had the lectures been compulsory, given that they are not too successful and the lectures are optional?

In this paper, we will be interested only in the first two tasks: prediction (conditioning) and intervention. Let us stress the difference between them. While prediction gives evidence only about the students from universities where attendance at lectures is compulsory, intervention speaks about the success of all students when we make sure that lectures are compulsory at all universities. In what follows, we will show how to answer these questions with the help of compositional causal models. It will appear that, in contrast to causal diagrams, where the required results are computed in two totally dissimilar ways, we can use the same apparatus for causal compositional models. Syntactically, both these answers can be got as a composition of a model with a degenerated one-dimensional distribution. The only difference is just in a pair of parentheses. Unfortunately (but quite naturally), this pair of parentheses may bring some computational difficulties.

2 Compositional Models

In this paper we consider finite valued variables $A, E, C, T, X, Y, Z, W, ...$ that are denoted by upper case Latin characters. Boldface characters, such as $\mathbf{a}, \mathbf{b}, \mathbf{c}$ denote values, or combination of values of these variables. Sets of these variables are denoted by lower case characters $(x, y, ...)$, and their probability distributions are denoted using characters of a Greek alphabet, such as κ, λ, π. So, $\kappa(X_1, X_2, \ldots, X_n)$ denotes an n-dimensional probability distribution, and $\kappa(\mathbf{a})$ is a value of this distribution for state $\mathbf{a}$, which is a combination (vector) of values of variables $X_1, X_2, \ldots, X_n$. An $(n-1)$-dimensional marginal distribution of κ is denoted by κ^{-X_i}, or, denoting $x = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$, we use also the symbol $\kappa^{\downarrow x}$. The latter symbol $\kappa^{\downarrow y}$ can be used for any $y \subseteq \{X_1, \ldots, X_n\}$.

2.1 Operator of Composition

The term *compositional model* is derived from the fact that these multidimensional models (distributions) are *composed* (assembled) from a system of low-dimensional distributions using an *operator of composition* that was introduced first in [5]. For its rigorous mathematical definition and a detailed survey of its properties see [6]; here we will do just with the following brief summary.

Consider two (non-empty) sets of variables x and y (notice that they may be, but need not to be, disjoint, one may be a subset of the other). Let κ and λ be distributions defined for x and y, respectively, such that the marginal distribution $\lambda^{\downarrow x\cap y}$ dominates $\kappa^{\downarrow x\cap y}$, i.e.,

$$\forall \mathbf{a} \quad \lambda^{\downarrow x\cap y}(\mathbf{a}) = 0 \Longrightarrow \kappa^{\downarrow x\cap y}(\mathbf{a}) = 0.$$

In this case we can define composition of κ and λ by the formula (notice that we define $\frac{0\cdot 0}{0} = 0$)

$$\kappa \triangleright \lambda = \frac{\kappa \cdot \lambda}{\lambda^{\downarrow x\cap y}}.$$

Note that for disjoint x and y the marginal $\kappa^{\downarrow x\cap y} = \lambda^{\downarrow x\cap y} = 1$, and $\kappa \triangleright \lambda = \lambda \triangleright \kappa$ simplifies to a product of (independent) distributions. If $\lambda^{\downarrow x\cap y}$ does not dominate $\kappa^{\downarrow x\cap y}$ then the composition is

not defined. However, to avoid technical problems, let us assume in this paper that whenever we use the operator of composition we assume it is defined.

It is known that the composition of distributions $\kappa(x)$ and $\lambda(y)$ is always a distribution of variables $x \cup y$. In [6], many properties of the operator of the composition are proven. In this paper, we need mainly the following three that are formulated for $\kappa(x)$ and $\lambda(y)$.

$$(\kappa \triangleright \lambda)^{\downarrow x} = \kappa; \tag{1}$$
$$z \supseteq x \cap y \Longrightarrow (\kappa \triangleright \lambda)^{\downarrow z} = \kappa^{\downarrow x \cap z} \triangleright \lambda^{\downarrow y \cap z}, \tag{2}$$
$$x \supseteq y \Longrightarrow \kappa \triangleright \lambda = \kappa; \tag{3}$$
$$z \subseteq x \Longrightarrow \kappa^{\downarrow z} \triangleright \kappa = \kappa. \tag{4}$$

2.2 Causal Compositional Models

Consider variables $\{X_1, X_2, \ldots, X_n\}$. For each variable X_i let $\mathcal{C}(X_i)$ denote the set of the variables that are causes of X_i. Naturally, we assume that $X_i \notin \mathcal{C}(X_i)$. In accordance with Pearl [15], we say that the causal model is *Markovian* if there exists an ordering of variables (without loss of generality we will assume that it is the ordering $X_1, X_2, \ldots, X_n$) such that $\mathcal{C}(X_1) = \emptyset$, and for all $i = 2, 3, \ldots, n$, $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. It means that considering only Markovian models, we exclude feedback relations.

Denote $x_i = \mathcal{C}(X_i) \cup \{X_i\}$. If we have probability distributions $\mu_i(x_i)$ we can construct a compositional causal model (CCM) as

$$\kappa(X_1, \ldots, X_n) = \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n). \tag{5}$$

Therefore the proper causal models for the example from Section 1.1 (Fig. 1(b)) is just

$$\tau(T) \triangleright \omega(C, T) \triangleright \alpha(A, C) \triangleright \varepsilon(A, E, T).$$

To understand properly expression (5) (and therefore also the above-presented causal model for the example) we have to highlight some properties following from formulas (1)–(2). First of all, we have to realize that the composition is not commutative. The reader can get a simple counterexample when considering a composition of two distributions $\kappa(x)$, $\lambda(y)$ for which $\kappa^{\downarrow x \cap y} \neq \lambda^{\downarrow x \cap y}$. Though the absence of commutativity is often considered a drawback, for causal models it can be an advantage, because causal relation is also asymmetric. As we will see later, the operation of composition is neither associative. Therefore, to make expression (5) unambiguous, we have to make an agreement that if not specified by parentheses otherwise, in all formulas the operator is performed from left to right, i.e.,

$$\begin{aligned}&\mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)\\ &\quad = (\ldots((\mu_1(x_1) \triangleright \mu_2(x_2)) \triangleright \mu_3(x_3)) \triangleright \ldots \triangleright \mu_{n-1}(x_{n-1})) \triangleright \mu_n(x_n).\end{aligned}$$

There are several theorems in [6] saying under what conditions one can change the ordering of distributions in a compositional model without influencing the resulting joint distribution. It is important to stress that for causal models, most of such transformations are forbidden. For causal models, we can consider only the orderings guaranteeing their Markovianity, i.e., for which $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. And it is the result of Kratochvíl that says that all these orderings define the same joint probability distribution $\kappa(X_1, \ldots, X_n)$ (see [10]).

In what follows, we will need the following three properties (for the respective proofs see [6]). For distributions $\mu_1(x_1), \mu_2(x_2), \mu_3(x_3)$

$$x_1 \supseteq x_2 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright \mu_3 \triangleright \mu_2; \tag{6}$$
$$x_1 \supseteq x_2 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3); \tag{7}$$
$$x_2 \supseteq x_1 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3). \tag{8}$$

The reader certainly realized that all the above introduced properties, including property (6), describe Markovianity preserving modifications. It means that, for example, if $\mu_1 \triangleright \mu_2 \triangleright \mu_3$ is Markovian CCM then $x_1 \supseteq x_2 \cap x_3$ guarantees that $\mu_1 \triangleright \mu_3 \triangleright \mu_2$ is also Markovian (it follows from the fact that under this assumption $x_3 \cap (x_1 \cup x_2) = x_3 \cap x_1$).

2.3 Perfectization Procedure

Consider an arbitrary compositional model $\kappa(x_1 \cup \ldots \cup x_n) = \pi_1(x_1) \triangleright \ldots \triangleright \pi_n(x_n)$. Property 1 guarantees that $\pi_1 = \kappa^{\downarrow x_1}$ but generally $\pi_j \neq \kappa^{\downarrow x_j}$ for $j > 1$. This may be in some situations disadvantageous. For example, if we want to know probability of variable $X \in x_1$ then we can get it directly from π_1 without necessity to consider multidimensional model κ. However, if $X \notin x_1$ the only way how to ascertain the probability of X one has to consider $\kappa^{\downarrow x_1 \cup \ldots \cup x_j} = \pi_1 \triangleright \ldots \triangleright \pi_j$ such that $X \in x_1 \cup \ldots \cup x_j$. This disadvantage can easily be circumvented when restricting our attention to perfect sequences, for which, as it was proved in [6], all π_j are marginals of the multidimensional model κ.

We say that a compositional model $\kappa = \pi_1 \triangleright \ldots \triangleright \pi_n$ is *perfect* if

$$\begin{aligned}
&\pi_1 \triangleright \pi_2 = \pi_2 \triangleright \pi_1,\\
&\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_3 \triangleright (\pi_1 \triangleright \pi_2),\\
&\qquad \vdots\\
&\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n = \pi_n \triangleright (\pi_1 \triangleright \ldots \triangleright \pi_{n-1}).
\end{aligned}$$

Perhaps, it is not visible for the first sight, but it is not difficult to prove a stronger property than that mentioned above. In [6], it is proved that a compositional model $\kappa = \pi_1 \triangleright \ldots \triangleright \pi_n$ is perfect *if and only if* all distributions π_j are marginals of the multidimensional model κ. Therefore, considering that low-dimensional distributions π_k are carriers of local information, the constructed multidimensional model – if it is perfect – represents global information faithfully reflecting all of the local input.

It is important to realize that restricting our attention only to perfect models is not at the loss of generality. This is because any compositional model $\pi_1 \triangleright \ldots \triangleright \pi_n$ can be transformed into a perfect model $\kappa_1 \triangleright \ldots \triangleright \kappa_n$ using the following procedure:

$$\begin{aligned}
\kappa_1 &= \pi_1,\\
\kappa_2 &= \kappa_1^{\downarrow K_2 \cap K_1} \triangleright \pi_2,\\
\kappa_3 &= (\kappa_1 \triangleright \kappa_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} \triangleright \pi_3,\\
&\vdots\\
\kappa_n &= (\kappa_1 \triangleright \ldots \triangleright \kappa_{n-1})^{\downarrow K_n \cap (K_1 \cup \ldots K_{n-1})} \triangleright \pi_n.
\end{aligned}$$

In this case, namely, $\kappa_1 \triangleright \ldots \triangleright \kappa_n$ is perfect and $\pi_1 \triangleright \ldots \triangleright \pi_n = \kappa_1 \triangleright \ldots \triangleright \kappa_n$. Moreover, since π_j and κ_j are defined for the same set of variables, it is evident that if $\pi_1 \triangleright \ldots \triangleright \pi_n$ is Markovian that also $\kappa_1 \triangleright \ldots \triangleright \kappa_n$ is Markovian.

Perhaps, we should mention that from the theoretical point of view the process of perfectization described above is simple. Unfortunately, it need not be valid from the point of view of computational complexity. The process requires marginalization that may be, in some situations, rather computationally expensive (for details on marginalization in compositional models see [1, 12]).

3 Conditioning and Intervention

Let us recall the difference between conditioning and intervention. Going back to our example, while conditioning by $C = 1$ gives evidence only about the universities, where attendance at lectures is compulsory, the intervention $do(C = 1)$ speaks about *all* universities under the assumption that we make sure (e.g., by a legislative act) that attendance at lectures is compulsory at all universities.

To simplify the following formulas, denote in the rest of this subsection $y = x_1 \cup \ldots \cup x_n$ and $z = y \setminus \{X\}$ for some $X \in x_1 \cup \ldots \cup x_n$. It means that considering a causal compositional model $\kappa(y) = \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)$ the conditioning by $X = \mathbf{a}$ leads to a distribution of variables z: $\kappa(z|X = \mathbf{a})$.

As shown in [2, 7], we can realize both the conditioning and intervention as a composition of the causal compositional model $\mu_1 \triangleright \ldots \triangleright \mu_n$ with a degenerated one-dimensional distribution $\delta_{\mathbf{a}}(X)$, which is a distribution of variable X achieving probability 1 for value $X = \mathbf{a}$, i.e.,

$$\delta_{\mathbf{a}}(X) = \begin{cases} 1 & \text{if } X = \mathbf{a},\\ 0 & \text{otherwise.} \end{cases}$$

Using this denotation we use the following simple formulas proven in [7]:

$$\kappa(z|X=\mathbf{a}) = (\delta_{\mathbf{a}}(X) \triangleright (\mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m))^{-X},$$

and

$$\kappa(z|do(X=\mathbf{a})) = (\delta_{\mathbf{a}}(X) \triangleright \mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m)^{-X}.$$

The goal of this paper is to illustrate the applicability of causal compositional models to business process modeling, and thus we do not go more deeply into theory of compositional models (for this the reader is referred to [6, 7, 8], nevertheless let us stress the importance of the pair of brackets in which the formulas above differ from each other. This difference arises from the fact that the operator of composition is not associative.

The reader familiar with Pearl's causal diagrams [15] has certainly noticed an advantage of the introduced compositional models. While here we can compute both conditioning and intervention from one causal compositional model, in causal diagrams we have to consider two different graphs. Conditioning is computed from the "full" causal diagram, from which, for the computation of intervention, we have to delete all the arrows heading to the intervention variable.

3.1 Anticipating Operator

In the next section we will illustrate on our example from Section 1.1 the possibility to compute results of intervention and conditioning even in the case of model with a hidden variable. For this task, it will appear advantageous to compensate the lack of associativity by introducing another operator, so called *anticipating operator*, defining a special type of composition. For a set of variables z and distributions $\kappa(x)$ and $\mu(y)$ it is defined by the formula:

$$\kappa \circledcirc_z \mu = (\mu^{\downarrow(z\setminus x)\cap y} \cdot \kappa) \triangleright \mu.$$

Its advantage is expressed by Theorem 9.4 in [6] saying that for $\pi(z)$, $\kappa(x)$ and $\mu(y)$

$$\pi(z) \triangleright \kappa(x) \triangleright \mu(y) = \pi(z) \triangleright (\kappa(x) \circledcirc_z \mu(y)). \tag{9}$$

3.2 Elimination of Hidden Variables

Let us now derive formulas to be used for computation of conditioning and intervention in the simple causal compositional model describing our example from Section 1.1. Assuming it is in a perfect form we can express it with the help of its marginals:

$$\kappa(A,C,E,T) = \kappa(T) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T).$$

Computation of the conditional $\kappa(E|C=\mathbf{a})$ is simple.

$$\begin{aligned}
\kappa(E|C=\mathbf{a}) &= \big(\delta_{\mathbf{a}}(C) \triangleright (\kappa(T) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T))\big)^{\downarrow\{E\}} \\
&= \Big(\big(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C,E,T)\big)^{-T}\Big)^{\downarrow\{E\}} \\
&\overset{(2)}{=} \big(\delta_{\mathbf{a}}(C) \triangleright (\kappa(A,C,E,T))^{-T}\big)^{\downarrow\{E\}} = \big(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C,E)\big)^{\downarrow\{E\}}.
\end{aligned}$$

Notice that during these computations we used the property (2) from Section 2.1. This is why the symbol (2) appears above the respective equality sign. This type of explanation will also be used in the subsequent computations.

To compute $\kappa(E|do(C=\mathbf{a}))$ we will need to apply most of the properties introduced above, as well as the anticipating operator defined in the preceding section. Since we cannot expect the reader is accustomed to the computations with the operator of composition we will perform just one elementary modification at each of the following steps. This is why the following computations look even more cumbersome than they really are.

$$\begin{aligned}
\kappa(E|do(C=\mathbf{a})) &= \big(\delta_{\mathbf{a}}(C) \triangleright \kappa(T) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T)\big)^{\downarrow\{E\}} \\
&\overset{(3)}{=} \big(\delta_{\mathbf{a}}(C) \triangleright \kappa(T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T)\big)^{\downarrow\{E\}} \\
&\overset{(6)}{=} \big(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \kappa(T) \triangleright \kappa(A,E,T)\big)^{\downarrow\{E\}} \\
&\overset{(9)}{=} \Big(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \Big(\kappa(T) \circledcirc_{\{C,A\}} \kappa(A,E,T)\Big)\Big)^{\downarrow\{E\}} \\
&\overset{(2)}{=} \Big(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \Big(\kappa(T) \circledcirc_{\{C,A\}} \kappa(A,E,T)\Big)^{-T}\Big)^{\downarrow\{E\}}
\end{aligned}$$

To express $\Big(\kappa(T) \circledcirc_{\{C,A\}} \kappa(A,E,T)\Big)^{-T}$ we will take advantage of the idea of extension used for this purpose by Pearl in [15]. The reader can realize that it is the way how to take into account the mutual dependence of variables A, C, E (notice that it plays the same role of what is realized by inheritance of parents in Shachter's edge reversal rule [17]).

$$\begin{aligned}
&\Big(\kappa(T) \circledcirc_{\{C,A\}} \kappa(A,E,T)\Big)^{-T} = \Big(\kappa(T) \circledcirc_{\{A\}} \kappa(A,E,T)\Big)^{-T} \\
&= \Big(\Big(\kappa(C,T) \circledcirc_{\{A\}} \kappa(A,E,T)\Big)^{-C}\Big)^{-T} = \big(\kappa(A) \cdot \kappa(C,T) \triangleright \kappa(A,E,T)\big)^{\downarrow\{A,E\}} \\
&= \big(\kappa(A) \cdot \kappa(C) \triangleright \kappa(C,T) \triangleright \kappa(A,E,T)\big)^{\downarrow\{A,E\}} \\
&\overset{(3)}{=} \big(\kappa(A) \cdot \kappa(C) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T)\big)^{\downarrow\{A,E\}} \\
&\overset{(7)}{=} \big(\kappa(A) \cdot \kappa(C) \triangleright (\kappa(C,T) \triangleright \kappa(A,C)) \triangleright \kappa(A,E,T)\big)^{\downarrow\{A,E\}} \\
&= \big(\kappa(A) \cdot \kappa(C) \triangleright \kappa(A,C,T) \triangleright \kappa(A,E,T)\big)^{\downarrow\{A,E\}} \\
&\overset{(8)}{=} \big(\kappa(A) \cdot \kappa(C) \triangleright (\kappa(A,C,T) \triangleright \kappa(A,E,T))\big)^{\downarrow\{A,E\}} \\
&= \big(\kappa(A) \cdot \kappa(C) \triangleright \kappa(A,C,E,T)\big)^{\downarrow\{A,E\}} \overset{(2)}{=} \big(\kappa(A) \cdot \kappa(C) \triangleright \kappa(A,C,E)\big)^{\downarrow\{A,E\}} \\
&= \Big(\kappa(C) \circledcirc_{\{A\}} \kappa(A,C,E)\Big)^{-C},
\end{aligned}$$

which eventually leads to

$$\kappa(E|do(C=\mathbf{a})) = \Big(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \Big(\kappa(C) \circledcirc_{\{A\}} \kappa(A,C,E)\Big)^{-C}\Big)^{\downarrow\{E\}}.$$

3.3 Hypothetical Example - Numerical Computations

Let us illustrate the above-derived formulas on data corresponding to the example from Section 1.1. For the sake of simplicity, we consider all the variables to be binary. The data concerning 200 students from 16 universities are presented together with corresponding percentages as estimates of corresponding probabilities (see Table 1).

From Table 1 we can immediately see that the total probability of a success at the first attempt is $\kappa(E=1) = 0.68$.

Let us first consider the inapt model without the hidden variable T (recall that it assumes the conditional independence of C and E given A). To construct it, we need three distributions, namely a distribution for variable C, a distribution for variables A, C and the last one for variables A, E. Then we compute these distributions from Table 1, and denote them $\kappa_c(C), \kappa_{ac}(A,C), \kappa_{ae}(A,E)$, respectively. Constructing the compositional model corresponding to the graph from Figure 1(a)

$$\kappa(A,C,E) = \kappa_c(C) \triangleright \kappa_{ac}(A,C) \triangleright \kappa_{ae}(A,E) = \kappa_{ac}(A,C) \triangleright \kappa_{ae}(A,E)$$

we cannot expect to get a perfect model (the reader can see that it is really not perfect from the last row in Table 1), and therefore keep in mind that the estimated distributions κ_c, κ_{ac} and κ_{ae} are not marginals of κ.

Table 1: Estimates of probabilities

C	0	0	0	0	1	1	1	1
A	0	0	1	1	0	0	1	1
E	0	1	0	1	0	1	0	1
frequencies	18	42	22	38	12	12	12	44
percentages	0.09	0.21	0.11	0.19	0.06	0.06	0.06	0.22
$\kappa_{ace} = \kappa_{ac} \triangleright \kappa_{ae}$	0.11	0.19	0.9	0.21	0.04	0.08	0.08	0.20

Now, it is an easy task to compute

$$\kappa(E|do(C=1)) = \kappa(E|C=1) = ((\delta_1(C) \triangleright \kappa_{ac}(A,C)) \triangleright \kappa_{ae}(A,E))^{\downarrow\{E\}},$$

from which we immediately get

$$\kappa(E=0|do(C=1)) = \kappa(E=0|C=1) = 0.31,$$
$$\kappa(E=1|do(C=1)) = \kappa(E=1|C=1) = 0.69.$$

Therefore, as it was said in Introduction, this model expects a slight increase of success after the intervention making attendance at lectures compulsory at all universities.

Consider, now, a more complex model including the hidden variable T, i.e.,

$$\lambda(A,C,E,T) = \lambda_t(T) \triangleright \lambda_{ct}(C,T) \triangleright \lambda_{ac}(A,C) \triangleright \lambda_{aet}(A,E,T).$$

As said above, since the type of a university is for us a hidden variable, for this model we can estimate only one distribution from the data, namely $\lambda_{ac} = \kappa_{ac}$.

Fortunately, as it was shown in the preceding section, in this very model we can eliminate the hidden variable getting

$$\lambda(E|C=1) = \big(\delta_1(C) \triangleright \lambda_{ace}(A,C,E)\big)^{\downarrow\{E\}},$$

and

$$\lambda(E|do(C=1)) = \left(\delta_1(C) \triangleright \lambda_{a,c}(A,C) \triangleright \left(\lambda_c(C) \circledcirc_{\{A\}} \lambda_{a,c,e}(A,C,E)\right)^{-C}\right)^{\downarrow\{E\}},$$

from which we get

$$\lambda(E=0|C=1) = 0.30,$$
$$\lambda(E=1|C=1) = 0.70,$$
$$\lambda(E=0|do(C=1)) = 0.33,$$
$$\lambda(E=1|do(C=1)) = 0.67.$$

4 Model verification

When discussing the hypothetical example, in Fig. 1 we introduced two possible causal models that were denoted κ and λ in Section 3.3. Each of them yielded different results for the intervention $do(C=1)$. In this section we are going to show that based on the data from Section 3.3 we should not have taken the simpler model from Fig. 1(a) (i.e., model κ) into consideration.

When setting up a model, one should always consider, and verify, all (or at least most of) the conditional independence relations that can be read from the model. For this, several tools were described in the literature. When the model is in a form of a Bayesian network, one can use either a *d-separation rule* or the rule based on separation in a *moralized ancestral graph* [11]. For the compositional models, *persegrams* were designed [10], from which one can read all the independence relations that must hold in the considered model.

Table 2: Conditional frequencies of C and E for $A = 0$ (left) and $A = 1$ (right)

	E=0	E=1
C=0	18	42
C=1	12	12

	E=0	E=1
C=0	22	38
C=1	12	44

We do not describe here the mentioned procedures. Just let us recall how the conditional independence relation is defined. Consider a distribution $\pi(x)$, and variables $X, Y \in x$, and a subset (possibly empty) of variables $y \subseteq x \setminus \{X, Y\}$. We say that for distribution π variables X and Y are *conditionally independent given set of variables* y (in symbol $X \perp\!\!\!\perp_\pi Y|x$) if

$$\pi^{\downarrow y\cup\{X,Y\}} \cdot \pi^{\downarrow y} = \pi^{\downarrow y\cup\{X\}} \cdot \pi^{\downarrow y\cup\{Y\}}.$$

Either of the above-mentioned tools yields the following systems of conditional independence relations:

- for model κ described in Fig. 1(a) it is just one relation: $C \perp\!\!\!\perp_\kappa E|A$;
- for model λ described in Fig. 1(b) there are two relations: $A \perp\!\!\!\perp_\lambda T|C$, and $E \perp\!\!\!\perp_\lambda C|\{A, T\}$.

In the latter model, both the independence relations contain the unobserved variable T, so we do not have a possibility to verify (or, better said, to exclude) this model.

But, having data from Table 1 we can test whether the conditional independence $C \perp\!\!\!\perp E|A$ holds true, which is necessary for model κ. The test of this relation can be performed using well-known Cochran-Mantel-Haenszel test or in a simplified manner using the sum of partial chi-squares for two conditional probabilities in Table 2. Now we get value of chi-square for $A = 0$ equal to $\chi^2_0 = 2.987$ and for $A = 1$ equal to $\chi^2_1 = 3.246$ both with one degree of freedom. Therefore we yield summary test statistics $\chi^2 = 6.233$ which is with two degrees of freedom above the five percent critical value, now equal to 5.99. And thus we should had rejected the null hypothesis of conditional independence of C and E given A, and should not had considered this causal model.

5 Conclusions

On a slightly paradoxical example we tried to convince the reader that considering causal influence in business process models may be not only useful but sometimes inevitable, and that having such a causal model one should do their best to verify the model (or more precisely, to test whether the model is not excluded by the data). The necessity of having a simple example led us to illustrate the proposed approach on artificially generated data. However, we tried to generate realistically looking (non-degenerated) data and this is why the numerical results may seem not so convincing. Naturally, when more mutually dependent effects are taken into consideration, or when considering a more complex model with a greater number of variables, the results obtained from descriptive and causal models can be substantially different.

Another purpose of this paper was to show that causal models need not be only causal diagrams. The employment of causal compositional models can be used to describe causal relations among considered variables. The composition of model distributions with a special (degenerated) distribution surprisingly provides a possibility to evaluate both conditional probabilities and the effects of causal interventions using a unifying apparatus of compositional models.

Acknowledgement
The research was partially supported by GA ČR under the project number 15-00215S. The second author wants to express his sincere thanks to Judea Pearl for the motivation to start employing causal models in applications.

References

[1] V. Bína, R. Jiroušek. Marginalization in Multidimensional Compositional Models. *Kybernetika,* 42, 4, 405–422, 2006.

[2] V. Bína, R. Jiroušek. On computations with causal compositional models. *Kybernetika,* 51, 3, 525–539, 2015.

[3] A. Feugas, S. Mosser, L. Duchien. A Causal Model to predict the Effect of Business Process Evolution on Quality of Service. In *Proceedings of the 9th International ACM Sigsoft Conference on the Quality of Software Architectures,* pages 143–152. ACM, New York, 2013.

[4] Y. Hagmayer, S. A. Sloman, D. A. Lagnado, M. R. Waldmann. Causal reasoning through intervention. In *Causal learning: Psychology, philosophy, and computations,* Alison Gopnik and Laura Schulz (eds.) Oxford University Press, Oxford, New York, 2007.

[5] R. Jiroušek. Composition of Probability Measures on Finite Spaces. In *13th Conf. on Uncertainty in Artificial Intelligence,* pages 274–281. Morgan Kaufman, San Francisco, 1997.

[6] R. Jiroušek. Foundations of compositional model theory. *Int. J. General Systems,* 40, 6, 623–678, 2011.

[7] R. Jiroušek. On Causal Compositional models: simple examples. In *Proc. of the 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems,* A. Laurent et al. (eds.), Part I, pages 517–526, Springer International Publishing, CCIS 442, Switzerland, 2014.

[8] R. Jiroušek, V. Kratochvíl. Foundations of compositional models: structural properties. *Int. J. General Systems* 44, 1, 2–25, 2015.

[9] C. W. Kirgwood. *System Dynamics Methods: A Quick Introduction.* Arizona State University, 2013. [http://www.public.asu.edu/~kirkwood/sysdyn/SDIntro/SDIntro.htm]

[10] V. Kratochvíl. Probabilistic Compositional Models: solution of an equivalence problem. *Int. J. Approximate Reasoning,* 54, 5, 590–601, 2013.

[11] S. L. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

[12] F. M. Malvestuto. Marginalization in Models Generated by Compositional Expressions. *Kybernetika* 51, 4, 541–570, 2015.

[13] S. Morgan, C. Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge University Press, New York, 2007.

[14] R. J. Murnane, J. B. Willett. *Methods Matter: Improving Causal Inference in Educational and Social Science Research.* Oxford University Press, Oxford, 2011.

[15] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, New York, 2009.

[16] M. Ryall. Causal Ambiguity, Complexity, and Capability-Based Advantage, *Management Science,* 55, (3), 389–403, 2009.

[17] R. D. Shachter. Evaluating Influence Diagrams. *Operations Res.,* 34, 871–882, 1986.

[18] L. Váchová, V. Bína. A Note on Computation of Causal Effect in Economic Models. In *Proceedings of the 31st Mathematical models in Economy,* pages 974–979, College of Polytechnics, Jihlava, 2013.

Fuzzy Neural Networks on Embedded platforms

Petr CERMAK[1] and Jiri MARTINU[2]

[1] *Silesian University in Opava*
Bezruc sq. 13, Opava, 74601
Czech Republic
petr.cermak@fpf.slu.cz
[2] slu.martinu@gmail.com

Abstract

Fuzzy modeling is the method that describes a behavior of real systems using the fuzzy logic and the fuzzy reasoning. However, in cases when the need for real-time control of the process in embedded systems behavior arises, a standard HW platforms such as personal computers or the ARM platforms are not suitable regarding their limited performance. Additionally, there are many cases in which the conventional approaches fail due to nonlinear system behavior. The afore mentioned is the reason of involving state-of-the-art technologies such as the FPGAs and the Fuzzy Neural Networks into the chain of modeling. The Takagi-Sugeno fuzzy non-linear regression model is also one of the suitable Artificial Intelligence means for fuzzy modeling.

Keywords: Fuzzy, Neural network, Takagi-Sugeno, FUZNET-FPGA, HDL, FPGA, SMMDPU, SoC.

1 Introduction

The T-S fuzzy non-linear regression models (Takagi-Sugeno models) appear as one of the suitable tools used in complex systems modeling technique, as in [1, 2, 3]. Identification of fuzzy models can be divided into two steps: a structure identification and a parameter estimation. The structure identification depends on the relevant input variables, number of rules, and partitioning of the input space. Parameter estimation phase, the model parameters are determined. Parameter estimation process can be relatively easily automated by different optimizing techniques with the help of a neural network learning process. In my experience, a structural identification takes high influence to a quality of modeling. Procedures of the structural and parametric identification was implemented into the FUZNET Fuzzy neural network development system [3]. Implementations allow fuzzy reasoning, parametric and structural identification on different platforms, pure application software implementation on a particular OS, mixed software implementation on a bare metal with the Scheme programming language. Whenever we need faster response, we must select implementation/synthesis on a Field Programmable Gate Arrays (FPGA). All those implementations will be further discussed.

2 Brief background of TS fuzzy neuro modeling

This chapter briefly introduces the TS fuzzy neuro modeling. T-S fuzzy nonlinear regression model is described by a system of the following rules:

$$\begin{array}{l} IF\ (x_1\ is\ A_{1,1})\ AND\ \ldots\ (x_n\ is\ A_{n,1}) \\ \quad THEN\ (y_2\ =\ f\,(x_1,x_2,x_3,...,x_n)) \\ IF\ (x_1\ is\ A_{1,r})\ AND\ \ldots\ (x_n\ is\ A_{n,r}) \\ \quad THEN\ (y_r\ =\ f\,(x_1,x_2,x_3,...,x_n)) \end{array} . \tag{1}$$

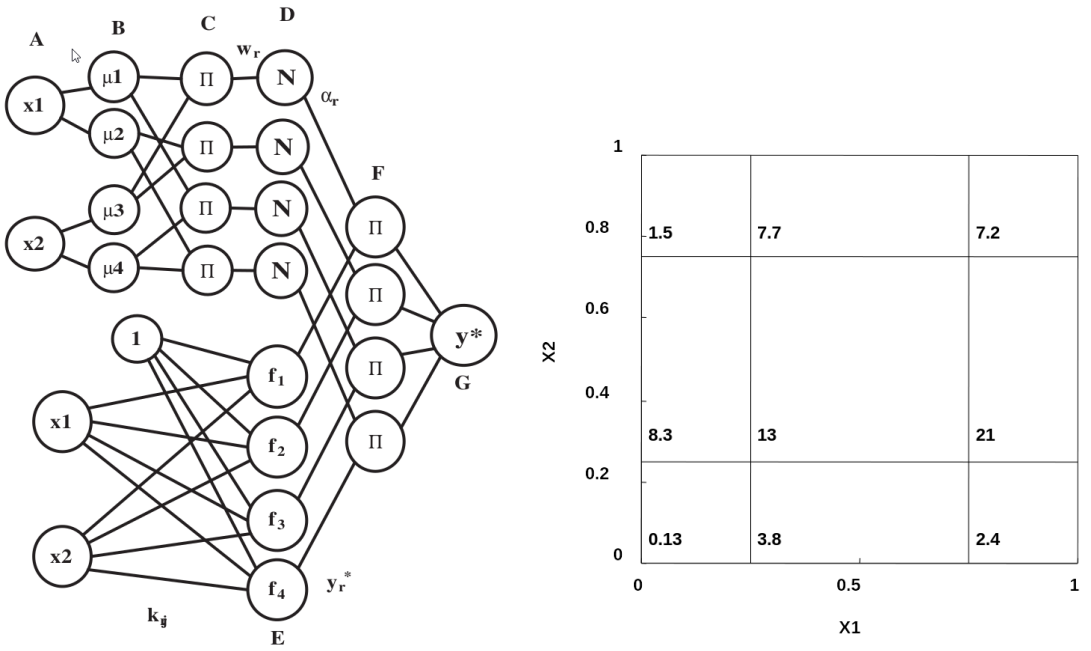


Figure 1: Structure of Fuzzy Neural Network and Space Diversification with corresponding Error values [2, 3]

In Fig. 1, we can see FNN of TSK system with 4 rules [3]. The antecedent parts of the rules are in the layer (C) and consequent part is in the layer (E). The Layer (B) computes values of membership functions according to Layer (A) inputs:

$$\mu\left(x, s_1, s_2, c\right) = \begin{cases} x < c & e^{s_1(x-c)^2} \\ x \geq c & e^{s_2(x-c)^2} \end{cases}. \tag{2}$$

The Layer (D) realizes normalized antecedent output of all rules:

$$\alpha_r = \frac{w_r}{\sum_{s=1}^{n} w_s}. \tag{3}$$

The final value is computed using the sum and multiply operators between the normalized antecedent outputs and the outputs of regression functions in the Layer (E):

$$y^* = \sum_{r=1}^{n} \alpha_r f_r\left(X_1, X_2\right). \tag{4}$$

Second part of Fig. 1 shows diversification of the Input space. There are 9 Rules, each with a calculated error in subspace.

The k index specifies all trainings values belonging to *r-th* Rule subspace:

$$E_r = \sum_{k=1}^{K} \alpha_{k,r} \left|y_k - y_k^*\right| = \sum_{k=1}^{K} \alpha_{k,r} E_k \tag{5}$$

The value of the Rule Subspace Error allows to determine which Rules and Term must be expanded or merged. The Rules with a lower Error can be merged and complexity decreases. Rules with the higher Error can be expanded, complexity increases and the error of the whole model decreases. The maximal number of rules is set by the Init. Both shapes of the Membership functions and the linear regression coefficients are adjusted during an on-line learning process. Please refer to the [1]for further description.

3 FUZNET

The proposed FUZNET program tool was developed based on fuzzy neural network technology including the new identifying procedures [1, 2, 3]. This fuzzy neural network environment appears to be suitable for fuzzy non-linear regression model identification as well as for the FUZNET program including its application in real system modeling.

This system is extension of the neural network (NN) development system based on the Back Propagation algorithm. The Figure 2 shows the basic FUZNET structure.

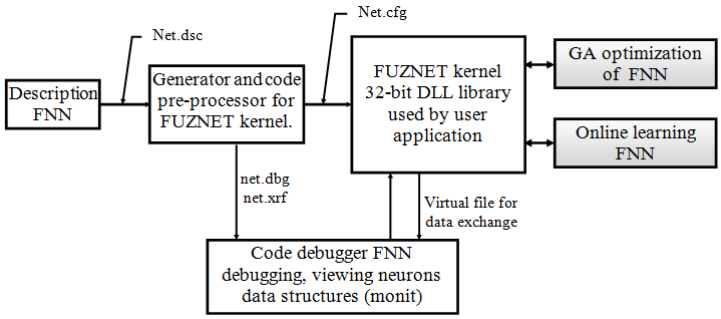


Figure 2: The FUZNET structure [1, 2, 3]

3.1 FNN description

For the FUZNET Kenel and its Fuzzy-Neural network the number of inputs with the language values, determination of layers (an its neurons), initial values, weights and the membership functions types must be specified. The algorithms, data structures and interfaces for the FNN learning process and its following computations create the FUZNET Kernel. Included are also procedures (scripts) for communication with debugging GUI tools and an interface for the Dynamic Link Libraries (DLL) for exporting the interface function to the Matlab m-files. The scripts allows monitoring of the FNN structure and step-by-step debugging.

3.2 Additional algorithms for optimization of FNN

For optimization of the FNN, the genetic optimization algorithm was selected [2]. We can separate process of GA optimization on two subtasks. The first - a structural model identification, determines how many linguistic values will be used for each fuzzy variables belonging to certain input. The configuration of inputs and fuzzy variables with membership functions also determines max number of rules in FNN. The second one - a parametric identification, indicates that the shapes of membership functions are determined. In the case of TSK system, it also provides setting of the regression coefficients in the consequent parts. Those coefficients are in fact the weights of the linear perceptrons in the FNN. The GA determines also the parameters of of the FNN learning process, namely the learning coefficients. The GA use some advanced genetic operators, namely [2]:

- Crossover operator
- Mutation operator
- Sexual reproduction
- Life time of gene
- Shaded zone

A dynamic structure adaptation is performed by the on-line learning functionality [1].

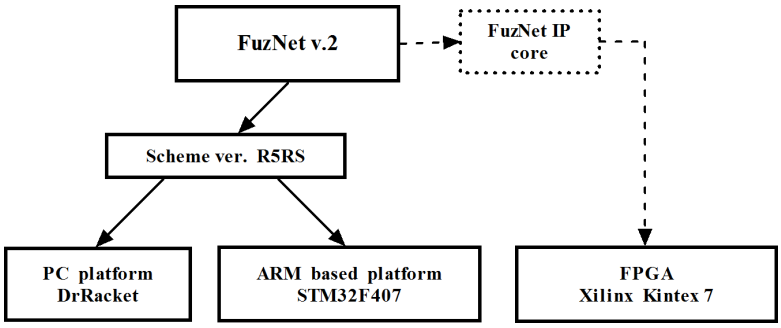


Figure 3: The FUZNET implementation on different HW platforms [3], modified

According the Figure 3 scheme, the following implementation options were considered:

- FuzNet on the PC or the ARM platform with the Scheme interpreter installed (e.g. Dr. Racket, Armpit Scheme [4])
- FPGA with the FuzNet dedicated IP Core

The first option proved the functionality but its processing was relatively slow mainly on the ARM-based platform (STM32F407). The Scheme implementation structure programing includes the following main steps:

1. Creating of all objects (neurons, layers, etc.)
2. Creating all FNN connections between layers
3. Providing initializations of needed parameters
4. Storing of the state after initialization to the data-tree

The regression coefficients, shapes of the membership functions and other configuration parameters are stored in the tree-data structure. This allows to store and reload a whole configuration of the FNN at any time. There is also a possibility of using additional algorithms for the FNN optimization by means of Genetic Algorithms (GA).

The FUZNET IP core on the FPGA was chosen for further implementation in need of fast processing. This option is further discussed in the following chapter.

4 FUZNET on the Field Programmable Gate Arrays

In the real engineering, we need to model complex systems working in Real Time in time critical applications. The Time Response and Transport Delay must be in the order of milliseconds and there exist also some space restrictions. In such situations, we must use the FPGA or ASICs (in production). Computation on processor based systems (Intel, ARM, etc.) could be distributed on each of the cores. But overhead of OS and the bottleneck - access to RAM - it is a serious problem. We need an amount of a distributed RAM [5] as storage for the Fuzzy Neural Networks coefficients and for the set of quite simple operations like comparison, multiplexing signals and DSP [6] (Digital Signal Processor block). Because operations, distributed memory and BRAM are hardware blocks, we need to define connections between them. We can use some languages for electronic design automation. The major software systems for a HDL design supports two languages, namely VHDL and Verilog. The following code is our example of the HDL harware description for the Linear regression computation f_r:

```
LIBRARY ieee;
USE ieee.std_logic_1164.all;
USE ieee.numeric_std.all;
USE work.arr_pkg.all;

entity LinReg is Port(
```

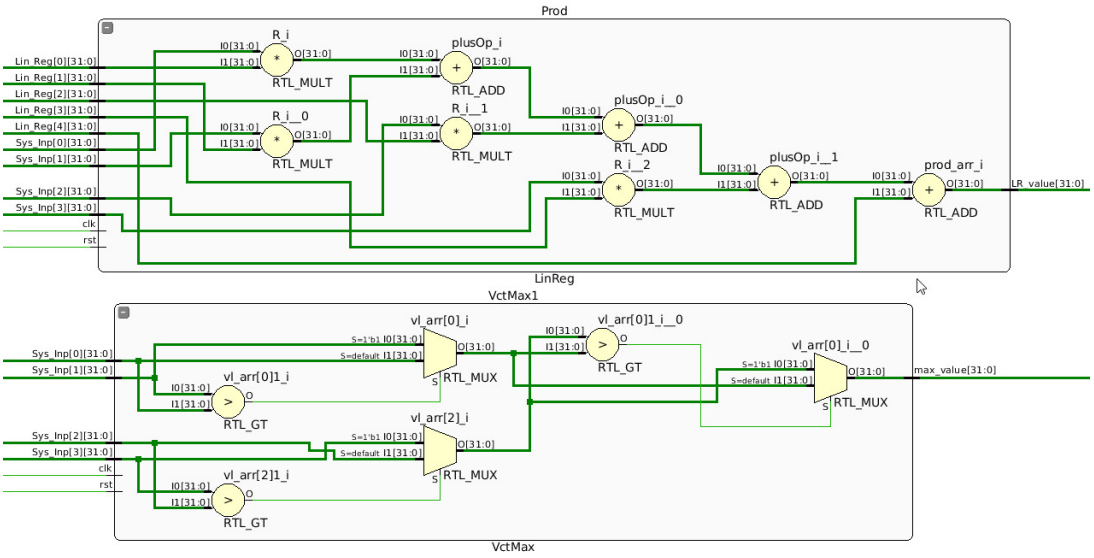


Figure 4: Synthesized functions LinReg and MaxVct

```
clk : in boolean;
rst : in boolean;
Sys_Inp : in tvl_array;
Lin_Reg : in tcoef_array;
LR_value : out signed (31 downto 0) );
end LinReg;

architecture Behavioral of LinReg is
        function prod_arr( Inputs: tvl_array; Coefs: tcoef_array)
        return signed (31 downto 0) is
                variable LR_v : signed (31 downto 0) :=
                                        (others => '0');
        begin
                for Rct in 0 to Inputs'right loop
                        LR_v := LR_v + to_integer( Inputs( Rct) *
                                        Coefs( Rct));
                end loop;
                return LR_v + Coefs( Coefs'right);
        end prod_arr;

begin
        LR_value <= prod_arr( Sys_Inp, Lin_Reg);
end Behavioral;
```

This example describes a new *LinReg* entity block with the *entity LinReg is Port* connector port and with the description of the *architecture Behavioral of LinReg is* behavior. Behavior defined between *begin* and *end* seems like programming construct for loop. But in comparison to classical program code execution, each operation in loop is separated/synthesized to a different piece of hardware. This allows parallel computation during the one clock cycle. The output diagram after implementation/synthesis on the FPGA logic hardware is shown on Fig. 4.

The *Sys_Inp* Inputs are the 32-bit signed values used for computation. The *Lin_Reg* coefficients are the linear regression coefficients and together with the *Sys_Inp* inputs they compute the value of function in the consequent layer (E). The *VctMax* in the second example shows how to determine a maximal value from the *Sys_Inp* input vector. This example can realize the *VctMin* - minimum, product $\prod$ layer (C) functions with minor modification. It is obvious that we used some basic blocks like:

- RTL_ADD - adder included in DSP,
- RTL_MULT - multiplier included in DSP,
- RTL_GT - comparison great than,
- RTL_MUX - multiplexer,
- input connectors,
- output connector,
- paths as "wires".

Because the Fuzzy Neural Network in Fig. 1 is in its nature a distributed computational network, we can describe computational functions of all layers and neurons by a simple operation in HDL. We do not mention the learning process yet. In need of an on-line Fuzzy Neural Network adaptation the situation appears to be more complicated. We have two ways to realize the on-line adaptation process:

- to implement/synthesize the whole proces in HDL,
- to program softcore processor/s and use them for the on-line adaptation.

The HD (Hardware Description) and synthesis of the whole adaptation process could be very hard task and also utilization of resources on the FPGA could be terrible, however it could be the speediest way for computation. In the case we would like to change the adaptation process, we must to redesign the whole HD.

The second case uses the softcore processors with the low level gate utilization. Low level gate utilization decreases the time delay of the gates paths and increases the value of a maximum processor frequency. The Adaptation algorithm is realized using firmware or software. Software processor is shortly described in the next chapter.

4.1 Softcore SMMDPU

SYSTEM MEDIA Multiple Data Processor Unit SMMDPU [7] is low number gate stack based multiprocessor with a compile and debug support on the FPGA fabric. Processor has still a fixed arithmetic. The *SMMDPU_Proto* covers HD for max. 24 data processors configurable to use a data width from the 16 to the 128 bits. In the case of synthesis for the FUZNET, the data width is set to the 64 bits. It is essential to convert numbers from textual to binary form in block *IntCnv_B_D_H* for compiling and debuging. The binary, hexadecimal and decimal formats are allowed. It is possible to define the custom instruction names, new functions and variable names stored in block *Vocabulary*. The keyword searching HD in the Block *Vocabulary* we can synthesize as sequential or fully parallel. Parallelized version of the Vocabulary HD consumes a quite huge amount of hardware logic. In the case of the maximum optimization, the *IntCnv_B_D_H* and *Vocabulary* blocks can operate only in a few processor clock cycles. Compile and debug support is realized as state machine in the *CompileDebug* block.

4.2 FUZNET-FPGA structure

Let the programmable logic realizing forward computation and the Softcore processors realizing on-line adaptation to join to the new proposed tool - FUZNET-FPGA . The Fig. 6 shows the proposed block diagram of the FUZNET-FPGA with TS configured as follows: 2 Inputs, 1 Output, 3 Membership functions per Input. The maximum number of rules is set to 9. Computation proces consists of steps given by single or multiple clocks on Programmable Logic.

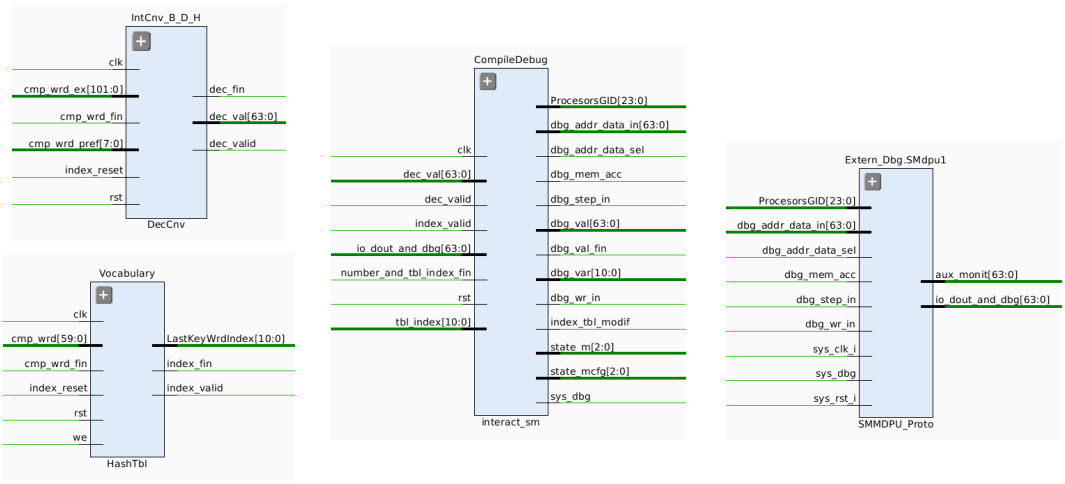


Figure 5: Parts of the Interactive SMMDPU Processor [7]

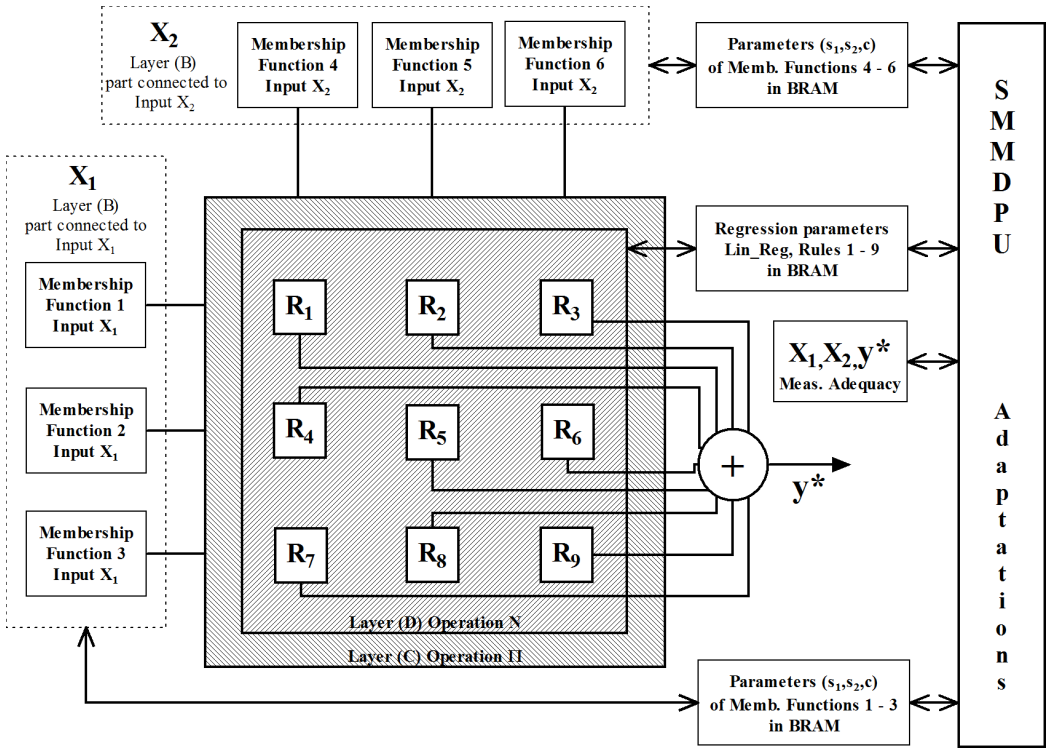


Figure 6: Block schema of the FUZNET-FPGA for certain TS configuration.

Computation, forward propagation process :

1. Acquire Inputs in Layer (A)
2. Compute Output of Membership functions, layer (B) in parallel way and propagate to Layer (C)
3. Compute Agregation operation $\prod$ in Layer (C), Results send to inputs of Rules
4. Normalize w_r in Layer (D), propagate α_r parameters to Rules
5. Compute Regression Functions of the Rules in parallel way using LinReg Fig. 4. Layer(E), multiply with α_r in Layer (F)

6. Result as sum of in y^*, Layer (G).

Sharing of the Regression Coefficients and Shape Parameters of the Membership functions makes it possible using double ported memory BRAMs. On Figure we can see as BRAM blocks.

Process of the On-Line adaptation :

1. Measure adequacy, Measure Error
2. Adequacy is ok? If Yes goto 1
3. Accessing to BRAMs. Backpropagate error
4. Adaptation shapes of Membership functions,(s_1, s_2, c) and adaptation of Regression Coefficients for LinReg Fig. 4., store into BRAMs
5. Goto 1.

5 Conclusion

Major contribution of this paper is description of the FUZNET implementation on the Field Programmable Gate Arrays. Decomposition of the T-S fuzzy Neural Network is provided. Each Computational Blocks are described in Hardware Description Language. Such kind of decomposition and optimized hardware synthesis allows forward computation (T-S model) in massive parallel way within less than millisecond. The on-line adaptation needs to use the Softcore Data Multiprocessor SMMDPU and it is slightly slower. Access to the Hardware blocks is realized using dual port BRAM memory. In the frame of the current state of the project the synthesized designs test and performance tests on different FPGA chips of the Xilinx Series 7 Kintex platform are performed [8].

Usage of the FUZNET-FPGA is supposed in the areas of the real time complex systems modeling, prediction of the time series and control with adaptation. Possible usage areas are the following:

- Automotive industry
- Industry 4.0
- Control of VTOL and UAVs

Acknowledgement
The authors would like to thank for support Project SGS/3/2016 "Embedded Image processing methods for medical and UAV use" and company SYSTEM MEDIA for granted licence of the SMMDPU IP Core.

References

[1] P. Cermak. Online Learning of Neural Takagi-Sugeno Fuzzy Model. In *24rd International Conference of the North American Fuzzy Information Processing Society*, pages 478-483, Ann Arbor, Michigan, U.S.A., 2005.

[2] P. Cermak, P. Chmiel. Parameters optimization of fuzzy-neural dynamic models, In *23rd International Conference of the North American Fuzzy Information Processing Society,* pages 762-767, Banff, Alberta, Canada, 2004.

[3] P. Cermak, M. Pokorny. The fuzzy-neuro development system FUZNET, In *18th International Conference on Methods Models in Automation Robotics (MMAR)*, pages 75-80, Miedzyzdroje, Poland, 2013.

[4] H. Montas. A Scheme Interpreter for ARM Microcontrollers, Fischell Department of Bioengineering of the A.J. Clark School of Engineering at the University of Maryland at College Park, Maryland, U.S.A., http://armpit.sourceforge.net/, 2012.

[5] Xilinx. 7 Series FPGAs Memory Resources, User Guide, UG473 (v1.12) September 27, 2016. https://www.xilinx.com/support/documentation/user_guides/ug473_7Series_Memory_Resources.pdf.

[6] Xilinx. 7 Series DSP48E1 Slice, User Guide, UG479 (v1.9) September 27, 2016. https://www.xilinx.com/support/documentation/user_guides/ug479_7Series_DSP48E1.pdf.

[7] SYSTEM MEDIA. SMMDPU IP Core, *User Guide*, Opava, Czech Republic, 2016.

[8] Xilinx. 7 Series FPGAs Data Sheet: Overview, DS180 (v2.4) March 28, 2017. https://www.xilinx.com/support/ documentation/data_sheets/ds180_7Series_Overview.pdf.

Hidden Auto-Conflict in the Theory of Belief Functions

Milan DANIEL[1] and Václav KRATOCHVÍL[2]

[1] *Jan Becher - Karlovarská Becherovka, a.s., Pernod Ricard group*
Luxembourg Plaza, Přemyslovská 43, Prague 3, CZ - 130 00
Czech Republic
milan.daniel@pernod-ricard.com

[2] *Institute of Information Theory and Automation, Czech Academy of Sciences*
Pod vodárenskou věží 4, Prague 8, CZ - 182 08
Czech Republic
velorex@utia.cas.cz

Abstract

Hidden conflicts of belief functions in some cases where the sum of all multiples of conflicting belief masses being equal to zero were observed. Relationships of hidden conflicts and auto-conflicts of belief functions are pointed out. We are focused on hidden auto-conflicts here — on hidden conflicts appearing when three or more numerically same belief functions are combined. Hidden auto-conflict is a kind of internal conflict. Degrees of hidden auto-conflicts and full non-conflictness are defined and analysed. Finally, computational issues of hidden auto-conflicts and non-conflictness are presented.

Keywords: Belief functions; Dempster-Shafer theory; Uncertainty; Conflicting belief masses; Internal conflict; Conflict between belief functions; Auto-Conflict; Hidden conflict; Hidden auto-conflict; Full non-conflictness.

1 Introduction

When combining belief functions (BFs) by the conjunctive rules of combination, some conflicts often appear (they are assigned either to $\emptyset$ by non-normalised conjunctive rule $\textcircled{\cap}$ or distributed among other belief masses by normalisation in Dempster's rule of combination $\oplus$). Combination of conflicting BFs and interpretation of their conflicts are often questionable in real applications. Thus a series of papers related to conflicts of BFs was published, e.g. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. A new interpretation of conflicts of belief functions was introduced in [11]: important distinction of internal conflicts of individual BFs (due to their inconsistency) from conflicts between BFs (due to conflict/contradiction of evidences represented by the BFs) was introduced there. Note that zero sum of all multiples of conflicting belief masses (denoted by $m_{\textcircled{\cap}}(\emptyset)$) is usually considered as non-conflictness of the belief functions in all these approaches.

When analyzing the conflict between belief functions based on their non-conflicting parts[1] defined by Daniel in [4] a positive value of conflict was observed even in a situation when sum of all multiples of conflicting belief masses equals to zero. This arose a series of new questions: how to interpret the sum of conflicting masses, is the conflict based on non-conflicting parts of belief functions correct? These questions are studied in [13]. The answers are positive in favour of the conflict based on non-conflicting parts. This led to a definition of a hidden conflict of BFs there.

Different levels / degrees of hidden conflicts are defined and investigated there. In correspondence to the degrees of hidden conflict, there are studied different degrees of non-conflictness, including full non-conflictness and conditions, under which belief functions are fully non-conflicting. In accordance with the original Daniel's approach from [11], there are observed and investigated not only hidden conflicts between two belief functions, but also internal hidden conflicts of individual BFs. The research covers also computational aspects of hidden conflict.

[1] Conflicting and non-conflicting parts of belief functions originally come from [12].

By investigating a hidden conflict, we have noticed a hidden auto-conflict. Auto-conflict is a term describing hidden conflict defined by Martin's et al. in [14, 15, 8]. It is a sum of multiples of conflicting belief masses when two or more numerically same BFs are conjunctively combined. An idea of auto-conflict of any positive order was defined and briefly presented in 2006 [14] and further studied in [15] two years later. Our current contribution is focused to investigation of hidden auto-conflict and its relation to original Martin's et al. results.

2 Preliminaries

We assume classic definitions of basic notions from theory of *belief functions* [16] on finite exhaustive frames of discernment $\Omega_n = \{\omega_1, \omega_2, ..., \omega_n\}$.

A *basic belief assignment (bba)* is a mapping $m : \mathcal{P}(\Omega) \longrightarrow [0,1]$ such that $\sum_{A\subseteq\Omega} m(A) = 1$; the values of the bba are called *basic belief masses (bbm)*. $m(\emptyset) = 0$ is usually assumed. $\mathcal{P}(\Omega) = \{X|X \subseteq \Omega\}$ is *power-set* of Ω. A *belief function (BF)* is a mapping $Bel : \mathcal{P}(\Omega) \longrightarrow [0,1]$, $Bel(A) = \sum_{\emptyset\neq X\subseteq A} m(X)$. A *plausibility function* $Pl(A) = \sum_{\emptyset\neq A\cap X} m(X)$. Because there is a unique correspondence among m and corresponding Bel and Pl thus we often speak about m as of belief function.

A *focal element* is a subset of the frame of discernment $X \subseteq \Omega$, such that $m(X) > 0$. In the case of $0 < |X| < n$ it is a proper focal element. If all the focal elements are *singletons* (i.e. one-element subsets of Ω), then we speak about a *Bayesian belief function* (BBF); in fact, it is a probability distribution on Ω. If there are only focal elements such that $|X| = 1$ or $|X| = n$ we speak about *quasi-Bayesian BF* (qBBF). In the case of $m(\Omega) = 1$ we speak about *vacuous BF* (VBF) and otherwise about *a non-vacuous BF*; in the case of the only focal element $\emptyset \neq X \subset \Omega$, i.e., if $m(X) = 1$, we speak about *a categorical BF*. If all focal elements have a non-empty intersection, we speak about *a consistent BF*; and if all of them are nested, about *a consonant BF*.

Dempster's (normalized conjunctive) rule of combination $\oplus$ is given as

$$(m_1 \oplus m_2)(A) = \sum_{X\cap Y=A} K m_1(X) m_2(Y)$$

for $A \neq \emptyset$, where $K = \frac{1}{1-\kappa}$, $\kappa = \sum_{X\cap Y=\emptyset} m_1(X)m_2(Y)$, and $(m_1 \oplus m_2)(\emptyset) = 0$, see [16]. Putting $K = 1$ and $(m_1 \circledcirc m_2)(\emptyset) = \kappa$ we obtain the *non-normalized conjunctive rule of combination* $\circledcirc$, see e. g. [17].

Smets's *pignistic probability* is given by $BetP(\omega_i) = \sum_{\omega_i\in X\subseteq\Omega} \frac{1}{|X|}\frac{m(X)}{1-m(\emptyset)}$, see e.g. [17]. *Normalized plausibility of singletons*[2] of Bel is a probability distribution Pl_P such that $Pl\_P(\omega_i) = \frac{Pl(\{\omega_i\})}{\sum_{\omega\in\Omega} Pl(\{\omega\})}$ [18, 19].

A conflict of BFs Bel', Bel'' *based on their non-conflicting parts* is defined by the expression $Conf(Bel', Bel'') = (m'_0 \circledcirc m''_0)(\emptyset)$, where non-conflicting part Bel_0 (of a BF Bel) is unique consonant BF such that $Pl\_P_0 = Pl\_P$ (normalized plausibility of singletons corresponding to Bel_0 is the same as that corresponding to Bel). For an algorithm to compute Bel_0 see [4].

The auto-conflict of order s of a belief function Bel given by bba m if defined by

$$a_s(m) = (\textcircled{\scriptsize\cap}_{i=1}^{s} m)(\emptyset),$$

where $s \geq 1$ and $\circledcirc$ is the non-normalized conjunctive combination; and simply $a_2(m) = (m \circledcirc m)(\emptyset)$, [15, 14].

The basic properties of auto-conflict are the following[15, 14]:

$$a_s(m) \leq a_{s+1}(m),$$

and

$$a(m) = a_2(m) > 0 \quad \text{implies} \quad lim_{s\to\infty} a_s(m) = 1.$$

[2] A plausibility of singletons is called a *contour function* by Shafer in [16], thus $Pl\_P(Bel)$ is a normalization of a contour function in fact.

3 Hidden Conflicts of Belief Functions

3.1 An Introductory Example

Let us suppose two simple consistent belief functions Bel' and Bel'' on a three-element frame of discernment $\Omega_3 = \{\omega_1, \omega_2, \omega_3\}$ given by the bbas $m'(\{\omega_1, \omega_2\}) = 0.6$, $m'(\{\omega_1, \omega_3\}) = 0.4$, and $m''(\{\omega_2, \omega_3\}) = 1.0$. Then $(m' \circledcirc m'')(\emptyset) = 0$ what seems — and it is usually considered — to be a non-conflictness of m' and m'', but there is positive conflict based on non-conflicting parts $Conf(Bel', Bel'') = (m'_0 \circledcirc m''_0)(\emptyset) = 0.4 > 0$. (This holds true despite of Theorem 4 from [4] which should be revised in a future).

We can easily verify this situation: the only focal element of m'' has a non-empty intersection with both focal elements of m', thus $(m' \circledcirc m'')(\emptyset) = \sum_{X \cap Y = \emptyset} m'(X) m''(Y) = (\text{empty sum}) = 0$; Bel'' is already consonant itself, thus $Bel''_0 = Bel''$, $m''_0 = m''$, $Pl'(\{\omega_1\}) = 1$, $Pl'(\{\omega_2\}) = 0.6$, $Pl'(\{\omega_3\}) = 0.4$, thus $m'_0(\{\omega_1\}) = 0.4$, $m'_0(\{\omega_1, \omega_2\}) = 0.2$, $m'_0(\{\omega_1, \omega_2, \omega_3\}) = 0.4$, hence $Conf(Bel', Bel'') = (m'_0 \circledcirc m''_0)(\emptyset) = m'_0(\{\omega_1\}) m''_0(\{\omega_2, \omega_3\}) = 0.4 \cdot 1 = 0.4$.

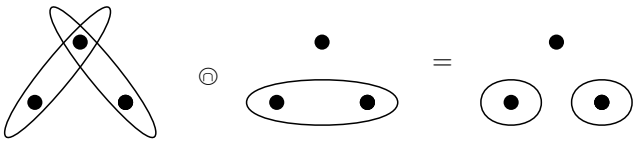


Figure 1: Introductory Example: focal elements of m', m'', and of $m' \circledcirc m''$.

3.2 Observation of Hidden Conflict

The following questions arise: Does $(m' \circledcirc m'')(\emptyset) = 0$ really represent non-conflictness of respective BFs? Is the definition of conflict based on non-conflicting parts correct? Are m' and m'' conflicting or non-conflicting? What does $(m' \circledcirc m'')(\emptyset) = 0$ mean?

We can formalize our assumptions from [13] as follows:

Assumption As1 Conjunctive combination of two mutually non-conflicting BFs Bel' and Bel'' is mutually non-conflicting with any of the individual BFs Bel' and Bel''.

By induction, Assumption **As1** can be extended as follows:

Assumption As1* Conjunctive combination of two mutually non-conflicting BFs Bel' and Bel'' is mutually non-conflicting with any number of combinations with individual BF Bel' and also with any number of combinations with the other individual BF Bel'', thus with $\textcircled{\cap}_1^k Bel'$ and with $\textcircled{\cap}_1^l Bel'$, for any $k, l \geq 1$.

Supposing symmetry of conflictness / non-conflictness we obtain **As1**** as it follows:

Assumption As1** Conjunctive combination of two mutually non-conflicting BFs Bel' and Bel'' is mutually non-conflicting with any number of combinations of any of both the individual BFs Bel' and/or Bel'', thus with $\textcircled{\cap}_1^k Bel' \circledcirc \textcircled{\cap}_1^l Bel'$, for any $k, l \geq 0$, $k + l \geq 1$.

Thus Assumption **As1*** is just a reformulation of **As1** and supposing symmetry of conflictness / non-conflictness also Assumption **As1**** is just a reformulation of **As1**.

Based on Assumption **As1**, we can easily show that the BF from Introductory Example are not non-conflicting: Let us suppose that Bel' and Bel'' are non-conflicting now. Thus their combination $Bel' \circledcirc Bel''$ should be also non-conflicting with both of them. Does this hold for BFs from our example? This holds true when we combine $m' \circledcirc m''$ with m'' one more time. It follows from the idempotency of categorical m'': $m' \circledcirc m'' \circledcirc m'' = m' \circledcirc m''$ and therefore $(m' \circledcirc m'' \circledcirc m'')(\emptyset) = 0$ again. On the other hand, we obtain positive $(m' \circledcirc m'' \circledcirc m')(\emptyset) = (m' \circledcirc m' \circledcirc m'')(\emptyset) = 0.48$. See Table 1 and Figure 2. When m'' and m' are combined once, then we observe $m_{\circledcirc}(\emptyset) = 0$. When combining m'' with m' twice then $m_{\circledcirc}(\emptyset) = 0.48$. We observe some kind of a hidden conflict. And we have an argument for correctness of positive value of $Conf(Bel', Bel'')$.

Analogous result was obtained using a decisional interpretation of the BFs [13]. Hence $(m' \circledcirc m'')(\emptyset)$ really does not mean non-conflictness of the BFs. It means a simple or partial compatibility of respective focal elements only.

Table 1: Hidden conflict in the Introductory Example

X :	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	Ω_3	$\emptyset$
$m'(X)$:	0.0	0.0	0.0	0.60	0.40	0.00	0.00	—
$m''(X)$:	0.0	0.0	0.0	0.00	0.00	1.00	0.00	—
$(m'\circledcirc m'')(X)$:	0.00	0.60	0.40	0.00	0.00	0.00	0.00	0.00
$(m'\circledcirc m''\circledcirc m'')(X)$:	0.00	0.60	0.40	0.00	0.00	0.00	0.00	0.00
$(m'\circledcirc m''\circledcirc m')(X)$:	0.00	0.36	0.16	0.00	0.00	0.00	0.00	0.48
$(m'\circledcirc m''\circledcirc m'\circledcirc m'')(X)$:	0.00	0.36	0.16	0.00	0.00	0.00	0.00	0.48

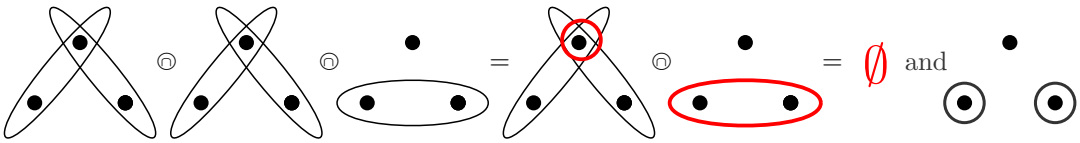


Figure 2: Arising of a hidden conflict between BFs in the Introductory Example: focal elements of m', m', m'' — $m'\circledcirc m', m''$ and of $(m'\circledcirc m')\circledcirc m''$.

3.3 A Simple Definition of Hidden Conflict and its Relationship to Auto-Conflict

Definition 1 *Let us suppose two BFs Bel', Bel'' defined by bbas m', m'', such that $(m'\circledcirc m'')(\emptyset) = 0$. If there further holds $(m'\circledcirc m''\circledcirc m')(\emptyset) > 0$ or $(m'\circledcirc m''\circledcirc m'')(\emptyset) > 0$ we say that there is* a hidden conflict *of the BFs.*

Observation 1 *A condition $(m'\circledcirc m''\circledcirc m')(\emptyset) > 0$ or $(m'\circledcirc m''\circledcirc m'')(\emptyset) > 0$ from Definition 1 is equivalent to the following condition $(m'\circledcirc m''\circledcirc m'\circledcirc m'')(\emptyset) > 0$.*

We have to note that a hidden conflict is quite a new phenomenon first time defined in [13], it is qualitatively different from the other referred approaches, even different form ideas of all Daniel's works on conflict of belief functions from previous years. Till now, it was supposed that $m_{\circledcirc}(\emptyset)$ includes both all conflicts between BFs and also all internal conflicts of individual BFs. Thus conflict between BFs was supposed to be less or equal to $m_{\circledcirc}(\emptyset)$. Here, we deal with a situation of a positive conflict between BFs while $m_{\circledcirc}(\emptyset) = 0$. The presented approach is new and different from all previous ones, but it is definitely not against all of the previous approaches, especially not against the conflict between BFs based on their non-conflicting parts, which has enabled observation of hidden conflict and which is supported by the existence of a hidden conflict.

We have already observed that $m_{\circledcirc}(\emptyset) = 0$ does not mean full non-conflictness of BFs and that the condition $(m'\circledcirc m''\circledcirc m'\circledcirc m'')(\emptyset) > 0$ together with $(m'\circledcirc m'')(\emptyset) = 0$ defines a hidden conflict. What about the condition $(m'\circledcirc m''\circledcirc m'\circledcirc m'')(\emptyset) = 0$? Is this condition sufficient for full non-conflictness of BFs Bel' and Bel''? May some conflict be still hidden there?

There are repeated combinations of m' and m'' which resembles Martin's auto-conflict. There is $a(m') = (m'\oplus m')(\emptyset) = 0$ and also $a(m'') = (m''\oplus m'')(\emptyset) = 0$, thus there is no auto-conflict of the individual input BFs. But $a(m'\oplus m'') = ((m'\oplus m'')\oplus(m'\oplus m'')(\emptyset) = (m'\circledcirc m''\circledcirc m'\circledcirc m'')(\emptyset) > 0$. The simple definition of the hidden conflict is: $(m'\circledcirc m'')(\emptyset) = 0$ and $a(m'\oplus m'') > 0$. Thus the question from the previous example may be reformulated: Is the condition $a(m'\oplus m'') = 0$ sufficient for full non-conflictness of BFs Bel' and Bel''?

We have shown that validity of the condition $(m'\circledcirc m''\circledcirc m'\circledcirc m'')(\emptyset) = 0$ is sufficient for full non-conflictness of BFs Bel' and Bel'' only on Ω_3 - on the frame of discernment of the Introductory Example. It is not sufficient in general. To solve the question in general, we have to consider a larger frame of discernment.

3.4 Little Angel Example

For Ω_5 one can find the following Little Angel Example (see Table 2 and Figure 3). Similarly to Introductory Example, we have two consistent BFs Bel^i and Bel^{ii} with disjoint sets of max-

Table 2: Hidden Conflict in the Little Angel Example

X :	$A=\{\omega_1,\omega_2,\omega_5\}$	$B=\{\omega_1,\omega_2,\omega_3,\omega_4\}$	$C=\{\omega_1,\omega_3,\omega_4,\omega_5\}$	$X=\{\omega_2,\omega_3,\omega_4,\omega_5\}$	$\emptyset$
$m^i(X)$:	0.1	0.30	0.60	0.00	—
$m^{ii}(X)$:	0.0	0.00	0.00	1.00	—

X :	$A\cap X$	$B\cap X$	$C\cap X$	$A\cap B\cap X$	$A\cap C\cap X$	$B\cap C\cap X$	$\emptyset$
$(m^i\circledcirc m^{ii})(X)$:	0.1	0.3	0.6	0.0	0.0	0.0	0.000
$(m^i\circledcirc m^{ii}\circledcirc m^{ii})(X)$:	0.10	0.30	0.60	0.00	0.00	0.00	0.00
$(m^i\circledcirc m^i\circledcirc m^{ii})(X)$:	0.01	0.09	0.36	0.06	0.12	0.36	0.00
$(m^i\circledcirc m^i\circledcirc m^{ii}\circledcirc m^{ii})(X)$:	0.01	0.09	0.36	0.06	0.12	0.36	0.00
$(m^i\circledcirc m^{ii}\circledcirc m^{ii}\circledcirc m^{ii})(X)$:	0.010	0.090	0.360	0.060	0.120	0.360	0.000
$(m^i\circledcirc m^i\circledcirc m^i\circledcirc m^{ii})(X)$:	0.001	0.027	0.216	0.036	0.126	0.486	0.108
$m^*(X)$:	0.001	0.027	0.216	0.036	0.126	0.486	0.108

where $m^* = m^i\circledcirc m^i\circledcirc m^i\circledcirc m^{ii}\circledcirc m^{ii}\circledcirc m^{ii}$.

plausibility elements and where zero condition $(m^i\circledcirc m^{ii})(\emptyset)=0$ holds true. Moreover here holds also $(m^i\circledcirc m^{ii}\circledcirc m^i\circledcirc m^{ii})(\emptyset)=0$ (see Table 2) while $Conf(Bel^i,Bel^{ii})=0.1$ is positive again. Positiveness of the $Conf$ value can be easily seen from the fact that sets of max-plausibility elements are disjoint for Pl^i and Pl^{ii}. Numerically, we have again $Bel_0^{ii}=Bel^{ii}$, and $Pl\_P^i=(^{10}/_{39},{}^{4}/_{39},{}^{9}/_{39},{}^{9}/_{39},{}^{7}/_{39})$. We obtain $m_0^i(\{\omega_1\})=0.1, m_0^i(\{\omega_1,\omega_3,\omega_4\})=0.2, m_0^i(\{\omega_1,\omega_3,\omega_4,\omega_5\})=0.3, m_0^i(\{\Omega_5\})=0.4$, and $Conf(Bel^i,Bel^{ii})=m_0^i(\{\omega_1\})m^{ii}(X)=0.1$.

Analogous arguments hold true for the positive $Conf$ and hidden conflict again (of the 2nd degree this time): (i) $BetP^i=(0.2583,0.1083,0.225,0.225,0.1833)$ which is not numerically the same as Pl_P^i, but both prefer ω_1, whereas $BetP^{ii}=Pl\_P^{ii}=(0.00,0.25,0.25,0.25,0.25)$.
(ii) Despite to the assumption As1* there is $((m^i\circledcirc m^{ii})\circledcirc(m^i\circledcirc m^i))(\emptyset)=0.108>0$, see Table 2.

In language of auto-conflicts there is $a_2(m^i\circledcirc m^{ii})=0$ and $a_3(m^i\circledcirc m^{ii})>0$.

For an existence of a hidden conflict, it is the structure of focal elements that is important — not their belief masses. Belief masses are important for the size of a conflict. In general, we can take $m^i(A)=a$, $m^i(B)=b$, $m^i(C)=c$, for A,B,C defined in Table 2 and for any $a,b,c>0$, such that $a+b+c=1$ and we obtain $m(\emptyset)=6abc$ as a hidden conflict of the 2nd degree (in our numeric case there is $6abc=6\cdot 0.1\cdot 0.3\cdot 0.6=0.108$). For a graphical representation of the Little Angel Example, see Figure 3.

Degrees of hidden conflict, its maximal value, and the issue of full non-conflictness will be briefly introduced in the following subsection.

3.5 Degrees of Hidden Conflict and Full Non-conflictness

When analyzing examples from the previous section, we have observed different levels where conflicts were hidden. We can formalize degrees of hidden conflict:

Definition 2 *Assume two BFs Bel^i, Bel^{ii} defined by bbas m^i, m^{ii}, such that for some $k>0$ $(\textcircled{\cap}_{j=1}^{k}(m^i\circledcirc m^{ii}))(\emptyset)=0$. If there further holds $(\textcircled{\cap}_{j=1}^{k+1}(m^i\circledcirc m^{ii}))(\emptyset)>0$ we say that there is* a hidden conflict of the k-th degree *of BFs Bel^i and Bel^{ii}.*

Analogously to a particular degree of hidden conflict, there are degrees of non-conflictness as well. Particular degree of non-conflictness is not important. However, there is an important question whether there is some hidden conflict or not, i.e. whether or not the BFs in question are fully non-conflicting.

Definition 3 *We say that BFs Bel^i and Bel^{ii} are* fully non-conflicting *if there is no hidden conflict of any degree. I.e. if $(\textcircled{\cap}_{j=1}^{k}(m^i\circledcirc m^{ii}))(\emptyset)=0$ for any $k>0$.*

Thus there is a question of how many times we have to combine $(m^i\circledcirc m^{ii})$, i.e., for which k value of $(\textcircled{\cap}_{j=1}^{k}(m^i\circledcirc m^{ii}))(\emptyset)$ shows whether there is some hidden conflict of the BFs Bel^i and Bel^{ii} or not.

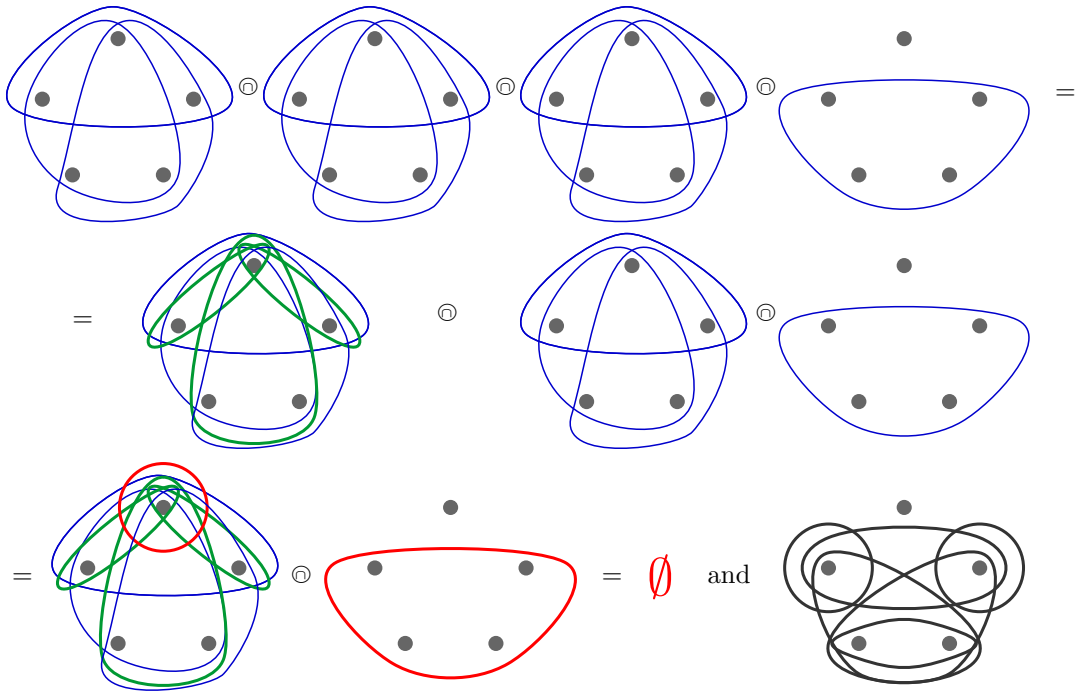


Figure 3: Arrising of a hidden conflict between BFs in the Little Angel Example. Focal elements of m^i, $m^i \circledcap m^i$, $m^i \circledcap m^i \circledcap m^i$ and of $(m^i \circledcap m^i \circledcap m^i) \circledcap m^{ii}$. Red-colored focal elements are those responsible for creation of the empty-set in the last step.

Theorem 1 (maximal degree of hidden conflict) *For any non-vacuous BFs Bel^i and Bel^{ii} on any frame Ω_n it holds that*

$$(\textcircled{\cap}_{j=1}^{n-1}(m^i \circledcap m^{ii}))(\emptyset) = 0 \quad \textit{iff} \quad (\textcircled{\cap}_{j=1}^{k}(m^i \circledcap m^{ii}))(\emptyset) = 0 \tag{1}$$

for any $k > n-2$.

Considering the notion of auto-conflict we obtain the following theorem:

Theorem 2 *For any two BFs Bel^i and Bel^{ii} defined by bbas m^i and m^{ii} the following holds:*
(i) There is a hidden conflict of the k-th degree *of BFs Bel^i and Bel^{ii} if and only if* $a_k(m^i \circledcap m^{ii}) = 0$ & $a_{k+1}(m^i \circledcap m^{ii}) > 0$.
(ii) Bel^i and Bel^{ii} are fully non-conflicting if and only if auto-conflict of any order of their conjunctive combination is zero, i.e., if and only if $a_k(m^i \circledcap m^{ii}) = 0$ for any $k > 0$.
(iii) If Bel^i and Bel^{ii} are non-vacuous BFs on any finite frame of discernment Ω_n it holds that

$$a_{n-1}(m^i \circledcap m^{ii}) = 0 \quad \textit{iff} \quad a_k(m^i \circledcap m^{ii}) = 0 \ \textit{ for any } k > n-2.$$

4 Hidden Auto-Conflict of a Belief Function

4.1 Examples and Definition of Hidden Auto-Conflict

Let us return to properties of auto-conflict. In general, the following holds true:

$$a_s(m) \leq a_{s+1}(m) \ \ \& \ \ a_2(m) > 0 \Rightarrow lim_{s\to\infty} a_s(m) = 1$$

by [14]. It is stated in [15] that $lim_{s\to\infty} a_s(m) = 1$ holds true (without any assumption). This is not correct generally, (any consistent BF is a counter-example) but under non-explicitly stated assumption of BFs considered in [15]. I.e. a special subclass of quasi-Bayesian BFs, BFs with all singletons plus Ω_n as their focal elements. I.e. qBBFs with exactly $n+1$ focal elements.

Such BFs have always $a_2(m) > 0$. On the other hand for all the consistent BFs it hods that $a(m) = a_2(m) = 0 = a_k(m)$, for any $k > 0$. Even $lim_{k\to\infty} a_k(m) = 0$.

For any other BFs it holds that $a(m) \geq 0$ and $a_s(m) \leq a_{s+1}(m)$. Moreover it holds that $a_{k-1}(m) < a_k(m)$ implies $a_k(m) < a_{k+1}(m)$. This follows the number and cardinalities of focal elements of $a_k(m)$; $a_k(m)$ has all focal elements of $a_{k-1}(m)$, plus possibly some additional focal elements (defined by intersection of focal elements of $a_{k-1}(m)$). Furthermore, if $a_{k-1}(m)$ and $a_k(m)$ have the same focal elements, this holds true also for any $a_s(m)$ where $s > k$. Specially, whenever $\emptyset$ is a focal element of $a_k(m)$ it is also a focal element of any $a_s(m)$ where $s > k$ and $a_s(m) < a_{s+1}(m)$.

If $a(m) = a_s(m) = 0$ for a non-consistent BF Bel and some $s > 1$, then there exists some $k > s$, such that $a_k(m) > 0$. We can call such an auto-conflict as hidden auto-conflict. There was not hidden auto-conflict in our Introductory example of hidden conflict, but there is a hidden auto-conflict in the Little Angel Example:

Hidden auto-conflict in the Little Angel Example: Let us compute conjunctive combination of the input BFs Bel^i and Bel^{ii}: $m = m^i \oplus m^{ii}$. We obtain $m(\{\omega_2, \omega_5\}) = 0.1$, $m(\{\omega_2, \omega_3, \omega_4\}) = 0.3$, $m(\{\omega_3, \omega_4, \omega_5\}) = 0.6$, $m(\emptyset) = 0$ (See line $(m^i \oplus m^{ii})(X)$ in Table 2.)

There is no auto-conflict: $a(m) = 0$ (See line $(m^i \textcircled{\cap} m^i \textcircled{\cap} m^{ii} \textcircled{\cap} m^{ii})(X)$ in Table 2.) There is a (positive) auto-conflict of degree 3: $a_3(m) = 0.108$ (See line $(m^*(X)$ in Table 2.) Thus the positive auto-conflict of the order 3 $a_3(m) = 0.108$ is hidden by zero auto-conflict $a(m)$.

For focal elements and arising of the hidden auto-conflict of $m = m^i \oplus m^{ii}$ see Figure 4.

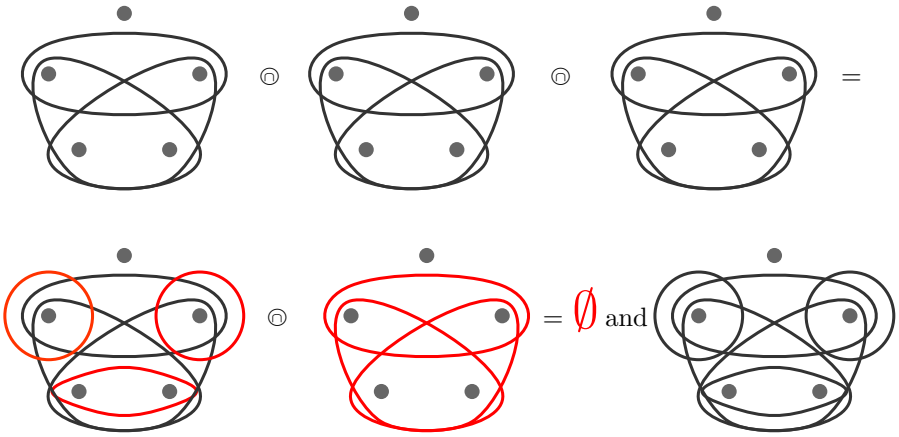


Figure 4: Arising of a hidden auto-conflict of BF $Bel^i \textcircled{\cap} Bel^{ii}$, focal elements of $Bel^i \textcircled{\cap} Bel^{ii}$, $\textcircled{\cap}_1^2(Bel^i \textcircled{\cap} Bel^{ii})$, and $\textcircled{\cap}_1^3(Bel^i \textcircled{\cap} Bel^{ii})$.

Definition 4 *We say that BF Bel defined by bba m has a* hidden auto-conflict*, if its auto-conflict* $a(m) = a_2(m) = 0$ *and if* $a_s(m) > 0$ *for some* $s > 0$.

We say that BF Bel defined by bba m have hidden auto-conflict of degree s *if* $a_{s+1}(m) = 0$ *and* $a_{s+2}(m) > 0$.

Example 1 *We can find an example of hidden auto-conflict even on* Ω_3 *frame of discernment. Let us suppose BF given by bba* m^{iii} *from Table 1; we obtain* $a(m^{iii}) = 0$ *and* $a_3(m^{iii}) = 0.18$ *there. Hence again, the positive auto-conflict of the order 3* $a_3(m^{iii}) = 0.18$ *is hidden by zero auto-conflict* $a(m^{iii})$*. For focal elements and arising of the hidden auto-conflict of* $m = m^i \oplus m^{ii}$ *see Figure 6.*

Example 2 (A general example) *We can take conjunctive sum of any two BFs with hidden conflict of 2nd degree, thus* $Bel = Bel^i \textcircled{\cap} Bel^{ii}$ *: there is* $(m^i \textcircled{\cap} m^i \textcircled{\cap} m^{ii} \textcircled{\cap} m^{ii})(\emptyset) = 0 = a(m^i \textcircled{\cap} m^{ii})$ *and* $a_3(m^i \textcircled{\cap} m^{ii}) = (m^i \textcircled{\cap} m^i \textcircled{\cap} m^i \textcircled{\cap} m^{ii} \textcircled{\cap} m^{ii} \textcircled{\cap} m^{ii})(\emptyset) > 0$*. Thus the positive* $a_3(m)$ *is hidden by* $a(m) = 0$.

4.2 Maximal Degree of Hidden Auto-Conflict

The following theorem provides a solution to the question whether BF Bel on Ω_n has any auto-conflict or whether it is completely non-conflicting.

Table 3: Example of a hidden auto-conflict on Ω_3

X :	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	Ω_3	$\emptyset$
$m^{iii}(X)$:	0.0	0.0	0.0	0.50	0.30	0.20	0.00	—
$(m^{iii}\textcircled{\cap}m^{iii})(X)$:	0.30	0.20	0.12	0.25	0.09	0.04	0.00	0.00
$(m^{iii}\textcircled{\cap}m^{iii}\textcircled{\cap}m^{iii})(X)$:	0.36	0.21	0.09	0.125	0.027	0.008	0.00	0.18

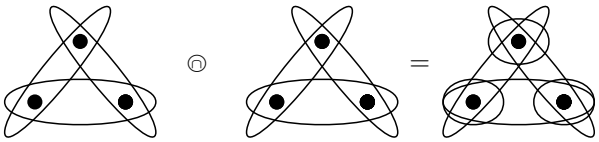


Figure 5: Focal elements of m^{iii} and $m^{iii}\textcircled{\cap}m^{iii}$; no auto-conflict of non-consistent BF Bel^{iii}: $a_2(m^{iii}) = 0$.

Theorem 3 *For any belief function Bel on Ω_n the following holds:*

$$a_n(Bel) = 0 \quad \textit{iff} \quad a_k(Bel) = 0 \ \textit{ for all } \ k \geq n. \tag{2}$$

Thus, maximal degree of hidden auto-conflict is equal to $n-2$.

Idea of proof. Number of focal elements (f.e.) is decreased until it is fixed. Thus there is at most $n-1$ decreases (creations of less f.e.); n-times Bel, $(n-1)$-times $\textcircled{\cap}$.

For a given BF Bel it may be sufficient to compute auto-conflict of even lesser degree:

Theorem 4 *For a given belief function Bel on Ω_n the following holds:*

$$a_{s+1}(Bel) = 0 \quad \textit{iff} \quad a_k(Bel) = 0 \ \textit{ for all } \ k \geq s+1, \tag{3}$$

where s is maximal cardinality of a proper focal element of Bel (focal element different from Ω_n). Thus Bel may have hidden auto-conflict up to degree $s-1$.

Idea of proof. Analogously to the proof of Theorem 3, there is at most s decreases for $s < n$.

According to this theorem we can see that a special class of qBBFs with just $n+1$ focal element has maximal cardinality of a proper focal element 1 and maximal degree of hidden auto-conflict 0. Thus there is no hidden auto-conflict on the special class of BFs on which auto-conflicts were studied by Martin et al. [15]. Similarly, according to Theorem 3, we can see that there is no hidden auto-conflict of any BF on Ω_2.

For a general BF on Ω_n, there are proper focal elements up to cardinality $s = n-1$. Thus $s+1 = n$ and we can consider Theorem 3 to be a special case of Theorem 4.

We have an upper bound for a degree of hidden conflict. And we can ask the following questions: May be this upper bound reached? How the BFs with maximal degree of auto-conflict looks like?

Example 3 (A general example) *Similarly to hidden conflicts in general, structure of focal elements is important for existence and particular degree of hidden auto-conflicts (not values of belief masses - values are important for resulting value of hidden auto-conflicts). Thus we can present a BF $Bel^{(n)}$ on Ω_n, having all subsets of Ω_n of cardinality $n-1$ as its focal elements. For simplicity we may assume $m^{(n)}(X) = 1/n$ for all $|X| = n-1$. We obtain $a_{n-1}(Bel^{(n)}) = 0$ and $a_n(Bel^{(n)}) > 0$. There is a hidden conflict of $n-2$ degree.*

Example 4 *[Specific examples] (i) Simple example of maximal degree of hidden auto-conflict on Ω_3 (thus maximal degree is 1 there) was already presented in Example 1. There is $a(m^{iii}) = 0$ and $a_3(m^{iii}) = 0.18$.*
(ii) Having $Bel^{(3)}$ from Example 3 with $m^{(3)}(X) = 1/3$, we obtain $a(m^{(3)}) = 0$ and $a_3(m^{(3)}) = 2/9 = 0.22\overline{2}$. ($(\textcircled{\cap}_1^3 m^{(3)})(\{\omega_i\}) = 2/9$ and $(\textcircled{\cap}_1^3 m^{(3)})(X) = 1/27$ for $|X| = 2$).

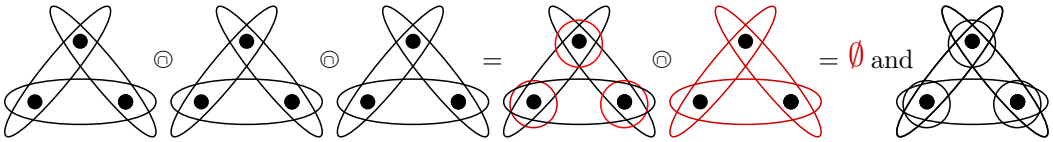


Figure 6: Arising of a hidden auto-conflict of BF Bel^{iii} on Ω_3; focal elements of m^{iii}, $m^{iii} \textcircled{\cap} m^{iii}$ and $m^{iii} \textcircled{\cap} m^{iii} \textcircled{\cap} m^{iii}$.

(iii) Let us suppose $Bel^{(16)}$ *on* Ω_{16} *now:* $m^{(16)}(X) = 1/16 = 0.0625$ *for all* $|X| = 15$ *and* $m^{(16)}(X) = 0$ *otherwise. We obtain* $a_{15}(Bel^{(16)}) = 0$ *and* $a_{16}(Bel^{(16)}) = 1.134227 \cdot 10^{-6}$. *Thus there is very small hidden auto-conflict of* 14^{th} *degree. Value of the auto-conflict rapidly grows up with order:* $a_{18}(Bel^{(16)}) = 4.428801 \cdot 10^{-5}$, $a_{24}(Bel^{(16)}) = 5.460487 \cdot 10^{-3}$, $a_{36}(Bel^{(16)}) = 0.1480689$.
(iv) Supposing m^{xvi} *with bbms* $m^{xvi}(X)$*:* 0.005, 0.010, 0.015, 0.020, 0.030, 0.040, 0.050, 0.060 0.065, 0.075, 0.085, 0.095, 0.105, 0.110, 0.115, 0.120 *for* $|X| = 15$ *and* $m^{xvi}(X) = 0$ *otherwise we obtain:* $a_{15}(Bel^{xvi}) = 0$ *and* $a_{16}(Bel^{xvi}) = 7.089108 \cdot 10^{-9}$. *Thus even less value of auto-conflict of* 14^{th} *degree than in (iii). Value of the auto-conflict rapidly grows up with degree again:* $a_{18}(Bel^{xvi}) = 2.789497 \cdot 10^{-7}$, $a_{24}(Bel^{xvi}) = 4.029782 \cdot 10^{-5}$, $a_{36}(Bel^{xvi}) = 0.001939817$.

Theorem 5 *Belief functions on* Ω_n *with maximal degree* $n-2$ *of hidden auto-conflict are just BF from one of the following categories:*
(i) $Bel^{(n)}$ *with structure as in Example 3*
(ii) $Bel^{(n)}$ *with structure as in Example 3 extended with focal element* Ω_n
There are no other BFs on Ω_n *with hidden auto-conflict of degree* $n-2$.

Idea of proof. Both the classes are obvious. Decreasing number of focal elements removes the hidden auto-conflict; decreasing cardinality of any of the focal elements decreases degree of hidden conflict/auto-conflict.

Remark We can compare very similar equations (1) and (2) in Theorems 1 and 3 differing just in one order of auto-conflict. Is this difference correct? YES, it is. From Theorem 5 we can see that for obtaining maximal degree of a hidden auto-conflict, we need combination of n pieces of BF $Bel^{(n)}$, thus we need multiples of n focal elements of cardinality $(n-1)$. The same we need in the situation from Theorem 1: for maximal degree of hidden conflict all focal elements X of both the BFs have to satisfy $|X| \geq n-1$; but in this case $(m^i \oplus m^{ii})$ is an argument of $\textcircled{\cap}_{j=1}^{n-1}$, thus, (at least) one of the focal elements s.t. $X = |n-1|$ is a focal element of let us say Bel^i and therefore it is enough to combine only from $n-1$ pieces of BF Bel^{ii} to obtain the intersection of all n focal elements s.t. $X = |n-1|$. Thus there is really less order of auto-conflict sufficient in Theorem 1.

4.3 Characterisation and Overview of Properties of Auto-Conflict and of Hidden Auto-Conflict

Simply, any consistent belief function has no auto-conflict of any order and therefore no hidden auto-conflict as well. (Any intersection of any number of focal elements of a consistent BF is non-empty, thus $(\textcircled{\cap}_1^k m)(\emptyset) = 0$ for any $k > 0$).

For non-consistent belief function Bel on Ω_n there are possible auto-conflicts of some degrees. From the examples shown earlier in this section, we have seen that auto-conflict $a(Bel)$ may be equal to zero, thus that there is a hidden auto-conflict. From Theorem 3 we know that degree of hidden auto-conflict is at most $n-2$, i.e., that any non-consistent BF has (positive) auto-conflict of order n.

Auto-conflict is utilized by Martin et al. as an alternative measure of internal conflict for belief functions. But, values of auto-conflicts of higher orders have no reasonable interpretation and a way how to compare them with values of auto-conflict (of 2nd order). Hidden auto-conflict may be considered as an extension of this measure of internal conflict. Unfortunately, analogously to values of higher order auto-conflicts, we have no procedure to compare the values of hidden auto-conflicts of different degrees.

Theorem 6 *For auto-conflicts of a belief function Bel on Ω_n, the following holds:*
(i) Auto-conflict of a consistent BF of any order is zero (there is no auto-conflict in fact). Also $lim_{s\to\infty} a_s(Bel) = 0$ for any consistent BF Bel.
(ii) Auto-conflict of order 1 is e equal to $m(\emptyset)$, hence it is zero for any normalised BF Bel.
(iii) Auto-conflicts of orders from 2 to $k+1$ are zero for any non-consistent BF, where $k \leq n-2$ is a degree of hidden auto-conflict; specially,
for $k = 0$ $a_2(Bel) > 0$ and also $a_s(Bel) > 0$ for all $s > 2$,
for $k = 1$ $a_2(Bel) = 0$ and $a_s(Bel) > 0$ for all $s > 2$,
for $k = 2$ $a_2(Bel) = a_3(Bel) = 0$ and $a_s(Bel) > 0$ for all $s > 3$, etc...
(iv) Positive auto-conflict is increasing: if $a_s(Bel) > 0$ then $a_s(Bel) < a_{s+1}(Bel)$, thus $a_1(Bel) = a_2(Bel) = ... = a_{k+1}(Bel) = 0 < a_{k+2}(Bel) < a_{k+3}(Bel) < ...$, for $0 \leq k \leq n-2$. And $lim_{s\to\infty} a_s(Bel) = 1$ for any non-consistent BF.

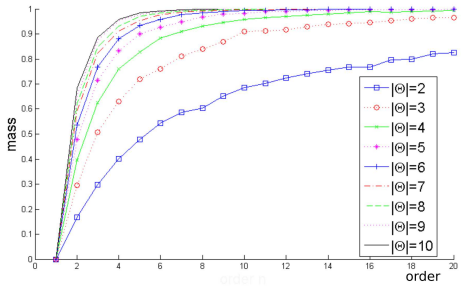
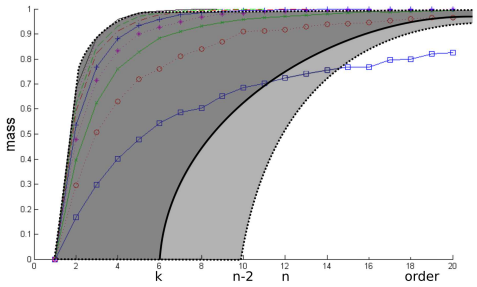


Figure 7: Average auto-conflicts of qBBFs from [15], $|\Theta| = |\Omega| = 2 - 10$. ($|X| = 1$ for all proper focal elements).



.Figure 8: Auto-conflicts of general non-consistent .BFs and of non-consistent BFs with proper focal elements X up to $|X| = k$ (darker) on $|\Omega_{12}| = 12$

Hidden auto-conflicts of qBBFs from [15] are just around the curves from $[1, 0]$ to $[\infty, 1]$ see Figure 7; for proper focal elements holds $|X| = 1$ there. Whereas hidden auto-conflicts of general non-consistent BFs are in the entire grey area on Figure 8, BF with proper focal elements X up to $|X| = k$ in darker grey area above the curve from $[k, 0]$ to $[\infty, 1]$, and zero auto-conflicts of consistent BFs are on straight line $[0, 1]$ to $[\infty, 0]$.

Yes, auto-conflict is an intrinsic property of belief functions as it is stated in [15] but, $a(Bel) \geq 0$ in general. Hence it may be equal to zero while there may be no or some hidden auto-conflict.

We have seen, that $a(Bel) > 0$ is an intrinsic property of a class of quasi-Bayesian BFs with just $n+1$ focal elements, which was studied in [15]. More generally, it is also an intrinsic property of a class of non-consistent BFs with two or more disjunctive focal element e.g. $X \cap Y = \emptyset$; there always holds that $a(Bel) > 0$.

5 Computational Issues

Based on Definition 4 and Theorem 3, the complexity of computation of the degree of hidden auto-conflict of BF Bel is — on a general Ω_n — $O(n)$ of $\circledcirc$ operations. In the case of checking existence of a hidden auto-conflict of a BF we obtain the complexity $O(log_2(n))$ of $\circledcirc$ operations utilizing a simplification of computation based on $\circledcirc_{j=1}^{2k} m = \circledcirc_{j=1}^{k} m \circledcirc \circledcirc_{j=1}^{k} m$. Note that the complexity of $\circledcirc$ operation depends on the number and structure of focal elements.

During our analysis of hidden conflicts a series of example computations was performed on frames of discernment of cardinality from 5 up to 16. A number of focal elements rapidly grows up to $|\mathcal{P}(\Omega)| = 2^{|\Omega|} - 1$ when conjunctive combination $\circledcirc$ is repeated; see e.g. 65534 focal elements of the presented BFs $Bel^{(16)}$ and Bel^{xvi} on Ω_{16} in Example 4. Because the degree of the hidden auto-conflict and existence of the hidden auto-conflict depends on the number and the structure of focal elements not on their bbms, we have usually used same bbms for all focal elements of a BF in our computations on frames of cardinality greater than 10.

All our experiments were performed in R Language and Environment [20] using R Studio [21]. We are currently developing an R package for dealing with belief functions on various frames of

discernment. It is based on a relational database approach — nicely implemented in R in package called data.table [22].

6 Several Important Remarks

We have to underline that neither hidden conflict of belief functions nor hidden auto-conflict of a belief function are a new measure of conflict. These notions are intended for understanding conflictness / non-conflictness, they enable to point out the conflict also in situations where conflicts had not been expected, in situations where $m_{\textcircled{\cap}}(\emptyset) = 0$; hence to point out and to help to understand the conflicts which are hidden by $m_{\textcircled{\cap}}(\emptyset) = 0$.

The values either of hidden conflict or hidden auto-conflict have not yes any enough reasonable interpretation. We are only interesting whether they are zero (thus no hidden conflict is there) or whether they are positive (thus hidden conflicts appear there).

Both degrees of hidden conflict and degrees of hidden auto-conflict do not present any size or a strength of the conflict. They present the level / degree how the conflict is hidden. Thus they are rather degrees of hiddenness of the conflicts / auto-conflicts. The higher degree, the higher hiddenness, thus less conflict and less strength of the same value. We continue with notation from [13] here. It seems that conflict / auto-conflict hidden in degree k is better formulation than k-th degree of hidden (auto-)conflict or (auto-)conflict of degree k.

Repeating application of conjunctive combination $\oplus$ of a BF with itself is used here to simulate situation where different independent believers have numerically the same bbm. Thus this has nothing to do with idempotent belief combination (where no conflict between two BFs is possible).

There is brand new idea of hidden conflicts in [13] and in this contribution. The assumption of non-conflictness when $m_{\textcircled{\cap}}(\emptyset) = 0$ was relaxed, due to observation of conflict even in the cases where $m_{\textcircled{\cap}}(\emptyset) = 0$. Both these studies want to point out the existence of hidden (auto-)conflicts in situations where no conflict was expected till now. Thus the definitions of hidden conflict and hidden auto-conflict are not anything against the previous Daniel's research and results on conflict of belief functions e.g. [11, 2, 4]. Of course, same parts of the previous approaches should be updated to be fully consistent with the new presented results on hidden (auto-)conflicts.

7 Conclusion

Following recently observed and analysed hidden conflicts of belief functions [13], a relationship of hidden conflicts of BFs and auto-conflict of a BF has been pointed out and analysed in this study. Further hidden auto-conflict of a belief function has been defined and analysed here. New results has been compared with Martin's et al. results on auto-conflict.

These qualitatively new phenomena of conflicts of BFs moves us to better understanding of nature of conflicts of belief functions in general and bring a challenge to elaborate and update existing approaches to conflicts of BFs.

Acknowledgement

This work was supported by the Czech Science Foundation (project 16-12010S).

References

[1] Thomas Burger. Geometric views on conflicting mass functions: From distances to angles. *International Journal of Approximate Reasoning*, 70:36–50, 2016.

[2] Milan Daniel. Properties of plausibility conflict of belief functions. In *International Conference on Artificial Intelligence and Soft Computing*, pages 235–246. Springer, 2013.

[3] Milan Daniel and Jianbing Ma. Conflicts of belief functions: continuity and frame resizement. In *International Conference on Scalable Uncertainty Management*, pages 106–119. Springer, 2014.

[4] Milan Daniel. Conflict between belief functions: a new measure based on their non-conflicting parts. In *International Conference on Belief Functions*, pages 321–330. Springer, 2014.

[5] Sébastien Destercke and Thomas Burger. Toward an axiomatic definition of conflict between belief functions. *IEEE transactions on cybernetics*, 43(2):585–596, 2013.

[6] Eric Lefèvre and Zied Elouedi. How to preserve the conflict as an alarm in the combination of belief functions? *Decision Support Systems*, 56:326–333, 2013.

[7] Weiru Liu. Analyzing the degree of conflict among belief functions. *Artificial Intelligence*, 170(11):909–924, 2006.

[8] Arnaud Martin. About conflict in the theory of belief functions. In *Belief Functions: Theory and Applications*, pages 161–168. Springer, 2012.

[9] Johan Schubert. The internal conflict of a belief function. In *Belief Functions: Theory and Applications*, pages 169–177. Springer, 2012.

[10] Philippe Smets. Analyzing the combination of conflicting belief functions. *Information fusion*, 8(4):387–412, 2007.

[11] Milan Daniel. Conflicts within and between belief functions. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 696–705. Springer, 2010.

[12] Milan Daniel. Non-conflicting and conflicting parts of belief functions. In *ISIPTA*, volume 11, pages 149–158. Citeseer, 2011.

[13] Milan Daniel and Václav Kratochvíl. Hidden conflict of belief functions. In *In preparation*, 2017.

[14] Christophe Osswald and Arnaud Martin. Understanding the large family of dempster-shafer theory's fusion operators — a decision-based measure. In *International Conference on Information Fusion*, 2006.

[15] Arnaud Martin, Anne-Laure Jousselme, and Christophe Osswald. Conflict measure for the discounting operation on belief functions. In *International Conference on Information Fusion*, pages 1003–1010, 2008.

[16] Glenn Shafer. *A mathematical theory of evidence*, volume 1. Princeton University Press Princeton, 1976.

[17] Philippe Smets. Decision making in the tbm: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005.

[18] Barry R Cobb and Prakash P Shenoy. On the plausibility transformation method for translating belief function models to probability models. *International journal of approximate reasoning*, 41(3):314–330, 2006.

[19] Milan Daniel. Probabilistic transformations of belief functions. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 539–551. Springer, 2005.

[20] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[21] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015.

[22] Matt Dowle and Arun Srinivasan. *data.table: Extension of 'data.frame'*, 2016. R package version 1.10.0.

The two-sided (max/min, plus) problem is NP-complete

Martin GAVALEC[1] and Daniela PONCE[1] and Karel ZIMMERMANN[1,2]

[1] *University of Hradec Králové*
FIM UHK, Rokitanského 62, 50003 Hradec Králové
Czech Republic
martin.gavalec@uhk.cz, daniela.ponce@uhk.cz

[2] *Charles University Prague*
MFF UK, Malostranské nám. 25, 11800 Praha 1
Czech Republic
karel.zimmermann@mff.cuni.cz

Abstract

The two-sided (max/min,+) linear system of equations

$$\max_{j\in J}(a_{ij}+x_j)=\min_{k\in J}(b_{ik}+y_k),\ i\in I$$

with real coefficients is considered. Recognizing the solvability of the system is shown to be an NP-complete problem. An application of the problem to the synchronization of cyclically repeated groups of activities with deterministic processing times is briefly discussed.

Keywords: (max, plus) algebra; (min, plus) algebra; (max, plus) linear system; (min, plus) linear system; two-sided (max/min, plus) linear system; NP-completeness.

1 Introduction

The systems we are going to study have $(\max,+)$ linear functions on one side and $(\min,+)$ linear functions on the other side. Such systems of equation can be applied to synchronizing the release and completion times of activities with given processing times.
]

Example 1.1 The following example gives a motivation for the research presented in this paper. We use the notation $I=\{1,\dots,m\}$ and $J=\{1,\dots,n\}$, for natural numbers m,n.

We assume that passengers should be transported from places $P_j,\ j\in J$ to destinations $D_k,\ k\in K$ via transit points $T_i,\ i\in I$. The passengers are supposed to change for other transport means at the transit points. The distances between T_i and P_j are equal to $a_{ij}>0$ and the distances between T_i and D_k are equal to $c_{ik}>0$. The components of $x=(x_1,\dots,x_n)^T$ denote the departure times at $P_j,\ j\in J$ and the components of $y=(y_1,\dots,y_n)^T$ denote the arrival times to destinations $D_k,\ k\in K$. We assume that the passengers can continue their journey from T_i to D_k only after all passengers who need to continue via T_i have arrived there. If the passengers from P_j do not continue through T_i, we set $a_{ij}=-\infty$. The departure times from T_i to D_k are, under our assumptions, equal to y_k-c_{ik}. The earliest feasible departure times from a fixed transit point T_i to a fixed D_k must satisfy the inequality

$$\max_{j\in J}(x_j+a_{ij})\le y_k-c_{ik}.$$

The departure times of the passengers from T_i coordinated with their latest arrival times to T_i must satisfy the relations

$$\max_{j\in J}(x_j+a_{ij})\le \min_{k\in K} y_k-c_{ik}.$$

We may investigate the question whether there exist $x,\ y$ satisfying the system of equations

$$\max_{j\in J}(x_j + a_{ij}) = \min_{k\in K}(y_k - c_{ik}),\ i \in I.$$

If x, y satisfy the system of equations, no unnecessary delays between arrivals and departures in the transit points will occur.

In the rest of this paper we assume $J = K$. No loss of generality will be caused by this assumption because the missing entries can be filled in by coefficients $-\infty$ on the left and ∞ on the right hand side. For computing with infinite values in $(\max, +)$-linear and $(\min, +)$-linear functions, see, e.g., [1, 3, 6].

2 Problem formulation and solvability

The following notation is used: R is the set of real numbers, $I = \{1, \dots, m\}$, $J = \{1, \dots, n\}$, A, B are matrices with elements $a_{ij}, b_{ij} \in R$, $i \in I$, $j \in J$ (the superscript T denotes transposition)

$$(A \mathbin{\mathrm{o}} x)_i = \max_{j\in J}(a_{ij} + x_j), \quad (B \mathbin{\mathrm{o}'} y)_i = \min_{j\in J}(b_{ij} + y_j), \quad \text{for } i \in I,$$

$$A \mathbin{\mathrm{o}} x = \big((A \mathbin{\mathrm{o}} x)_1, \dots, (A \mathbin{\mathrm{o}} x)_m\big)^T, \quad B \mathbin{\mathrm{o}'} y = \big((B \mathbin{\mathrm{o}'} y)_1, \dots, (B \mathbin{\mathrm{o}'} y)_m\big)^T.$$

We will study the system of equations with vector variables x, y

$$A \mathbin{\mathrm{o}} x = B \mathbin{\mathrm{o}'} y \tag{1}$$

which was applied in the motivating example. The set of all solutions (x, y) of (1) will be denoted $M(A, B)$.

Assume $A, B \in R^{m\times n}$. For any $x, y \in R^n$, (x, y) belongs to $M(A, B)$ if and only if

$$\max_{j\in J}(a_{ij} + x_j) = \min_{k\in J}(b_{ik} + y_k),\ i \in I. \tag{2}$$

System (2) is equivalent to the conjunction of the following conditions.

$$(\forall i \in I)(\forall j \in J)(\forall k \in J) \quad a_{ij} + x_j \le b_{ik} + y_k, \tag{3}$$
$$(\forall i \in I)(\exists j \in J)(\exists k \in J) \quad a_{ij} + x_j = b_{ik} + y_k, \tag{4}$$

or, after an easy modification, to

$$(\forall i \in I)(\forall j \in J)(\forall k \in J) \quad x_j - y_k \le b_{ik} - a_{ij}, \tag{5}$$
$$(\forall i \in I)(\exists j \in J)(\exists k \in J) \quad x_j - y_k = b_{ik} - a_{ij}, \tag{6}$$

For $i \in I$, $j, k \in J$, put $q_{jk}^{(i)} = b_{ik} - a_{ij}$ and define the square matrix $Q = (q_{jk}) \in R^{n\times n}$ by putting $q_{jk} = \min\limits_{i\in I} q_{jk}^{(i)}$. Also write

$$W_i = \left\{(j, k) \in J \times J \mid q_{jk}^{(i)} = q_{jk}\right\}. \tag{7}$$

Theorem 2.1 For any $x, y \in R^n$, $(x, y) \in M(A, B)$ if and only if

$$(\forall j \in J)(\forall k \in J) \quad x_j - y_k \le q_{jk}, \tag{8}$$
$$(\forall i \in I)(\exists (j, k) \in W_i) \quad x_j - y_k = q_{jk}. \tag{9}$$

Proof

The assertions follow directly from (3), (4) and (7).

3 NP-completeness

The solvability of the system of equations (1) will be be investigated in the rest of this paper. This problem will be referred to as the 'two-sided (max/min, plus) problem' (TSMMP) .

The main result says that the recognition version of the problem is NP-complete. In particular, it will be shown in Section 4 that the so-called consistent choice (CC) problem polynomially transforms to TSMMP. As CC is NP-complete, so is TSMMP.

3.1 Consistent choice

For a set M denote by $\mathcal{P}_2(M)$ the set of all two-element subsets of M. Let m, n be natural numbers and let $\mathcal{M} = (M_i; i \in N)$ be a system of non-empty disjoint sets, each M_i having cardinality $\leq m$. We will call the number m the *width of* $\mathcal{M}$.

Furthermore, put $M = \bigcup(M_i; i \in N)$. Let $P \subseteq \mathcal{P}_2(M)$ be such that

$$(\forall i, j \in N)(\forall x \in M_i, y \in M_j)\big[\,\{xy\} \in P \Rightarrow i \neq j\,\big]\ . \tag{10}$$

A subset $C \subseteq M$ is called a *consistent choice* in $(\mathcal{M}, P)$ if

$$(\forall i \in N)\,|M_i \cap C| = 1, \tag{11}$$

$$(\forall x, y \in C)\,\big[\,x \neq y \Rightarrow \{xy\} \in P\,\big]. \tag{12}$$

By the *consistent m-choice* (CC_m) problem we understand the problem of recognizing whether there exists a consistent choice in $(\mathcal{M}, P)$ for every system $\mathcal{M}$, of width m, consisting of non-empty disjoint sets and every $P \subseteq \mathcal{P}_2(\bigcup \mathcal{M})$, satisfying (10). CC_m has been studied in [5] under the description of the problem of compatible representatives, among other similar problems.

3.2 Polynomial-time transformations

We say that a recognition problem $\mathcal{A}_1$ *polynomially transforms* to another recognition problem $\mathcal{A}_2$ if, given any string x, we can construct a string y within polynomial time (depending on $|x|$) such that x is a *yes* instance of $\mathcal{A}_1$ if and only if y is a *yes* instance of $\mathcal{A}_2$. If $\mathcal{A}_1$ is NP-complete and polynomially transforms to $\mathcal{A}_2$, then $\mathcal{A}_2$ is NP-complete as well [2, 4]. The following result, with slightly modified notation, was published in [5].

Theorem 3.1 Let $m \geq 2$ be a natural number. Then

(i) the satisfiability problem SAT_m polynomially transforms to CC_m,

(ii) CC_m is NP-complete for $m > 2$.

4 NP-completeness of TSMMP

The main result of this paper is proved in this section.

Theorem 4.1

(i) CC_3 polynomially transforms to TSMMP,

(ii) TSMMP is NP-complete.

Proof (i) Suppose $(\mathcal{M}, P)$ is an instance of CC_3. That is, $\mathcal{M}$ is a system of non-empty disjoint sets $(M_h; h \in H)$, each M_h having cardinality ≤ 3.

Without any loss of generality, we may assume that every M_h is a three-element set of natural numbers. Thus, we have $M = \bigcup\limits_{h \in H} M_h = \{1, 2, \ldots, 3p\}$, where $H = \{1, 2, \ldots, p\}$ and $M_h = \{3h-2, 3h-1, 3h\}$ for every $h \in H$, Furthermore, P is a set of two-element subsets of M such that

$$(\forall h, h' \in H)(\forall s \in M_h,\, s' \in M_{h'})\big[\{s, s'\} \in P \Rightarrow h \neq h'\big]\ . \tag{13}$$

We say that $(\mathcal{M}, P)$ is a *yes* instance of CC_3 if there is a consistent choice in $(\mathcal{M}, P)$, i.e., $C \subseteq M$ with

$$(\forall h \in H)\, |M_h \cap C| = 1, \tag{14}$$

$$(\forall s, s' \in C)\, \big[\, s \neq s' \Rightarrow \{s, s'\} \in P \,\big]. \tag{15}$$

In the opposite case, $(\mathcal{M}, P)$ is called a *no* instance of CC_3.

Our aim is to construct an instance of TSMMP, (A, B), which is solvable if and only if $(\mathcal{M}, P)$ is a *yes* instance of CC_3. The construction of (A, B) has to be done in polynomial time depending on $|(\mathcal{M}, P)|$.

The rows of A, B will be indexed by $i \in I = \{1, 2, \ldots, m\}$, and the columns by $j \in J = \{1, 2, \ldots, n\}$, where $p = |H|$, $m = 1 + 4p$ and $n = 3p$.

Using the notation from Section 2, we write $q_{jk}^{(i)} = b_{ik} - a_{ij}$ for every $i \in I$, $j, k \in J$, and define the square matrices $Q^{(i)} = \left((q_{jk}^{(i)}\right) \in R^{n\times n}$. Furthermore, we define $Q = \big(q_{jk}\big) \in R^{n\times n}$ by putting $q_{jk} = \min\limits_{i\in I} q_{jk}^{(i)}$. We also define, for every $i \in I$,

$$W_i = \left\{(j, k) \in J \times J \,|\, q_{jk}^{(i)} = q_{jk}\right\}. \tag{16}$$

By Theorem 2.1, (A, B) is solvable if and only if there are $x, y \in R(n)$ such that

$$(\forall j \in J)(\forall k \in J) \quad x_j - y_k \leq q_{jk}, \tag{17}$$

$$(\forall i \in I)(\exists (j, k) \in W_i) \quad x_j - y_k = q_{jk}. \tag{18}$$

System (A, B), the auxiliary matrices $Q^{(i)}$, $i \in I$, and the goal matrix Q are constructed in the reversed order, starting from the given instance $(\mathcal{M}, P)$. The construction is done and verified in seven steps. Steps S1–S4 describe the conditions to be satisfied by Q, $Q^{(i)}$ and W_i, while S5 shows how to choose the entries in A and B in order that the described conditions will be satisfied. S6 and S7 show that (A, B) constructed in S5 is solvable (it is a *yes* instance of TSMMP) if and only if the given $(\mathcal{M}, P)$ is a *yes* instance of CC_3.

Step 1. The goal matrix $Q \in R^{n\times n}$ encoding $(\mathcal{M}, P)$ will be computed as follows: if Q is expressed in 3×3 blocks S_{gh}, $g, h \in H$, then

(a) the diagonal blocks have the form

$$S_{hh} = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}, \tag{19}$$

(b) the entries of every non-diagonal block S_{gh}, $g \neq h$ have values 0 or -1, according to the following rule: if $j \in M_g$, $k \in M_h$ then

$$q_{jk} = \begin{cases} 0 & \text{if } \{j, k\} \in P \\ -1 & \text{otherwise.} \end{cases} \tag{20}$$

The zero entries in the diagonal block S_{hh} encode the elements of the input set M_h, while the zero entries of the non-diagonal block matrix S_{gh} encode the consistency relation P between the elements of M_g and M_h. The entries -1 in a non-diagonal block S_{gh} encode the inconsistent pairs $(j, k) \notin P$ with $j \in M_g$, $k \in M_h$.

Step 2. The matrices $Q^{(i)} \in R(n, n)$, for $i \in M$ will be constructed in S5 by choosing the ith rows in A and B in such a way that the following conditions will be satisfied (using 3×3 blocks $S_{gh}^{(i)}$, similarly as in S1).

(a) If $i = 1$, then all entries in $Q^{(1)}$ are 0. Put $i_a = 1$.

(b) If $i = i_a + h \in \{i_a + 1, i_a + 2, \ldots, i_a + p\}$ and $h \in H$, then the entries in the diagonal block $S_{hh}^{(i_a+h)}$ are 0, while the entries in the remaining blocks $S_{g'h'}^{(i_a+h)}$ with $(g', h') \neq (h, h)$ have positive values. Put $i_b = i_a + p$.

(c) If $i = i_b + j \in \{i_b + 1, i_b + 2, \ldots, i_b + 3p\}$ and $3h - 2 \leq j \leq 3h$ with $h \in H$, then the entries in the jth row of $Q^{(i_b+j)}$ have the values 0 or -1, as determined by (19) and (20). Moreover, all three diagonal entries in the diagonal block matrix $S_{hh}^{(i_b+j)}$ are equal to 0, in accordance with (19). All remaining entries in $Q^{(i_b+j)}$ are non-negative.

Step 3. Compute $Q^M = \min_{1 \leq i \leq 1+4p} Q^{(i)}$, the minimum of the matrices defined in S2, and verify that it is equal to Q defined in S1.

First, $q_{jk}^M \leq 0$ for every $j, k \in J$, in view of S2(a). Therefore, the non-negative values in S2(b) and S2(c) do not influence the entries in Q^M.

Second, q_{jk}^M is equal to 0 or -1, in accordance with (19) and (20). This is guaranteed individually in the jth row of $Q^{(i_b+j)}$, for every $j \in J$, in view of S3(c). Hence, $Q^M = Q$.

Step 4. Compute W_i for every $i \in I$.

(a) If $i = 1$, then all entries in $Q^{(1)}$ are 0. Therefore, $W_1 = \{(j, k) \in J \times J \,|\, q_{jk} = 0\}$. In particular, $(j, j) \in W_1$ for every $j \in J$.

(b) If $i = i_a + h \in \{i_a + 1, i_a + 2, \ldots, i_a + p\}$ and $h \in H$, then the only entries in $Q^{(i_a+h)}$ with value q_{jk} are the diagonal entries in the diagonal block $S_{hh}^{(i_a+h)}$ which are equal to 0. The other entries have the minimal value either -1, or positive. Hence every W_{i_a+h} consists of the three diagonal positions in S_{hh}.

(c) If $i = i_b + j \in \{i_b + 1, i_b + 2, \ldots, i_b + 3p\}$ and $3h - 2 \leq j \leq 3h$ with $h \in H$, then W_{i_b+j} contains the jth row in $Q^{(i_b+j)}$, and, moreover, the diagonal positions in the diagonal block matrix $S_{hh}^{(i_b+j)}$ (one of these positions is also contained in the jth row). Depending on P, there may exist further positions contained in W_{i_b+j}, but this has no influence on the result.

Step 5. Compute the ith rows, A_i and B_i, for every $i \in I$, so that the properties described in S2 would be satisfied.

(a) If $i = 1$, then all entries in $Q^{(1)}$ are 0. This condition is fulfilled when $a_{ij} = b_{ik} = 2$. Clearly, any constant value can be used instead of 2.

(b) If $i = i_a + h \in \{i_a + 1, i_a + 2, \ldots, i_a + p\}$ and $h \in H$, then the entries in the diagonal block $S_{hh}^{(i_a+h)}$ are 0, while the entries in the remaining blocks $S_{g'h'}^{(i_a+h)}$ with $(g', h') \neq (h, h)$ have positive values.

The conditions are satisfied when we put $a_{ij} = b_{ik} = 2$ for the positions (j, k) in the block S_{hh} (that is, for $j, k \in \{3h - 2, 3h - 1, 3h\}$), and put $a_{ij} = 0$, $b_{ik} = 4$ for the other (j, k).

(c) If $i = i_b + j \in \{i_b + 1, i_b + 2, \ldots, i_b + 3p\}$ and $3h - 2 \leq j \leq 3h$ with $h \in H$, then the entries in the jth row of $Q^{(i_b+j)}$ have the values 0, -1, as determined by (19) and (20). Moreover, all three diagonal entries in the diagonal block matrix $S_{hh}^{(i_b+j)}$ are equal to 0, in accordance with (19). All remaining entries in $Q^{(i_b+j)}$ are non-negative.

For fixed $j \in J$, the above conditions are satisfied if we put $a_{ij} = 3$, put $b_{ik} = 3$ ($b_{ik} = 2$) provided that the $q_{jk}^{(i)} = 0$ ($q_{jk}^{(i)} = -1$), and put $a_{ij'} = 2$ (respectively $a_{ij'} = 1$) provided that the row $j' \neq j$ intersects S_{hh} (respectively does not intersect S_{hh}).

Step 6. Assume that (A, B) is a solvable instance of TSMMP. Then, by Theorem 2.1, there are $x, y \in R(n)$ such that for every $i \in I$ there is a pair $(j, k) \in W_i$ such that $x_j - y_k = q_{jk}$. For every $h \in H$, W_{i_a+h} consists of three diagonal positions in S_{hh}, according to Step 4(b). Thus, in particular we have $(j_h, j_h) \in W_{i_a+h}$ where $j_h \in M_h = \{3h - 2, 3h - 1, 3h\}$.

We verify that $C = \{j_h \,|\, h \in H\}$ is a consistent choice in $(\mathcal{M}, P)$. Assume $g, h \in H$, $g \neq h$. Then we have $x_{j_g} - y_{j_g} = q(j_g, j_g) = 0$, i.e., $x_{j_g} = y_{j_g}$, and similarly, $x_{j_h} = y_{j_h}$.

By an easy computation we get $q(j_g, j_h) + q(j_h, j_g) = (x_{j_g} - y_{j_h}) + (x_{j_h} - y_{j_g}) = (x_{j_g} - x_{j_h}) + (x_{j_h} - x_{j_g}) = 0$. In view of Step 1, all entries in Q are non-positive. Therefore, $q(j_g, j_h) + q(j_h, j_g) = 0$ implies $q(j_g, j_h) = q(j_h, j_g) = 0$. That is, $\{j_g, j_h\} \in P$, by (20). We have proved that $(\mathcal{M}, P)$ is a *yes* instance of CC_3.

Step 7. Conversely, assume that given $(\mathcal{M}, P)$ is a *yes* instance of CC_3. That is, there is a P-consistent choice $C = \{j_h \,|\, h \in H\}$ with $j_h \in M_h = \{3h - 2, 3h - 1, 3h\}$. We define vectors $x, y \in R(n)$ as follows. Put $x_j = 2$ if $j = j_g$ for some $g \in H$, and put $y_k = 2$ if $k = j_h$ for some $h \in H$. Furthermore, put $x_j = 0$, $y_k = 4$ for the remaining $j, k \in J$.

It is easy to verify that conditions (17), (18) are satisfied. Hence, (A, B) is solvable and (x, y) is its solution.

The computation in steps S1–S7 has been done in $O(p^2) = O(|(\mathcal{M}, P)|^2)$ time. That is, we have proved that CC_3 polynomially transforms to TSMMP.

(ii) The NP-completeness of TSMMP follows from (i) and from Theorem 3.1 (ii).

5 Example

The polynomial transformation of a given $(\mathcal{M}, P)$ to (A, B) is illustrated in this section. We start with a simple situation with a consistent choice from two sets.

Assume $H = \{1, 2\}$, $M_1 = \{1, 2, 3\}$, $M_2 = \{4, 5, 6\}$ and

$$P = \Big\{\{1, 6\}, \{2, 4\}, \{2, 5\}\Big\}.$$

Step 1.

We have $p = 2$, $m = |I| = 1 + 4p = 9$, $n = |J| = 3p = 6$. $A, B \in R^{9\times 6}$ will be constructed in such a way that the goal matrix has the form

$$Q = \left[\begin{array}{rrr|rrr} 0 & -1 & -1 & -1 & -1 & 0 \\ -1 & 0 & -1 & 0 & 0 & -1 \\ -1 & -1 & 0 & -1 & -1 & -1 \\ \hline -1 & 0 & -1 & 0 & -1 & -1 \\ -1 & 0 & -1 & -1 & 0 & -1 \\ 0 & -1 & -1 & -1 & -1 & 0 \end{array}\right]. \tag{21}$$

In block notation, we have

$$Q = \left[\begin{array}{cc} S_{11} & S_{12} \\ S_{21} & S_{22} \end{array}\right]. \tag{22}$$

The zero diagonals in the block matrices S_{11} (respectively S_{22}) are the sets W_1 (respectively W_2) encoding M_1 (respectively M_2) from the given CC_3 instance $(\mathcal{M}, P)$. In view of (18), at least one zero element must be chosen from every one of the above mentioned diagonals. The consistency of the chosen elements is ensured by the block matrices S_{12} and S_{21}, in which the zeros encode the consistency relation P, while the -1s correspond to inconsistent pairs. In fact, each matrix is the transpose of each other, as a consequence of the symmetry of Q.

Step 2.

For $i \in I$, the matrices $Q^{(i)}$ with entries $q^{(i)}_{jk} = b_{ik} - a_{ij}$ are computed in such a way that $Q = \min\limits_{i\in I} Q^{(i)}$. The matrix minimum is computed independently for each pair of indices (j, k). In the pictures, the first additional column in $Q^{(i)}$ contains the values a_{ij}, while the first row contains the values b_{ik}.

All entries in $Q^{(1)}$ are equal to 0. In the figure below, $Q^{(2)}$, respectively $Q^{(3)}$, were computed from the second rows $A_2 = (2, 2, 2, 0, 0, 0)$, $B_2 = (2, 2, 2, 4, 4, 4)$ and from the third rows $A_3 = (0, 0, 0, 2, 2, 2)$, $B_2 = (4, 4, 4, 2, 2, 2)$, respectively, in accordance with S2(b) in the proof of Theorem 4.1.

$$\left[\begin{array}{c|ccc|ccc} a_{2j}|b_{2k} & 2 & 2 & 2 & 4 & 4 & 4 \\ \hline 2 & 0 & 0 & 0 & 2 & 2 & 2 \\ 2 & 0 & 0 & 0 & 2 & 2 & 2 \\ 2 & 0 & 0 & 0 & 2 & 2 & 2 \\ \hline 0 & 2 & 2 & 2 & 4 & 4 & 4 \\ 0 & 2 & 2 & 2 & 4 & 4 & 4 \\ 0 & 2 & 2 & 2 & 4 & 4 & 4 \end{array}\right], \quad \left[\begin{array}{c|ccc|ccc} a_{3j}|b_{3k} & 4 & 4 & 4 & 2 & 2 & 2 \\ \hline 0 & 4 & 4 & 4 & 2 & 2 & 2 \\ 0 & 4 & 4 & 4 & 2 & 2 & 2 \\ 0 & 4 & 4 & 4 & 2 & 2 & 2 \\ \hline 2 & 2 & 2 & 2 & 0 & 0 & 0 \\ 2 & 2 & 2 & 2 & 0 & 0 & 0 \\ 2 & 2 & 2 & 2 & 0 & 0 & 0 \end{array}\right]. \tag{23}$$

Similarly, we compute $Q^{(4)}$, $Q^{(5)}$, $Q^{(6)}$, $Q^{(7)}$, $Q^{(8)}$, and $Q^{(9)}$ from the corresponding rows in A and B. These matrices induce the zeros and -1s prescribed in the corresponding row of the goal matrix Q, and also the three diagonal zeros in the corresponding block matrices S_{11}, S_{22}.

$$\left[\begin{array}{c|ccc|ccc} a_{4j}|b_{4k} & 3 & 2 & 2 & 2 & 2 & 3 \\ \hline 3 & 0 & -1 & -1 & -1 & -1 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 2 & 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 & 1 & 1 & 2 \end{array}\right], \quad \left[\begin{array}{c|ccc|ccc} a_{5j}|b_{5k} & 2 & 3 & 2 & 3 & 3 & 2 \\ \hline 2 & 0 & 1 & 0 & 1 & 1 & 0 \\ 3 & -1 & 0 & -1 & 0 & 0 & -1 \\ 2 & 0 & 1 & 0 & 1 & 1 & 0 \\ \hline 1 & 1 & 2 & 1 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 & 2 & 2 & 1 \end{array}\right],$$

$$\left[\begin{array}{c|ccc|ccc} a_{6j}|b_{6k} & 2 & 2 & 3 & 2 & 2 & 2 \\ \hline 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 0 & 0 \\ 3 & -1 & -1 & 0 & -1 & -1 & -1 \\ \hline 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 & 1 \end{array}\right], \quad \left[\begin{array}{c|ccc|ccc} a_{7j}|b_{7k} & 2 & 3 & 2 & 3 & 2 & 2 \\ \hline 1 & 1 & 2 & 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 & 2 & 1 & 1 \\ \hline 3 & -1 & 0 & -1 & 0 & -1 & -1 \\ 2 & 0 & 1 & 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 & 1 & 0 & 0 \end{array}\right],$$

$$\left[\begin{array}{c|ccc|ccc} a_{8j}|b_{8k} & 2 & 3 & 2 & 2 & 3 & 2 \\ \hline 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ \hline 2 & 0 & 1 & 0 & 0 & 1 & 0 \\ 3 & -1 & 0 & -1 & -1 & 0 & -1 \\ 2 & 0 & 1 & 0 & 0 & 1 & 0 \end{array}\right], \quad \left[\begin{array}{c|ccc|ccc} a_{9j}|b_{9k} & 3 & 2 & 2 & 2 & 2 & 3 \\ \hline 1 & 2 & 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 & 1 & 1 & 2 \\ \hline 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 & 0 & 0 & 1 \\ 3 & 0 & -1 & -1 & -1 & -1 & 0 \end{array}\right].$$

As the result of these computations, we get

$$A = \left[\begin{array}{ccc|ccc} 2 & 2 & 2 & 2 & 2 & 2 \\ \hline 2 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 2 \\ \hline 3 & 2 & 2 & 1 & 1 & 1 \\ 2 & 3 & 2 & 1 & 1 & 1 \\ 2 & 2 & 3 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 3 & 2 & 2 \\ 1 & 1 & 1 & 2 & 3 & 2 \\ 1 & 1 & 1 & 2 & 2 & 3 \end{array}\right], \quad B = \left[\begin{array}{ccc|ccc} 2 & 2 & 2 & 2 & 2 & 2 \\ \hline 2 & 2 & 2 & 4 & 4 & 4 \\ 4 & 4 & 4 & 2 & 2 & 2 \\ \hline 3 & 2 & 2 & 2 & 2 & 3 \\ 2 & 3 & 2 & 3 & 3 & 2 \\ 2 & 2 & 3 & 2 & 2 & 2 \\ \hline 2 & 3 & 2 & 3 & 2 & 2 \\ 2 & 3 & 2 & 2 & 3 & 2 \\ 3 & 2 & 2 & 2 & 2 & 3 \end{array}\right].$$

For every consistent choice C consisting of $j_1 \in M_1$, $j_2 \in M_2$ with $\{j_1, j_2\} \in P$, there is a solution $x, y \in R(n)$, which can be found by the rules shown in S7 of the proof of Theorem 4.1.

For example, if $C = \{1, 6\}$, then we get $x = (2, 0, 0, 0, 0, 2)^T$ and $y = (2, 4, 4, 4, 4, 2)^T$. Other possible solutions are computed with $C = \{2, 4\}$ (when we get $x = (0, 2, 0, 2, 0, 0)^T$ and $y = (4, 2, 4, 2, 4, 4)^T$), or with $C = \{2, 5\}$ ($x = (0, 2, 0, 0, 2, 0)^T$ and $y = (4, 2, 4, 4, 2, 4)^T$).

Conclusions

The solvability and computational complexity of the two-sided (max/min, plus) problem (TSMMP) were investigated. Necessary and sufficient conditions for the solvability of a given instance have been described.

Moreover, it has been proved that the recognition version of TSMMP is NP-complete. In particular, it has been proved that the consistent choice problem of width 3 (CC_3) polynomially transforms to TSMMP. As CC_3 is known to be NP-complete, so is TSMMP.

Acknowledgement
The support of the Grant Agency of Excellence UHK FIM is gratefully acknowledged.

References

[1] Butkovič, P. *Max-linear Systems: Theory and Algorithms*, Springer-Verlag, Berlin, 2010.

[2] Cook, S.A. The Complexity of Theorem Proving Procedures. In *Proc. 3rd ACM Symp. on the Theory of Computing*, pages 151–158, ACM, New York, 1971.

[3] Cuninghame-Green, R.A. *Minimax Algebra*. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1979.

[4] Garey, M.R., Johnson, D.S. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman & Comp., San Francisco, 1979.

[5] Knuth, D.E., Raghunathan, A., The problem of compatible representatives, *SIAM J. on Discrete Mathematics Archive*, 5(3):422–427, 1992.

[6] Krivulin, N.K. *Methods of Idempotent Algebra in Modelling and Analysis of Complex Systems*. St. Petersburg University Press, St. Petersburg, 2009 (in Russian).

F-transform based numerical solution of partial differential equations

Michal HOLČAPEK[1] and Radek VALÁŠEK[1]

[1] *University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, NSC IT4Innovations, 30. dubna 22, 701 03 Ostrava 1, Czech Republic*
Michal.Holcapek@osu.cz, Radek.Valasek@osu.cz

Abstract

In this paper, we propose a numerical method based on the F-transform for solving a certain type of partial differential equations (PDEs) with homogeneous Dirichlet boundary conditions and initial conditions. We show how the PDEs, after the application of the Crank-Nicolson scheme for time discretization, can be approximated by a system of linear equations with the direct F-transform components as their variables. The numerical solution of PDEs is obtained by solving the system of linear equations and the application of the inverse F-transform. The proposed method is then adjusted to the three-dimensional boundary value problem.

Keywords: Partial differential equations, multivariate F-transform, Black-Scholes equation, Financial modeling.

1 Introduction

The *fuzzy transform* (*F-transform*, for short) was introduce by Perfilieva in [9] (see also [10]) as a soft computing method that is used for the approximation of functions. A generalization to two-dimensional case was proposed in [11] (see also [12]). The F-transform has two phases: *direct* and *inverse*. The direct F-transform transforms a bounded continuous (integrable) real function to a finite vector of real numbers that are called the components of F-transform. The inverse F-transform sends the latter vector back to a continuous function as the result that approximates the original one.

The F-transform technique has been already applied, among others, in solving (partial) differential equations. In [9], a numerical method based on the F-transform has been proposed for an ordinary Cauchy problem. The proposed technique which generalized the Euler method showed its potential for solving differential equations in comparison with numerous classical techniques. In [11] (see also [12]), the F-transform technique was used to solve special types of partial differential equations (PDEs). In [2], a novel algorithm based on the F-transform has been proposed to obtain an approximate solution for a class of second-order ordinary differential equations with classical initial conditions. In [7], new numerical methods based on the higher degree F-transform for solving the Cauchy problem have been presented. The results of the proposed methods outperformed the second order Runge-Kutta method.

The aim of this paper is to propose a numerical solution of a type of PDEs that satisfies the homogeneous Dirichlet boundary conditions and an initial condition with the help of n-variate F-transform. This type of PDEs is very important in financial mathematics, particularly, in option pricing (see [1, 3, 8]).

The structure of this paper is as follows: Section 2 introduces the concept of multidimensional uniform fuzzy partition. Section 3 contains the definition of the direct and inverse multivariate F-transform. A numerical solution of a type of PDEs is proposed in Section 4. Section 5 illustrates the proposed method on an example of 3D boundary valued problem. Section 6 is a conclusion.

2 Multidimensional uniform fuzzy partition

Let $\mathbb{Z}$ and $\mathbb{R}^n$ denote the set of integers and the set of real vectors, respectively. A uniform fuzzy partition can be defined using a generating function K which may be modified by a parameter h specifying the required bandwidth. Each basic function of the uniform fuzzy partition is then obtained by a shift of the modified generating function K, where the uniformity for all shifts is supposed (see, [4]). In the multidimensional case, we need to extend the generating function of one variable to a function of n variables. We propose the following straightforward extension of Definition 2.1 in [4].

Definition 1 *A function $K : \mathbb{R}^n \to [0,1]$ is said to be a* n-variate generating function *if K is an even Lebesgue integrable function (fuzzy relation) for which*

$$K(\boldsymbol{x}) \begin{cases} > 0, & \text{if } \boldsymbol{x} \in (-1,1)^n; \\ = 0, & \text{otherwise.} \end{cases} \tag{1}$$

and the function of one variable

$$K(c_1, \ldots, c_{i-1}, x, c_{i+1}, \ldots, c_n)$$

is non-increasing in $[0,1]$ for any choice of $i = 1, \ldots, n$ and constants $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n \in (-1,1)$. A n-variate generating function K is said to be normal *if $K(0, \ldots, 0) = 1$.*

We say that K *is a generating function* if K is a univariate generating function. In this paper, we restrict ourselves to fuzzy partitions of $\mathbb{R}^n$ that are constructed by the product of n generating functions (cf., [6]).[1] The following lemma has been proved in [5].

Lemma 1 *Let $K_1, \ldots, K_n$ be generating functions. Then,*

$$K(x_1, \ldots, x_n) = K_1(x_1) \cdots K_n(x_n)$$

defines an n-variate generating function.

In the next example, we introduce the most popular uniform generating functions in the theory and applications of the F-transform.

Example 1 *Functions $K_T, K_C : \mathbb{R} \to [0,1]$ defined by*

$$K_T(x) = \max(1 - |x|, 0), \tag{2}$$

$$K_C(x) = \begin{cases} \frac{1}{2}(1 + \cos(\pi x)), & -1 \le x \le 1; \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

for any $x \in \mathbb{R}$, are called the triangle *and* raised cosine *generating functions.*

Using an n-variate generating function, we can construct a uniform fuzzy partition of $\mathbb{R}^n$ as follows (cf., Definition 3.1 in [4]).

Definition 2 *Let K be an n-variate generating function, and let $\boldsymbol{h} \in \mathbb{R}^n_+$ and $\boldsymbol{c} \in \mathbb{R}^n$ be vectors of positive real numbers and real numbers, respectively. A system of fuzzy relations $\mathbb{A} = \{A_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{Z}^n\}$ defined by*

$$A_{\boldsymbol{k}}(x_1, \ldots, x_n) = K\left(\frac{x_1 - c_1}{h_1} - k_1, \cdots, \frac{x_n - c_n}{h_n} - k_n\right)$$

is said to be an n-dimensional uniform fuzzy partition *of $\mathbb{R}^n$ determined by the triplet $(K, \boldsymbol{h}, \boldsymbol{c})$ if a la* Ruspini condition *is satisfied, i.e.,*

$$\sum_{\boldsymbol{k} \in \mathbb{Z}^n} A_{\boldsymbol{k}}(x_1, \ldots, x_n) = 1 \tag{4}$$

for any $(x_1, \ldots, x_n) \in \mathbb{R}^n$. The parameters $\boldsymbol{h}$ and $\boldsymbol{c}$ are called the bandwidth, shift *and* central node*, respectively. The functions $A_{\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{Z}^n$, are called* basic functions.

[1] An example of bivariate generating function, which cannot be constructed by the product of two univariate generating functions can be found in [5].

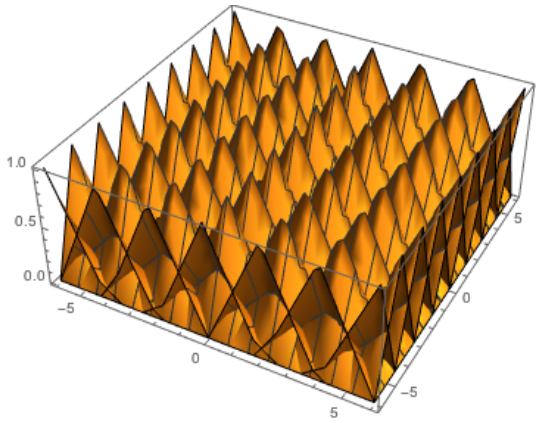


Figure 1: Triangle 2D fuzzy partition of the square $[-6,6] \times [-6,6]$ determined by the triplet $(K_{TT}, (2,1), (-6,-6))$.

Note that we restrict ourselves here to uniform fuzzy partitions where only two functions in each direction can be overlapped. The reason is a simpler expression of boundary and initial conditions by the F-transform components as will be seen in the next section. An application of generalized fuzzy partitions (for the formal definition, we refer to []) is the subject of our future research.

We use $c_{\boldsymbol{k}} = (c_{\boldsymbol{k}1}, \ldots, c_{\boldsymbol{k}n}) = (c_1 + k_1 h_1, \ldots, c_n + k_n h_n)$ to denote the $\boldsymbol{k}$-th *node* of an n-dimensional uniform fuzzy partition determined by $(K, \boldsymbol{h}, \boldsymbol{c})$. The *basic function* $A_{\boldsymbol{k}}$ can be then expressed as

$$A_{\boldsymbol{k}}(x_1, \ldots, x_n) = K\left(\frac{x_1 - c_{\boldsymbol{k}1}}{h_1}, \ldots, \frac{x_n - c_{\boldsymbol{k}n}}{h_n}\right). \tag{5}$$

For n-variate generating functions which are the products of univariate ones is useful the following lemma (see).

Lemma 2 *Let $K = \prod_{s=1}^{n} K_s$ be a multivariate generating function. Then, $(K, \boldsymbol{h}, \boldsymbol{c})$ determines an n-dimensional fuzzy partition if and only if each triplet (K_s, h_s, c_s), $s = 1, \ldots, n$, determines a 1-dimensional uniform fuzzy partition.*

For the purpose of this paper, we restrict ourselves to the uniform fuzzy partitions of an n-orthotope $\Phi = \prod_{s=1}^{n} [a_s, b_s]$. Generally, if Φ is an n-orthotope and $\mathbb{A} = \{A_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{Z}^n\}$ is a uniform fuzzy partition of $\mathbb{R}^n$ determined by the triplet $(K, \boldsymbol{h}, \boldsymbol{c})$, then $\mathbb{A}|_\Phi \subset \mathbb{A}$ is said to be a uniform fuzzy partition of Φ determined by $(K, \boldsymbol{h}, \boldsymbol{c})$ provided that for any $A_{\boldsymbol{k}} \in \mathbb{A}|_\Phi$ there exists $\boldsymbol{x} \in \Phi$ such that $A_{\boldsymbol{k}}(\boldsymbol{x}) > 0$, and $\sum_{A_{\boldsymbol{k}} \in \mathbb{A}|_\Phi} A(\boldsymbol{x}) = 1$ for any $\boldsymbol{x} \in \Phi$. Intuitively, the family $\mathbb{A}|_\Phi$ consists of all basic functions from $\mathbb{A}$ that have a non-empty intersection with Φ.

For the sake of simplicity, we consider the following definition of a uniform fuzzy partition of an n-orthotope Φ. For an n-variate generating partition K and natural numbers $M_s > 2$, $s = 1, \ldots, n$, each uniform fuzzy partition of Φ is determined by a triplet $(K, \boldsymbol{h}, \boldsymbol{c})$ such that

$$\boldsymbol{h} = (h_1, \ldots, h_n) = \left(\frac{b_1 - a_1}{M_1}, \ldots, \frac{b_n - a_n}{M_n}\right) \text{ and } \boldsymbol{c} = (a_1, \ldots, a_n). \tag{6}$$

Denote by $\mathbb{K} = \prod_{s=1}^{n} \{0, \ldots, M_s\}$ the set of indexes. Then, for any $\boldsymbol{k} = (k_1, \ldots, k_n) \in \mathbb{K}$, the $\boldsymbol{k}$-th basic function of the uniform fuzzy partition of Φ has the following form:

$$A_{\boldsymbol{k}} = K\left(\frac{x_1 - a_1}{h_1} - k_1, \ldots, \frac{x_n - a_n}{h_n} - k_n\right). \tag{7}$$

An example of 2-dimensional uniform fuzzy partition of the square $[-6,6] \times [-6,6]$ determined by $(K_{TT}, (2,1), (-6,-6))$ is depicted in Fig. 1.

3 Multivariate F-transform

Let $\mathbb{A}|_\Phi = \{A_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{K}\}$ be a fixed n-dimensional fuzzy partition of an n-orthotope Φ determined by a triplet $(K, \boldsymbol{h}, \boldsymbol{c})$ such that condition (6) is satisfied, where $\mathbb{K} = \prod_{s=1}^{n}\{0, \dots, M_s\}$. The direct n-variate F-transform of a given function f with the domain Φ with respect to a uniform fuzzy partition $\mathbb{A}|_\Phi$ is defined as the weighted average of this function, where the weights are determined from the basic functions of the uniform fuzzy partition. In this paper, we consider the following definition (cf. [10, 11]).

Definition 3 *Let $\mathbb{A}|_\Phi = \{A_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{K}\}$ be a uniform fuzzy partition of an n-orthotope Φ, and let f be a bounded and continuous function on Ω. We say that the family $\mathbf{F} = \{F_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{K}\}$ of real numbers given by*

$$F_{\boldsymbol{k}} = \frac{\int_\Phi f(x_1, \dots, x_n) A_{\boldsymbol{k}}(x_1, \dots, x_n)\, dx_1 \dots dx_n}{\int_\Phi A_{\boldsymbol{k}}(x_1, \dots, x_n)\, dx_1 \dots dx_n}, \quad \boldsymbol{k} \in \mathbb{K},$$

is the direct (integral) n-variate F-transform *of f with respect to $\mathbb{A}|_\Phi$. The real value $F_{\boldsymbol{k}}$ of the family $\mathbf{F}$ is called the $\boldsymbol{k}$-th* component of the direct n-variate F-transform with respect to $\mathbb{A}|_\Phi$.

One can see that the original continuous function f is now expressed by a family of multivariate F-transform components, where each component carries a local information about the original function f. Denote by $<$ the lexicographical ordering of the index set $\mathbb{K}$ and arrange the elements of $\mathbb{K}$ into the sequence $\boldsymbol{k}_1 < \dots < \boldsymbol{k}_P$, where $P = |\mathbb{K}|$. Then, the family $\mathbf{F}$ can be interpreted as a P-dimensional vector of components of multidimensional F-transform. In the next part, we use the previously introduced linear ordering of the F-transform components.

The following definition provides the way how the original function can be approximated from the F-transform components. This phase is called the inverse (n-dimensional) F-transform with respect to a uniform fuzzy partition.

Definition 4 *$\mathbb{A}|_\Phi = \{A_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{K}\}$ be a uniform fuzzy partition of an n-orthotope Φ, and let f be a bounded and continuous function on Φ. Let $\mathbf{F} = \{F_{\boldsymbol{k}} \mid \boldsymbol{k} \in \mathbb{K}\}$ be the direct n-variate F-transform of f with respect to $\mathbb{A}|_\Phi$. The* inverse n-variate F-transform *of f with respect to $\mathbb{A}|_\Phi$ is*

$$\mathbf{F}[f](x_1, \dots, x_n) = \sum_{\boldsymbol{k} \in \mathbb{K}} F_{\boldsymbol{k}} A_{\boldsymbol{k}}(x_1, \dots, x_n), \tag{8}$$

for any $(x_1, \dots, x_n) \in \Phi$.

Obviously, using the lexicographical ordering of $\mathbb{K}$, expression (8) can be written as follows ($P = |\mathbb{K}|$):

$$\mathbf{F}[f](x_1, \dots, x_n) = \sum_{r=1}^{P} F_{\boldsymbol{k}_r} A_{\boldsymbol{k}_r}(x_1, \dots, x_n). \tag{9}$$

4 Numerical solution of PDEs by multivariate F-transform

4.1 Formulation of problem

Let $\Omega = \prod_{s=1}^{n}(a_i, b_i)$. In this paper, we are interested in the following type of partial differential equations (PDEs):

$$\frac{\partial f}{\partial t} = \sum_{\substack{i,j=1 \\ i \le j}}^{n} \alpha_{ij} \frac{\partial f}{\partial x_i \partial x_j} + \sum_{i=1}^{n} \alpha_i \frac{\partial f}{\partial x_i} + \alpha f + g, \tag{10}$$

where $f, g : \Omega \times (0, T) \to \mathbb{R}$ are bounded and continuous functions $\Omega \times (0, T)$ (f satisfies the respective differentiability conditions) and $\alpha_{ij}, \alpha_i, \alpha : \Omega \to \mathbb{R}$, and the following homogeneous Dirichlet boundary condition and the initial condition are assumed:

$$f(x_1, \dots, x_n, t) = 0, \quad (x_1, \dots, x_n, t) \in \partial\Omega \times (0, T),$$
$$f(x_1, \dots, x_n, 0) = \varphi(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in \Omega.$$

We use t to denote the time variable, which appears very often as a variable in real boundary value problems, for example, in physics or financial mathematics. Therefore, we choose its domain $(0,T)$. Nevertheless, our results remains true if one replaces the time variable t by an arbitrary variable y with a domain (a,b). In the next subsection, we provide a numerical solution of (10) with the help of multivariate F-transform.

4.2 Numerical solution based on the F-transform technique

The first step in our proposal is based on the discretization of the time domain by means of the Crank-Nicolson method that is often used for the numerical solution of PDEs. For the sake of simplicity, we put $\boldsymbol{x} = (x_1, \ldots, x_n)$ and write

$$f(\boldsymbol{x}, t) = f(x_1, \ldots, x_n, t),$$
$$\alpha_{ij}(\boldsymbol{x}) = \alpha_{ij}(x_1, \ldots, x_2)$$

and similarly for the functions g, α_i and α. Let $L > 2$ be a natural number and $h_t = \frac{T}{L}$. Denote $t_\ell = \ell \cdot h_t$ for $\ell = 0, \ldots, L$. Then, for any $\boldsymbol{x} \in \Omega$, we have

$$\frac{f(\boldsymbol{x}, t_{\ell+1}) - f(\boldsymbol{x}, t_\ell)}{h_t} = \frac{1}{2}\Bigg(\sum_{\substack{i,j=1 \\ i \le j}}^{n} \alpha_{ij}(\boldsymbol{x}) \frac{\partial f}{\partial s_i \partial s_j}(\boldsymbol{x}, t_{\ell+1}) + \sum_{i=1}^{n} \alpha_i(\boldsymbol{x}) \frac{\partial f}{\partial s_i}(\boldsymbol{x}, t_{\ell+1}) + \alpha(\boldsymbol{x}) f(\boldsymbol{x}, t_{\ell+1})$$
$$+ g(\boldsymbol{x}, t_{\ell+1}) + \sum_{\substack{i,j=1 \\ i \le j}}^{n} \alpha_{ij}(\boldsymbol{x}) \frac{\partial f}{\partial s_i \partial s_j}(\boldsymbol{x}, t_\ell) + \sum_{i=1}^{n} \alpha_i(\boldsymbol{x}) \frac{\partial f}{\partial s_i}(\boldsymbol{x}, t_\ell) + \alpha(\boldsymbol{x}) f(\boldsymbol{x}, t_\ell) + g(\boldsymbol{x}, t_\ell) \Bigg).$$

Similarly to the method of finite differences (FMD), the second step in our proposal is a discretization of the domain Ω to get a family of nodes over which (including time) the previous equation can be transformed into a system of linear equations. In contrast to FMD, we use the F-transform components that represent the original function f and its partial derivatives and also the function g at the given nodes. Since the F-transform does not preserve the product of two function, we do not consider the F-transform components for the functions α_{ij}, α_i and α, instead we use their function values at the given nodes. The procedure is detailed described in the next part.

Let $\mathbb{A}|_\Phi = \{A_{\boldsymbol{k},\ell} \mid \boldsymbol{k} \in \mathbb{K},\ \ell = 0, \ldots, L\}$ be an $(n+1)$-dimensional fuzzy partition of $\Phi = \overline{\Omega} \times [0,T]$, where $\overline{\Omega}$ is the topological closure of Ω and $\mathbb{K} = \prod_{s=1}^{n} \{0, \ldots, M_s\}$, determined by a triplet $(K, \boldsymbol{h}, \boldsymbol{c})$ such that $\boldsymbol{h} = (h_1, \ldots, h_n, h_t)$ and $\boldsymbol{c} = (a_1, \ldots, a_n, 0)$ with

$$h_i = \frac{b_s - a_s}{M_s},\ s = 1, \ldots, n, \ \text{ and } \ h_t = \frac{T}{L}.$$

Denote $c_{sk_s} = a_s + k_s h_s$ for $k_s = 0, \ldots, M_s$ and $s = 1, \ldots, n$. Note that $c_{s0} = a_s$ and $c_{sM_s} = b_s$. Finally, put $\mathbb{K}^- = \prod_{i=1}^{n} \{1, \ldots, M_i - 1\}$. Then, for any $\boldsymbol{k} \in \mathbb{K}^-$ and $\ell = 0, \ldots, L-1$, we can rewrite the previous equation with the help of the F-transform components as follows:

$$\frac{F_{\boldsymbol{k},\ell+1} - F_{\boldsymbol{k},\ell}}{h_t} = \frac{1}{2}\Bigg(\sum_{\substack{i,j=1 \\ i \le j}}^{n} \alpha_{ij}(\boldsymbol{c_k}) F^{x_i x_j}_{\boldsymbol{k},\ell+1} + \sum_{i=1}^{n} \alpha_i(\boldsymbol{c_k}) F^{x_i}_{\boldsymbol{k},\ell+1} + \alpha(\boldsymbol{c_k}) F_{\boldsymbol{k},\ell+1} + G_{\boldsymbol{k},\ell+1}$$
$$+ \sum_{\substack{i,j=1 \\ i \le j}}^{n} \alpha_{ij}(\boldsymbol{c_k}) F^{x_i x_j}_{\boldsymbol{k},\ell} + \sum_{i=1}^{n} \alpha_i(\boldsymbol{c_k}) F^{x_i}_{\boldsymbol{k},\ell} + \alpha(\boldsymbol{c_k}) F_{\boldsymbol{k},\ell} + G_{\boldsymbol{k},\ell} \Bigg),$$

where $F^{x_i x_j}_{\boldsymbol{k},\ell}$, $F^{x_i}_{\boldsymbol{k},\ell}$, $F_{\boldsymbol{k},\ell}$ and $G_{\boldsymbol{k},\ell}$ are the F-transform components of the functions $\frac{\partial f}{\partial x_i \partial x_j}$, $\frac{\partial f}{\partial x_i}$, f and g at the node $(\boldsymbol{c_k}, t_\ell)$, where $\boldsymbol{c_k} = (c_{1k_1}, \ldots, c_{nk_n})$.

Denote by e_i the unit vector with the value 1 at the i-th position and the zero vector by 0 $(i = 1, \ldots, n)$. Further, let us express the values of multivariate F-transform components of partial

derivatives by the symmetric finite differences, i.e., for any $i, j = 1, \dots, n$, $i \neq j$, we have

$$\begin{aligned} F^{x_i x_j}_{\boldsymbol{k},\ell} &= \frac{F_{\boldsymbol{k}+e_i+e_j,\ell} - F_{\boldsymbol{k}+e_i+(-e_j),\ell} - F_{\boldsymbol{k}+(-e_i)+e_j,\ell} + F_{\boldsymbol{k}+(-e_i)+(-e_j),\ell}}{4h_i h_j}, \\ F^{x_i x_i}_{\boldsymbol{k},\ell} &= \frac{F_{\boldsymbol{k}+e_i,\ell} - 2F_{\boldsymbol{k},\ell} + F_{\boldsymbol{k}+(-e_i),\ell}}{h_i^2}, \\ F^{x_i}_{\boldsymbol{k},\ell} &= \frac{F_{\boldsymbol{k}+e_i,\ell} - F_{\boldsymbol{k}-e_i,\ell}}{2h_i}, \end{aligned}$$

where $u+v$ denotes the common addition of vectors u and v and $-u$ the opposite vector to u. For simplicity, we write $u - v$ instead of $u + (-v)$ for vectors u and v. Now, using the substitution of the previous expressions of the multivariate F-transform components of (partial) derivatives into the previous equation, and by a simple manipulation, we obtain the following equation:

$$\begin{aligned} &\sum_{\substack{i,j=1\\i<j}}^{n} C_{e_i+e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}+e_i+e_j,\ell+1} + \sum_{\substack{i,j=1\\i<j}}^{n} C_{e_i-e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}+e_i-e_j,\ell+1} + \sum_{\substack{i,j=1\\i<j}}^{n} C_{-e_i+e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}-e_i+e_j,\ell+1} + \\ &\sum_{\substack{i,j=1\\i<j}}^{n} C_{-e_i-e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}-e_i-e_j,\ell+1} + \sum_{i=1}^{n} C_{e_i}(\boldsymbol{c_k}) F_{\boldsymbol{k}+e_i,\ell+1} + C_0(\boldsymbol{c_k}) F_{\boldsymbol{k},\ell+1} + \sum_{i=1}^{n} C_{-e_i}(\boldsymbol{c_k}) F_{\boldsymbol{k}-e_i,\ell+1} \\ &= \sum_{\substack{i,j=1\\i<j}}^{n} -C_{e_i+e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}+e_i+e_j,\ell} + \sum_{\substack{i,j=1\\i<j}}^{n} C^{\star}_{e_i-e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}+e_i-e_j,\ell} + \sum_{\substack{i,j=1\\i<j}}^{n} C^{\star}_{-e_i+e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}-e_i+e_j,\ell} \\ &+ \sum_{\substack{i,j=1\\i<j}}^{n} C^{\star}_{-e_i-e_j}(\boldsymbol{c_k}) F_{\boldsymbol{k}-e_i-e_j,\ell} + \sum_{i=1}^{n} C^{\star}_{e_i}(\boldsymbol{c_k}) F_{\boldsymbol{k}+e_i,\ell} + C^{\star}_0(\boldsymbol{c_k}) F_{\boldsymbol{k},\ell} + \sum_{i=1}^{n} C^{\star}_{-e_i}(\boldsymbol{c_k}) F_{\boldsymbol{k}-e_i,\ell} + D_{\boldsymbol{k},\ell} \end{aligned}$$

where, for any $i, j = 1, \dots, n$ with $i < j$, we have:

$$\begin{aligned} C_{e_i+e_j}(\boldsymbol{c_k}) &= C_{-e_i-e_j}(\boldsymbol{c_k}) = -\frac{\alpha_{ij}(\boldsymbol{c_k})}{8h_i h_j} h_{t_\ell}, \\ C_{-e_i+e_j}(\boldsymbol{c_k}) &= C_{e_i-e_j}(\boldsymbol{c_k}) = -C_{e_i+e_j}, \\ C_{e_i}(\boldsymbol{c_k}) &= -\left(\frac{\alpha_{ii}(\boldsymbol{c_k})}{h_i^2} + \frac{\alpha_i(\boldsymbol{c_k})}{2h_i}\right)\frac{h_{t_\ell}}{2}, \\ C_{-e_i}(\boldsymbol{c_k}) &= C_{0,e_i} + \frac{\alpha_i(\boldsymbol{c_k})}{2h_i} h_{t_\ell}, \\ C_0(\boldsymbol{c_k}) &= 1 - \left(\alpha(\boldsymbol{c_k}) - \sum_{i=1}^{n} \frac{2\alpha_{ii}(\boldsymbol{c_k})}{h_i^2}\right)\frac{h_{t_i}}{2}, \end{aligned} \tag{11}$$

$$\begin{aligned} C^{\star}_{e_i+e_j}(\boldsymbol{c_k}) &= C^{\star}_{-e_i-e_j)}(\boldsymbol{c_k}) = -C_{e_i+e_j}(\boldsymbol{c_k}), \\ C^{\star}_{-e_i+e_j}(\boldsymbol{c_k}) &= C^{\star}_{e_i-e_j)}(\boldsymbol{c_k}) = -C_{-e_i+e_j}(\boldsymbol{c_k}), \\ C^{\star}_{e_i}(\boldsymbol{c_k}) &= -C_{e_i}(\boldsymbol{c_k}), \\ C^{\star}_{-e_i}(\boldsymbol{c_k}) &= -C_{-e_i}(\boldsymbol{c_k}), \\ C^{\star}_0(\boldsymbol{c_k}) &= 2 - C_0(\boldsymbol{c_k}) \end{aligned} \tag{12}$$

$$D_{\boldsymbol{k},\ell} = (G_{\boldsymbol{k},\ell+1} + G_{\boldsymbol{k},\ell})\,\frac{h_t}{2}. \tag{13}$$

To express the previous equation in a more compact way, denote by $E = \{e_i \mid i = 1, \dots, n\} \cup \{0\}$ the set of all unit vectors e_i and the zero vector 0, and define, for any $\boldsymbol{k} \in \mathbb{K}^-$ and $\boldsymbol{j} \in \mathbb{Z}^n$,

$$C_{\boldsymbol{j}}(\boldsymbol{c_k}) = \begin{cases} C_0(\boldsymbol{x_k}), & \text{if } \boldsymbol{j} = 0; \\ C_{u+v}(\boldsymbol{c_k}), & \text{if } |u|, |v| \in E,\ |v| \neq |u|,\ \boldsymbol{j} = u + v; \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

where $|u| = (|u_1|, \dots, |u_n|)$, and analogously $C^\star_{\boldsymbol{j}}(\boldsymbol{k})$. For example, if $n = 2$, then $C_{(1,-1)}(\boldsymbol{c_k}) = C_{(1,0)+(0,-1)}(\boldsymbol{c_k})$ or $C_{(2,3)}(\boldsymbol{c_k}) = 0$. Using the ordered sequence $\boldsymbol{k}_1 < \cdots < \boldsymbol{k}_N$ of all indexes of $\mathbb{K}^-$, where $N = |\mathbb{K}^-|$ and $<$ is the lexicographical ordering, we can express the previous equation in a more compact form as follows:

$$\sum_{j=1}^{N} C_{\boldsymbol{k}_j - \boldsymbol{k}_i}(\boldsymbol{c_{k_i}}) F_{\boldsymbol{k}_j,\ell+1} = \sum_{j=1}^{N} C^\star_{\boldsymbol{k}_j - \boldsymbol{k}_i}(\boldsymbol{c_{k_i}}) F_{\boldsymbol{k}_j,\ell} + D_{\boldsymbol{k}_i,\ell}. \tag{15}$$

Let us consider two $N \times N$ matrices $\mathbf{K}$ and $\mathbf{K}^\star$ defined by

$$\mathbf{K}_{ij} = C_{\boldsymbol{k}_j - \boldsymbol{k}_i}(\boldsymbol{c_{k_i}}), \quad \mathbf{K}^\star_{ij} = C^\star_{\boldsymbol{k}_j - \boldsymbol{k}_i}(\boldsymbol{c_{k_i}}),$$

and two vectors

$$\mathbf{F}^-_\ell = (F_{\boldsymbol{k}_1,\ell}, \dots, F_{\boldsymbol{k}_N,\ell})^T, \quad \mathbf{D}_\ell = (D_{\boldsymbol{k}_1,\ell}, \dots, F_{\boldsymbol{k}_N,\ell})^T.$$

Then, equation (15) can be written in the matrix form as follows:

$$\mathbf{K}\mathbf{F}^-_{\ell+1} = \mathbf{K}^\star \mathbf{F}^-_\ell + \mathbf{D}_\ell. \tag{16}$$

Note that the both matrices $\mathbf{K}$ and $\mathbf{K}^*$ are sparse. Recall that

$$\mathbf{F}^-_0 = (\varphi(\boldsymbol{c_{k_1}}), \dots, \varphi(\boldsymbol{c_{k_N}}))^T$$

according to the initial condition, i.e. for $t = t_0$. Moreover, using the $(n+1)$-variate F-transform of the function g, one can simply compute the vector

$$\mathbf{D}_\ell = \frac{h_t}{2}(G_{\boldsymbol{k}_1,\ell+1} + G_{\boldsymbol{k}_1,\ell}, \dots, G_{\boldsymbol{k}_N,\ell+1} + G_{\boldsymbol{k}_N,\ell}))^T, \quad \ell = 0, \dots, L-1.$$

Since $\mathbf{F}^-_0$ and $\mathbf{D}_0$ are known, we can compute the vector $\mathbf{F}^-_1$ of the multivariate F-transform components for $t = t_1$ as a solution of equation (16). Repeating this procedure, one can find the components of the $(n+1)$-variate F-transform for all the indexes from $\mathbb{K}^- \times \{0, \dots, L\}$. Using the homogeneous boundary condition, one can complete the family $\mathbf{F} = \{F_{\boldsymbol{k},\ell} \mid \boldsymbol{k} \in \mathbb{K},\ \ell = 0, \dots, L\}$ of the multivariate F-transform components of f by adding

$$F_{\boldsymbol{k},\ell} = 0 \quad \text{for} \quad \boldsymbol{k} \in \mathbb{K} \setminus \mathbb{K}^- \text{ and } l = 0, \dots, L.$$

Let $\boldsymbol{k}_1 < \cdots < \boldsymbol{k}_P$, where $P = |\mathbb{K}|$, be the linear ordering of the indexes of $\mathbb{K}$ introduced in Section 3. The approximate solution of the PDEs in (10) with a homogeneous Dirichlet boundary condition and an initial condition is obtained by the inverse $(n+1)$-variate F-transform, i.e.,

$$\mathbf{F}[f](x_1, \dots, x_n, t) = \sum_{j=1}^{P} \sum_{\ell=0}^{L} F_{\boldsymbol{k}_j,\ell} A_{\boldsymbol{k}_j,\ell}(x_1, \dots, x_n, t). \tag{17}$$

for any $(x_1, \dots, x_n, t) \in \Phi = \overline{\Omega} \times [0, T]$, where $\overline{\Omega}$ denotes the topological closure of Ω. One could see that $\mathbf{F}[f]$ satisfies the homogeneous boundary and initial conditions.

5 3D boundary value problem

In this section, we present the real procedure used in the numerical solution of 3D PDEs with the help of fuzzy transform technique. A numerical solution of 2D PDEs can be found in [].

Let $\Omega = (a_1, b_1) \times (a_2, b_2)$, and let $f, g : \Omega \times (0, T) \to \mathbb{R}$ be bounded and continuous functions (f satisfies the respective differentiability conditions), and let $\alpha_{ij}, \alpha_i, \alpha : \Omega \to \mathbb{R}$, $i, j = 1, 2$ $(i \le j)$, be real functions. The task is to solve numerically the partial differential equation

$$\frac{\partial f}{\partial t} = \sum_{\substack{i,j=1 \\ i \le j}}^{2} \alpha_{ij}(x_1, x_2) \frac{\partial f}{\partial x_i \partial x_j} + \sum_{i=1}^{2} \alpha_i(x_1, x_2) \frac{\partial f}{\partial x_i} + \alpha(x_1, x_2) f + g, \tag{18}$$

under the homogeneous Dirichlet boundary condition and the initial condition:

$$f(x_1, x_2, t) = 0, \quad (x_1, x_2, t) \in \partial\Omega \times (0, T),$$
$$f(x_1, x_2, 0) = \varphi(x_1, x_2), \quad (x_1, x_2) \in \Omega.$$

Put $\Phi = \prod_{s=1}^{2}[a_s, b_s] \times [0, T]$. Let L, M_1, M_2 be natural numbers greater than 2 and $\mathbb{A}|_\Phi = \{A_{\boldsymbol{k},\ell} \mid \boldsymbol{k} \in \prod_{s=1}^{2}\{0, \dots, M_s\},\ \ell = 0, \dots, L\}$ be a 3-dimensional fuzzy partition of Φ determined by a triplet $(K, (h_1, h_2, h_t), (a_1, a_2, 0))$, where

$$h_s = \frac{b_s - a_s}{M_s},\ s = 1, 2, \quad \text{and} \quad h_t = \frac{T}{L}.$$

Put $c_{sk_s} = a_s + k_s h_s$ for any $k_s = 0, \dots, M_s$ and $s = 1, 2$. For a 3-dimensional boundary values problem, we set

$$\mathbb{K}^- = \prod_{s=1}^{2}\{1, \dots, M_s - 1\},$$

i.e., for a fixed time t_ℓ we must compute $(M_1 - 2) \cdot (M_2 - 2)$ F-transform components $F_{\boldsymbol{k},\ell}$. From (11), one can simply derive the coefficients of the matrix $\mathbf{K}$, particularly, we have

$$\begin{aligned}
C_{(1,1)}(c_{\boldsymbol{k}}) &= C_{(-1,-1)}(c_{\boldsymbol{k}}) = -\frac{\alpha_{1,2}(c_{\boldsymbol{k}})}{8h_1h_2}h_t,\\
C_{(-1,1)}(c_{\boldsymbol{k}}) &= C_{(1,-1)}(c_{\boldsymbol{k}}) = \frac{\alpha_{1,2}(c_{\boldsymbol{k}})}{8h_1h_2}h_t,\\
C_{(1,0)}(c_{\boldsymbol{k}}) &= -\left(\frac{\alpha_{1,1}(c_{\boldsymbol{k}})}{h_1^2} + \frac{\alpha_1(c_{\boldsymbol{k}})}{2h_1}\right)\frac{h_t}{2},\\
C_{(0,1)}(c_{\boldsymbol{k}}) &= -\left(\frac{\alpha_{2,2}(c_{\boldsymbol{k}})}{h_2^2} + \frac{\alpha_2(c_{\boldsymbol{k}})}{2h_2}\right)\frac{h_t}{2},\\
C_{(0,0)}(c_{\boldsymbol{k}}) &= 1 - \left(\alpha(c_{\boldsymbol{k}} - \frac{2\alpha_{1,1}(c_{\boldsymbol{k}})}{h_1^2} - \frac{2\alpha_{2,2}(c_{\boldsymbol{k}})}{h_2^2}\right)\frac{h_t}{2},\\
C_{(-1,0)}(c_{\boldsymbol{k}}) &= -\left(\frac{\alpha_{1,1}(c_{\boldsymbol{k}})}{h_1^2} - \frac{\alpha_1(c_{\boldsymbol{k}})}{2h_1}\right)\frac{h_t}{2},\\
C_{(0,-1)}(c_{\boldsymbol{k}}) &= -\left(\frac{\alpha_{2,2}(c_{\boldsymbol{k}})}{h_2^2} - \frac{\alpha_2(c_{\boldsymbol{k}})}{2h_2}\right)\frac{h_t}{2}.
\end{aligned}$$

According to (12), the (principally non-zero) coefficients of the matrix $\mathbf{K}^*$ are

$$\begin{aligned}
C^\star_{(1,1)}(\boldsymbol{c_k}) &= C^\star_{(-1,-1)}(\boldsymbol{c_k}) = -C_{(1,1)}(\boldsymbol{c_k}),\\
C^\star_{(-1,1)}(\boldsymbol{c_k}) &= C^\star_{(1,-1)}(\boldsymbol{c_k}) = -C_{(-1,1)}(\boldsymbol{c_k}),\\
C^\star_{(0,1)}(\boldsymbol{c_k}) &= -C_{(0,1)}(\boldsymbol{c_k}),\\
C^\star_{(0,-1)}(\boldsymbol{c_k}) &= -C_{(0,-1)}(\boldsymbol{c_k}),\\
C^\star_{(0,0)}(\boldsymbol{c_k}) &= 2 - C_{(0,0)}(\boldsymbol{c_k}).
\end{aligned}$$

Finally, we can compute $D_{\boldsymbol{k},\ell} = (G_{\boldsymbol{k},\ell+1} + G_{\boldsymbol{k},\ell})\,\frac{h_t}{2}$ from (13), where $G_{\boldsymbol{k},\ell}$ denote the coefficient of the multivariate F-transform of g for $\boldsymbol{k} \in \mathbb{K}^-$, which is $\boldsymbol{k}$-component of the vector $\mathbf{D}_\ell$. Introducing $\mathbf{F}_0^-$ from the initial condition, we can iteratively solve the equation in (16) to get the multivariate F-transform components for all the indexes from $\mathbb{K}^- \times \{0, \dots, L\}$. Completing the family $\mathbf{F} = \{F_{\boldsymbol{k},\ell} \mid \boldsymbol{k} \in \mathbb{K},\ \ell = 0, \dots, L\}$ using the homogeneous boundary condition and applying the inverse 3-variate F-transform we find the numerical solution of the PDEs in (18).

In the end of this section, we provide an illustrative example.

Example 2 *Solve the following partial differential equation:*

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial x \partial y} + \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} + \frac{160}{(20x + 1) \cdot (8y + 1)} \cdot f + g \tag{19}$$

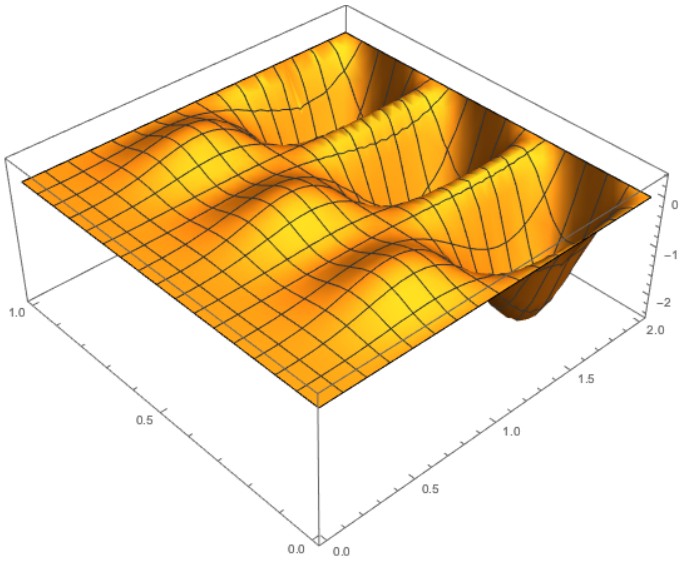


Figure 2: Solution of the PDEs from Example 2 for $t = 1$.

where

$$g(x,y,t) = t\Big(\sin(\pi x)\Big(\sin^2(3\pi y)\Big(x^2\Big(-\frac{160t}{(20x+1)(8y+1)} + 19\pi^2 t + 2\Big) - 2xt - 2t\Big) - 3\pi x(x+2)t\sin(6\pi y)\Big) - 18\pi^2 x^2 t\sin(\pi x)\cos^2(3\pi y) - \pi xt\cos(\pi x)\sin(3\pi y)\Big((x+4)\sin(3\pi y) + 6\pi x\cos(3\pi y)\Big)\Big)$$

and the boundary and initial conditions is as follows:

$$f(x,y,t) = 0, \quad (x,y,t) \in \partial((0,2)\times(0,1))\times(0,1),$$
$$f(x,y,0) = 0, \quad (x,y) \in (0,2)\times(0,1).$$

The analytical solution of the PDEs is as follows:

$$f(x,y,t) = t^2x^2\sin(\pi x)\sin^2(3\pi y), \quad (x,y,t) \in [0,2]\times[0,1]\times[0,1]. \tag{20}$$

In Fig. 2, we depict a part of the previous solution at the time $t = 1$. For simplicity, we use the triangle 3D fuzzy partition of the 3-orthotope $[0,2]\times[0,1]\times[0,1]$ determined by the triplet $(K_{TTT}, (h,h,1/64), (0,0,0))$ for different settings of the bandwidth h. The time discretization is adjusted to $L = 64 = 2^6$ and we put $h_t = 1/64$. The axis x and y are discretized by $M_m = M_{x,m} = M_{y,m} = 2^m$ for $m = 6,\dots,10$, and we put $h_{4,m} = 4/M_m$ and $h_{8,m} = 8/M_m$. The multivariate F-transforms are computed with respects to different uniform fuzzy partitions with the vector of bandwidths $(h_{r,m}, h_{r,m}, 1/64)$ for $r = 4, 8$ and $m = 6,\dots,10$. A comparison of the numerical solution of PDEs with the help of the F-transform for different uniform fuzzy partitions is depicted in Tables 1 and 2. Particularly, we compare the analytical and numerical solution of PDEs at the time $t = 1$ (the last step of iteration) with respect to the following error:

$$Err_{norm} = \frac{||f_{as} - f_{ns}||}{||f_{ns}||} \tag{21}$$

and with respect to times that are used for the computations. For a comparison, we include the finite difference method (FDM). For simplicity, we use FT-4 and FT-8 to denote the F-transforms that are derived from bandwidths $h_{4,m}$ and $h_{8,m}$ ($m = 6,\dots,10$), respectively. Note that FT-4 (FT-8) deals with 16 (64) times less sparse matrices $\mathbb{K}$ and $\mathbb{K}^*$; therefore, the computation time is significantly shorter for higher levels of discretization. One can see from tables that the F-transform

based method can obtain nearly the same results as the FDM with less effort (smaller sparse matrices) and consequently we obtain shorter computational times. Note that the computational times depicted in Table 2 include also the computational times used for the derivation of the F-transform components of G. Since the applied algorithm can be improved by, e.g., parallelization, the resulting computational times can be significantly decreased.

Discretization	FDM	FT-4	FT-8
32	0.02124		
64	0.00052	0.02265	0.07871
128	0.00130	0.00599	0.02259
256	0.00033	0.00156	0.00598
512	0.00008	0.00042	0.00155
1024	0.00002	0.00014	0.00042
2048		0.00007	0.00014

Table 1: Comparison of errors of the numerical solution of PDEs from Example 2 at the time $t = 1$ with respect to different levels of discretization.

Discretization	FDM	FT-4	FT-8
32	0.045		
64	0.25	0.24	0.19
128	2.14	0.91	0.75
256	26.5	3.6	3
512	403	15.3	12
1024	6241	76.1	50.2
2048		570	218

Table 2: Comparison of computational times for the finite difference method (FDM) and the F-transform based method (FT-4 and FT-8).

6 Conclusion

In this paper, we proposed a method for a numerical solution of a type of partial differential equations with homogeneous Dirichlet boundary conditions and initial conditions, which is based on the multivariate F-transform. We derived a system of linear equations whose solution provides the components of the direct F-transform. The approximate solution of PDEs was introduced by the inverse F-transform. The proposed method of numerical solution of PDEs by means of the F-transform was demonstrated on an artificial example 3D-boundary valued problems.

Acknowledgment

This work was supported by the project LQ1602 IT4Innovations excellence in science. The additional support was provided by the Czech Science Foundation through the project of No.16-09541S.

References

[1] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659, 1973.

[2] W. Chen and Y. Schen. Approximate solution for a class of second-order ordinary differential equations by the fuzzy transform. *Intelligent and Fuzzy Systems*, 27:73–82, 2014.

[3] Espen Gaarder Haug. *The Complete Guide to Option Pricing Formulas.* McGraw-Hill Education, 2007.

[4] Michal Holčapek, Vilém Novák, Irina Perfilieva, and Vladik Kreinovich. Necessary and sufficient conditions for generalized uniform fuzzy partitions. *Fuzzy Sets and Systems (2014), http://dx.doi.org/10.1016/j.fss.2014.10.017*, 2014.

[5] Michal Holčapek and Tomáš Tichý. Discrete multivariate f-transform of higher degree. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2015) (IEEE, 2015)*, pages 1–8.

[6] Petra Hoďáková. *Fuzzy (F-) transform of functions of two variables and its applications in image processing.* PhD thesis, University of Ostrava, 2014.

[7] A. Khastan, I. Perfilieva, and Z. Alijani. A new fuzzy approximation method to cauchy problems by fuzzy transform. *Fuzzy sets syst.*, 288:75–89, 2016.

[8] R. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4:141–183, 1973.

[9] Irina Perfilieva. Fuzzy transform: Application to the reef growth problem. In R.V. Demicco and G.J. Klir, editors, *Fuzzy Logic in Geology*, chapter 9, pages 275–300. Academic Press, Amsterdam, 2003.

[10] Irina Perfilieva. Fuzzy transforms and their applications to image compression. Bloch, Isabelle (ed.) et al., Fuzzy logic and applications. 6th international workshop, WILF 2005, Crema, Italy, September 15–17, 2005. Revised selected papers. Berlin: Springer. Lecture Notes in Computer Science 3849. Lecture Notes in Artificial Intelligence, 19–31 (2006), 2006.

[11] Martin Štěpnička and Radek Valášek. Fuzzy transforms and their application to wave equation. *J. Electr. Eng.*, 55(12):7–10, 2004.

[12] Martin Štěpnička and Radek Valášek. Numerical solution of partial differential equations with help of fuzzy transform. In *Proc. of the 2005 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2005)*, pages 1153–1162, 2005.

Maximizing Minimum Range Methods for Interval Weight Estimation from a Given Pairwise Comparison Matrix

Masahiro INUIGUCHI[1] and Shigeaki INNAN[1]

[1] *Graduate School of Engineering Science, Osaka University*
1-3 Machikaneyama-cho, Toyonaka, Osaka 560-8531
Japan
inuiguti@sys.es.osaka-u.ac.jp, innan@inulab.sys.es.osaka-u.ac.jp

Abstract

In order to express the vagueness of human judgement, methods for interval weight estimation from a crisp pairwise comparison matrix were proposed in Interval AHP. The interval weights estimated by the original method do not reflect well the vagueness of human judgement existing in the given pairwise comparison matrix. Then β-relaxation of minimum widths and γ-relaxation of minimum weighted widths are proposed for better interval weight estimation methods. However, their qualities depend on the selection of parameters β and γ. To overcome this shortcoming, a parameter-free interval weight estimation method has been proposed. In this paper, we further investigate parameter-free interval weight estimation methods and examine their usefulness by numerical experiments. We show that the parameter-free methods have similar performances to β- and γ-relaxation methods with appropriate parameters although their accuracy scores are a little worse.

Keywords: AHP, Interaval Weight Estimation, Linear Programming.

1 Introduction

AHP (Analytic Hierarchy Process) [1] is a useful tool for multiple criteria decision making. Under a predetermined hierarchical structure of criteria and alternatives, the decision maker gives pairwise comparison matrices showing the relative importance between alternatives and between criteria. From a pairwise comparison matrix between alternatives in view of each criterion, the scores of alternatives in the criterion are obtained by the maximum eigenvalue method or by the geometric mean method [2]. Similarly, from a pairwise comparison matrix between criteria, the weights of criteria are obtained also by the maximum eigenvalue method or by the geometric mean method. Using the weighted sum of scores, the holistic scores of alternatives are obtained. Because human judgement is often imperfect, the obtained pairwise comparison matrices are often inconsistent. In the conventional AHP, the consistency index is defined to evaluate the consistency and if the consistency index is in a certain range, the estimated scores and weights are accepted and used for the decision analysis.

On the other hand, from the viewpoint that the inconsistency of pairwise comparison is caused by the human vague evaluation on scores and weights, Interval AHP [3] has been proposed. In Interval AHP, weights are estimated as intervals reflecting the vagueness of human evaluation. Such interval weights are estimated by minimizing the sum of widths of interval weights under the reasonability of the given pairwise comparison matrix and the normality of interval weights. The estimation problem is reduced to a linear programming problem. However, recently, it is shown that the interval weights estimated by the conventional Interval AHP is not adequate [4, 5].

Several approaches [4, 5] to interval weight estimation have been proposed. β-relaxation method [5] considers all suboptimal interval weights by relaxing the optimality of the conventional interval weight estimation problem. The logarithmic conversion method [4] is proposed for overcoming the tendency of the conventional method that the width of a larger interval weight

becomes smaller than that of smaller interval weight. Although the logarithmic conversion method evaluates the widths of interval weights fairly but the normality condition cannot be introduced without losing the linearity of the reduced estimation problem. Then weights are introduced to the objective function for improving the unbalanced influence of the widths among interval weights [6], This approach replaces the sum of widths of interval weights with the weighted sum of widths of interval weights in the objective function of the estimation problem. The concept of β-relaxation method can be applied to this modified problem. It is called γ-relaxation method.

The results of numerical experiments show that β- and γ-relaxation method work well when parameter β and γ are chosen appropriately. Especially, β-relaxation method with appropriate β performs best. Tuning parameters β and γ may require some effort. Parameter-free methods for interval weight estimation with good performances are much-needed.

Under the circumstances, a parameter-free approach to interval weight estimation has been proposed by the authors [6]. In the parameter-free method, to estimate the interval weight of the i^{th} item (criterion or alternative), we first minimize the sum of widths of interval weights of items except the i^{th} one. Then, we maximize (resp. minimize) the upper (resp. lower) bound of interval weight of the i^{th} item with keeping the sum of widths of interval weights of items except the i^{th} one at the minimum. The maximum (resp. minimum) value defines the upper (resp. lower) bound of the estimated i^{th} interval weight. In this way, all possible fluctuations to the i^{th} item are considered in the estimated i^{th} interval weight. It is shown that this parameter-free estimation method works as well as β-relaxation method with appropriate β by numerical experiments with real-valued pairwise comparison matrices.

In this paper, we apply the idea of the previous parameter-free method to minimization of weighted sum of widths of interval weights. Moreover, we examine the performances of these parameter-free methods by numerical experiments. Although real-valued pairwise comparison matrices are used for numerical experiments in the previous paper [6], in real world applications, we do not often come across real-valued pairwise comparison matrices because human evaluation cannot be so accurate. In the conventional AHP, considering this fact, components of pairwise comparison matrices are selected from integers 1 to 9 or their reciprocals. Therefore, in this paper, we use discretized pairwise comparison matrices whose components take values from integers 1 to 9 or their reciprocals. Two numerical experiments are executed in this paper. One evaluates to what extent estimated interval weights are similar to true interval weights by a coincidence index. The other evaluates to what extent the dominance relations obtained from estimated interval weights are similar to that obtained from true interval weights. The previous interval weight estimation method in Interval AHP, β- and γ-relaxed methods and the proposed two parameter-free methods are compared in the first experiment. On the other hand, those methods as well as the eigen value method in the conventional AHP are compared in the second experiment.

This paper is organized as follows. In Section 2, the conventional Interval AHP as well as β- and γ-relaxation methods are briefly reviewed. In Section 3, two parameter-free interval weight estimation methods are proposed. In Section 4, numerical experiments and their results are described. The property and advantage of the proposed methods are demonstrated. In Section 5, some concluding remarks are given.

2 Previous Approaches in Interval AHP

2.1 The conventional method

We briefly review Interval AHP proposed in [3]. For the sake of simplicity, we define $N = \{1, 2, \ldots, n\}$ and $N\backslash j = N\backslash\{j\} = \{1, 2, \ldots, j-1, j+1, \ldots, n\}$ for $j \in N$.

In AHP, a weight vector $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)^{\mathrm{T}}$ for criteria or for alternatives is estimated from a pairwise comparison matrix A.

$$A = \begin{bmatrix} 1 & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{n1} & \cdots & 1 \end{bmatrix}, \tag{1}$$

where we assume the reciprocity, i.e., $a_{ij} = 1/a_{ij}$, $i, j \in N$. Because the $(i,j)^{\text{th}}$ component a_{ij} of A shows the relative importance of the i^{th} item over the j^{th} item. Theoretically, we have $a_{ij} = w_i/w_j$,

$i, j \in N$ for weights w_i and w_j of i^{th} and j^{th} items. However, because of the vagueness of human judgement, we assume only $a_{ij} \approx w_i/w_j$, $i, j \in N$, where $\approx$ stands for "approximately equals to". Then w_i, $i \in N$ are estimated so as to minimize the sum of differences between a_{ij} and w_i/w_j, $i, j \in N$ $(i < j)$ in the conventional AHP (see [7]).

In Interval AHP, we assume an interval weight vector $\boldsymbol{W} = (W_1, W_2, \ldots, W_n)^{\mathrm{T}}$ reflecting the vagueness of human evaluation instead of a weight vector $\boldsymbol{w}$, where $W_i = [w_i^{\mathrm{L}}, w_i^{\mathrm{R}}]$, $i \in N$ and $w_i^{\mathrm{L}} \leq w_i^{\mathrm{R}}$, $i \in N$. Accordingly, we assume that a_{ij} is obtained as w_i/w_j with randomly chosen $w_i \in W_i$ and $w_j \in W_j$. Therefore, $\boldsymbol{W}$ should satisfy $a_{ij} \in [w_i^{\mathrm{L}}/w_j^{\mathrm{R}}, w_i^{\mathrm{R}}/w_j^{\mathrm{L}}], i, j \in N,\ i < j$. Let $\mathcal{W}(A)$ be the set of all interval weight vectors $\boldsymbol{W}$ satisfying this condition. Moreover, corresponding to the normality condition of $\boldsymbol{w}$ in the conventional AHP, we require the interval weight vector $\boldsymbol{W}$ to satisfy the normality condition, i.e., $\sum_{j \in N \setminus i} w_j^{\mathrm{R}} + w_i^{\mathrm{L}} \geq 1,\ i \in N$ and $\sum_{j \in N \setminus i} w_j^{\mathrm{L}} + w_i^{\mathrm{R}} \leq 1,\ i \in N$. This condition ensures that for any $w_i^{\circ} \in W_i$, there exist $w_j \in W_j$, $j \in N \setminus i$ such that $\sum_{j \in N \setminus i} w_j + w_i^{\circ} = 1$. Let $\mathcal{W}^{\mathrm{N}}$ be the set of all interval weight vectors $\boldsymbol{W}$ satisfying the normality condition.

In Interval AHP [3], interval weights W_i, $i \in N$ are estimated by solving the following linear programming problem:

$$\underset{\boldsymbol{W}}{\text{minimize}} \{ d(\boldsymbol{W}) \mid \boldsymbol{W} \in \mathcal{W}(A) \cap \mathcal{W}^{\mathrm{N}}, \epsilon \leq w_i^{\mathrm{L}} \leq w_i^{\mathrm{R}},\ i \in N \}, \tag{2}$$

where ϵ is a small positive number and $d : \mathcal{W}^{\mathrm{N}} \to [0, +\infty)$ is defined by

$$d(\boldsymbol{W}) = \sum_{i \in N} (w_i^{\mathrm{R}} - w_i^{\mathrm{L}}). \tag{3}$$

$d(\boldsymbol{W})$ shows the sum of widths of interval weights W_i, $i \in N$ and it has been considered that the smaller $d(\boldsymbol{W})$ the better estimation.

Let $\mathcal{W}^{\mathrm{DM}}$ and $\hat{d}$ be the set of optimal solutions and the optimal value to problem (2), respectively.

Once an interval weight vector $\boldsymbol{W}$ is obtained, we define a dominance relation between alternatives under the assumption that utility values $u_i(o_p)$ of alternatives o_p in view of each criterion are given. We use dominance relation [8] defined by

$$o_p \succsim o_q \Leftrightarrow \forall \boldsymbol{w} \in \boldsymbol{W}, \mathbf{e}^{\mathrm{T}} \boldsymbol{w} = 1; \sum_{i \in N} w_i (u_i(o_p) - u_i(o_q)) \geq 0, \tag{4}$$

where $\mathbf{e} = (1, 1, \ldots, 1) \in \mathbf{R}^n$. $o_p \succsim o_q$ implies that o_p certainly dominates o_q. This dominance relation is only a preorder (reflexive and transitive) because of interval weights. From $\succsim$, we obtain a strong dominance relation $\succ$ by $o_p \succ o_q \Leftrightarrow o_p \succsim o_q$ and $o_q \not\succsim o_p$. (4) is rewritten as

$$o_p \succsim o_q \Leftrightarrow \min \left\{ \sum_{i \in N} w_i (u_i(o_p) - u_i(o_q)) \,\middle|\, \boldsymbol{w} \in \boldsymbol{W}, \mathbf{e}^{\mathrm{T}} \boldsymbol{w} = 1 \right\} \geq 0. \tag{5}$$

2.2 β- and γ-Relaxation Methods

It is shown that the conventional method for interval weight estimation does not reflect the vagueness of evaluation [5]. Some alternative approaches have been proposed and examined their usefulness by numerical experiments [5]. The results of numerical experiments [5] show that β- and γ-relaxation methods work well. Especially, β-relaxation method performs best among the alternative approaches. Then we describe β- and γ-relaxation methods in this subsection.

It is demonstrated that interval weights estimated by the conventional estimation method tend to be narrow [5, 6]. The minimization of the sum of the widths of interval weights can make the estimated interval weights too narrow. Therefore, we enlarge the widths of estimated interval weights by relaxing its minimality. Namely, we consider all interval weight vectors $\boldsymbol{W}$ such that $d(\boldsymbol{W}) \leq \beta \cdot \hat{d}$, where $\beta > 1$ shows the degree of relaxation. Consequently, the interval weight vector

$\boldsymbol{W}$ is estimated by solving the following two linear programming problems for each $j \in N$:

$$\begin{aligned} \underset{\boldsymbol{W}}{\text{maximize}}\{w_j^{\mathrm{R}} \mid \boldsymbol{W} \in \mathcal{W}(A) \cap \mathcal{W}^{\mathrm{N}},\ d(\boldsymbol{W}) \le \beta \cdot \hat{d}, \\ \epsilon \le w_i^{\mathrm{L}} \le w_i^{\mathrm{R}},\ i \in N\}, \end{aligned} \tag{6}$$

$$\begin{aligned} \underset{\boldsymbol{W}}{\text{minimize}}\{w_j^{\mathrm{L}} \mid \boldsymbol{W} \in \mathcal{W}(A) \cap \mathcal{W}^{\mathrm{N}},\ d(\boldsymbol{W}) \le \beta \cdot \hat{d}, \\ \epsilon \le w_i^{\mathrm{L}} \le w_i^{\mathrm{R}},\ i \in N\}. \end{aligned} \tag{7}$$

Then the estimated interval weights $W_j = [w_j^{\mathrm{L}}, w_j^{\mathrm{R}}]$, $j \in N$ are obtained by the optimal values of those linear programming problems. The interval weight vectors $\boldsymbol{W}$ satisfy the normality condition as show in [9].

Estimated interval weights by the conventional method has the tendency that the width of a larger interval weight becomes smaller than that of smaller interval weight. To overcome this property, weights are introduced to the objective function of the conventional interval weight estimation problem. The weights λ_i, $i \in N$ are determined by $\lambda_i = \left(\sqrt[n]{\prod_{j \in N} a_{ij}}\right)^{-1}$, $i \in N$, which are related to the estimated weights in the geometric mean method in the conventional AHP. The sum of the widths of interval weights $d(\boldsymbol{W})$ of (3) is replaced with the weighted sum of the widths of interval weights,

$$d^{\lambda}(\boldsymbol{W}) = \sum_{i \in N} \lambda_i (w_i^{\mathrm{R}} - w_i^{\mathrm{L}}). \tag{8}$$

Let $\hat{d}^{\lambda}$ be the optimal value of problem (2) with replacement of $d(\boldsymbol{W})$ with $d^{\lambda}(\boldsymbol{W})$. Then the interval weights $W_j = [w_j^{\mathrm{L}}, w_j^{\mathrm{R}}]$, $j \in N$ are estimated by (6) and (7) with replacements of $d(\boldsymbol{W})$ and $\hat{d}$ with $d^{\lambda}(\boldsymbol{W})$ and $\hat{d}^{\lambda}$. This estimation method is called γ-relaxation method.

The results of numerical experiments reported in [5] show the advantage of β-relaxation of minimum widths with a suitable parameter β over several other estimation methods. However, β-relaxation of minimum widths needs a selection of suitable parameter β. Tuning of β may require some effort and the best β depends on the problem setting.

3 The Proposed Method

Considering the parameter dependency of β- and γ-relaxation methods, we propose a parameter-free estimation methods for the interval weights. The interval weights estimated by the conventional method tend to be narrow. Then β- and γ-relaxation methods are proposed to relax the minimality of the sum and weighted sum of widths of interval weights, respectively. In the proposed parameter-free methods, we plan to make the widths of interval weights wider. The wider interval weights are, the more easily they satisfy constraints of problem (2). Therefore, maximizing the widths of interval weights with no constraints other than those of problem (2) does not work well. Then the idea of the proposed methods is to consider all possible variation of the k^{th} interval weights under the sum or weighted sum of widths of other interval weights is kept at the minimum when the k^{th} interval weight is estimated. In this way, considering the poorness of data, we give all possible fluctuation to it when we estimate the k^{th} interval weight.

Let us define the sum and weighted sum of widths of interval weights except the k^{th} one by

$$d_{\bar{k}}(\boldsymbol{W}) = \sum_{i \in N \setminus k} (w_i^{\mathrm{R}} - w_i^{\mathrm{L}}),\ k \in N, \tag{9}$$

$$d_{\bar{k}}^{\lambda}(\boldsymbol{W}) = \sum_{i \in N \setminus k} \lambda_i (w_i^{\mathrm{R}} - w_i^{\mathrm{L}}),\ k \in N. \tag{10}$$

Then the proposed method using the sum of interval weights is described as follows.

⟨1⟩ Solve the following linear programming problem for each $k \in N$:

$$\underset{\boldsymbol{W}}{\text{minimize}}\{d_{\bar{k}}(\boldsymbol{W}) \mid \boldsymbol{W} \in \mathcal{W}(A) \cap \mathcal{W}^{\mathrm{N}}, \epsilon \le w_i^{\mathrm{L}} \le w_i^{\mathrm{R}},\ i \in N\}. \tag{11}$$

Let $\hat{d}_{\bar{k}}$ be the optimal value to problem (11).

Table 1: Four settings of true interval priority weights

item	A	B	C	D
X_1	[0.27, 0.33]	[0.23, 0.37]	[0.23, 0.37]	[0.25, 0.35]
X_2	[0.21, 0.29]	[0.19, 0.31]	[0.20, 0.30]	[0.20, 0.30]
X_3	[0.15, 0.25]	[0.15, 0.25]	[0.17, 0.23]	[0.15, 0.25]
X_4	[0.09, 0.21]	[0.11, 0.19]	[0.10, 0.20]	[0.10, 0.20]
X_5	[0.03, 0.17]	[0.07, 0.13]	[0.03, 0.17]	[0.05, 0.15]

⟨2⟩ Solve the following two linear programming problems for each $k \in N$:

$$\underset{\boldsymbol{W}}{\text{maximize}}\{w_k^{\mathrm{R}} \mid \boldsymbol{W} \in \mathcal{W}(A) \cap \mathcal{W}^{\mathrm{N}},\ d_{\bar{k}}(\boldsymbol{W}) = \hat{d}_{\bar{k}}, \epsilon \le w_i^{\mathrm{L}} \le w_i^{\mathrm{R}},\ i \in N\}, \tag{12}$$

$$\underset{\boldsymbol{W}}{\text{minimize}}\{w_k^{\mathrm{L}} \mid \boldsymbol{W} \in \mathcal{W}(A) \cap \mathcal{W}^{\mathrm{N}},\ d_{\bar{k}}(\boldsymbol{W}) = \hat{d}_{\bar{k}}, \epsilon \le w_i^{\mathrm{L}} \le w_i^{\mathrm{R}},\ i \in N\}. \tag{13}$$

Let $\hat{w}_i^{\mathrm{L}}(k)$ and $\hat{w}_i^{\mathrm{R}}(k)$, $i \in N$ be values of w_i^{L} and w_i^{R}, $i \in N$, respectively, at the obtained optimal solution.

⟨3⟩ The interval weights $\check{W}_j = [\check{w}_j^{\mathrm{L}}, \check{w}_j^{\mathrm{R}}]$, $j \in N$ are estimated by the following equations:

$$\check{w}_j^{\mathrm{R}} = \max_{k \in N} \hat{w}_j^{\mathrm{R}}(k), \tag{14}$$

$$\check{w}_j^{\mathrm{L}} = \min_{k \in N} \hat{w}_j^{\mathrm{L}}(k). \tag{15}$$

Because we have $([\hat{w}_1^{\mathrm{L}}(k), \hat{w}_1^{\mathrm{R}}(k)], [\hat{w}_2^{\mathrm{L}}(k), \hat{w}_2^{\mathrm{R}}(k)], \ldots, [\hat{w}_n^{\mathrm{L}}(k), \hat{w}_n^{\mathrm{R}}(k)]) \in \mathcal{W}^{\mathrm{N}}$, $k \in N$, we obtain $([\check{w}_1^{\mathrm{L}}, \check{w}_1^{\mathrm{R}}], [\check{w}_2^{\mathrm{L}}, \check{w}_2^{\mathrm{R}}], \ldots, [\check{w}_n^{\mathrm{L}}, \check{w}_n^{\mathrm{R}}]) \in \mathcal{W}^{\mathrm{N}}$ (see [8, 9]).

We note the following facts:

1. Problems (11), (14) and (15) has bounded solution owing to the normality condition.
2. we do not always have $\check{w}_k^{\mathrm{L}} = \hat{w}_k^{\mathrm{L}}(k)$, $\check{w}_k^{\mathrm{R}} = \hat{w}_k^{\mathrm{R}}(k)$, $k \in N$ because the constraints of problems (12) and (13) depend on $k \in N$. This is why we take step ⟨3⟩.

When we use the weighted sum of interval weights instead of the sum of interval weights, $d_{\bar{k}}(\boldsymbol{W})$ and $\hat{d}_{\bar{k}}$ are replaced with $d_{\bar{k}}^{\lambda}(\boldsymbol{W})$ and $\hat{d}_{\bar{k}}^{\lambda}$, respectively in the procedure composed of ⟨1⟩ to ⟨3⟩, where $\hat{d}_{\bar{k}}^{\lambda}$ is the optimal value to problem (11) with replacement of $d_{\bar{k}}(\boldsymbol{W})$ with $d_{\bar{k}}^{\lambda}(\boldsymbol{W})$.

The good performances of the proposed method with sum of interval weights have been observed in numerical experiments with real-valued pairwise comparison matrices [6]. However, considering the human ability, it is hard for the decision maker to give precise relative importance degrees for components of pairwise comparison matrices. As indicated in the conventional AHP, each component of a pairwise comparison matrix would be given a number from 1, 2,..., 9 or their reciprocals. In this paper, we conduct the numerical experiments with pairwise comparison matrices whose components are numbers from 1, 2,..., 9 and their reciprocals. Namely, each component of a pairwise comparison matrix is assumed to be given by the closest number to the relative importance among 1, 2,..., 9 and their reciprocals. Moreover, to confirm the necessity of interval weight estimation, we compare the performance of the proposed method in preference estimation with a crisp weight estimation method of the conventional AHP.

4 Numerical Experiments

4.1 Generation of pairwise comparison matrices

In this section, we examine the performances of the proposed parameter-free interval estimation methods. by numerical experiments. In the experiments, we set $n = 5$ (number of criteria) and assume the decision maker has true interval weights $T_i = [t_i^{\mathrm{L}}, t_i^{\mathrm{R}}]$, $i \in N$ of criteria, unconsciously. Moreover, the decision maker is assumed to perceive randomly selected priority weights $v_i \in T_i$ and $v_j \in T_j$ when s/he makes a pairwise comparison. Therefore, we generate v_i and v_j by random

Table 2: Coincidence degrees of the estimated interval weights

Setting A						
Estimation method	X_1	X_2	X_3	X_4	X_5	average
Conventional method	0.104	0.240	0.423	0.400	0.457	0.325
β-relaxation method	0.354	0.499	0.490	0.639	0.707	0.538
γ-relaxation method	0.307	0.444	0.515	0.587	0.571	0.485
The proposed method (1)	0.372	0.426	0.538	0.579	0.623	0.508
The proposed method (2)	0.264	0.324	0.435	0.489	0.544	0.411
Setting B						
Estimation method	X_1	X_2	X_3	X_4	X_5	average
Conventional method	0.134	0.242	0.418	0.358	0.506	0.331
β-relaxation method	0.594	0.609	0.534	0.553	0.469	0.552
γ-relaxation method	0.618	0.578	0.561	0.551	0.590	0.580
The proposed method (1)	0.684	0.607	0.565	0.491	0.403	0.550
The proposed method (2)	0.603	0.532	0.516	0.432	0.347	0.486
Setting C						
Estimation method	X_1	X_2	X_3	X_4	X_5	average
Conventional method	0.102	0.240	0.298	0.391	0.477	0.302
β-relaxation method	0.584	0.567	0.322	0.585	0.694	0.551
γ-relaxation method	0.571	0.507	0.335	0.565	0.583	0.512
The proposed method (1)	0.664	0.517	0.335	0.521	0.626	0.533
The proposed method (2)	0.521	0.398	0.265	0.429	0.557	0.434
Setting D						
Estimation method	X_1	X_2	X_3	X_4	X_5	average
Conventional method	0.132	0.241	0.416	0.372	0.500	0.332
β-relaxation method	0.514	0.580	0.519	0.619	0.628	0.572
γ-relaxation method	0.489	0.529	0.535	0.584	0.617	0.550
The proposed method (1)	0.580	0.534	0.552	0.542	0.546	0.551
The proposed method (2)	0.460	0.427	0.478	0.471	0.471	0.461

numbers obeying uniform distribution on T_i and T_j, respectively. Namely, components a^0_{ij} and a^0_{ji} of a pairwise comparison matrix $A^0 = (a^0_{ij})$ are evaluated by v_i/v_j and v_j/v_i, respectively, for $i, j \in N$ such that $i < j$. We assume $a^0_{ii} = 1$, $i \in N$. By this way, we get a real-valued pairwise comparison matrix A^0. However, because of the limit of human recognition ability, a^0_{ij} is discretized to a_{ij} by

$$a_{ij} = \begin{cases} \min(round(a^0_{ij}), 9), & \text{if } a^0_{ij} \geq 1, \\ \dfrac{1}{\min(round(a^0_{ji}), 9)}, & \text{if } a^0_{ij} < 1, \end{cases} \tag{16}$$

where *round* is the round-off function. By this way, we obtain a discretized pairwise comparison matrix $A = (a_{ij})$. Using A, we estimate the interval priority weights $W_i = [w^{\mathrm{L}}_i, w^{\mathrm{R}}_i]$, $i \in N$ by several estimation methods. Then we evaluate the similarity between T_i and W_i, $i \in N$. To make the comparisons unbiased, we evaluate their similarities for 1,000 pairwise comparison matrices generated in the same way. In the following numerical experiments, we prepare four settings of true interval weights T_i, $i \in N$. They are given in Table 1. Each setting of true interval priority weights has its character. In setting A, the width of the true interval priority weight increases as its center value decreases. In setting B, the width of the true interval priority weight increases as its center value increases. In setting C, the width of the true interval priority weight increase as its center value leaves from 0.2. In setting D, the width of the true interval priority weight stays the same although its center value varies. In all settings, T_i satisfies $\sum_{i \in N}(t^{\mathrm{R}}_i + t^{\mathrm{L}}_i)/2 = 1$ and the normality condition, $\boldsymbol{T} = (T_1, T_2, \ldots, T_n)^{\mathrm{T}} \in \mathcal{W}^{\mathrm{N}}$.

4.2 Coincidence degrees

In order to measure the accuracy of estimated interval priority weights, we evaluate the degrees of coincidence between true interval priority weights T_i, $i \in N$ and the estimated interval priority

Table 3: Estimation performance of dominance relation between alternatives

	Setting A			Setting B		
C.I.	average	≤ 0.1	≤ 0.15	average	≤ 0.1	≤ 0.15
	0.0520	931	992	0.0278	1,000	1,000
Estimation method	risk	mismatch	reverse	risk	mismatch	reverse
Conventional method	2.853	3.082	0.004	3.192	3.494	0.009
β-relaxation method	0.416	2.974	0	0.613	2.669	0
γ-relaxation method	0.696	2.597	0	0.78	2.149	0.003
The proposed method (1)	0.470	2.701	0	0.893	3.472	0.004
The proposed method (2)	0.297	3.755	0	0.563	4.086	0
AHP (crisp estimation)	5.023	5.023	0.023	5.009	5.009	0.009
AHP with minimal difference	2.425	2.839	0	2.206	2.603	0
	Setting C			Setting D		
C.I.	average	≤ 0.1	≤ 0.15	average	≤ 0.1	≤ 0.15
	0.0525	937	991	0.0354	999	1,000
Estimation method	risk	mismatch	reverse	risk	mismatch	reverse
Conventional method	2.243	2.919	0.054	2.314	2.813	0.034
β-relaxation method	0.212	3.677	0.003	0.230	3.209	0.001
γ-relaxation method	0.291	2.926	0.006	0.343	2.563	0.002
The proposed method (1)	0.389	3.684	0.008	0.429	3.504	0.006
The proposed method (2)	0.219	4.631	0.005	0.256	4.41	0.004
AHP (crisp estimation)	4.099	4.099	0.099	4.041	4.041	0.041
AHP with minimal difference	1.698	2.537	0.005	1.585	2.325	0.001

weights W_i, $i \in N$. We calculate the following index:

$$P_i = \frac{s(T_i \cap W_i)}{s(T_i) + s(W_i) - s(T_i \cap W_i)}, \tag{17}$$

where we define $s([x^{\mathrm{L}}, x^{\mathrm{R}}]) = x^{\mathrm{R}} - x^{\mathrm{L}}$ and $s(\emptyset) = 0$. Index P_i shows a coincidence degree of the true interval priority weight with the estimated interval priority weight. Index P_i takes values in $[0, 1]$. The larger values of P_i, the better estimation.

In this experiments, we compare two proposed methods with the conventional estimation method, β-relaxation method and γ-relaxation method. The parameters β and γ are selected as $\beta = 1.2$ and $\gamma = 1.1$ which are performed well in the preliminary experiments. The averages of P_i in 1,000 pairwise comparison matrices and their average are shown in Table 2. In Table 2, the proposed method (1) stands for the proposed method using $d_{\bar{k}}$ while the proposed method (2) stands for the proposed method using $d^{\lambda}_{\bar{k}}$. In the estimation problem, for each $i \in N$, only five data about w_i is given in the form of ratios w_i/w_j between two variables. It is hard to estimate the interval weights because of a poor setting. Therefore, P_i values are not very high as shown in Table 2. Nevertheless, the estimated interval weights of some estimation methods take P_i values more than 0.5. In all settings, P_i values in the conventional estimation method are rather low and distributed widely. Thus, it is hard to say that the estimated interval weights are coincide with the true interval weights. On the other hand, β-relaxation method performs very well. Between the proposed methods, the method using $d_{\bar{k}}$ (the proposed method (1)) works better. Its performance is comparable to γ-relaxation method. γ-relaxation method also works well although its performance is less than β-relaxation method. The proposed method using $d^{\lambda}_{\bar{k}}$ (the proposed method (2)) performs less than β-, γ-relaxation and the proposed method using $d_{\bar{k}}$. Nevertheless, its performance is much better than the conventional estimation method. Considering the needlessness of parameter tuning, the proposed methods would be useful.

4.3 Estimation accuracy of dominance relation

In this experiment, we evaluate the estimation accuracy of dominance relation between alternatives. We compare the five interval estimation methods considered in the previous experiment and the eigen vector methods in the conventional AHP without and with a minimal difference threshold.

Table 4: Marginal utility function values of five alternatives

	criteria				
alternative	X_1	X_2	X_3	X_4	X_5
o_1	0.25	0.30	0.10	0.15	0.20
o_2	0.20	0.25	0.30	0.10	0.15
o_3	0.15	0.20	0.25	0.30	0.10
o_4	0.10	0.15	0.20	0.25	0.30
o_5	0.30	0.10	0.15	0.20	0.25

The method with a minimal difference threshold implies the confirmation of dominance relation between alternatives only when the total utility difference is not less than the given threshold. The introduction of the minimal difference threshold comes from the fact that the estimated dominance relation becomes a weak order in the conventional AHP unlike the true dominance relation. The threshold given in this example is 0.01. In the conventional AHP, the estimated dominance relation is accepted when the following consistency index (C.I.) is not larger than 0.1 or 0.15:

$$\text{C.I.} = \frac{\mu_{\max} - 1}{n - 1}, \tag{18}$$

where $\mu_{\max}$ is the maximal eigen value of the pairwise comparison matrix A. For reference, in each setting, the average value of C.I. and the number of pairwise comparison matrices whose C.I. is not greater than 0.1 and 0.15 are shown in the C.I. row of Table 3. Those values show that the evaluation by the conventional AHP is accepted with high probability. Therefore, the comparison with the conventional AHP is significant and important in knowing the necessity of the interval weight estimations.

In order to evaluate the dominance relation, we need marginal utility function values $u_i(o_p)$ of the i^{th} criterion. In this numerical experiment, those values are given as in Table 4. For each of 1,000 pairwise comparison matrices generated under each setting, we estimate the interval weights W_i, $i \in N$ and then the dominance relation between alternatives using (5). We compare the dominance relation estimated from W_i, $i \in N$ with the relation estimated from T_i, $i \in N$. In each pairwise comparison matrix, we count pairs of alternatives satisfying each of the following cases:

risk: A strong dominance is estimated from W_i, $i \in N$ when no dominance is confirmed in T_i, $i \in N$, or the reverse strong dominance is estimated from W_i, $i \in N$ when a strong dominance is confirmed in T_i, $i \in N$.
mismatch: Different results are obtained between evaluations by T_i, $i \in N$ and by W_i, $i \in N$.
reverse: The reverse strong dominance is estimated from W_i, $i \in N$ when a strong dominance is confirmed in T_i, $i \in N$.

If the reverse strong dominance is estimated from W_i, $i \in N$ when a strong dominance is confirmed in T_i, $i \in N$, the dominance evaluation by W_i, $i \in N$ is devastating. The dominance evaluation by W_i, $i \in N$ includes a fallacy if 'reverse' count is positive. If a strong dominance is estimated from W_i, $i \in N$ when no dominance is confirmed in T_i, $i \in N$, the dominance evaluation by W_i, $i \in N$ is audacious. The dominance evaluation by W_i, $i \in N$ includes a risk or possibly a fallacy if 'risk' count is positive. If different results are obtained between evaluations by T_i, $i \in N$ and by W_i, $i \in N$, the dominance evaluation by W_i, $i \in N$ is errant. The dominance evaluation by W_i, $i \in N$ includes a doubt if 'mismatch' count is positive.

Using 1,000 data about risk, mismatch and reverse, we obtain frequency distributions of risk, mismatch and reverse over the number of alternative pairs ranged from 0 to 10. We note that we have 10 possible alternative pairs because we have five alternatives. From the distributions, we can calculate the expected risk, mismatch and reverse. The expected risk, mismatch and reverse of each estimation method in each setting are shown in the lower part of Table 3. The smaller those expected values, the better the estimation.

From Table 3, we find that the performances of the conventional interval weight estimation method and the conventional AHP are not very good. The true dominance relation obtained from T_i, $i \in N$ is only a preorder. Indeed, no dominance is confirmed in five alternative pairs in settings A and B and in four alternative pairs in settings C and D. The fact that the estimated dominance relation becomes a weak order in the conventional AHP downgrades its evaluation. As in Table 3,

'risk' and 'mismatch' counts take at least five in settings A and B and at least four in settings C and D. To compensate this disadvantage, we introduced minimal difference threshold to the conventional AHP. The minimal difference threshold improves the performance well. Nevertheless, 'risk' count is still higher than those in the other four methods. The conventional AHP with the minimal difference threshold almost outperforms the conventional interval weight estimation method.

β- and γ-relaxation methods as well as the proposed methods (1) and (2) share a similar property that 'risk' counts are smaller than the others. Therefore, these methods can evaluate the dominances keeping 'risk' count small without a great deterioration of 'mismatch' count. Namely, for avoid a risk, applications of those interval weight estimation methods are suitable. Among those four interval weight estimation methods, β-relaxation method could be the best. However, a parameter tuning is necessary. Actually, we selected parameters β and γ so as to maximize the coincidence degrees P_i, $i \in N$ using the true interval weights. In real world applications, we do not know true interval weights so that a good selection of parameters β and γ would be difficult. In this sense, the proposed methods (1) and (2) are useful because they do not require any parameter tuning and the performances are not much worse than β-relaxation method.

5 Concluding Remarks

In this paper, we proposed two parameter-free interval weight estimation methods and examined their performances by numerical experiments. We discretized randomly generated pairwise comparison matrices considering the limit of human ability as in the conventional AHP. In a numerical experiment about coincidence degrees to true interval weights, we compared the proposed two methods with three previous methods: the conventional interval weight estimation method, β- and γ-relaxation methods. It is shown that the proposed methods are a little worse than but comparable to β- and γ-relaxation methods which perform very well. Moreover, in a numerical experiment about estimation accuracy of dominance relation, we compared the proposed two methods with formerly described three methods and the conventional AHP as well as the conventional AHP with a minimal difference threshold. It is shown that β- and γ-relaxation methods as well as the proposed two methods give risk-averse evaluations. The performances of the proposed methods are a little worse than β- and γ-relaxation methods with appropriate parameters. In the sense that the proposed methods do not require any parameter tuning, the proposed methods are useful.

The further confirmation of the usefulness of the proposed methods by numerical experiments with various settings as well as a further improvement of the proposed methods with consideration of their properties are necessary. Good estimation methods for interval weights from an interval pairwise comparison matrix should be investigated. Moreover, the proposed methods should be applied to both weight evaluations on alternatives and criteria in the hierarchy of the decision problem, and their performances should be evaluated. These would be a part of future research topics.

Acknowledgment

This work was supposed by JSPS KAKENHI Grant Number 26350423.

References

[1] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, 1980.

[2] T. L. Saaty and C. G. Vargas, Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios, *Mathematical Modelling*, Vol.5, pp.309–324, 1984.

[3] K. Sugihara and H. Tanaka: Interval evaluations in the analytic hierarchy process by possibilistic analysis, *Computational Intelligence*, Vol.17, pp.567–579, 2001.

[4] M. Inuiguchi and S. Innan, Logarithmic conversion approach to the estimation of interval priority weights from a pairwise comparison matrix, *Integrated Uncertainty in Knowledge Modelling and Decision Making: 4th International Symposium*, *IUKM 2015*, *Proceedings*, pp.89–101, 2015.

[5] S. Innan and M. Inuiguchi, Modified interval weight estimation methods for interval AHP and their comparisons, *Proceedings of ISIS 2015*, *The 16th International Symposium on Advanced Intelligent Systems*, pp.1410–1417, 2015.

[6] S. Innan and M. Inuiguchi, Improvement of interval weight estimation in Interval AHP, *Proceedings of SCIS-ISIS 2016*, *Joint 8th International Conference on Soft Computing and Intelligent Systems and 17th International Symposium on Advanced Intelligent Systems*, pp.552–557, 2016.

[7] Y. Kato, *Exposition of AHP: Fundamentals and applications* (in Japanese), Minerva Shobo, Kyoto, 2013.

[8] T. Entani and M. Inuiguchi, Pairwise comparison based interval analysis for group decision aiding with multiple criteria, *Fuzzy Sets and Systems*, Vol.274, pp.79–96, 2015.

[9] H. Tanaka, K. Sugihara and Y. Maeda, Non-additive measures by interval probability functions, *Information Sciences*, Vol.164, pp.209–227, 2004.

Annual planning in badminton using a fuzzy approach

Šárka KŘÍŽKOVÁ[1] and Martin GAVALEC[2]

University of Hradec Králové, Faculty of Informatics and Management, Czech Republic
[1] sarka.krizkova@uhk.cz, [2] martin.gavalec@uhk.cz

Abstract

This article enhances the importance of annual planning in sport, focusing specifically on badminton. This article describes the basic concept of annual planning, while reviewing the main strategies of annual planning. It also describes the basic concept of the periodisation which is used for annual planning. A great annual plan is based on good multi-criteria decision making thus in the article, it is briefly proposed usaging fuzzy theory for annual planning.

Keywords: periodisation, annual planning, badminton, decision making, fuzzy theory.

1 Introduction

Annual planning is the tool used by athletes to reach their goals. Especially while society is goal-oriented with sport being no exception. Nowadays the main aim in sports is to be better no matter how the athletes will achieve that. The athletes are pushed by their club, country or sponsors to achieve the best results. In order to achieve such great results, coaches help their athletes execute favorite sport training and reach required goals through planning and adjusting their trainings [1]. Thus the coach has to plan and schedule the entire program appropriately, so that the skills, bio-movement capabilities and mental characteristic improvements can be followed consequently and logically [2]. Badminton is one of the most complex sports in the world and at the same time is a great competition for players and coaches at various levels. Therefore, it is necessary to have an annual periodisation in order to create psychological and physiological adaptations required for athletes [2]. Despite badmintons long history in the world, there has not been yet any ideas using fuzzy theory for annual planning.

2 Literature review

The book published by [3] provides the first serious summary of scientific and empiric concept of periodisation in the training process. The recent work of [4] describes training programmes designed as an organisational/ management problem. On one hand periodisation can be defined as planned distribution or variation in training methods and means on a cyclic or periodic basis [5]. The basic goals are to exploit complementary training effects at optimal times, manage fatigue, and prevent stagnation or overtraining [6]. Corresponding decisions should be made with respect to several aspects, including biological response to training stimuli, the athletes development status and specific demands of badminton [4]. Stated differently, periodisation is a problem in the game theory and the periodisation is a form of coopetition [7]. In terms relating to game theory, periodisation is the use of planned unpredictability to manipulate or outmanoeuvre another player [8]. Paper by [9] describes that in the coaching the right amount of unpredictability should not be left to chance. The risk/reward trade off is also useful analogy. In fact, the available tactics are so numerous that the issue really becomes of avoiding haphazard strategies.

3 Annual planning in badminton

Planning a workout is methodical and a scientific procedure. Planning helps coaches to eliminate randomness as well as aimless approaches used in training. Appropriate planning considers an

athletes potential, rate of development, the facilities and equipment. An annual training program is necessary to maximise performance. In principle, this means that athletes must train continually for 11 months, then reduce the amount of work during the last month. This work should vary from regular training to facilitate physiological, psychological, and CNS rest and regeneration before beginning another year of training. Suitable planning is based on the athletes performance (in competitions or during tests), progress in training factors and competition schedule.

Annual planning is the tool used by athletes for a decade. Annual planning is in fact the roof of training procedure during the period which is divided into smaller cycles listed in the following subsections. Planning workout is methodical and scientific procedure. Planning helps coaches to eliminate the random as well as the aimless approach used in training. Appropriate planning consider an athletes potential, rate of development, the facilities and equipment. An annual training program is necessary to maximize performance. In principle, this means that athletes must train continually for 11 months, then reduce the amount of work during the last month. This work should vary from regular training to facilitate physiological, psychological, and CNS rest and regeneration before beginning another year of training. Suitable planning is based on the athletes performance (in competitions or during tests), progress in training factors and competition schedule. Annual planning is the tool used by athletes for a decade. The annual planning is in the fact the roof of training procedure during the period which is divided into smaller cycles listed in the following subsections [10].

Planning is based on the concept of the periodisation. Such scheduling increases the training organisation and provides possibility of orderly and scientific conducting the periodisation [2]. Before athletes or coaches start planning cycles they need to set a goal. Goals can be set in the long term plan as well as in the short term plan. The goal is set according the player vision or mission. The goal has to be measurable and clearly defined, because it helps to evaluate the plan.

In training methodology, one of the most challenging and complex problems is peaking athletic shape on the planned date. Often, athletes peak before the main competition due to being pushed to reach a high level without adequately alternating work with short regeneration phases. It is also common for athletes to peak after the top competition, the result of deficient preparation or an inadequate load or demand. A typical example of poor planning occurs in gymnastics when routines are finalised just before an important competition.

3.1 Sections of the annual plan

- Weekly cycles

  Usually the plan is organised into weekly sections.

- Competitions

  This section allows targeted competitions to be identified at the start of the season. Usually the tournaments are divided to three main categories. "A" tournaments are targeted tournaments, where athletes is aiming to play at their optimal level and if they do so they could win their event. On the "B" tournament the player would hope to compete well, but would not be expected to be contesting the later stages of that event. On the "C" tournament athletes tested their strategies and the result is not important at all.

- Training camps

  This section allows for the strategically placing training camps thus the player develop their ability in the different training environment.

- Testing/review

  This section helps identify the timing of the tests, formal review and evaluation.

- Activity volume

  The total amount of the time spent each week in the training and competition. It relates to the total quantity of work performed in the training session/phase. In sports like running and cycling volume can be represented as time or duration, or distance covered, whereas in the weight room volume or volume load can be reps x sets x weight lifted for each resistance exercise. For sprints, throws, and lower body jumps can be number of repetitions. Volume can

be increased through increasing the volume (method above) during the session or increasing the density (frequency) of training, or both.

- Activity intensity

  Intensity is a measure of how hard the activity - training and competition is. Intensity relates to power output, opposing force, or velocity of progression. It requires an increase in neuromuscular activation and can be expressed as meters per second (speed), kg (force), or Watts (power). The complexity of the badminton training and competition makes this very difficult to measure in the reality. Thus Borg defines the RPE scale as is shown in table1, which helps coach and players evaluate each sessions and plan each session with respect to their planned goal [11]. Two RPE scales are commonly used:

  – The original Borg scale or category scale (6 to 20 scale)
  – The revised category-ratio scale (0 to 10 scale)

  The original scale was developed in healthy individuals to correlate with exercise heart rates (e.g., RPE 15 would approximate a HR of 150 bpm), and to enable subjects to better understand terminology [11]. The category ratio scale was later developed and has since also been modified to more specifically record symptomatic breathlessness (Modified Borg Dyspnoea Scale) shown in the table 1. RPE scales are particularly valuable when HR measures of exercise intensity are inaccurate or dampened, such as in patients on beta blocker medication. This is due to the scales ability to capture the perceived exertion from central cardiovascular,respiratory and central nervous system functions.

- Activity load

  Activity load is the effective activity volume x activity intensity. In theory than, an adult internationally player should have a training load maximum of 300 i.e. 30 hours per week x 10 on the intensity scale. Of course the real number is much lower, as maximal intensity could not be reached for all the 30 hours.

Table 1: Borg Rating of Perceived Exertion Scale

Activity intensity	**Description of the difficulty of the training units**
0	Nothing at all
0.5	Very, very light
1	Very light
2	Fairly light
3	Moderate
4	Somewhat hard
5	Medium
6	Hard
7	Very hard
8	Very hard
9	Very, very hard
10	Maximal

3.2 Periodisation in the annual planning

A key factor for annual planning is the periodisation. This concept divides planning into easy manage segments called phases of training. In badminton there are two main phases - preparatory and competitive. After these two phases is the transition phase. In some literature the transition phase is included as well between preparatory and competitive phase. Preparatory phase has two subsections - general preparation and specific preparation. Competitive phase has two subsections - precompetitive and competitive.

3.3 Cycle of the annual planning

- Long term training cycle

  A period of training which ends with the peak performance. The aim is usually to win or achieve a particular level of performance on that tournament. Long term cycles can be from three months till four years length depend on the level of the athlete.

- Medium term training cycle

  As was mentioned in section periodisation we have several phases. Each of long term cycle is split into medium cycle of transition, preparation and competition.

- Short term training cycle

  Short term cycles (between 1 and 12 weeks) are sub-divisions of the medium term cycles. The transition cycle is often short in the duration so is not usually broken down into short term blocks. The preparation cycle is usually split into short term cycles called general preparation and specific preparation. General preparation tends to involve an emphasis on:

  - Basic fitness (strength, flexibility, endurance and speed)
  - Technical development (in a tactical context)

  Specific preparation involves a shift towards:

  - Sports-specific fitness (elastic strength, speed-endurance, agility etc.)
  - Tactics (employing new/refined technical skills)

  The competition cycle is usually split into short term cycles called competition development and priority competition. Activities in the competition cycle are characterized by:

  - Maintenance of fitness levels
  - High intensity, short duration work
  - Tactical emphasis

  Where a short term cycle consists of a number of weeks, it may necessary to split that down into shorter blocks (e.g. 3 to 6 weeks). This can be useful to refresh the training, giving a different emphasis and hence potentially supporting greater training adaptation. The example of annual plan is shown in Figure 1 [12] .

4 Decision making while annual planning

While coach prepares annual plan there is a lot of key factors but as well there is a lot other factors which needs to be included, but often they are forgotten. These factors can make different as being top national player and be an international player. Usually the coaches create the plan in the excel according the goal setting for the season. The top international players aiming for medals at the Olympics usually have four-years plan with several micro-cycle. The international players with goal to reach exact position in the ranking usually plan for one year according the tournament calendar.

Is it shown in Figure 2 [12] the example of annual plan with two main peak. During the season in the badminton there are two separate competitive seasons such as national championship and European championship. Because there are two distinct competitive phases such a plan is called a bicycle (bi in Latin means two). Figure 2 illustrates a bicycle that incorporates the following training phases:

- Preparatory phase I which should be the longer preparatory phase
- Competitive phase I
- Short transition (12 weeks) linked with a preparatory phase II. The unloading transition phase is for recovery

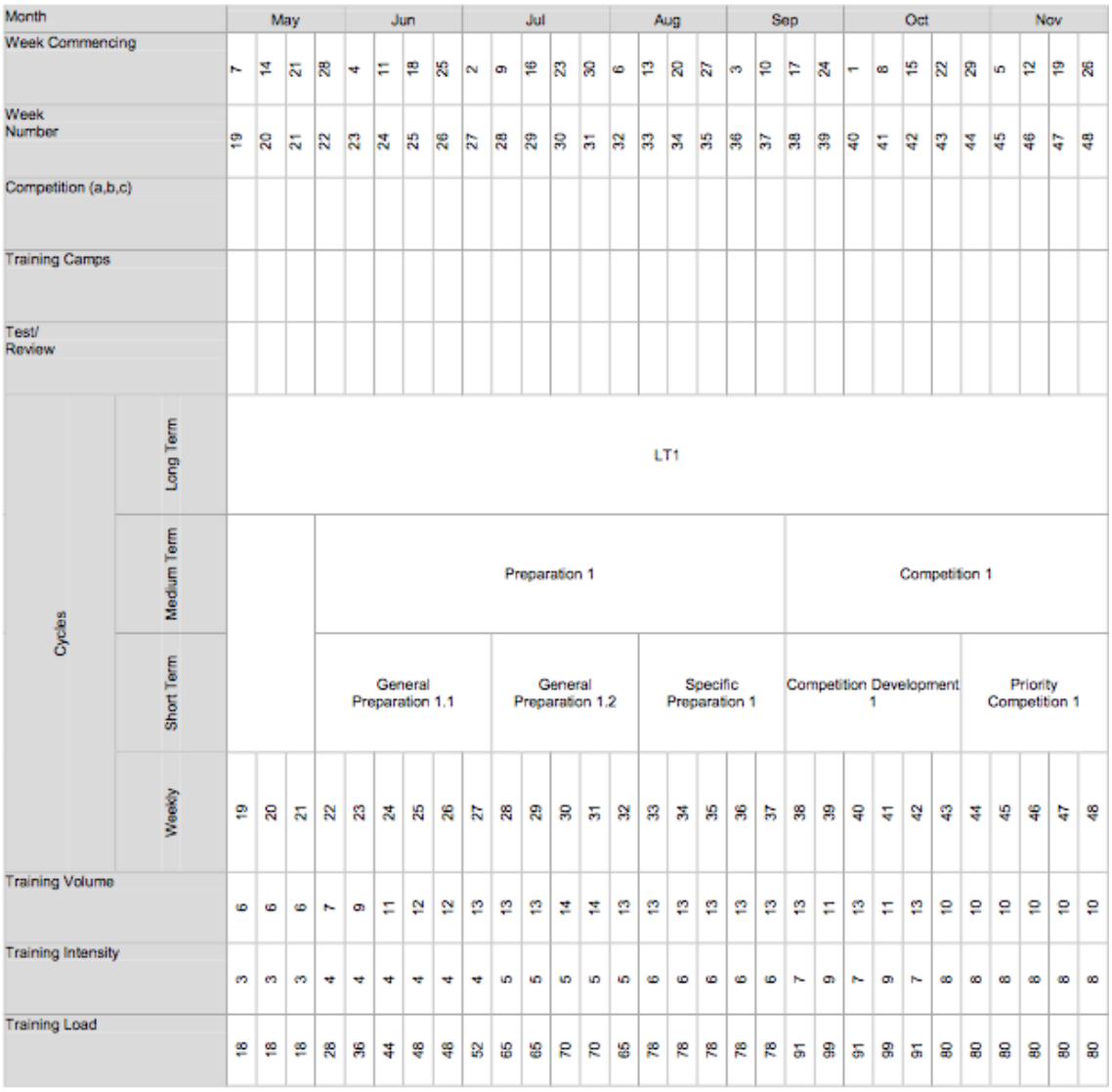


Figure 1: Example of the long cycle term

- Competitive phase II
- Transition phase

Thus this kind of planning not reflect all factors influence performance and it is very subjective despite of the coaches or athletes experiences and knowledges. In this article is proposed to used fuzzy theory to built up more objective annual plan, because the annual plan is about decision making with the aim to fulfill goals.

4.1 Fuzzy approach in decision making

Fuzzy sets theory has been broadly used for modelling decision making processes based on imprecise and vague information such as judgment of decision makers. Qualitative aspects are represented by means of linguistic variables, which are expresses qualitatively by linguistic terms and quantitatively by fuzzy set in the universe of discourse and respective membership function [13]. Thus creating the annual plan using fuzzy approach, coaches need to define important criteria for planning and evaluate them with respect to importance for the players. In the second step sub-criteria and sub sub-criteria are defined.Then the coaches need to structure hierarchical model and priori-

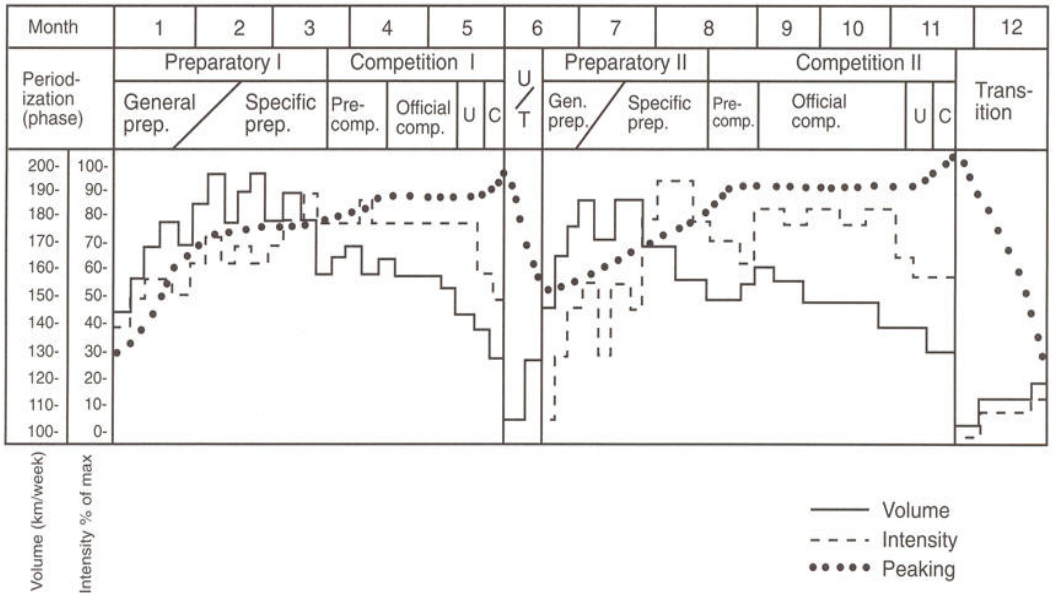


Figure 2: Example of the year plan

tise the order of criteria. Some of the key factors involved in decision making are fitness, fatigue, intensity, volume, specificity, variation, strength, endurance, periodisation and programming.

5 Conclusion

Badminton coaches are under real pressure to make a great annual plan for their athlets. They usually do not make it for one player only but for the whole club including different level of the players with different ambitions. The coaches must attempt to prepare an effective and comprehensive annual periodisation for promoting the qualitative and quantitative levels among the athletes. Thus the players develop all required factors at the championship level including physical, mental, technical and tactical readiness. These key factors are usually hardly measurable thus using fuzzy theory while designing annual plan give coaches additional support. One of the limitations of this paper is small number of articles describe decision making in the badminton.

Acknowledgement
The support of the Specific Research Grant #2107 of UHK FIM is gratefully acknowledged.

References

[1] Jay Hoffman. *Physiological aspects of sport training and performance.* Champaign, IL: Human Kinetics, 2 edition, 1961.

[2] Tudor O. Bompa. *Periodization Training for Sports.* Human Kinetics, 1 edition, 1999.

[3] Lev P Matveyev. *Fundamentals of Sports Training.* FiS Publisger, 1 edition, 1964.

[4] Mel C. Siff. *Supertraining.* Supertraining Institute, 6 edition, 2000.

[5] F. W. Dick. *Planning the programme: Sport training principle.* A and C Black, London, 3 edition, 1997.

[6] J. Graham. Periodization research and an example application. *Strength and Conditioning Journal*, 24(6):62–70, 2002.

[7] A. M.Brandenburg and B. J.Nalebuff. The right game: Use game theory to shape strategy. *Harvard Business Review*, 73(4):57–71, July/August 1995.

[8] M. D. Davis. *Game theory.* Dover, NY, 1 edition, 1997.

[9] Avinash K. Dixit and Barry J. Nalebuff. *Thinking Strategically.* W. W. Norton, NY, 1 edition, 1991.

[10] E. Alijani. *Textbook of Science of training.* Chakameh Publisher, 1 edition, 2005.

[11] G. Borg. *Borg's perceived exertion and pain scales.* Human Kinetics, 1 edition, 1998.

[12] Mike Woodward. Badminton coach education level 2, 2013.

[13] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Syst. Man Cybernet*, 3:28–44, 1973.

Fuzzy quantifiers on fuzzy four-fold tables

Jiří KUPKA[1] and Pavel RUSNOK[2]

Institute for Research and Applications of Fuzzy Modeling
Centre of Excellence IT4Innovations - Division of University of Ostrava
30. dubna 22, 701 03 Ostrava, Czech Republic
jiri.kupka@osu.cz[1] pavel.rusnok@osu.cz[2]

Abstract

We propose a generalization of four-fold table known from classical statistics to the case when fuzzy sets represent attributes in data. We define fuzzy quantifiers on the generalized four-fold table and investigate various classes of such quantifiers. We present theorems that allow constructing new fuzzy quantifiers with the help of fuzzy logic connectives. Fuzzy confirmation measure, which is used in fuzzy association analysis, is shown as an example of the implicative class of fuzzy quantifiers.

Keywords: Association analysis, Fuzzy quantifiers, Confidence, Four-fold table.

1 Introduction

Association rules and Apriori algorithm, for searching such rules, were introduced in [1, 2]. Association analysis was investigated already in a more general framework of Observational Calculi in [3]. For more recent work and detailed description of the relationship between Observational Calculi and Association rules see [4].

More theoretically founded research of association rules that preceded the research in 90's was already performed in late 60's, see [3]. Many fuzzifications of association rule method's were proposed, but many of them were ad-hoc and even repetitive. Fuzzy association analysis is still lacking more detailed theoretical analysis as it was done in the case of classical associations in [3, 4]. For a good survey of various approaches to fuzzy associations see [5] and reference therein.

The (fuzzy) associational analysis is not used purely for finding patterns in the data but is also applied in the wider context of machine learning. The classical association analysis was applied to the task of classification (see [6, Chapter 6.8] and references therein). On the other hand, also prediction and regression of variables were successfully performed with the help of fuzzy association analysis (see [7, 8, 9, 10]).

We mentioned the lack of theoretical investigation in the context of fuzzy association analysis. We think that it is needed to show which approaches are more meaningful from a theoretical point of view than just empirical evidence, that is changing from a study to study. However, there are already some investigations for example of suitable operators in support measure of fuzzy association rules in [11]. But the authors investigate only a few types of measures, and more general investigations are needed. We contribute to a wider type of association analysis by generalizing classical four-fold tables to so-called fuzzy four-fold tables. We define a fuzzy quantifier and investigate various classes of them. We follow the approach of Ivánek, in [12, 13], who showed some ways of constructing quantifiers in Observational Calculi via fuzzy connectives. We generalize his results to fuzzy four-fold table and also show the relationship to known fuzzy association analysis.

The article is structured in the following way. In Section 2, we recall some basic definitions and also we define various classes of fuzzy quantifiers. In Section 3, we present the construction theorems for fuzzy quantifiers. We conclude with Section 4.

2 Definitions

In general we consider an $(m+1)$-dimensional table D of the following form:

Table 1: Tables representing an initial dataset D and data consisting of fuzzy attributes φ, ψ, etc. that are derived from the dataset D.

	X_1	X_2	...	X_m	Y
o_1	e_{11}	e_{21}	...	e_{m1}	f_1
o_2	e_{12}	e_{22}	...	e_{m2}	f_2
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
o_N	e_{1N}	e_{2N}	...	e_{mN}	f_N

$\rightarrow$

	φ	ψ	...
o_1	$\varphi(o_1)$	$\psi(o_1)$	...
o_2	$\varphi(o_2)$	$\psi(o_2)$	...
$\vdots$	$\vdots$	$\vdots$	...
o_N	$\varphi(o_n)$	$\psi(o_N)$	...

Table 2: Example of fuzzy four-fold table $E(\varphi, \psi, \otimes, \neg)$, where $\otimes$ is the minimum t-norm and $\neg$ is involutive negation $\neg a = 1 - a$.

	φ	ψ
o_1	0.3	1
o_2	0.6	0.3

	ψ	$\neg\psi$
φ	0.6	0.6
$\neg\varphi$	1	0.4

Rows o_i are called *objects* and columns X_i and Y are *attributes* or *variables* in this paper. Our task is to get some knowledge from the data set and then to predict value of the *dependent* variable Y from a new m-dimensional vector.

Attributes φ, ψ are called *fuzzy attributes* if they are constructed with the help of a system of fuzzy sets (e.g. representing evaluative linguistic expressions). For each fuzzy attribute φ one can get $\varphi(o_i) \in [0, 1]$ for any $o_i \in D$.

Recall, that a *t-norm* (triangular norm) is a map $\otimes : [0, 1]^2 \mapsto [0, 1]$, which is commutative, associative, non-decreasing in both arguments and its neutral element is 1, i.e. $a \otimes 1 = a$ for all $a \in [0, 1]$. A t-conorm $\oplus$ satisfies all the axioms of t-norm except that the neutral element is 0, i.e. $0 \oplus a = a$ for all $a \in [0, 1]$. Finally, a negation is a non-increasing operation $\neg : [0, 1] \mapsto [0, 1]$ such that $\neg 0 = 1$ and $\neg 1 = 0$.

For any two fuzzy attributes φ, ψ, a generalized fuzzy four-fold table $E(\varphi, \psi, \otimes, \neg)$ can be constructed, i.e.

$$E := \begin{array}{c|c|c} & \psi & \neg\psi \\ \hline \varphi & a & b \\ \hline \neg\varphi & c & d \end{array}, \tag{1}$$

where

$$\begin{aligned} a &= \sum_{o_i \in D} \varphi(o_i) \otimes \psi(o_i), \\ b &= \sum_{o_i \in D} \varphi(o_i) \otimes \neg\psi(o_i), \\ c &= \sum_{o_i \in D} \neg\varphi(o_i) \otimes \psi(o_i), \\ d &= \sum_{o_i \in D} \neg\varphi(o_i) \otimes \neg\psi(o_i). \end{aligned}$$

Example 1 *Let us consider an example of trivial data in Table 2 consisting of two objects with two fuzzy attributes defined on the objects. The fuzzy four-fold table $E(\varphi, \psi, \otimes, \neg)$ with the minimum t-norm and the involutive negation is calculated next to it.*

2.1 Fuzzy quantifiers

A *fuzzy quantifier* q is a map defined on the set of the fuzzy four-fold tables, i.e. $q : (\mathbb{R}_0^+)^4 \to [0, 1]$, $(a, b, c, d) \mapsto e$. To simplify the notation, we occasionally write fuzzy quantifiers as maps of fewer variables, when they do not depend on all variables. There are many classes of classic (non-fuzzy) quantifiers, for a survey of them we refer to [14].

A fuzzy quantifier q is *implicational* if and only if it satisfies a *monotonicity condition*: $a' \geq a, b' \leq b$ implies

$$q(a', b') \geq q(a, b). \tag{2}$$

A fuzzy quantifier q is *double implicational* if and only if $a' \geq a$, $b' \leq b$ and $c' \leq c$ implies

$$q(a', b', c') \geq q(a, b, c).$$

A fuzzy quantifier q is *equivalence* if and only if $a' \geq a$, $b' \leq b$, $c' \leq c$ and $d' \geq d$ implies

$$q(a', b', c', d') \geq q(a, b, c, d).$$

A fuzzy quantifier q is *ratio-implicational* if and only if $a'b \geq ab'$ implies

$$q(a', b') \geq q(a, b).$$

It follows easily from definitions that every equivalence quantifier is double implicational and every double implicational quantifier is implicational as well. Moreover, every ratio-implicational quantifier is also an implicational one.

2.2 Examples

Here we present examples of every class of fuzzy quantifier introduced in previous section.

A quantifier

$$q(a, b) = \frac{a}{a + b}$$

is *implicational* and known from association rule mining as the confidence measure.

A quantifier

$$q(a, b, c) = \frac{a}{a + b + c}$$

is *double implicational* and is known after the name Jaccard measure.

A quantifier

$$q(a, b, c, d) = \frac{a + d}{a + b + c + d}$$

is *equivalence.*

And finally, a quantifier

$$q(a, b) = \frac{a}{a + \theta \cdot b} \qquad \theta > 0$$

is ratio-implicational.

3 Construction theorems

Results mentioned in this section are highly motivated by results obtained by J. Iványek in [12, 13]. Iványek proved seven construction theorems for quantifiers defined on classical four-fold tables. We generalize his results for fuzzy quantifiers on fuzzy four-fold tables.

3.1 From implicational quantifiers equivalency ones

It is possible to define a double implicational quantifier from implicational one as the following lemma shows.

Lemma 1 *Let q be an implicational quantifier and let $\otimes : [0, 1]^2 \to [0, 1]$ be a map nondecreasing in each argument. Then the quantifier $\bar{q}$ constructed by*

$$\bar{q}(a, b, c) := \otimes(q(a, b), q(a, c))$$

is double implicational. If $\otimes$ is a t-norm then $\bar{q}$ satisfies the inequality

$$\bar{q}(a, b, c) \leq \min(q(a, b), q(a, c))$$

for all generalized four-fold tables (1).

Furthermore, we can get from double implicational quantifier, by the following construction, the equivalence quantifier.

Lemma 2 *Let q be a double implicational quantifier and $\oplus : [0, 1]^2 \to [0, 1]$ be a map nondecreasing in each argument. Then the quantifier $\tilde{q}$ constructed by*

$$\tilde{q}(a, b, c, d) := \oplus(q(a, b, c), q(d, b, c))$$

is an equivalence. Moreover, if $\oplus$ is a t-conorm, $\tilde{q}$ satisfies the inequality

$$\tilde{q}(a, b, c, d) \geq \max(q(a, b, c), q(d, b, c))$$

for all generalized four-fold tables (1).

We have shown how to construct the double implicational quantifiers from implicational quantifiers first and then from there the equivalency quantifiers. The recipes for constructing implicational quantifiers themselves are shown in the following section.

3.2 From fuzzy implications to fuzzy implicational quantifiers

Below, we consider *boundary conditions* for implicational quantifiers - $q(\infty,\infty) = q(0,0) = 1$ and $q(0,\infty) = 0$. To shorten our notation, fuzzy implicational quantifiers can be seen as maps $q : (\mathbb{R}_0^+)^2 \to [0,1]$, $(a,b) \mapsto e$, satisfying boundary and the monotonicity condition (2).

Remark 1 *In the original definition ([12]), any implicational quantifier is a map defined on pairs of natural numbers (i.e. on $(\mathbb{N} \cup \{0\})^2$) preserving the monotonicity condition. Our boundary condition (and also dealing with fuzzy attributes) requires extension of this notion to $\bar{\mathbb{R}}^2$ where $\bar{\mathbb{R}}$ denotes the extended real half-line $[0,\infty]$. Boundary conditions are simple natural restrictions coming from the notion of a confidence of an association rule used in data mining.*

A *fuzzy implication* is a binary operation $I : [0,1]^2 \to [0,1]$ for which $I(0,0) = I(1,1) = 1$, $I(1,0) = 0$ and $x' \leq x$, $y' \geq y$ implies $I(x',y') \geq I(x,y)$.

Lemma 3 *Let I be fuzzy implication and $\varphi_1, \varphi_0 : \bar{\mathbb{R}} \to [0,1]$ be functions such that, for $i = 1,2$,*

1. *φ_i is nonincreasing,*
2. *$\varphi_i(0) = 1$, and*
3. *$\varphi_i(\infty) = 0$.*

Then $q_I : \bar{\mathbb{R}}^2 \to [0,1]$ defined by $q_I(a,b) := I(\varphi_1(a), \varphi_0(b))$ is a fuzzy implicational quantifier.

Note that the fuzzy implicational quantifiers fulfill boundary conditions automatically due to the construction in Lemma 3. However, we need to impose the boundary conditions on q for the construction in the following Lemma 4 to work.

3.3 From fuzzy implicational quantifiers to fuzzy implications

It is possible to go the other way around and construct fuzzy implications from fuzzy implicational quantifiers.

Lemma 4 *Let q be a fuzzy implicational quantifier and $\psi_1, \psi_0 : [0,1] \to \bar{\mathbb{R}}$ be functions such that, for $i = 1,2$,*

1. *ψ_i is nonincreasing,*
2. *$\psi_i(1) = 0$, and*
3. *$\psi_i(0) = \infty$.*

Then $I_q : [0,1]^2 \to [0,1]$ defined by $I_q(a,b) := q(\psi_1(a), \psi_0(b))$ is a fuzzy implication.

Remark 2 *In Lemmas above one can take simply any rotation $\tilde{\varphi} : \bar{\mathbb{R}}^2 \to [0,1]^2$ (or its inverse) for which*

$$\tilde{\varphi}(0,0) = (1,1),\ \tilde{\varphi}(\infty,0) = (0,1),\ \tilde{\varphi}(0,\infty) = (1,0),\ \tilde{\varphi}(\infty,\infty) = (0,0),$$

and which reverses monotonicity of each argument. Special cases of such rotations are maps in Lemmas 3 and 4 above.

3.4 Fuzzy ratio-implicational quantifiers

Ratio-implicational quantifiers have certain properties summarized in the following lemma.

Lemma 5 *Let $q(a,b)$ be a ratio-implicational fuzzy quantifier. Then:*

(i) $q(a,b)$ *is implicational,*

(ii) if $a'b = ab'$ *then* $q(a',b') = q(a,b)$,

(iii) there are numbers m^*, $M^* \in [0,1]$ *such that*

$m^* = q(0,b)$ *for all* $b > 0$,
$M^* = q(a,0)$ *for all* $a > 0$,
$m^* \leq q(a,b) \leq M^*$ *for all* $a,b > 0$.

(iv) there is a non-decreasing function f^* *defined on non-negative real numbers and* ∞ *such that*

$q(a,b) = f^*(\frac{a}{b})$.

Lemma 5 allows us to prove the following two theorems, which link together the ratio-implicational quantifiers, fuzzy implications and ratios $\frac{b}{a+b}$ and $\frac{a}{a+b}$.

Theorem 1 *Let I be a fuzzy implication then $q_I(a,b) = I\left(\frac{b}{a+b}, \frac{a}{a+b}\right)$ is a ratio-implicational quantifier with $m^* = 0$ and $M^* = 1$.*

Theorem 2 states very strong statement, that every ratio-implicational quantifier, which fulfills the conditions $q(0,b) = 0$ for all $b > 0$ and $q(a,0) = 0$ for all $a > 0$, can be expressed as fuzzy implication.

Theorem 2 *Let q be a ratio-implicational quantifier with $m^* = 0$ and $M^* = 1$. Then there exists a fuzzy implication I_q such that*

$$q(a,b) = I_q\left(\frac{b}{a+b}, \frac{a}{a+b}\right).$$

4 Conclusion

We have presented a few possible ways of constructing fuzzy quantifiers on fuzzy four-fold tables. Some of the constructions can be used in post-processing of fuzzy association rules, which are mined from data, while its confidence (implicational fuzzy quantifier) is calculated for them. With the help of constructions presented in this paper, we may mine double implications and equivalences in the sense of Section 2.1.

Our general definition of fuzzy four-fold table allows us to investigate other approaches to fuzzy association rules and compare them. In our future work, we will concentrate on this.

Acknowledgment

This work was supported by the NPU II project LQ1602 IT4Innovations excellence in science.

References

[1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216, 1993.

[2] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, pages 487–499, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.

[3] Petr Hájek and Tomáš Havránek. *Mechanizing Hypothesis Formation (Mathematical Foundations for a General Theory)*. Springer-Verlag, 1978.

[4] Jan Rauch. *Observational calculi and association rules*, volume 469. Springer, 2012.

[5] Martin Ralbovský. *Fuzzy GUHA*. PhD thesis, University of Economics, Prague, 2009.

[6] Jiawei Han. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[7] Martin Štěpnička, Antonín Dvořák, Viktor Pavliska, and Lenka Vavříčková. A linguistic approach to time series modeling with the help of f-transform. *Fuzzy Sets and Systems*, 180(1):164–184, 2011.

[8] Martin Štepnička, Michal Burda, and Lenka Štepničková. Fuzzy rule base ensemble generated from data by linguistic associations mining. *Fuzzy Sets and Systems*, 285:140–161, 2016.

[9] Jiří Kupka and Pavel Rusnok. On regression methods based on linguistic descriptions. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–7. IEEE, 2015.

[10] Pavel Rusnok. Linguistic descriptions as a modeling tool for multivariate time series. In *Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology*, 2015.

[11] Didier Dubois, Eyke Hüllermeier, and Henri Prade. A systematic approach to the assessment of fuzzy association rules. *Data Mining and Knowledge Discovery*, 13(2):167–192, 2006.

[12] Jiří Ivánek. Construction of implicational quantifiers from fuzzy implications. *Fuzzy sets and systems*, 151(2):381–391, 2005.

[13] Jiří Ivánek. Combining implicational quantifiers for equivalence ones by fuzzy connectives. *International journal of intelligent systems*, 21(3):325–334, 2006.

[14] Jan Rauch. Classes of association rules: an overview. In *Data Mining: Foundations and Practice*, pages 315–337. Springer, 2008.

Transductive Learning Using Feature Spaces Provided by Semidefinite Programmings

Yoshifumi KUSUNOKI, Keigo IWASA and Keiji TATSUMI

Graduate School of Engineering, Osaka University
2-1, Yamadaoka, Suita, Osaka 565-0871
Japan
kusunoki@eei.eng.osaka-u.ac.jp

Abstract

In this paper, we study transductive learning, which is a non-inductive version of semi-supervised learning. Under assumption that data in the same cluster have the same label, we exploit the cluster structure to estimate the labels of unlabeled data. To that end, we focus feature vectors obtained by semidefinite-programming-based clustering methods. Such feature vectors emphasize the cluster structure in the data. That is, data in the same cluster are mapped to similar feature vectors. Hence, we expect that a classifier for feature vectors of labeled data can correctly classify unlabeled data. The above idea is realized by simultaneously optimization of semidefinite programmings for clustering and the max-margin problem of the support vector machine.

Keywords: Transductive learning, Support vector machine, Semidefinite programming.

1 Introduction

In this paper, we propose a new method to solve transductive learning, which is a machine learning problem like supervised learning. Supervised learning is to construct a classifier from a data set consisting of labeled objects, i.e., input and output (class label) pairs. On the other hand, transductive learning is another problem to assign labels to unlabeled objects using (a few) labeled objects and (many) labeled objects. We can use the information of the input vectors of unlabeled objects in the transductive learning unlike in the supervised learning. Additionally, in the transductive learning, we need not to obtain classifiers for unseen objects. When there are a lot of data but assigning labels takes much cost, the transductive learning is effective.

Since we know unlabeled object in transductive learning, we can exploit distribution and/or cluster structures of data for the label prediction. In this study, under the assumption that objects in the same cluster are grouped into the same class, we consider a transductive learning method based on clustering techniques. In semidefinite-programming-based clustering methods, such as max-cut [4], we solve an optimization problem where objects which may belong to the same cluster are mapped to similar feature vectors by dealing with the Gram matrix of feature vectors as a variable to be optimized. We combine that idea with SVM (Support Vector Machine) [1], which is one of major methods of supervised learning. In other words, we simultaneously optimize the margin maximization problem of SVM for classification of labeled objects and the clustering problem to find suitable feature vectors. As a result, we expect that unlabeled objects are correctly classified.

The rest of this paper is organized as follows. In Section 2, we describe settings of transductive learning. Additionally, we introduce S^3VM [2] which is an SVM-based method for transductive learning. In Section 3, we provide the detail of the proposed method. First, we reformulate the optimization of SVM to a semidefinite programming. Remarking that we can deal with the Gram matrix of objects as a variable, we add objective functions and constraints for the Gram matrix to the semidefinite programming. The labels of the unlabeled objects are determined by its optimal solution. In Section 4, to examine the effectiveness of the proposed method, we show results of numerical experiments. We argue that the proposed method reduces classification error for the unlabeled objects comparing with S^3VM and SVM only for the labeled objects.

2 Transductive Learning

2.1 Settings

In this section, we explain problem settings in transductive (semi-supervised) learning. Let $\mathcal{X} \subseteq \mathbf{R}^n$ be the set of input vectors which are n-dimensional real valued, and let $C = \{1, \ldots, c\}$ be the set of labels. A data set S is given by m pairs of input vectors and labels: $S = ((x^1, y_1), \ldots, (x^m, y_m))$, where $(x^i, y_i) \in \mathcal{X} \times C \cup \{0\}$ for $i = 1, \ldots, m$. Here $y_i = 0$ means that the i-th label is unknown. The aim of transductive learning is to determine the labels of the unlabeled objects using the whole data set S. Remark that, contrary to supervised learning, we can exploit input vectors x^i of unlabeled objects. On the other hand, *inductive* semi-supervised learning is a problem to obtain a classifier d which assigns a label to each input vectors in $\mathcal{X}$.

2.2 Semi-supervised Support Vector Machine

Now, we introduce S^{3}VM (semi-supervised support vector machine) studied in [2]. S^{3}VM is a extension of SVM (support vector machine), which is a method for supervised learning. S^{3}VM can be applied to the inductive semi-supervised learning.

S^{3}VM models a classifier by c discriminant functions $f_p : \mathcal{X} \to \mathbf{R}$, $p = 1, \ldots, c$. They are real-valued linear functions: $f_p(x) = (w^p)^\top x + b_p$. The label of $x \in \mathcal{X}$ is determined by a classifier $d : \mathcal{X} \to C$, which is based on the linear discriminant functions:

$$d(x) = \operatorname*{argmax}_{p \in C} f_p(x).$$

The parameters $((w^1, b_1), \ldots, (w^c, b_c))$ are obtained by the one-against-all approach. That is, for each label $p \in C$ we consider a 2-class problem associated with S, and obtain the p-th parameter (w^p, b_p) by binary S^{3}VM. The data set of the p-th 2-class problem, denoted by $S_p = ((x^1, y_1^p), \ldots, (x^m, y_m^p))$, is defined as follows.

$$y_i^p = \begin{cases} 1 & \text{if } y_i = p, \\ -1 & \text{if } y_i \in C \setminus \{p\}, \\ 0 & \text{if } y_i = 0. \end{cases}$$

In the rest of this section, we consider S^{3}VM for 2-class problems. Let $C = \{-1, 1\}$ be the set of labels. We deal with one linear discriminant function $f(x) = w^\top x + b$. In S^{3}VM, the following optimization problem is solved to obtain the parameter (w, b) and the labels $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_m)$ of objects.

$$\begin{aligned} \underset{w,b,\tilde{y}}{\text{minimize}} \quad & \frac{1}{2}\|w\|^2 + \frac{\mu_l^2}{2} \sum_{i:y_i \neq 0} L(\tilde{y}_i, f(x^i)) + \frac{\mu_u^2}{2} \sum_{i:y_i = 0} L(\tilde{y}_i, f(x^i)) \\ \text{subject to} \quad & \tilde{y}_i = y_i, \ i : y_i \neq 0, \\ & \sum_{i:y_i=0} \tilde{y}_i = 2r - 1, \end{aligned} \tag{1}$$

where, L is the ℓ_2 hinge loss function defined as follows.

$$L(y, y') = (\max\{(0, 1 - yy')\})^2. \tag{2}$$

The first term in the objective function of (1) is a regularization term. The second and third terms in the objective function is losses of classification for the labeled and unlabeled objects, respectively. μ_l and μ_u are weights for the losses.

The first constraint of (1) fixes the labels of labeled objects. The second constraint balances the numbers of objects in class -1 and class 1, r is the constant adjusting the balance. In this study, we set $r = \left\lceil m_u \frac{|\{i \mid y_i = 1\}|}{m_l} \right\rceil$, where m_l and m_u are the numbers of labeled and unlabeled objects, respectively.

To solve the problem (1), we adopt S^{3}VM^{light} [2]. S^{3}VM^{light} is an algorithm based on a label exchange local search. First, the problem (1) without the third term in the objective function

and the second constraint is solved, and the parameter (w, b) of f is obtained. b is modified to satisfy the second constraint of balancing classes. Label variables $\tilde{y}_i$ of unlabeled objects i are fixed according to the discriminant function f.

$$\tilde{y}_i = \begin{cases} 1 & \text{if } f(x^i) \geq 0 \\ -1 & \text{if } f(x^i) < 0 \end{cases} \tag{3}$$

Next, with fixing the variables $\tilde{y}_i$, $i = 1, \ldots, m$, the problem (1) is solved, and obtain the discriminant function f. With fixing f, we consider a pair (i, j) satisfying the following condition.

$$y_i = y_j = 0,\ \tilde{y}_i \neq \tilde{y}_j,\ L(\tilde{y}_i, f(x^i)) + L(\tilde{y}_j, f(x^j)) > L(-\tilde{y}_i, f(x^i)) + L(-\tilde{y}_j, f(x^j)). \tag{4}$$

This condition means that the loss L is decreased if the objects i and j are exchanged. Hence, we update their labels as $\tilde{y}_i = -\tilde{y}_i$ and $\tilde{y}_j = -\tilde{y}_j$.

As shown above, to solve the problem (1), S^{3}VM^{light} iteratively and alternately solves (1) with fixed $\tilde{y}$ and updates $\tilde{y}$ with the fixed discriminant function f. Furthermore, by gradually increasing the weight μ_u, an effect like simulated annealing is provided, and prevents solutions from trapping local optimal points.

3 Feature Space Obtained by Semidefinite Programming

In this study, we propose a method for transductive learning using feature vectors reflecting cluster structure. First, we reformulate the optimization problem of SVM to a semidefinite programming.

We consider SVM for supervised learning, i.e., there are no unlabeled objects. An optimization problem of SVM with the one-against-all approach is formulated by the following convex quadratic programming.

$$\begin{aligned} \underset{w,b}{\text{minimize}} \quad & \sum_{p=1}^{c} \|w^p\|^2 \\ \text{subject to} \quad & y_i^p((w^p)^\top x^i + b_p) \geq 1,\ i = 1, \ldots, m,\ p = 1, \ldots, c. \end{aligned} \tag{5}$$

Let $X = [x^1\ \cdots\ x^m]$ be the matrix of the input vectors, and $W = [w^1\ \cdots\ w^c]$ be that of normal vectors. Define $\tilde{X} = [X\ W]$. The Gram matrix of $\tilde{X}$ is

$$\tilde{X}^\top \tilde{X} = \begin{bmatrix} X^\top X & X^\top W \\ W^\top X & W^\top W \end{bmatrix}. \tag{6}$$

Furthermore, let $U = X^\top W$, $S = X^\top X$, $T = W^\top W$. The problem (5) is reformulate the following semidefinite programming.

$$\begin{aligned} \underset{T,U,b}{\text{minimize}} \quad & \mathbf{tr}(T) \\ \text{subject to} \quad & y_i^p(u_{ip} + b_p) \geq 1,\ i = 1, \ldots, m,\ p = 1, \ldots, c, \\ & S = X^\top X, \\ & \begin{bmatrix} S & U \\ U^\top & T \end{bmatrix} \succeq O, \end{aligned} \tag{7}$$

where $\mathbf{tr}$ means the trace of a matrix, and $A \succeq 0$ means A is a semidefinite matrix. u_{ip} is the ip-th element of the matrix U.

The Gram matrix S is fixed to $X^\top X$ in (7), however, we can deal with S as a variable. Consider a nonlinear mapping φ for the input vectors $x^1, \ldots, x^m$, and their feature vectors $\varphi(x^1), \ldots, \varphi(x^m)$. Suppose that S is the Gram matrix of the feature vectors. The problem (7) is considered as SVM with the nonlinear mapping φ. Since providing the feature vectors $\varphi(x^1), \ldots, \varphi(x^m)$ is equivalent to providing their Gram matrix S, we can deal with the nonlinear mapping as a variable in the optimization problem. By considering appropriate objective functions and constraints for S, we can simultaneously obtain optimal feature vectors and discriminant functions in the corresponding feature space.

3.1 Proposed Method

To realize the above idea, we consider suitable feature vectors for transductive learning. In this study, we assume that objects belonging to the same cluster in the input space have the same label. Under that assumption, if we can concentrate objects which can be grouped as clusters in the input space, we can attain correct classification of unlabeled objects by classifying labeled objects.

We consider objective functions and constrains, added to (7), to obtain suitable feature vectors. Let G be the Gram matrix of the input vectors or a kernel matrix induced by some kernel function, and let S be a matrix variable representing the Gram matrix of feature vectors. First, we propose maximization of the inner product of matrices $\langle G, S\rangle = \sum_{i=1}^{m}\sum_{j=1}^{m} g_{ij}s_{ij}$, where g_{ij} and s_{ij} are the ij-th element of G and S, respectively. To maximizing $\langle G, S\rangle$, the inner product s_{ij} in the feature space becomes large if g_{ij} in the input space is large, in other words, i and j are put closely in feature space if they are close in the input space. Furthermore, we expected that the closeness in the input space is emphasized in the feature space since we maximize linear function $\langle G, S\rangle$.

However, only by maximizing of $\langle G, S\rangle$, we may obtain a totally different distribution of feature vectors from that of input vectors. To avoid such unsuitable cases, we fix the inner products of some pairs (i, j) to $s_{ij} = g_{ij}$. This constraint moderates deviation of feature vectors from input vectors.

Let $E \in \{0, 1\}^{m\times m}$ be a matrix representing unfixed pairs. That is, letting e_{ij} be the ij-th element of E, for ij such that s_{ij} is free, we set $e_{ij} = 1$, and for ij such that $s_{ij} = g_{ij}$, we set $e_{ij} = 0$. The diagonal elements e_{ii} of E is 0, and E is symmetric, namely $s_{ij} = s_{ji}$.

Provided E and a kernel matrix G, we consider the following optimization problem.

$$\begin{array}{ll} \underset{S,T,U,b}{\text{minimize}} & -\displaystyle\sum_{(i,j):e_{ij}=1} g_{ij}s_{ij} + \lambda\,\mathbf{tr}(T) \\ \text{subject to} & y_i^p(u_{ip} + b_p) \geq 1,\ i : y_i \neq 0,\ p = 1, \ldots, c, \\ & s_{ij} = g_{ij},\ (i, j) : e_{ij} = 0, \\ & \begin{bmatrix} S & U \\ U^\top & T \end{bmatrix} \succeq O, \end{array} \tag{8}$$

where $\lambda > 0$ is a parameter balancing maximizing $\langle G, S\rangle$ and the objective function of SVM, and the first constraint is to correctly classify the labeled objects. Let U, b be an optimal solution of (8). The unlabeled objects are classified by the following decision rule.

$$d_i = \underset{p=1,\ldots,c}{\operatorname{argmax}}\{u_{ip} + b_p\}. \tag{9}$$

By increasing λ, the margins of linear discriminant functions for the feature vectors of labeled objects become larger. On the other hand, the effect of maximizing $\langle G, S\rangle$ becomes stronger.

Consider the problem (8) without the first constraint, and consider a sub-matrix $G' = (g_{ij})_{i,j=k_1,\ldots,k_l}$ of G and a sub-matrix $E' = (e_{ij})_{i,j=k_1,\ldots,k_l}$ of E. Suppose that all elements of G' are non-negative and all non-diagonal elements of E' is 1. Then, all feature vectors corresponding to $k_1, \ldots, k_l$ become parallel. Roughly speaking, the objective function of maximizing $\langle G, S\rangle$ has an effect to concentrate feature vectors. In contrast, the first constraint of (8) separates objects in different classes.

When all elements of E is 0, the optimization problem (8) coincides with SVM for supervised learning.

The dual problem of (8) is expressed as follows.

$$\begin{aligned} \underset{R,V}{\text{minimize}} \quad & \sum_{(i,j):e_{ij}=0} g_{ij} r_{ij} + 2 \sum_{i:y_i \neq 0} \sum_{p=1}^{c} y_i^p v_{ip} \\ \text{subject to} \quad & y_i^p v_{ip} \leq 0,\ i : y_i \neq 0,\ p = 1, \ldots, c, \\ & \sum_{i:y_i \neq 0} v_{ip} = 0,\ p = 1, \ldots, c, \\ & v_{ip} = 0,\ i : y_i = 0,\ p = 1, \ldots, c, \\ & r_{ij} = g_{ij},\ (i,j) : e_{ij} = 1, \\ & \begin{bmatrix} R & V \\ V^\top & \lambda I \end{bmatrix} \succeq O, \end{aligned} \tag{10}$$

where r_{ij} and v_{ip} express elements of R and V, respectively.

We solve semidefinite programmings by the primal-dual interior point method. Let k be the size of the semidefinite matrix and let l be the number of constraints in the standard form of a semidefinite programming. Each iteration of the interior point method needs computational cost $O(k^3 + l^3)$ [3]. Hence, when the number of the elements ij such that $e_{ij} = 1$ is small, solving (10) as the standard form reduces the computational cost. On the other hand, the number of the elements ij such that $e_{ij} = 1$ is large, we solve (8) as the standard form.

3.2 Procedure

We describe a concrete procedure of the proposed method. First, the standard SVM is applied to only labeled objects, and obtain $u_{ip} = (w^p)^\top \phi(x^i) + b_p$, $i = 1, \ldots, m$, $p = 1, \ldots, c$, where ϕ is a feature mapping corresponding to some kernel function. Let G be the kernel matrix induced by ϕ. For each object i, we define $d_i = \mathrm{argmax}_{p=1,\ldots,c}\{u_{ip} + b_p\}$. Consider unlabeled objects which are correctly classified by the SVM, namely, objects i such that $y_i = 0$ and $u_{id_i} + b_{d_i} - \max_{p \neq d_i}\{u_{ip} + b_p\} \geq 2$. The labels of the correctly classified objects are fixed, namely, $y_i = d_i$. We use them as labeled objects.

Next, we determine the matrix E expressing unfixed inner products. For each pair (i, j) of unlabeled objects, we set $e_{ij} = 0$. That is, we fix $s_{ij} = g_{ij}$ for all pairs of unlabeled objects. For the other pairs (i, j), we determine $e_{ij} = 0$ or 1 using a given threshold g_{th}. If $g_{ij} \geq g_{\mathrm{th}}$ then we set $e_{ij} = 1$, otherwise $e_{ij} = 0$. That is, the inner products of pairs with small values (or large distances) are fixed. By solving (8), we obtain u_{ip}, b_p, $i = 1, \ldots, m$, $p = 1, \ldots, c$, and by (9).

4 Numerical Experiments

To evaluate generalization capability of the proposed method, we conduct numerical experiments comparing with S^3VM. We also compare our method with SVM for labeled data sets, i.e., only labeled objects are used to obtain classifiers. We use benchmark data sets: "iris", "wine", "dermatology", "balance" obtained from UCI Machine Learning Repository [5]. We adopt RBF (Radial Basis Function) kernel for the kernel matrix G, which is defined by for $x, y \in \mathcal{X}$,

$$\kappa(x, y) = \exp\left(-\frac{\|x - y\|^2}{\gamma^2}\right), \tag{11}$$

where γ, called a kernel parameter, controls the scale of the input space. We select $\gamma = 1, 2, 5, 10, 20, 50$.

For a data set, $\omega\%$ of objects are randomly selected and they are used for labeled ones, and the rest of objects are used for unlabeled ones. A method is applied to the data set. Using the obtained classifier, we calculate classification error for the unlabeled objects (namely, ratio of the number of misclassified objects to that of total unlabeled objects). This evaluation is conducted 10 times for each data set and each method with each parameter. We select $\omega = 5, 10, 20$.

In proposed method, we choose the threshold as $g_{\mathrm{th}} = 0, 0.5, 0.9, 0.99$. We set $\lambda = 1$ in (8). For the parameters of S^3VM, we set $\mu_l = \infty$ and $\mu_u = 10^3$.

Table 1: Comparison of classification errors of SDPSVM, S^{3}VM, SVM

Label rate	5%			10%			20%		
Method	SDPSVM	S^{3}VM	SVM	SDPSVM	S^{3}VM	SVM	SDPSVM	S^{3}VM	SVM
Iris	**11.8** ± 3.4	**11.8** ± 9.8	13.2 ± 6.6	6.9 ± 3.1	**6.7** ± 3.8	7.9 ± 4.9	4.0 ± 1.6	4.7 ± 2.9	**3.9** ± 1.8
	(2,0)	2	2	(2,0.5)	2	2	(10,0.99)	2	20
Wine	**4.2** ± 1.2	6.9 ± 4.2	8.0 ± 4.2	**3.3** ± 1.1	3.5 ± 1.2	5.4 ± 1.7	**2.4** ± 0.6	**2.4** ± 1.2	3.5 ± 1.3
	(5,0)	2	5	(20,0.5)	2	5	(50,0.9)	2	20
Derm	**4.3** ± 1.6	6.0 ± 2.3	7.5 ± 2.2	**3.3** ± 0.9	4.1 ± 1.6	5.0 ± 1.5	**2.7** ± 0.4	2.8 ± 0.7	3.2 ± 0.9
	(10,0)	10	10	(5,0)	5	10	(10,0)	5	5
Balance	13.8 ± 2.5	14.3 ± 4.0	**13.6** ± 2.1	**9.1** ± 2.3	10.3 ± 1.8	9.5 ± 2.1	**5.8** ± 1.9	7.1 ± 3.5	**5.8** ± 2.0
	(20,0.99)	50	10	(10,0.99)	5	10	(10,0.99)	10	10

We show results of the numerical experiments in Table 1. Each row shows the means and the standard deviations of classification errors of the three methods for the benchmark data set appeared in the leftmost column. From left to right, we show the results in the case of $\omega = 5, 10, 20$. For each value of ω, the corresponding results of the proposed method, S^{3}VMand SVM are ordered from left to right. Since the errors of the three methods are changed by their parameters, we show the best (smallest) results in the table. The set of numbers appeared below each error is the selected parameters for the corresponding setting. From the results of numerical experiments, in most cases, we can see that the proposed method is better than S^{3}VM and SVM.

5 Concluding Remarks

In this study, we proposed a method for transductive learning, which simultaneously optimizes feature vectors and linear discriminant functions by a semidefinite programming. For future work, we extend the proposed method so that it can be applied to inductive semi-supervised learning.

References

[1] S. Abe: *Support Vector Machines for Pattern Classification*, Springer, New York, 2010.

[2] O. Chapelle, V. Sindhwani, S. S. Keerthi: Optimization Techniques for Semi-Supervised Support Vector Machines. *Journal of Machine Learning Research*, Vol. 9, pp. 203–233, 2008.

[3] K. Fujisawa, M. Kojima, K. Nakata: Exploiting Sparsity in Primal-dual Interior-point Methods for Semidefinite Programming. *Mathematical Programming*, Vol. 79, pp. 235-253, 1997.

[4] M. X. Goemans and D. P. Williamson: Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the Association for Computing Machinery*, Vol. 42, No. 6, pp. 1115–1145, 1995.

[5] M. Lichman: UCI Machine Learning Repository [`http://archive.ics.uci.edu/ml`]. Irvine, CA: University of California, School of Information and Computer Science, 2013.

Inconsistency of Fuzzy Grey Analytic Hierarchy Process Pairwise Comparison Matrix

Karel MLS

University of Hradec Králové
Rokitanského 62, 500 03 Hradec Králové
Czech Republic
E-mail: karel.mls@uhk.cz

Abstract

Pairwise comparisons of objects represent a popular method for capturing the decision maker's subjective preferences. For many reasons, these comparisons organised in the pairwise comparison matrix are generally inconsistent. In this paper, several approaches to quantification of the inconsistency of pairwise matrices are discussed. Namely, the original Saaty's Consistency Index and Consistency Ratio based on the maximal eigenvector, and Geometric Consistency Index based on the row geometric mean procedure are mentioned. Moreover, Fuzzy Consistency Index based on fuzzy maximal eigenvector of fuzzy comparison matrix constructed with triangular fuzzy numbers is introduced. Referring to the previous research on decision making in high uncertainty situations with small or incomplete discrete data, grey comparison matrix and Grey Consistency Index are proposed.

Keywords: Fuzzy Grey AHP, Consistency, Pairwise Comparison Matrix.

1 Introduction

The discussion on the importance of absolute consistency of pairwise comparison matrix A has a long tradition. It was mostly focused on how sensitive the preferences weight vector w to particular inconsistencies in A is. While the most authors consider inconsistencies in human estimates of local preferences as mistakes or imperfections, which are to be eliminated [1, 2], there are also voices advocating a positive role of inconsistency in decision making models. According to some authors, intransitive decisions may not be necessarily irrational, and in fact, may even better represent the preferences of decision makers [3, 4]. In general, pairwise comparisons allow inconsistent or intransitive preferences to arise. The well-known and widely used Analytic Hierarchy Process (AHP) [5], based on pairwise comparisons for eliciting preferences, is an example of the methods, that has been deeply criticised for allowing some inconsistencies or even intransitivities to happen.

In this paper, representative examples of existing methods benefiting from pairwise comparisons (Fuzzy and Fuzzy Grey AHP) are firstly mentioned. These were developed in pursuit to improve the traditional AHP approach in processing decision makers' uncertainty, or the situations when there is a lack of data, respectively. Then, the appropriate inconsistency indices are described. Finally the inconsistency measure for the Fuzzy Grey AHP is proposed.

2 Fuzzy Grey Analytic Hierarchy Process

The fundamental scale for pairwise comparisons has been carefully designed to enable the wide range of (nonlinear) evaluations with numerical and verbal intensities and meanings (from 1 - equal importance to 9 - extreme importance) [6]. Nevertheless, it only enables to express the cognitive uncertainty of the decision maker.

The pairwise comparison matrix A for the crisp case can be set as:

$$A = \begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & 1 & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & 1 \end{bmatrix} \tag{1}$$

The matrix A is reciprocal with positive elements a_{ij}, and $a_{ij} = \frac{1}{a_{ji}}$ for all $i, j = 1, 2, ..., n$.

In this case - in so called classical AHP, the formula $Aw = \lambda_{max} w$ has been proposed by Saaty to compute the weights of compared items. There λ_{max} is the maximal eigenvalue of the matrix A and w is the corresponding eigenvector.

In practice, human thinking and reasoning is usually fuzzy and contextually sensitive, considering the possibility of partial membership of the element in a set which is more acceptable than the crisp option. Concepts like Young-Old, Hot-Cold or Rich-Poor are typical representatives of a fuzzy set. The formal fuzzification of the AHP (FAHP) has been introduced by Laarhoven and Pedrycz [7] soon after the original Saaty's method definition in 1981. The usual approach is based on the transformation of real elements in the comparison matrix into the fuzzy numbers or into fuzzy linguistic variables. Then, after specific fuzzy operations, alternatives' rating is obtained.

Using triangular fuzzy numbers $\tilde{a} = (a^L; a^M; a^U)$ as elements of the fuzzy comparison matrix $\widetilde{A}$, where a^L, a^M, and a^U stand for the lower, the middle and the upper values of the fuzzy number, results in the following record [8]:

$$\widetilde{A} = \begin{bmatrix} (1;1;1) & (a_{12}^L; a_{12}^M; a_{12}^U) & \dots & (a_{1n}^L; a_{1n}^M; a_{1n}^U) \\ (a_{21}^L; a_{21}^M; a_{21}^U) & (1;1;1) & \dots & (a_{2n}^L; a_{2n}^M; a_{2n}^U; \\ \vdots & \vdots & \ddots & \vdots \\ (a_{n1}^L; a_{n1}^M; a_{n1}^U) & (a_{n2}^L; a_{n2}^M; a_{n2}^U) & \dots & (1;1;1) \end{bmatrix}, \tag{2}$$

where $\tilde{a}_{ij} = (a_{ij}^L; a_{ij}^M; a_{ij}^U) = (\frac{1}{a_{ji}^U}; \frac{1}{a_{ji}^M}; \frac{1}{a_{ji}^L})$ for all $i, j = 1, 2, ..., n$.

In this case, fuzzified Saaty's formula is $\widetilde{A}\widetilde{W} = \widetilde{\lambda}_{max}\widetilde{W}$. For $\widetilde{\lambda}_{max} = (\lambda_{max}^L; \lambda_{max}^M; \lambda_{max}^U)$ and $\widetilde{W} = (W^L; W^M; W^U)$.

To obtain the lower, the middle, and the upper normalised maximal eigenvectors W^L, W^M, and W^U, the appropriate formulas were defined:

$$W^L : A^L W^L = \lambda_{max}^L W^L, W^L = (w_1^L, ..., w_n^L), \sum_{i=1}^{n} w_i^L = 1, \tag{3}$$

$$W^M : A^M W^M = \lambda_{max}^M W^M, W^M = (w_1^M, ..., w_n^M), \sum_{i=1}^{n} w_i^M = 1, \tag{4}$$

$$W^U : A^U W^U = \lambda_{max}^U W^U, W^U = (w_1^U, ..., w_n^U), \sum_{i=1}^{n} w_i^U = 1. \tag{5}$$

More detailed analysis of the proper computation of the fuzzy eigenvalue as well as the fuzzy eigenvector is discussed in [9].

Specific field of FAHP analysis is dedicated to the consistency of the fuzzy model.

To cover situations with partial information as well, interval or more general grey system extension of the standard AHP method is needed. Fuzzy Grey AHP (FGAHP) was introduced in [10]. Grey System Theory (GST) has been designed for solving high uncertainty problems with small and incomplete discrete data sets [11]. GST includes fuzziness, but while fuzzy mathematics needs some previous information (usually based on cognitive experiences), GST handles with objective data. Actually, it does not require any previous information other than explicit data from measurements or observations.

Grey numbers are characterised by their degree of greyness $g^\circ = \frac{|\overline{g} - \underline{g}|}{|d_{max} - d_{min}|}$, where d_{min}, d_{max} are minimum and maximum values of the domain D of the grey number $\otimes g$, $\underline{g}, \overline{g} \in D$ are the lower and upper limits of the grey number.

There is $g^\circ = 0$ iff $\overline{g} = \underline{g}$ for white number and $g^\circ = 1$ iff $\underline{g} = d_{min}$ and $\overline{g} = d_{max}$ for black number.

Adding some new information, the degree of greyness of a grey number could be reduced. Whereas getting all such information to get a white number is not possible in real situations, whitenisation function to convert a grey number into a white number is introduced [12].

Grey comparison matrix $\otimes A$ is composed of the grey numbers representing expert's imperfect knowledge about pairs of elements $\otimes a_{ij}$ of the hierarchy:

$$\otimes A = \begin{bmatrix} 1 & \otimes a_{12} & \otimes a_{13} & \dots & \otimes a_{1n} \\ \otimes a_{21} & 1 & \otimes a_{23} & \dots & \otimes a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \otimes a_{n1} & \otimes a_{n2} & \otimes a_{n3} & \dots & 1 \end{bmatrix} \tag{6}$$

where $\otimes a_{ij} = \frac{1}{\otimes a_{ji}}$ and $\underline{a}_{ij}, \overline{a}_{ij}$ are the lower and the upper limits of $\otimes a_{ij}$, respectively. Concurrently, with respect to Saaty's fundamental scale, $d_{min} = \frac{1}{9}$ and $d_{max} = 9$.

The local grey priority vector $\otimes w$ was in [10] formulated according to the Row Geometric Mean Method (RGMM) as

$$\otimes w = (\otimes w_1, \otimes w_2, ..., \otimes w_n)^T, \otimes w_i \in \Big[\Big(\prod_{m=1}^{n} \underline{a}_{im}\Big)^{\frac{1}{n}}, \Big(\prod_{m=1}^{n} \overline{a}_{im}\Big)^{\frac{1}{n}}\Big]. \tag{7}$$

3 Inconsistency Measures

The inconsistency of a crisp pairwise comparison matrix A is measured by the Consistency Index CI_n as

$$CI_n = \frac{\lambda_{max} - n}{n-1} \tag{8}$$

where λ_{max} is the principle eigenvalue of A and $CI_n \geq 0$.

If A is an $n \times n$ positive reciprocal matrix, then A is consistent iff $CI_n = 0$. To compensate the dependency of the inconsistency measure CI on the order n of the matrix A, Saaty proposed the Consistency Ratio (CR). This is obtained by taking the ratio between CI_n to its expected value over a large number of positive reciprocal matrices of order n Random Index RI, which entries are chosen randomly [5].

For the prioritisation procedure based on geometric mean - RGMM - the Geometric Consistency Index (GCI) was proposed and studied in [13], with an interpretation analogous to that considered for Saaty's CR.

$$GCI = \frac{2}{(n-1)(n-2)} \sum_{i<j} \log^2 e_{ij} \tag{9}$$

where $e_{ij} = a_{ij}\frac{w_j}{w_i}$ is the error obtained where the ratio $\frac{\omega_i}{\omega_j}$ is approximated by a_{ij}.

There are several more inconsistency measures or indices of pairwise comparisons, introduced by different authors so far. The summary of these measures and the analysis of their properties with respect of the axiomatic system were introduced in [14]. It has been shown, that CI and GCI are the only two measures fulfilling all the proposed axioms.

Consistency Index CI and Consistency Ratio CR were also fuzzified. The formula (8) can be fuzzified by replacing the maximal eigenvalue λ_{max} with the fuzzy maximal eigenvalue $\tilde{\lambda}_{max}$. Then, considering triangular fuzzy numbers, Fuzzy Consistency Index $\widetilde{CI}$ can be defined with the formula

$$\widetilde{CI} = \frac{\tilde{\lambda}_{max} - n}{n-1} = \Big(\frac{\lambda^L_{max} - n}{n-1}; \frac{\lambda^M_{max} - n}{n-1}; \frac{\lambda^U_{max} - n}{n-1}\Big) \tag{10}$$

Center-of-area (COA) defuzzification method was applied on the $\widetilde{CR}$ to get the crisp number. Similarly to CR, fuzzy pairwise comparison matrix is said to be acceptably inconsistent if $COA_{\widetilde{CR}} < 0.1$ [9].

4 Inconsistency in Fuzzy Grey AHP

In this section, a method for obtaining the inconsistency measure in fuzzy grey pairwise matrices is introduced. Firstly, suppose the grey pairwise matrix $\otimes A$ from (6). Then, the grey analogy of the crisp original Saaty's formula is

$$\otimes A \otimes w = \otimes\lambda_{max} \otimes w, \tag{11}$$

where $\otimes\lambda_{max}$ is the grey maximal eigenvalue with the lower limit $\underline{\lambda}_{max}$ and the upper limit $\overline{\lambda}_{max}$

$$\underline{\lambda}_{max} = \min\left\{\max\left\{\lambda; |A-\lambda I| = 0, A = \{a_{ij}\}_{i,j=1}^{n}\right\}; a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}], i,j = 1,...,n\right\}, \tag{12}$$

$$\overline{\lambda}_{max} = \max\left\{\max\left\{\lambda; |A-\lambda I| = 0, A = \{a_{ij}\}_{i,j=1}^{n}\right\}; a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}], i,j = 1,...,n\right\}. \tag{13}$$

Then, considering discrete grey numbers, the Grey Consistency Index $\otimes CI$ is defined with the formula

$$\otimes CI = \frac{\otimes\lambda_{max} - n}{n-1}. \tag{14}$$

The lower and the upper limits of the $\otimes CI$ are computed as

$$\begin{aligned} \underline{CI} &= \frac{\underline{\lambda}_{max} - n}{n-1}, \\ \overline{CI} &= \frac{\overline{\lambda}_{max} - n}{n-1}. \end{aligned} \tag{15}$$

Grey Consistency Ratio $\otimes CR$ is defined as

$$\otimes CR = \frac{\otimes CI}{RI}, \tag{16}$$

with the lower and the upper limits

$$\begin{aligned} \underline{CR} &= \frac{\underline{CI}}{RI} = \frac{\underline{\lambda}_{max} - n}{RI(n-1)}, \\ \overline{CR} &= \frac{\overline{CI}}{RI} = \frac{\overline{\lambda}_{max} - n}{RI(n-1)}. \end{aligned} \tag{17}$$

Finally, applying some whitenisation function F to the $\otimes CR$ leads to the white (crisp) number $\nabla \otimes CR$. Analogously to the classical AHP method, the grey pairwise comparison matrix $\otimes A$ is said to be acceptably consistent if $\nabla \otimes CR < 0.1$.

5 Conclusions and Future Research

In this paper, the method for computing the inconsistency degree of a fuzzy grey pairwise comparison matrix is introduced. The proposed formula is taking into account the reciprocity of pairwise comparisons, the key property of all pairwise comparison matrices in AHP. Grey Consistency Index $\otimes CI$ and Grey Consistency Ratio $\otimes CR$ computed from the grey consistency matrix $\otimes A$ are estimated from the corresponding lower and upper limits.

Future research will be focused on two main domains: first, the sensitivity of $\otimes CR$ to the degree of greyness of the input grey pairwise comparison matrix $\otimes A$ will be analysed in more detail, and second, application of grey numbers and grey consistency indices in group decision, conflict resolution and negotiational situations will be proposed.

Acknowledgement
The support of the Grant Agency of Excellence, University of Hradec Králové, is greatly acknowledged.

References

[1] Tien-Chin Wang and Yueh-Hsiang Chen. Applying fuzzy linguistic preference relations to the improvement of consistency of fuzzy ahp. *Information Sciences*, 178(19):3755–3765, 2008.

[2] Alessio Ishizaka and Markus Lusti. An expert module to improve the consistency of ahp matrices. *International Transactions in Operational Research*, 11(1):97–105, 2004.

[3] Pedro Linares. Are inconsistent decisions better? an experiment with pairwise comparisons. *European Journal of Operational Research*, 193(2):492–498, 2009.

[4] Thomas L Saaty. Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 102(2):251–318, 2008.

[5] Roseanna W Saaty. The analytic hierarchy processwhat it is and how it is used. *Mathematical modelling*, 9(3-5):161–176, 1987.

[6] Patrick T Harker and Luis G Vargas. The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management science*, 33(11):1383–1403, 1987.

[7] PJM Van Laarhoven and Witold Pedrycz. A fuzzy extension of saaty's priority theory. *Fuzzy sets and Systems*, 11(1-3):229–241, 1983.

[8] Jaroslav Ramík and Petr Korviny. Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems*, 161(11):1604–1613, 2010.

[9] Jana Krejčí. Fuzzy eigenvector method for obtaining normalized fuzzy weights from fuzzy pairwise comparison matrices. *Fuzzy Sets and Systems*, 315:26–43, 2017.

[10] Karel Mls and Martin Gavalec. Fuzzy grey AHP. In Talasova, J and Stoklasa, J and Talasek, T, editor, *MATHEMATICAL METHODS IN ECONOMICS (MME 2014)*, pages 673–678, 2014. 32nd International Conference on Mathematical Methods in Economics (MME), Olomouc, CZECH REPUBLIC, SEP 10-12, 2014.

[11] Sifeng Liu and Yi Lin. *Grey information: theory and practical applications*. Springer Science & Business Media, 2006.

[12] Yingjie Yang and Robert John. Grey sets and greyness. *Information Sciences*, 185(1):249–264, 2012.

[13] Juan Aguaron and José Maria Moreno-Jiménez. The geometric consistency index: Approximated thresholds. *European journal of operational research*, 147(1):137–145, 2003.

[14] Matteo Brunelli and Michele Fedrizzi. Axiomatic properties of inconsistency indices for pairwise comparisons. *Journal of the Operational Research Society*, 66(1):1–15, 2015.

Proposal of Probabilistic risk Evaluation for System Development Project Based on Requirements Analysis

Shinji MOCHIDA

University of Marketing and Distribution Sciences/Faculty of Commerce
3-1 Gakuen Nishimachi Nishiku Kobe
Japan
mochida@red.umds.ac.jp

Abstract

The term "business system" refers to the methods which use to manage their information or business resource. Business system is build in order to earn benefit efficiently,or reduce cost in activities. All companies employ business systems during the development of new software, mechanical equipment or other business. In system development projects budget and duration are limited. Thus more efficient project management is needed in order to maintain schedules and reduce cost. But in most business system development project, costs and completion dates are difficult to predict. There are many risks for schedule delay and over cost. Narrowing down requirements to applicable specifications is effective means to shorten schedules and reduce costs.Also evaluating risk properly avoids schedule delay and over cost. This paper aims to evaluate risk of business system development project using requirement analysis with cost share rates[1] and probabilistic risk evaluation. Cost share rate is defined as the percentage of total cost assigned to each requirement. Cost share rate could indicates essentiality of requirements. It is effective to focus on efficient requirements in order to build system efficiently.Additionally this paper propose old Japanese craftsmen ships in order to avoid loss in system development projects. There are plus risks and minus risks against productivity of system development project. Also about benefit, even if the risk managements are planned ,unavoidable troubles remain. Then target profit of project must be over the benefit that is planned,this is old Japanese craftsman's idea.

Keywords: Cost Share Rate, Project Management, Requirements Analysis, β distribution, Bayesian analysis, Expected monetary value,Risk management.

1 Introduction

Meeting budget, finishing on schedule, and maintaining high quality are all important in project management. More efficient project management is needed in order to maintain schedules and reduce cost. In system development project, most large risk is in process that is translating requirements from clients into system specification by developers(see Figure 1). Risks are caused from misunderstanding requirements by system developer. Or requirements do not have enough items are needed by clients. Usually clients can not describe all items in requirements, because clients do not understand real purpose at first place or beginning of project. Thus finding essential and critical requirements is needed in order to reduce risk. Discussing and focusing on essential and critical requirements are needed in order to satisfy clients' demand. Thus this paper propose the method to find essential and critical requirements with cost share rate. Cost share rate is defined as the percentage of total cost assigned to each requirement. Cost share rate could indicates essentiality of requirements. If essential and critical requirements are specified, it is possible to narrow down requirements to applicable specifications.And possibility for success of system development project increase. Also this paper aims to evaluate risk of business system development project using with probabilistic risk evaluation. Because there are plus and minus risks in system development

projects, probabilistic risk evaluation is supposed to have possibility for evaluating risks properly. Although also even if the risk managements are planned ,unavoidable troubles remain. If planned value of system is just set according to requirements from clients, it does not meet value that clients expect because of unpredictable risks.Then to satisfy clients' demand, project value must be over the value that is expected by clients(see Figure 1). This is traditional Japanese craftsman's idea. On the other hand nowadays agile project management is needed in order to avoid uncertainty in projects. Also from agile management, evaluating requirements and risks properly are needed too. At the conclusion, this paper shows the potential to evaluate risk of business system development using cost share rate and probabilistic risk evaluation. Additionally ,because there are plus and minus risk in system development project ,this paper shows possibility to predict risk and benefit properly using Bayesian analysis. Bayesian analysis could visualize invisible plus risk to gain profit.

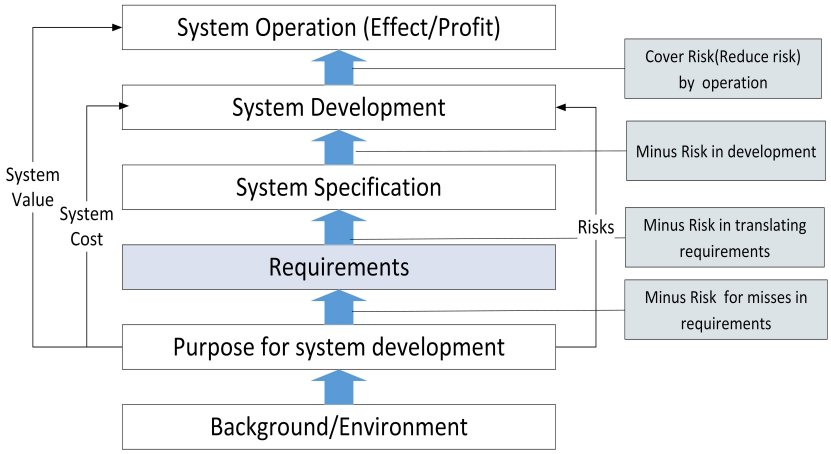


Figure 1: Risks in business system development project with customer and developer

2 Expected Value and Unexpected Risk

2.1 Japanese Craftsmen Ships and Benefit

Even if risk management is planned or scheduled in detail, unavoidable risk or trouble remain. Japanese traditional craftsman manage this issues well. They know relation ship with value, risk and idea (see Equation 1). Traditional Japanese craftsmen ship is consisted of two points. The first point is behavior pattern, and the second point is technical skill. The craftsmens behavior pattern is consisted of four points as follows:
(1) Dont want to do dull works.
(2) They know where benefit comes from.
(3) Customer is the first.
(4) Quality is satisfaction of the customer.
In the first item, dull work is work that doesnt need ideal or new technical skill. Dull work has lots of risks that does not satisfy clients. The situation that the second point means is as shown in Figure 2. Craftsman says I will show you what you have never seen. These behavior pattern gives craftsmen them identity and pride. From interview with the master of carpenter in Japan, master tells that staffs need to image of final construction and total costs. In old Japanese team, all staffs are allowed to work beyond their order. Thus in traditional Japanese work way staffs arent given perfect orders or detailed instruction. They know idea that clients do not expect give them profit. Also they know revolutionary idea cancels out unexpected or invisible risks.

$$Profit = Value - (Cost + PredictableRisk + UnpredictableRisk) + Idea(Value) \quad (1)$$

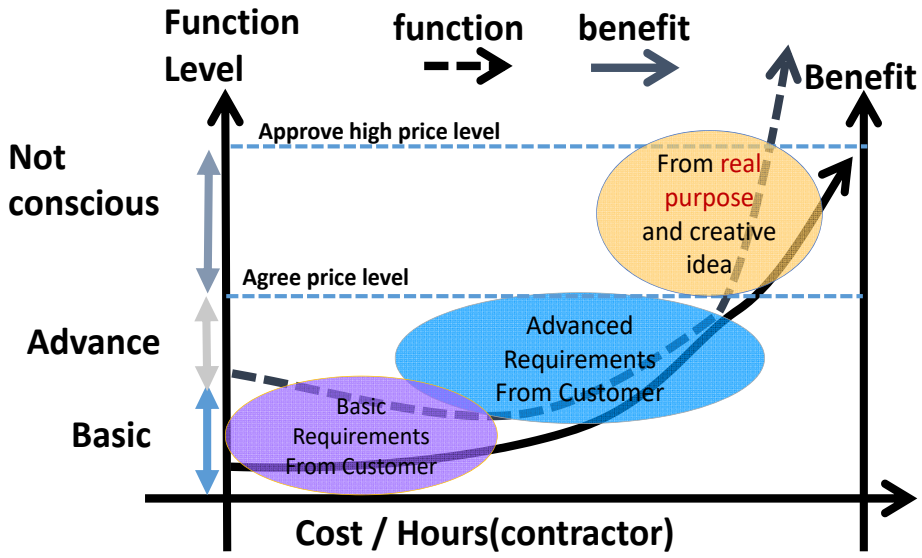


Figure 2: Japanese craftsmenship and benefit

2.2 Risk Management in Traditional Japanese Way

Japanese craftsmen know creative and original products give them benefit and identity. They need extra time to create idea. Thus, increasing productivity and saving time will be possible to make extra time in order to create original idea. At the results, developing productivity makes time to create idea, and provides benefit and identity to craftsmen. In traditional Japanese craftsmen ship, there two important points that lead project to success, one is idea that cover defects, another is increasing productivity,this is Kaizen. Finding a good transition way and exceed difficulties is essential point in traditional Japanese work way. Transition way is series of tasks of work. Finding efficient way is essential point to finish project under budgets and on schedule. In typical work way in US expected troubles are listed at first by risk management, and measures are settled for each trouble under risk management. On the other hand, in Japan team develops sub plans with their experience in order to exceed troubles when they meet troubles. Previously settled measures under risk management need much cost. But sub plan according to the situation at that time need less cost that previously settled measures need. Also it could be called Japanese concurrent engineering, because staffs think as they as can in their position parallel. Figure 3 shows the difference of Japanese way and US way to execute projects. In US way staffs are required to make actual products under the plan exactly. In Japanese way staffs aim to develop ideal products with Kaizen activity. Kaizen activity aims to realize ideal goal. Thus there is no limit or goal for spiral curves of Kaizen activity. Finding good transition way is needed in order to avoid schedule delay and over cost. In Japanese way staffs could choice transition way flexibly and effectively according to actual working situations or their skills. On the other hand, in US way tasks will be executed according to a plan that is planned previously under risk management. Good pint of US way is sourcing the cheapest manpower or parts from all over the world. But good point of Japanese way is finding better transition way flexibly and effectively according to actual working situations, and prioritization of tasks would be left to staffs. Thus in Japanese way, they could exceed troubles with small cost using their experiences. Finding good transition way is based on

Japanese Kaizen activity. Japanese Kaizen is based on Japanese old craftsmen ships. Japanese old craftsmen ships aim to develop products better than customer imagine. Japanese Kaizen does not only aim to reduce costs but also aim to develop ideal products. Japanese Kaizen aims to improve products that customer have not ever seen. Essential sprit of Japanese way is over specification from customer. Traditional Japanese craftsmen know that even if they avoid risk ,risks remain. Thus they know creative idea is needed to cover defects. Also experience is needed to exceed risks.

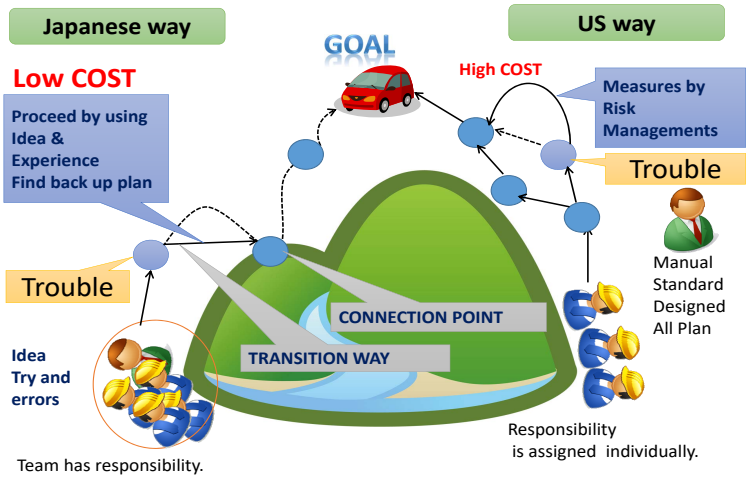


Figure 3: Good way for Troubleshooting

3 Project management and Risk management

3.1 Predicting All Risks

Business System development projects are challenging in that are many requirements demanded from customers even while these requirements are proposed with the same priority. Thus it is important to narrow down and prioritize requirements according to their essentiality and criticality to finish on schedule. Although system developers estimate according to the complexity of projects [2], but customers expect the cost to be based on the number of requirements they propose. Thus, this paper proposes cost share rate for business system development projects based on requirements analysis in order to estimate accurately. Cost share rate is defined as the percentage of total cost assigned to each requirement. Cost share rate could distinguish essential requirements. And requirements that has large cost share rate must have **large risk**, also should be under strict control. Because big change or modification for essential requirement give large impact to costs or schedule[3]. There are plus and minus risk in system development project ,Additionally, there predictable risks and unpredictable or invisible risks. To recognize plus risk and unpredictable risks are point in order to predict all risks in system development project.

3.2 Project Process and Risk

The processes that have high risk is planning and design process in project(see Table 1). In Table 1 survey, Preparation and design processes have high risk for schedule delay. On the other hand programming, test, document process have less risk except creating DB process. Because creating DB process must follow DB design process, creating DB process is delayed. Also in beginning of each task there are high risk. We collected the productivity data from the screen design work in a past

system development project, and apply to the Weibull distribution(see Equation 2)[4]. Thus the parameters are acquired by fitting productivity data into Weibull distribution. Productivity data consist of productivity according to time. Thus parameters(see Table **??**) and curve (see Figure 4) are acquired. Productivity at a certain time is obtained by f(t) as shown in Figure 4. Figure 4 shows that area of g(t):actual work days(colored blue area) is just match area of f(t):planned work days(area enclosed with square). It shows there is high risk at beginning of each task. Additionally in Figure 4 area S(area enclosed with square and Weibull curve) indicates probability distribution density of schedule delay, and gravity center of distribution density. In this possibility of schedule delay is 0.18(x=0.18). On the other hand in beginning project prior probability of schedule delay is considered 0.1, this is from past experience(see Table 4).Table 4 shows there are 10% for refinement(contingency). It is considered as prior probability for schedule delay.Thus these parameters as follows apply to Equation 3.
P(A)=0.1
P(B | A)=0.18
$P(B) = \frac{5}{16}$

$\therefore P(B \mid A) = \frac{0.18 \times 0.1}{\frac{5}{16}} = 0.0.6$

At the result it became the result of 0.6. it is less than that is in begging because of actual delay days. It is clear posterior probability is falling down, posterior probability is falling down too. Reducing uncertainty ai beginning of each task is needed to reduce risk for schedule delay of projects.

Table 1: Sample data from past project

Item	Pre	Design			Programing			Test	Doc	Sum
	*1	Dialog	Function	DB	Dialog	Function	DB		*2	
Start date:x	0	0.52	0.6	0.59	0.7	0.74	0.59	0.947	0.931	
Schejule day	11	29	27	27	9	58	39	3	13	216
Actual day	35	34	36	27	9	55	62	2	7	267
Delay day	24	5	9	0	0	-3	23	-1	-6	51

*1 Survey and Preparation *2 Document

$$f(t) = K\left(1 - exp(-(\frac{t+a}{\eta})^m)\right) + d \tag{2}$$

K : final productivity, a : preparation days, η : study resistance, m : efficiency, d : work repetition

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

4 Prediction of Cost Share Rate

Taking correct requirements is essential to estimate properly.And over cost or schedule delay is caused by missing evaluation of requirements. Otherwise over cost or schedule delay is caused by many remediation of requirements. Remediation for essential requirement and uncertain requirement has large risk.thus It is vital to distinguish the requirement which gives large impacts to specification or budget of project.This research propose the method to distinguish influential requirements that has large risks.This paper shows a method to distinguish an influential requirements with with linguistic analysis and cost share rate. At first overlapping keywords are extracted form each requirements with linguistic analysis.Overlapping keywords are words that appear in one requirement and groups of keywords that appear in each category in system development project(see Figure 2). Cost share rate is defined as the percentage of total cost assigned to

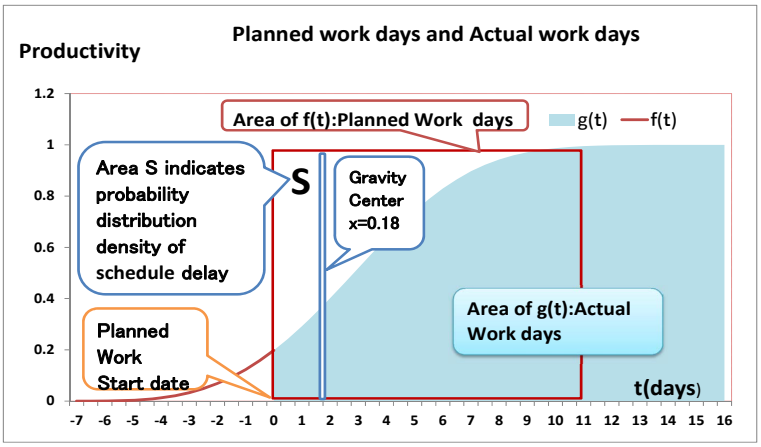


Figure 4: Planned work days and Actual work days

each requirement. This paper analyzes the requirements that were requested in past small system development projects by linguistic analysis. This small project is building a knowledge collecting system. In this project requirements were revised four times. Thus this paper analyzes requirement version one and version four. Cost share rate indicates importance of each requirement.Cost share rate is defined as the percentage of total cost assigned to each requirement(see Figure 5, Figure 6).

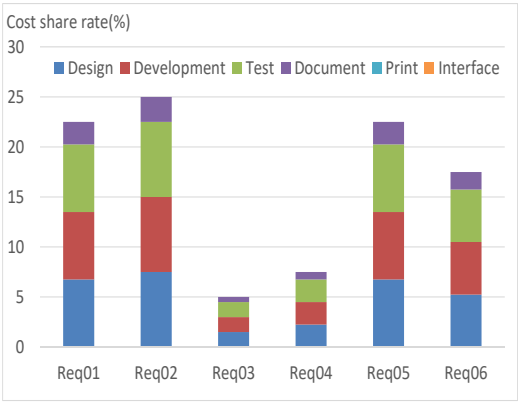


Figure 5: Cost share rate from requirement version one

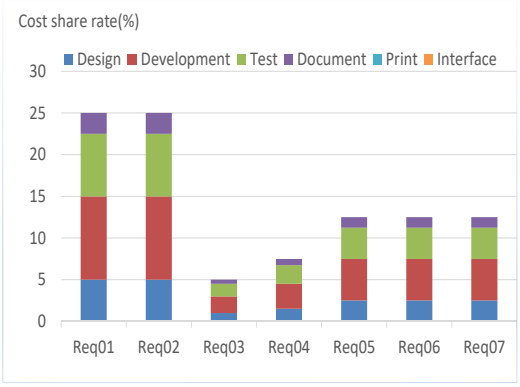


Figure 6: Cost share rate from requirement version four

5 Probability Prediction of schedule delay

This paper consider one of risk of system development project is schedule delay. Thus this research suppose probability of schedule delay follows β distribution(see Equation 4),Then parameters (see Table 3) are gained by curve fitting sample data into β distribution. Sample data in Table 1 are

Table 2: Overlap keywords from requirement version one and Relation with typical category

	Req01	Req02	Req03	Req04	Req05	Req06
Req01	0	3	3	2	2	2
Req02	3	0	4	3	3	2
Req03	3	4	0	7	3	3
Req04	2	3	7	0	2	2
Req05	2	3	3	2	0	3
Req06	2	2	3	2	3	0
Design	1	3	4	3	2	0
Development	3	2	5	4	3	4
Print	0	0	0	0	0	0
Check	0	0	0	0	0	0
Interface	0	0	0	0	0	0
Document	0	0	0	0	0	0
The number of over 4 relationship with requirement of each category are colored						

surveyed in past system development project; Rdf system for tool tracking in machine factory. In this analyzing process x is probability parameter that indicates start day for each task, and y is ratio for schedule delay against actual days. Table 1 shows that survey/preparation,design and programming process have risk of schedule delay. Although test and writing document process have no risk of schedule delay. Figure 7 shows β distribution curve in this case from parameters(see Table 3).

$$f(x) = c \times x^{\alpha\ -1}(1-x)^{\beta\ -1} \tag{4}$$

Table 3: Acquire Parameter for β distribution

C	α	β
0.839835	1.020625	3.047617

6 Risk Prediction for business system development project

Figure 7 shows that probability of schedule delay is 0.225. Thus this research assign probability of risk(0.225) to essential requirements that distinguished by requirements analysis.Actually Table 5 shows that requirement3 and requirement4 are about design and development, requirement6 is about development.These requirements are essential, and assigned probability of risk as 0.225. Requirement1 is about Main Design,and assigned probability of risk as 0.1, because in past three project contingency is set as10% (see Table 4). Contingency is set for refinement(deleting,adding or refactoring).Table 4 shows occupying cost rate for each work process from surveying of past system development project, that are medical record system,knowledge management system.Thus risk is calculated probability of risk $\times$ cost(see equation 5), distribution of cost is calculated by cost share rate. In estimated cost of requirement version one, estimated total cost is 109(see Table 5).And at completion actual total cost is 109.1(see Table 6). At the result these results are just match.

$$Risk(Expected\ Monetary\ Cost) = Probability \times Cost \tag{5}$$

Table 4: Cost Occupying Rate Data from Past Project

Project	Design	Development	Testing	Document	Refinement	Total
P1	3	4	1	1	1	10
P2	1	6	1	1	1	10
P3	1	6	1	1	1	10
Average	1.7	5.3	1	1	1	10

Table 5: Estimated budget from requirement version one with risk analysis

Requirement Version one	Attribute	Cost Share Rate Ver one(%)	Schdule delay Probability	Conditional Probability	Monitary Risk (Contengency)	Estimated Cost Ver one
Requirement1	M	22.5	0	0.1	2.25	24.75
Requirement2		25	0	0	0	25
Requirement3	D/De	5	0.225	0	1.125	6.125
Requirement4	D/De	7.5	0.225	0	1.6875	9.1875
Requirement5		22.5	0	0	0	22.5
Requirement6	De	17.5	0.225	0	3.9375	21.4375
Toatal		100			9	109

M:Main, D:design, De:Development

Table 6: Total cost at completion

Requirement Version four	Attribute	Cost Share Rate Ver four(%)	Increased cost Rate(%)	Total Cost At Completion
Requirement1	M	25	20	27
Requirement2		25	10	27.5
Requirement3	D/De	5	0	5
Requirement4	D/De	7.5	5	7.875
Requirement5		12.5	0	22.5
Requirement6	De	12.5	10	19.25
Requirement7		12.5		0
Toatal		100		109.125

M:Main, D:design, De:Development

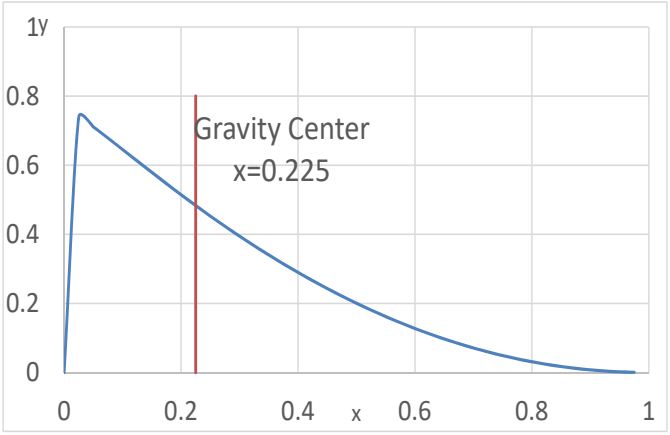


Figure 7: Beta distribution from past project

7 Conclusion

Meeting budget, finishing on schedule are all important in project management. However it is very difficult to meet budget and finish on fine in system development project. There are gaps between estimates by system developers and the estimates that customers are expecting, because there is a difference between system developers and customers in grouping the costs to estimate. Estimate miss by developer causes finally over cost and schedule delay. Also lack in requirement from clients cause over cost and schedule delay. As this paper shows, usually estimate is measured by amount of source code of system or complexity of system. It is not estimated based on essentiality or risk of requirements. Misunderstanding of requirement causes misses in estimate or schedule.Then, this paper proposes a method to find essential requirements in order to estimate accurately.Additionally, distinguishing requirements that have high risk is vital in order to finish projects fine. Usually on risk management risks are evaluated according to staffs' experience, not based on requirements analysis. Thus this paper show better results by using cost share rate from requirement analysis. Cost share rate could indicates essentiality of requirements. If essential or critical requirements could be distinguished with cost share rate, We could deal with essential requirements properly, and we could avoid risks. Additionally, this paper shows there high risk in beginning of project according to β distribution(see Figure 7). Also there are high risks in beginning of each tasks(see Figure 4), because there is low productivity in beginning of each tasks. However, this result was obtained by small case. Thus further research and study is needed to refine and improve this method to obtain cost share rate and risk more accurately.

8 Discussion

Customers and developers estimate costs differently, resulting in differing expectations for project cost. If estimates could be proposed according to the cost of customers requirements, there would be improved understanding between system developers and customers. Because in business system development projects there are many ways to implement requirements, there are large variability in translating user's requirement into system specification. Estimate differs greatly according to staff's skill. And it is one reason why estimate does not meet final cost or budget. Thus proper estimate by requirement analysis is needed in order to finish business system development project fine. Final estimate is build by adding contingency to base estimate. Also correct probability for risk is needed to build proper estimate. Requirements have invisible risks. There are plus risk

and minus risk. Minus risk gets prospect of profit worse, but plus risk gets prospect of profit well. But plus risk is not visible, plus risk is only in mind of staffs individually.This is one reason why experimental figure:10% of contingency budget is just match at completion actual extra cost. Actually estimated cost:109 match total cost:109(see Table 5,Table 6).Therefore there is potential to predict risk accurately using conditional probability or Bayesian analysis [6][7][8] in order to visualize plus risk.

Acknowledgement
This work was supported by JSPS KAKENHI Grant Number 17K00354.
The author would like to thank INTERLINK Inc. and TECHNO SOLUTION Inc. for data and advices.

References

[1] Shinji Mochida,James Righter, Joshua D. Summers: *A Study of Project Cost Management Based on Requirements Analysis, International Journal of BMFSA Vol. 21, No. 1* ,jp, 2016. 21-27

[2] Parviz F. Rad, Kou Itou, japanese translation: *PROJECT ESTIMATING AND COST MANAGEMENT* ,Seisansei Syutupan,jp, 2004

[3] Beshoy Morkos,James Mathieson,Joshua D. Summers: *Comparative analysis of requirements change pre-diction models: manual, linguistic, and neural net-work, Research in Engineering Design Vol. 25* ,Springer,US, 2014.139-156

[4] Shinji Mochida: *A Study on Method of Measuring Performance for Project Management, 20th ISPE International Conference on Concurrent Engineering* ,IOS Press Ebooks,US, 2013. 264-273

[5] Project Management Institute: *A Guide to the Project Management Body of Knowledge Fourth Ed. (JAPANESE TRANSLATION*,Project Management Institute,US, 2008

[6] Christian Robert : *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* ,Springer Texts in Statistics,US, 2007

[7] Paul G Hoel,Akira Asai,Masayasu Murakami japanese translation : *ELEMENTARY STATISTICS* ,Springer Texts in Statistics,jp, 2011

[8] David Blackwell,Kazuyoshi Mikuni japanese translation : *BASIC STATISTICS* ,Kogakusha,jp, 1972

Scheduling Problem on Operation Rooms

Shintaro MOHRI[1] and Hiroaki ISHII[2]

[1] *Faculty of Economics, KOBE GAKUIN University*
518 Arise, Ikawadani-cho, Nishi-ku, Kobe 651-2180
JAPAN
mohri@eb.kobegakuin.ac.jp

[2] *School of Science and Technology, KWANSEI GAKUIN University*
2-1 Gakuen, Sanda, Hyogo 669-1337
JAPAN
ishiihiroaki@kwansei.ac.jp

Abstract

Efficiency improvement in medical services is a important problem in recent years, and many researches are being considered. In this paper, the scheduling problem at medical service is considered that there is a due date to be operated each operation, and there is a constraint that a medical team has been determined. The due date of operation depends on the illness or patient's condition. For describing these conditions a fuzzy scheduling problem that describe as the degree of satisfying due date constraints is formulated. And a two-step procedure to solve this problem is introduced. First step, the integer programming problem is formulated to decide the assignment the jobs to each day. Next Parallel machine scheduling problem is considered as bin packing problem to decide the schedule for each day.

Keywords: operation room, scheduling, bin packing problem, Integer programming

1 Introduction

The importance of medical services is increasing due to decreasing birthrate and aging population in many countries including Japan in recent years. Also the efficient management of medical services has become necessary by increasing in social security costs and sophistication of medical care. Surgeries in the hospital are a large part of the profit and medical services. operative date scheduling was summarized by Cardon et al[1]. They classify problems by patient characteristics, performance measures, decision delineation and research methodology. Vijayakumar et al. [4] proposed the dual bin-packing approach to operation cases. They applied an integer programming and heuristic algorithm to their problem. Also the approximation algorithm for a fuzzy scheduling problem in the operation room with due date is considered by the authors[3]. In this paper, the scheduling of the operation for efficient management of the operating room and the medical team is considered and the solution procedure based on integer programming and parallel machine problem is developed.

2 Problem description

There are various constraints in accordance with the symptoms for each surgery. In particular, there are the cases that due date in surgery is very strict, and the lateness is not permitted. The other hand, the due date is not so strict in mild case. This constraint is expressed as fuzzy variable. Also the medical team that treats the operation has been determined. There is the constraint that one medical team can operate only one operation at a time. This constraint is satisfied by the algorithm that determine the schedule for each day.

2.1 Notation

In this section, the notation for the scheduling problem is defined. There is the surgery J^{R_l} with medical team R_l to operate. The operating room is M_1, M_2 i.e. $m = 2$. Due date is defined for each operation, and the urgency for exceeding due date is different depend on each operation. The solution of the problem is the schedule that the date of these operations are determined and determine the actual schedule of each operation on each day. The notation is defined as follows.

- Let $J_i^{R_l}$ $(i = 1, 2, \cdots, i_{\max})$ be the Job. This operation is operated by medical team R_lD
- Let $R_l \in R$ be the medical team. The medical team R_l cannot operate two operations simultaneously. The total operating time of the medical team is limited for each day.
- Let d_i $(i = 1, 2, \cdots, i_{\max})$ be the due date of the operation $J_i^{R_l}$. It is desirable that $J_i^{R_l}$ is operated by this due date.
- Let T_j $(j = 1, 2, \cdots, k_{\max})$ be the time slot, and $|T_j|$ be the length of time slot. It is the days in the schedule term.

The purpose of this problem is to determine the schedule of operations $J_i^{R_l}$ under satisfy constraints. The solution procedure of this problem is following two steps. The First step is considering the integer programming problem and assign dates of each operation. The next step is considering the parallel machine scheduling problem for each day.

2.1.1 Satisfaction degree for due date

Satisfaction degree of due date is given as a function of real number taking a value between 0 to 1 with the operation day as a parameter. Since each date is a discrete value, it is enough to decide the value for each day. Satisfaction degree with due date depends on a disease or a condition of a patient. Usually the satisfaction degree is 1 until the due date and it is non-incrementing value after the due date. The following examples are conceivable.

When the operative date of J_i is allowed to be delayed from d_i to $d_i + L$ days. For operation day T_d, the satisfaction degree $\mu_{J_i}(T_d)$ is defined as

$$\mu_{J_i}(T_d) = \begin{cases} 1 & (T_d \leq d_i) \\ \frac{(d_i+L)-T_d}{L} = 1 + \frac{d_i - T_d}{L} & (d_i < T_d \leq d_i + L) \\ 0 & (T_d > d_i + 5) \end{cases} \tag{1}$$

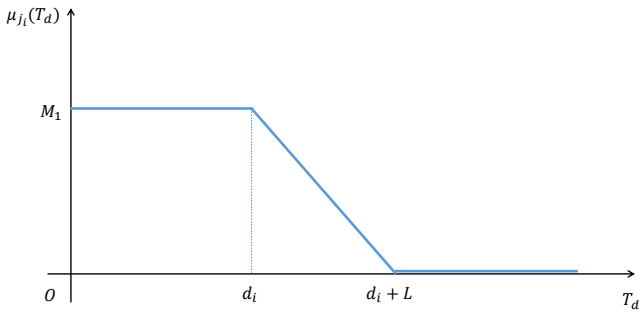


Figure 1: Definition of Satisfaction degree for due date

Example

When the operative date of J_5 is allowed to be delayed from $d_5 = 3$ to $3 + 4(L = 4)$ days. For operation day $T_d = 0, 1, 2, \cdots$, the satisfaction degree $\mu_{J_5}(T_d)$ is defined as

$$\mu(1), \mu(2), \mu(3), \mu(4), \mu(5), \mu(6), \mu(7), \mu(8), \cdots\} = \{1, 1, 1, 0.75, 0.5, 0.25, 0, 0, \cdots\} \tag{2}$$

In the case of severe surgery in which delay in treatment is life-threatening, satisfaction degree when delayed from the due date may be 0. That is

$$\mu_{J_i}(T_d) = \begin{cases} 1 & (T_d \leq d_i) \\ 0 & (T_d > d_i) \end{cases} \tag{3}$$

It can be defined individually depend on each case for operation.

2.2 Integer programming problem

In this section, the problem of assigning each operation to each schedule is considered by formulating an integer programming problem that maximizes the minimum value of the satisfaction degree that due date lateness of surgery operative date. To solve the problem the parameter for resource constraints to the formulation of integer programming problems is introduced .

$$R_{il} = \begin{cases} 1 & \text{If } R_l \text{ is required for processing job } J_i \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In this problem, the element which $R_{il} = 1$ is unique.

Example

There is the job $J_5^{R_3}$ which require the medical team R_3 for processing. The parameters of R_{5l} are defined as:

$$\{R_{51}, R_{52}, R_{53}, R_{54}, R_{55}, R_{55}, \cdots\} = \{0, 0, 1, 0, 0, 0, \cdots\} \tag{5}$$

Let L_{T_j} be the limit of total working time for day T_j, L_{R_l} be the limit of total working time for medical team R_l. Let L_μ be the minimum acceptable level of satisfaction degree.

Let the decision variable x_{ijk} be allocated the work J_i assigned to the schedule k, and maximize the minimum value of satisfaction that can accept the objective function.

The integer programming problem is formulated by using this parameter.

$$\text{Maximizing } \ L_\mu \tag{6}$$

$$x_{ijk} = \begin{cases} 1 & \text{Job } J_i \text{ is proceesed on } M_k \text{ at time interval } T_j \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } i, j, k, \tag{7}$$

$$\sum_{All j,k} x_{ijk} = 1 \text{ for all } i, \tag{8}$$

$$\sum_{All i,k} p_i \cdot x_{ijk} \leq 2L_{T_j} \text{ for all } j, \tag{9}$$

$$\sum_{All i,k} R_{il} \cdot p_i \cdot x_{ijk} \leq L_{R_l} \text{ for all } j, \tag{10}$$

$$\sum_{All i,k} R_{il} \cdot p_i \cdot x_{ijk} \leq L_{T_j} \text{ for all } j, \tag{11}$$

$$\sum_{All i,k} \mu_{ij} x_{ijk} > L_\mu \text{ for all } j, \tag{12}$$

The equation (7) is the intger constraint. The equation (8) is ensuring that all jobs are certainly processed just once. The inequality (9) and (10) are ensuring that total jobs can be processed for

each day. The inequality (11) is ensuring that total jobs can be processed for each day for each medical team. This ensure the feasibility of the schedule at the next step. The inequality (12) is the satisfaction degree constraint.

2.3 Bin packing problem

The jobs are assigned to each date by the previous formulation. Next step, the procedure is required that construct the actual schedule as Uniform Parallel Machine Scheduling Problem with resource constraints. From the result of the previous integer programming problem, it is guaranteed that the processing time for a each medical team on a day is shorter than the term on the day. Therefore, in order to solve the conflict of allocation of medical team, the jobs of the medical team is united as one work. This problem is $P2||C_{\max}$, which is a partition problem which divides the jobs into two. Here, since the partition problem is NP-Hard, approximate solution method is proposed. $P2||C_{\max}$ problem is transformed to a modified bin-packing problem. In the classical bin-packing problem is given a list of items with various sizes and bins with fixed capacity. The problem is to pack all items into a minimum number of bins without over the capacity. In this study, the problem is considered as that the capacity of bins is variable and all capacities are same. The problem is a number of bins is fixed, and minimizing capacity of bins. Our problem in the previous section is following. Jobs are supposed as items. Processing time is as the Processing jobs in an operating room m on the day l is equivalent to that packing items into each bin. Capacity is equal to the sum of the processing time to the work of the day if number of bins are fixed. The bin packing problem is NP-complete, exact algorithm does not exist. However the approximate solution procedure of the above problems is given by Ishii [2].

MFD

FFD(First Fit Decreasing) is an approximate algorithm for the classical bin packing problem. All items put into the bins in order as follows. The largest item which is not in any bins is put into the bin which is the smallest rest capacity. The algorithm MFD which solve our modified bin-packing problem is based on FFD. To apply FFD repeatedly with setting properly the upper bounds and lower bounds of the capacity. Seeking the smallest capacity by the Bisection method that FFD can put all items into the fixed number of bottles.

2.4 Heuristic Algorithm for improvement the solution

Assignment of the operations for each day by integer programming is supposed that the operations by a medical team is unity and preemption is allowed. However the schedule that obtained by reducing parallel machine problem have differ from this assumption.

2.4.1 Improvement for each day

If there is the day with an overtime, one operation room has an overtime and the other operation room have a margin, because the solution of ¶$2|pmtn|C_{\max}$ satisfy the constraints of business hour. If let t^o be overtime, the margin is greater than t^o. Let M_1 be the machine which have overtime and M_2 be the machine which have a margin. If there is the jobs group on M_1 which can be divided two parts such as it can improve the total processing time of the day by assigning them to both machines, assign one part to first on M_1 and assign the other part to last on M_2 respectively. The feasibility of the processing is satisfied.

2.4.2 Improvement through the days interval

If there is no date which can be improved the completion time by adjusting the jobs in the day, the solution can be improved by exchanging with the jobs assigned to another date.

3 Constraint of job release time

Here the constraint for the release time of operation is considered. For example, depending on the case of a patient, there the case that preparations or examination requires a certain amount of

time and it is impossible to start surgery immediately. The satisfaction degree μ^r of a release date is defined as maximizing the minimum value of this satisfaction degree.

The satisfaction degree of the start time is also defined similarly to the satisfaction degree of the due date. Satisfaction degree is given as a function of real number taking a value from 0 to 1 with the operation day of surgery as an argument. Since each schedule is a discrete value, you can decide the value for each schedule. Normally, the degree of satisfaction before release day becomes 0, and for after day, the satisfaction degree is set to 1. If the schedule can be adjusted to some extent, the satisfaction degree can be set to the non-decreasing value of $[0, 1]$.

Let J_i has the release date d_i^r and allowed to the advancement of a schedule for span A_i i.e. The operative date of J_i is allowed to the advancement of a schedule between d_i^r to A_i days. For operation day T_d, the satisfaction degree $\mu_{J_i}^r(T_d)$ is defined as

$$\mu_{J_i}^r(T_d) = \begin{cases} 0 & (T_d \leq d_i^r - A) \\ 1 + \frac{T_d - d_i^r}{A} & (d_i^r - A < T_d \leq d_i^r) \\ 1 & (d_i^r) \end{cases} \tag{13}$$

Example

When the operative date of J_5 is allowed to be start from $3(d_i^r - A = 3)$ to $7(d_i^r = 7)$ days, i.e. $d_i^r = 7$ and $A = 3$ For operation day $T_d = 0, 1, 2, \cdots$, the satisfaction degree $\mu_{J_5}^r(T_d)$ is defined as

$$\mu^r(1), \mu^r(2), \mu^r(3), \mu^r(4), \mu^r(5), \mu^r(6), \mu^r(7), \mu^r(8), \cdots\} = \{0, 0, 0, 0.25, 0.5, 0.75, 1, 1, \cdots\} \tag{14}$$

By introducing this value, the preceding integer programming problem is as follows.

$$\text{Maximizing } \mu^{\max} \tag{15}$$

$$x_{ijk} = \begin{cases} 1 & \text{Job } J_i \text{ is processed on } M_k \text{ at time interval } T_j \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

$$\sum_{All i,j,k} x_{ijk} = 1 \text{ for all } i, \tag{17}$$

$$\sum_{All\{j|T_j > d_i\}} x_{ijk} = 0 \text{ for all } j, \tag{18}$$

$$\sum_{T_j|T_j > d_i} p_i x_{ijk} \leq L_G \text{ for all } j, \tag{19}$$

$$\sum_{T_j, M_k} p_i x_{ijk} \leq T \text{ for all } j, \tag{20}$$

$$\mu_i j \cdot x_{ijk} \geq \mu^{\max} \text{ for all } i, \tag{21}$$

$$\mu_i^s j \cdot x_{ijk} \geq \mu^{\max} \text{ for all } i, \tag{22}$$

The inequality (22) is the satisfaction degree for release date constraint.

4 Conclusion

In this paper, the scheduling problem for operating rooms with due date constraint and medical team constraint is considered. The problem is formulated as a fuzzy scheduling problem, and develop the approximate algorithms based on a integer programming and bin-packing problem. In real situation, there are many types operation cases and many types of resource constraints. Especially to formulate the variety of resource constraint such as doctor, nurse, equipment, etc. is our further works .

References

[1] B. Cardon, E. demeulemeester, J. Beliën. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201:921–932, 2010.

[2] H. Ishii. Approximate algorithms for scheduling problem *Communications of the Operations Research Society of Japan*, 31:26–35, 1986, in Japanese.

[3] S. Mohei, H Ishii, Fuzzy Scheduling Problem for Surgical Operating Rooms In *Proc. of the ICBI2014*, Honolulu, Hawai, U.S.A., 2014.

[4] B. Vijayakumar et al. A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital. *European Journal of Operational Research*, 224, 583–591, 2013.

Linguistic Characterization of Natural Data by Applying Intermediate Quantifiers on Fuzzy Association Rules

Petra Murinová and Michal Burda

University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
NSC IT4 Innovations
30. dubna 22, 701 03 Ostrava 1, Czech Republic
petra.murinova@osu.cz, michal.burda@osu.cz

Abstract

The objective of this paper is to apply fuzzy natural logic together with the Fuzzy GUHA method for analysis and linguistic characterization of scientific data. Fuzzy GUHA is a tool for extracting linguistic association rules from data. Obtained associations are IF-THEN rules composed of evaluative linguistic expressions, which allow the quantities to be characterized with vague linguistic terms such as "very small", "big", "medium" etc. Originally, fuzzy GUHA provides several numerical indices of rule quality, which may not be easily understandable for domain experts that are not familiar with GUHA association rules. Therefore, we show in this paper that the theory of intermediate quantifiers (a constituent of fuzzy natural logic) can be applied to the results in an automatic manner in order to obtain natural linguistic summarization. We also present an idea of how the theory of generalized Aristotles's syllogisms can be used for a detailed data analysis.
Keywords: Fuzzy natural logic, Linguistic associations mining, Intermediate quantifiers, Generalized syllogisms, Fuzzy GUHA.

1 Introduction

The main goal of this paper is to put together theoretical results on intermediate quantifiers which were proposed in several papers (see e.g. [12, 14, 19, 17]) with the Fuzzy GUHA method [9], and to introduce a linguistic characterization of natural data using generalized intermediate quantifiers. The theory of intermediate quantifiers was introduced by Novák in [19] and now is a constituent of the theory of *Fuzzy Natural Logic* (FNL), which is a mathematical counterpart of the concept of *Natural Logic* introduced by Lakoff [11]. This theory is based on Łukasiewicz fuzzy type theory (Ł-FTT) [17], which is one of the existing higher-order fuzzy logics.

Fuzzy GUHA is a special method for automated search of association rules from numerical data. Generally, obtained associations are in the form $A \sim B$, which means that the occurrence of A is associated with the occurrence of B, where A and B are formulae created from objects' attributes. As proposed by Hájek et al. [9], the original GUHA method allowed only boolean attributes to be involved. Some parts of their approach was independently re-invented by Agrawal [1] many years later and is also known as the *mining of association rules* or *market basket analysis*. A detailed book on the GUHA method is [10], where one can find distinct statistically approved associations between attributes of given objects. Fuzzy GUHA is an extension of a classical GUHA method for fuzzy data. In this paper, we work with associations in the form of IF-THEN rules composed of evaluative linguistic expressions, which allow the quantities to be characterized with vague linguistic terms such as "very small", "big", "medium" etc.

To measure the interestingness of a rule, many numerical characteristics or indices have been proposed (see [29, 7] for a nice overview). As a supplement to them, we try to utilize the theory of intermediate quantifiers to characterize the intensity of association, which allows us to use linguistic characterizations such as "almost all", "most", "some", or "few". As a result, we may automatically obtain the following sentences from numerical bio-statistical data:

- *Almost all people, who suffer atopic tetter, live in an area affected by heavy industry and smoke, suffer from asthma.*
- *Most people who smoke and suffer from respiratory diseases also suffer from ischemic disease of leg.*

In the practice, it is often the case that some data are not available e.g. due the error in measures, missing results, or if the respondent is not willing to answer or has no opinion on the given subject. We can completely remove the cases with missing values to obtain clean data, but it can result in an excessive loss of information. Alternatively, we can handle missing values by using *fuzzy partial logics*, which were proposed by Běhounek and Novák in [5]. They provide formal apparatus for several types of missing information such as "unknown" or "undefined" (i.e. not meaningful) value. Basically, the semantics of these logics formed by algebras of truth values is extended by a special value "$*$".

The approach of this paper puts together results of several papers ([24, 26, 25, 27]). The main goal of this paper is to put together GUHA method, the theory of evaluative linguistic expressions and the theory of intermediate quantifiers and to bring new results from natural data using obtained information in sentences of natural language.

The rest of the paper is structured as follows: In the next section, we briefly overview the special methods of fuzzy natural logic (FNL) and we recall Fuzzy GUHA method. The main section of this paper is Section 3 where we show how the theory of intermediate quantifiers, which is one of three theories of fuzzy natural logic, can be applied together with Fuzzy GUHA method for an analysis of natural data. At the end of this paper, we propose an idea for future work, which will be based on the application of the theory of syllogistic reasoning.

2 Preliminaries

In this section, we give a very brief review of special methods of fuzzy natural logic and we recall Fuzzy GUHA method, which finds distinct associations between attributes of given objects.

By a fuzzy set, we denote a function $A : U \to [0, 1]$ where U is a universe and [0,1] is a support set of some standard algebra of truth values. The set of all fuzzy sets over U is denoted by $\mathscr{F}(U)$. If A is a fuzzy set in U, we will write $A \subsetsim U$.

2.1 Fuzzy natural logic

Fuzzy natural logic is developed using formal tools of the *fuzzy type theory* (FTT) that was in detail elaborated in [17]. The main objective of FNL is to develop a mathematical model of special human reasoning schemes that employ natural language. So, FNL also includes a model of the semantics of natural language. FNL is a formal mathematical theory that consists of three theories:

- The theory of evaluative linguistic expressions [18];
- The theory of fuzzy IF-THEN rules and approximate reasoning [21];
- The formal theory of intermediate quantifiers, generalized syllogisms and generalized square of opposition [12, 13].

The FNL has a very high potential for applications. In [20], we can find an application of fuzzy natural logic and fuzzy transform for analysis, forecasting and linguistic characterization of time series. In [30], linguistic associations are used to drive an ensemble of forecasters, and in [4] to adjust flood predictions. Some of the methods are available as a software package "lfl" (Linguistic Fuzzy Logic) [3] for the open-source R statistical environment [28].

2.2 Basic symbols

The *language* consists of variables $x_\alpha, \ldots$, special constants $c_\alpha, \ldots$ ($\alpha \in$ *Types*), the symbol λ, and brackets. The connectives (which are special constants) are *fuzzy equality/equivalence* $\equiv$, *conjunction* $\wedge$, *implication* $\Rightarrow$, *negation* $\neg$, *strong conjunction* $\&$, *strong disjunction* ∇, *disjunction* $\vee$, and *delta* Δ. We recall that by $T \vdash A_o$ we mean A_o is provable in T.

The truth values form an MV_Δ-algebra. Standard Łukasiewicz MV_Δ-algebra

$$\mathscr{L} = \langle [0,1], \vee, \wedge, \otimes, \to, 0, 1, \Delta \rangle \tag{1}$$

is a special case of MV_Δ-algebra, where

$$\begin{aligned} \wedge &= \text{minimum}, & \vee &= \text{maximum}, \\ a \otimes b &= 0 \vee (a + b - 1), & a \to b &= 1 \wedge (1 - a + b), \\ \neg a &= a \to 0 = 1 - a, & \Delta(a) &= \begin{cases} 1 & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The Łukasiewicz disjunction which interprets $\boldsymbol{\nabla}$ is defined as $a \oplus b = 1 \wedge (a + b)$.

A frame $\mathscr{M}$ is a *model* of T ($\mathscr{M} \models T$) if all axioms are true in the degree $\mathbf{1}$ in $\mathscr{M}$. If A_o is true in the degree $\mathbf{1}$ in all models of T then we write $T \models A_o$.

Interpretation of formulas in a frame $\mathscr{M}$ is defined w.r.t. an assignment p of elements from $\mathscr{M}$ to variables. Namely, p is a function from the set of all variables of the language J to elements from $\mathscr{M}$ in keeping with the corresponding types.

We will also need the following crisp predicate of sharp ordering of formulas from $Form_o$:

$$<_{(oo)o} \equiv \lambda x_o \, \lambda y_o \cdot \boldsymbol{\Delta}(x_o \Rightarrow y_o) \wedge \neg \boldsymbol{\Delta}(x_o \equiv y_o).$$

2.2.1 The theory of evaluative linguistic expressions

In the sequel, we suppose a special formal theory T^{Ev} in a language J^{Ev} of Ł-FTT where T^{Ev} provides formalization of a meaning of evaluative linguistic expressions. Evaluative linguistic expressions are expressions of natural language, such as *small, medium, big, very short, more or less deep, quite roughly strong*. This theory is a special theory of higher order fuzzy logic, which was introduced in [18], and it is based on on the standard Łukasiewicz MV-algebra which was defined in (1).

The meaning of an evaluative linguistic expression is constructed by means of a special formula that represents *intension*. The latter is interpreted in a model by a function from the set of possible worlds (in our theory, we prefer to speak about *contexts*) in a set of fuzzy sets. Intension determines, in each context, the corresponding extension that is a fuzzy set in a certain universe constructed as a *horizon*, which can be shifted along the universe.

By context in T^{Ev}, we understand a formula $w_{\alpha o}$, whose interpretation is a function $w : L \to M_\alpha$. Hence, the context determines in M_α a triple of elements,

$$\langle v_L, v_S, v_R \rangle, \quad v_L, v_S, v_R \in M_\alpha,$$

and $v_L = \mathscr{M}_p(w\bot)$, $v_S = \mathscr{M}_p(w\dagger)$, $v_R = \mathscr{M}_p(w\top)$.

For the theory of intermediate quantifiers, we consider only abstract expressions such as "very small" that contain no specification of "what is indeed small. For example, "very small animal" suggests considering various sizes of animals, depending on the context. Expressions of the form "⟨noun⟩ is $\mathscr{A}$" where $\mathscr{A}$ is an evaluative expression, are called *evaluative (linguistic) predications*. Consequently, these expressions have only one (abstract) context; as a result, their intension actually coincides with their extension. Therefore, we will omit the context in the formulas introduced below.

The meaning of evaluative expressions is obtained as an interpretation of special formulas in a model of T^{Ev}. The core idea are three horizons, which are defined as follows:

$$\begin{aligned} LH_{oo} &:= \lambda z_o \cdot \bot \sim z_o, \\ MH_{oo} &:= \lambda z_o \cdot \dagger \sim z_o, \\ RH_{oo} &:= \lambda z_o \cdot \top \sim z_o. \end{aligned}$$

The LH is a *left horizon* interpreted by a function assigning to each z_o a truth degree of the fuzzy equality with $\bot$; a similar construct applies to the *right* RH and *middle* MH horizons.

A ⟨linguistic hedge⟩ represents a class of adverbial modification that includes a class of *intensifying adverbs* such as "very", "roughly", "approximately", or "significantly". Two basic kinds of linguistic hedges can be distinguished in two groups as follows:

- *narrowing* hedges, e.g. "extremely", "significantly", "very";
- *widening* hedges, e.g. "more or less", "roughly", "very roughly".

We recall that, for example, "very small" is more precise than "small" which is more precise than "roughly small".

We introduce the following special linguistic hedges: $\{Ex, Si, Ve, ML, Ro, QR, VR\}$ (*extremely, significantly, very, more or less, roughly, quite roughly, very roughly*, respectively), which are ordered as follows:

$$Ex \preceq Si \preceq Ve \preceq \bar{\boldsymbol{\nu}} \preceq ML \preceq Ro \preceq QR \preceq VR \tag{2}$$

By $\preceq$, we denote a relation of the partial ordering of the hedges. It can be found in [18, page 23]. The $\bar{\boldsymbol{\nu}}$ is an empty hedge. Note that hedges *Ex*, *Si*, *Ve* have a narrowing effect, and *ML*, *Ro*, *QR*, *VR* have a widening effect with respect to the empty hedge. A special role in our theory is played by formulas $Sm\,\boldsymbol{\Delta}$, $Me\,\boldsymbol{\Delta}$, $Bi\boldsymbol{\Delta}$, where the connective $\boldsymbol{\Delta}$ has been used as a specific hedge that can be taken as the linguistic hedge "utmost" (or, alternatively, a "limit"). This construct makes it possible to also include classical quantifiers in our theory without the need to introduce them as special cases that are different from the rest of the theory.

2.3 Definition of intermediate quantifiers

Intermediate quantifiers are linguistic expressions, such as *most, many, almost all, a few, a large part of*, etc. which belong among generalized quantifiers whose theory has been studied already for many years (cf., for example,[32]). This theory was further generalized to the fuzzy approach (cf.[8]). The formal theory of *intermediate quantifiers* using the fuzzy type theory (a higher-order fuzzy logic) was introduced in [19].

The basic idea consists in the assumption that intermediate quantifiers are just classical quantifiers $\forall$ or $\exists$ but the universe of quantification is modified. This is achieved using the theory of evaluative linguistic expressions introduced in the previous subsection.This idea is characterized by the following definition:

Definition 1. *Let $A, B, z \in Form_{o\alpha}$ be formulas and $x \in Form_\alpha$ be variables where $\alpha \in \mathscr{S}$. An intermediate quantifier of a type $\langle 1, 1\rangle$ interpreting the sentence "$\langle$Quantifier$\rangle$ B's are A" is one of the following formulas:*

$$(Q^{\forall}_{Ev}\,x)(B, A) \quad := \quad (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\forall x)(z\,x \Rightarrow Ax)) \wedge Ev((\mu B)z)), \tag{3}$$

$$(Q^{\exists}_{Ev}\,x)(B, A) \quad := \quad (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\exists x)(zx \wedge Ax)) \wedge Ev((\mu B)z)) \tag{4}$$

where x represents elements, z, A, B are interpreted as fuzzy sets. The $(\mu B)z$ represents a measure of the fuzzy sets z w.r.t. B and Ev is an evaluative expressions.

By putting of the specific evaluative linguistic expression we obtain the definition of the concrete intermediate quantifier. By $Ex\,Bi$ we mean that the fuzzy set z is "extremely big" w.r.t. B, the formula $Bi\ Ve$ denotes the fact that the fuzzy set z is "very big" w.r.t. B and, finally, by $\neg(Sm\,\bar{\boldsymbol{\nu}})$ we understand that z is "not small" w.r.t. B.

Definition 2. *The following special intermediate quantifiers can be introduced – see Table 1.*

Theorem 1 (Valid implications; [19])**.** *Let $\boldsymbol{A}, \ldots, \boldsymbol{G}$ be the basic intermediate quantifiers defined above. Then the following sets of implications are provable in T^{IQ}:*

(a) $T^{IQ} \vdash \mathbf{A} \Rightarrow \mathbf{P}$, $\quad T^{IQ} \vdash \mathbf{P} \Rightarrow \mathbf{T}$, $\quad T^{IQ} \vdash \mathbf{T} \Rightarrow \mathbf{K}$.

(b) $T^{IQ} \vdash \mathbf{E} \Rightarrow \mathbf{B}$, $\quad T^{IQ} \vdash \mathbf{B} \Rightarrow \mathbf{D}$, $\quad T^{IQ} \vdash \mathbf{D} \Rightarrow \mathbf{G}$.

In a finite model, the measure $(\mu B)z$ can be given by

$$(\mu B)z = \begin{cases} 1 & \text{if } z = B, \\ \frac{|z|}{|B|} & \end{cases} \tag{5}$$

where $|z| = \sum_{u \in M_\alpha} z(u)$ and $|B| = \sum_{u \in M_\alpha} B(u)$. In this text, we restrict ourselves only on the quantifiers "Almost all", "Most", and "Many".

Table 1: A definition of special intermediate quantifiers

A: All B are $A := (Q^{\forall}_{Bi\boldsymbol{\Delta}}x)(B,A) \equiv (\forall x)(Bx \Rightarrow Ax)$,

E: No B are $A := (Q^{\forall}_{Bi\boldsymbol{\Delta}}x)(B,\neg A) \equiv (\forall x)(Bx \Rightarrow \neg Ax)$,

P: Almost all B are $A := (Q^{\forall}_{Bi\ Ex}x)(B,A) \equiv (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\forall x)(zx \Rightarrow Ax)) \wedge (Bi\ Ex)((\mu B)z))$,

B: Few B are $A := (Q^{\forall}_{Bi\ Ex}x)(B,\neg A) \equiv (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\forall x)(zx \Rightarrow \neg Ax)) \wedge (Bi\ Ex)((\mu B)z))$,

T: Most B are $A := (Q^{\forall}_{Bi\ Ve}x)(B,A) \equiv (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\forall x)(zx \Rightarrow Ax)) \wedge (Bi\ Ve)((\mu B)z))$,

D: Most B are not $A := (Q^{\forall}_{Bi\ Ve}x)(B,\neg A) \equiv (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\forall x)(zx \Rightarrow \neg Ax)) \wedge (Bi\ Ve)((\mu B)z))$,

K: Many B are $A := (Q^{\forall}_{\neg(Sm\,\bar{\boldsymbol{\nu}})}x)(B,A) \equiv (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\forall x)(zx \Rightarrow Ax)) \wedge \neg(Sm\,\bar{\boldsymbol{\nu}})((\mu B)z))$,

G: Many B are not $A := (Q^{\forall}_{\neg(Sm\,\bar{\boldsymbol{\nu}})}x)(B,\neg A) \equiv (\exists z)((\boldsymbol{\Delta}(z \subseteq B)\,\&(\forall x)(zx \Rightarrow \neg Ax)) \wedge \neg(Sm\,\bar{\boldsymbol{\nu}})((\mu B)z))$,

I: Some B are $A := (Q^{\exists}_{Bi\boldsymbol{\Delta}}x)(B,A) \equiv (\exists x)(Bx \wedge Ax)$,

O: Some B are not $A := (Q^{\exists}_{Bi\boldsymbol{\Delta}}x)(B,\neg A) \equiv (\exists x)(Bx \wedge \neg Ax)$.

2.4 Fuzzy GUHA

The classical GUHA method [9] deals with data in the form of Table 2, where $o_1, \ldots, o_n$ denote objects, $X_1, \ldots, X_m$ denote independent boolean attributes, Z denotes the dependent (explained) boolean attribute, and finally, symbols a_{ij} (or a_i) $\in \{0,1\}$ denote whether an object o_i carries an attribute X_j (or Z) or not.

Table 2: Standard GUHA Table.

	X_1	$\ldots$	X_m	Z
o_1	a_{11}	$\ldots$	a_{1m}	a_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
o_n	a_{n1}	$\ldots$	a_{nm}	a_n

The objective of the GUHA method is to search for associations of the form

$$\mathtt{A}(X_1, \ldots, X_p) \simeq \mathtt{B}(Z),$$

where $\mathtt{A}$, $\mathtt{B}$ are predicates containing only the connective AND and $X_1, \ldots, X_p$, for $p \leq m$, are all variables occurring in $\mathtt{A}$. The $\mathtt{A}$, $\mathtt{B}$ are called the *antecedent* and *consequent*, respectively.

The relationship between the antecedent and consequent is described by the so called *quantifier* $\simeq$. There are many quantifiers that characterize validity of the association in data [10]. For instance, the so-called *implicational quantifier* is defined as true if

$$\frac{a}{a+b} > \gamma \qquad \text{and} \qquad \frac{a}{n} > r,$$

where $\gamma \in [0,1]$ is a user-specified *degree of confidence* and $r \in [0,1]$ is a *degree of support*. Here a denotes the number of positive occurrences of $\mathtt{A}$ as well as $\mathtt{B}$ in the data; and b is the number of positive occurrences of $\mathtt{A}$ and of negated $\mathtt{B}$, i.e. of 'not $\mathtt{B}$'.

The aim of this paper is to replace the classical GUHA implicational quantifier with some of the *generalized intermediate quantifiers*. Namely, we restrict ourselves on the quantifier "Almost all" (**P**), "Most" (**T**), and "Many" (**K**) – see Definition 2. The algorithm for finding the truth value of the quantifier uses the truth degrees of both the antecedent and consequent evaluated on each object o_i and searches such z from Definition 2 by using a specific optimization technique, whose description is going to be published elsewhere. Contrary to the classical GUHA or fuzzy GUHA quantifiers, the intermediate quantifiers result in a non-crisp truth value from the interval $[0,1]$.

3 Algorithm using fuzzy natural logic

Let there be a model $\mathscr{M} \models T^{\mathrm{IQ}}$ which was constructed in [19] with the frame defined as follows:

$$\mathscr{M} = \langle (M_\alpha, =_\alpha)_{\alpha \in Types}, \mathscr{L}_\Delta \rangle.$$

Recall that $M_o = [0,1]$ is the support of the Łukasiewicz MV_Δ-algebra. The fuzzy equality $=_o$ is the Łukasiewicz biresiduation $\leftrightarrow$. Furthermore, $M_\epsilon = \{u_1, \ldots, u_n\}$ is a finite set with a fixed numbering of its elements and $=_\epsilon$ is a fuzzy equality defined by

$$[u_i =_\epsilon u_j] = (1 - \min\left(1, \frac{|i-j|}{s}\right))$$

for some fixed natural number $s \leq r$. This a separated fuzzy equality w.r.t. the Łukasiewicz conjunction $\otimes$. It can be verified that all the logical axioms of Ł-FTT and of the theory of evaluative linguistic expressions are true in the degree 1 in $\mathscr{M}$.

Let $A, B, z \in Form_{o\alpha}$ and the interpretation of μ be defined by (5). Then the general definition of the intermediate quantifier

$$(Q^{\forall}_{Ev} x)(B, A) \quad := \quad (\exists z)((\mathbf{\Delta}(z \subseteq B) \,\&\, (\forall x)(zx \Rightarrow Ax)) \wedge (Ev)((\mu B)z)), \quad (6)$$

can be rewritten in the finite model as follows:

$$\bigvee_{z \subseteq B} \left(\bigwedge_{u \in M_\alpha} \big(z(u) \to A(u)\big) \wedge f\left(\frac{|z|}{|B|}\right) \right), \quad (7)$$

where $\to$ is Łukasiewicz implication and $f : [0,1] \to [0,1]$ is a non-decreasing function representing Ev. Recall that $z, A, B \subsetneq M_\alpha$ in (7) are fuzzy sets which interpret formulas $z, A, B \in Form_{o\alpha}$ from (6). Later we will work with the semantical definition of intermediate quantifiers (7) and so by $z, A, B \subsetneq M_\alpha$ we will mean fuzzy sets. Note that $\mathbf{\Delta}(z \subseteq B)$ means that in the concrete model $\mathscr{M} \models T^{\mathrm{IQ}}$, $z(u) \leq B(u)$ holds for all $u \in M_\alpha$, $\wedge$ stands for minimum, $\vee$ for maximum, and $\to$ for Łukasiewicz implication. Therefore, we can rewrite the definition further as follows:

$$\sup_{z \subseteq B} \big\{ \min\{\mathrm{prop}(z), \mathrm{size}(z)\} \big\}, \quad (8)$$

where

$$\mathrm{prop}(z) := \min_{u \in M_\alpha} \big\{ z(u) \to A(u) \big\} == \min_{u \in M_\alpha} \big\{ 1, 1 - z(u) + A(u) \big\}, \mathrm{size}(z) := f\left(\frac{|z|}{|B|}\right). \quad (9)$$

Lemma 1. *Let $A, B \subsetneq M_\alpha$ be fuzzy sets.*
Let $p \in [\mathrm{prop}(B), 1]$ and let

$$z_p(u) := \min\big\{ B(u), 1 - p + A(u) \big\},$$

for all $u \in M_\alpha$. Then:

(a) $z_p \subseteq B$,

(b) $\mathrm{prop}(z_p) = p$,

(c) $\mathrm{size}(z_p) = \max\{\mathrm{size}(z') \mid \forall z' \subseteq B : \mathrm{prop}(z') \geq p\}$.

Proof. (a) Evidently, $z_p(u) \leq B(u)$, for all $u \in M_\alpha$.

(b) For all $u \in M_\alpha$, we obtain that

$$1 - z_p(u) + A(u) = \\ = 1 - \min\big\{ B(u), 1 - p + A(u) \big\} + A(u) \geq p. \quad (10)$$

Hence $\mathrm{prop}(z_p) \geq p$. Now we are going to prove that $\mathrm{prop}(z_p) = p$. Assume to the contrary that $\mathrm{prop}(z_p) > p$. Then for all $u \in M_\alpha$ we obtain that

$$\min\{1, 1 - \min\{B(u), 1 - p + A(u)\} + A(u)\} > p.$$

For $p = 1$ we obtain immediately a contradiction. For $p < 1$, we can derive:

$$1 - \min\big\{B(u), 1 - p + A(u)\big\} + A(u) > p,$$

and then

$$\begin{aligned} 1 - p + A(u) &> \min\big\{B(u), 1 - p + A(u)\big\}, \\ 1 - p + A(u) &> B(u), \\ 1 - B(u) + A(u) &> p, \\ \mathrm{prop}(B) &> p, \end{aligned}$$

which is again a contradiction.

(c) Assume to the contrary that there exists such $z' \subsetneq M_\alpha$, such that $z' \subseteq B$, $\mathrm{prop}(z') \geq p$ and $\mathrm{size}(z') > \mathrm{size}(z_p)$. Then for some $u \in M_\alpha$, $z'(u) > z_p(u)$, i.e.

$$z'(u) > \min\big\{B(u), 1 - p + A(u)\big\}.$$

Hence, either $z'(u) > B(u)$, which breaks the assumption that $z' \subseteq B$, or $z'(u) > 1 - p + A(u)$, which is in contradiction with $\mathrm{prop}(z') \geq p$. □

Lemma 2. *Let $\pi : p \mapsto \mathrm{prop}(z_p)$ and $\sigma : p \mapsto \mathrm{size}(z_p)$ be functions with domain $p \in [\mathrm{prop}(B), 1]$. Then*

$$\sup_{z \subseteq B} \big\{\min\{\mathrm{prop}(z), \mathrm{size}(z)\}\big\}$$

(same as (8)) equals to

$$\sup_{p \in [\mathrm{prop}(B), 1]} \big\{\min\{\pi(p), \sigma(p)\}\big\}.$$

Proof. Let $Z := \{z \subsetneq M_\alpha \mid z \subseteq B\}$ and let us define for each $p \in [\mathrm{prop}(B), 1]$, $Z_p := \{z \subsetneq M_\alpha \mid z \subseteq B \wedge \mathrm{prop}(z) = p\}$. Evidently, $Z_p \subseteq Z$, $Z_p \cap Z_{p'} = \emptyset$ for any $p \neq p'$, and

$$\bigcap_{p \in [\mathrm{prop}(B), 1]} Z_p = Z.$$

Now, let $t(z) := \min\{\mathrm{prop}(z), \mathrm{size}(z)\}$. It is clear that

$$\sup_{z \subseteq B} \big\{t(z)\big\} = \sup_{p \in [\mathrm{prop}(B), 1]} \Big\{ \sup_{z \in Z_p} \big\{t(z)\big\} \Big\}.$$

Note that $z_p \in Z_p$, hence,

$$\max_{z \in Z_p} \{\mathrm{size}(z)\} = \mathrm{size}(z_p) = \sigma(p)$$

(see Lemma 1), and $\mathrm{prop}(z) = \pi(p) = p$, for each $z \in Z_p$. Then

$$\sup_{z \in Z_p} \big\{t(z)\big\} = \sup_{z \in Z_p} \big\{\min\{\mathrm{prop}(z), \mathrm{size}(z)\}\big\} = \\ = \min\{\mathrm{prop}(z_p), \mathrm{size}(z_p)\} = \min\{\pi(p), \sigma(p)\}.$$

□

Algorithm 1 The Algorithm for Evaluation of an Intermediate Quantifier

function EVALUATEQUANTIFIER($A_{o\epsilon}, B_{o\epsilon}, f, \delta$)
  $d \leftarrow \sum_{u \in M_\epsilon} B(u)$
  $lo \leftarrow \min_{u \in M_\epsilon} \{1, 1 - B(u) + A(u)\}$
  $hi \leftarrow 1$
  repeat
    $\alpha \leftarrow \frac{lo+hi}{2}$
    $c \leftarrow \sum_{u \in M_\epsilon} \min\{B(u), 1 - \alpha + A(u)\}$
    $\beta \leftarrow f\left(\frac{c}{d}\right)$
    if $\alpha \leq \beta$ **then**
      $lo \leftarrow \alpha$
    end if
    if $\alpha \geq \beta$ **then**
      $hi \leftarrow \alpha$
    end if
  until $hi - lo \leq \delta$
  return α
end function

4 Experiment on Real Data

We illustrate the use of our proposal on a famous Edgar Anderson's Iris dataset [6, 2]. It is a multivariate data set that quantifies the morphologic variation of iris flowers of three related species. It contains four numeric columns with sepal and petal length and width, and a categorical column of iris species (setosa, versicolor, virginica).

The process of analysis was conducted as follows. First of all, numeric columns were transformed into membership degrees of fuzzy sets. From each numeric column, 5 fuzzy sets were extracted that model the following linguistic expressions: small, very small, medium, big, and very big. (E.g. the original numerical column "Sepal.Length" was used as a basis for the following fuzzy sets: "Sm.Sepal.Length", "VeSm.Sepal.Length", "Me.Sepal.Length", "Bi.Sepal.Length", and "VeBi.Sepal.Length".)

On the transformed data set, a fuzzy GUHA from the "lfl" package [3] of the R statistical environment [28] was executed. We have searched for all rules with at most four predicates in the antecedent. A detailed analysis of the iris data set is not the purpose of this paper – therefore, we provide only a few examples of the obtained rules that also illustrate the latter use of intemediate quantifiers. Among other, e.g. the following rule was obtained from the data set:

$$\text{Sm.Sepal.Length \& Sm.Petal.Length} \Rightarrow \text{Sm.Petal.Width},$$

$$\gamma = 0.995, r = 0.132,$$

which may be interpreted as:

> *"If an object has both sepal and petal of small length then it has petal of small width with support 0.132 and confidence 0.995."*

In order to provide a linguistic characterization of the intensity of the rule, we have then evaluated membership degrees of the three intermediate quantifiers "Almost all", "Most", and "Many". The above-mentioned rule thus can be interpreted as:

> *"**Almost all** irises with both sepal and petal of small length have its petal width small."*

For other rules see Table 3.

As can be seen, the most strict quantifier (among the three that were evaluated) is "Almost all". It corresponds roughly to the *confidence* around 0.98, although generally, there is not a functional dependency between confidence and quantifier's truth degree. (The investigation of how the truth degree of the quantifiers relates to the confidence or other interest measures is given for future work.) A slightly weaker is the quantifier "Most", which was still 1 although the truth degree of

Table 3: Example of obtained rules together with the degrees of support and confidence and membership degrees of quantifiers "almost all", "most", and "many"

	supp.	conf.	all	almost all	most	many	exists
Sm.Sepal.Len & Sm.Petal.Len ⇒ Sm.Petal.Width	0.132	0.995	0.954	1.000	1.000	1.000	1.000
Sm.Sepal.Len & Sm.Petal.Len ⇒ Me.Sepal.Width	0.119	0.902	0.000	0.682	0.902	1.000	1.000
Me.Sepal.Width ⇒ Sm.Petal.Width	0.218	0.350	0.000	0.174	0.255	1.000	1.000
Sm.Sepal.Len ⇒ Me.Sepal.Width	0.120	0.831	0.000	0.632	0.790	1.000	1.000
Sm.Sepal.Len ⇒ Sm.Petal.Len	0.132	0.915	0.491	0.802	0.934	1.000	1.000
Sm.Petal.Len ⇒ Me.Sepal.Width	0.222	0.687	0.000	0.324	0.469	1.000	1.000
Sm.Petal.Width ⇒ Me.Sepal.Width	0.218	0.686	0.000	0.329	0.469	1.000	1.000
VeSm.Sepal.Len ⇒ Sm.Petal.Width	0.064	1.000	0.999	1.000	1.000	1.000	1.000
VeSm.Sepal.Len ⇒ Me.Sepal.Width	0.054	0.852	0.021	0.532	0.776	1.000	1.000
Me.Sepal.Width ⇒ VeSm.Sepal.Len	0.054	0.087	0.000	0.142	0.200	0.886	1.000

"Almost all" decreased to 0.941. The weakest is the "many" quantifier that has a truth degree 0.962 even for a rule where the other two quantifiers decreased below 0.23.

As can be seen, the provided example shows that the fuzzy GUHA association rules may be interpreted using intermediate quantifiers that provide a linguistic description of the intensity of the found relationship. If missing information is present in the data, we can follow the principle of the theory of possibility and evaluate the quantifiers by considering the most pessimistic and optimistic case, which may result in an interval of possible membership degrees.

5 Application of generalized syllogisms

In the previous section, we showed that the theory of intermediate quantifiers can be applied for the interpretation of natural data using intermediate quantifiers. A very interesting question is: is it possible to infer other new results from linguistic expressions which were derived before? The main idea is to use the *theory of syllogistic reasoning*, which was in detail studied our previous papers (see [14, 15, 16]).

A *syllogism* (or *logical appeal*) is a special kind of logical argument in which the *conclusion* is inferred from two *premises*: the *major premise* (first) and *minor premise* (second). The syllogisms will be written as triples of formulas $\langle P_1, P_2, C\rangle$. The *generalized intermediate syllogism* (or simply, the *intermediate syllogism*) is obtained from any traditional syllogism (valid or not) when replacing one or more of its formulas by formulas containing intermediate quantifiers.

We say that a syllogism $\langle P_1, P_2, C\rangle$ is *valid* in a theory T if $T \vdash P_1 \,\&\, P_2 \Rightarrow C$, or equivalently, if $T \vdash P_1 \Rightarrow (P_2 \Rightarrow C)$. Note that this means that in every model $\mathscr{M} \models T$ the inequality

$$\mathscr{M}(P_1) \otimes \mathscr{M}(P_2) \leq \mathscr{M}(C) \tag{11}$$

holds true. Hence, the truth degree of both premises is equal to Łukasiewicz conjunction and it is less than or equal to the truth degree of the conclusion.

Generally, we may even consider the syllogisms as follows :

$$\begin{array}{l} Q_1\ A \text{ is } B \\ Q_2\ C \text{ is } D \\ \hline Q_C\ E \text{ is } F \end{array} \qquad \begin{array}{l} \text{Most good jokes are old} \\ \text{Most good jokes are funny} \\ \hline \text{Some funny jokes are old} \end{array} \tag{12}$$

where $A, B, C, D, E, F \in Form_{o\alpha}$ are formulas. A typical example of a syllogism with the intermediate quantifier "Most" in both premises is included in (12). Changing the formulas results in concrete examples of syllogisms including those of Figures 1–4, which were introduced and whose validity was syntactically proved in [14].

Below we introduce typical examples as follows:

Figure-II: **AEO**

All birds have wings.
No humans have wings.

Some humans are not birds.

or

Figure-I: **ATK**

All people who suffer from respiratory diseases suffer from asthma.
Most people who live in area affected by heavy industry
suffer from respiratory diseases.

Many people who live in area affected by heavy industry
suffer from suffer from asthma.

Typical examples of, so called, *non-trivial* syllogisms which consist of the intermediate quantifiers in the both premises are follows:

Figure-III: **TTI**

Most people who suffer from respiratory diseases suffer from asthma.
Most people who suffer from respiratory diseases
suffer from ischemic disease of leg.

Some people who suffer from asthma suffer from ischemic disease of leg.

5.1 Inferred generalized syllogisms

A practical application of syllogistic reasoning may be also in reducing of the size of the set of rules obtained from data. If some rule can be derived from the others then it is very likely not so much important or interesting because it does not carry any new knowledge. On the other side, if we obtain from data a rule that is more intense that the one that can be derived from the other rules, it may be an indicator that such rule may carry some surprising and thus potentially useful knowledge.

From (4) by the theory of syllogistic reasoning we can infer the following generalized syllogisms which give us the new information about studied natural data.

Figure-III: **PPI**

Almost all irises with both sepal and petal of small length have its petal width small.
Almost all irises with both sepal and petal of small length have its sepal width medium.

Some irises with medium sepal width have small petal width.

By the property of the monotonicity of all the quantifiers which was introduced in Theorem 1 we van infer other valid syllogisms, for example, **PTI**-III, **TPI**-III and **TTI**-III.

Similarly, we obtain the following syllogism of Figure-III:

Figure-III: **PPI**

Almost all irises with small sepal length have medium sepal width.
Almost all irises with small sepal length have small petal length.

Some irises with small petal length have medium sepal width.

Analogously as in previous syllogism by monotonicity property we infer other syllogisms.

Figure-I: **APP**

All irises with small petal width have medium sepal width.
Almost all irises with very small sepal length have small petal width.

Almost all irises with very small sepal length have medium sepal width.

By changing of the both premises we obtain the generalized syllogism of Figure-IV as follows:

Figure-IV: **KAI**

Many irises with very small sepal length have small petal width.
All irises with small petal width have medium sepal width.
Some irises with medium sepal width have very small sepal length.

By Theorem 1 we can infer the generalized syllogisms **TAI**-IV and **PAI**-IV.

6 Conclusion

In this paper, we have applied a standard data-mining method, namely the Fuzzy GUHA method, together with the theory of intermediate quantifiers, which is one of three main theories of *fuzzy natural logic*. We have found the linguistic associations in the form of IF-THEN rules composed of evaluative linguistic expressions. The theory of intermediate quantifiers was successfully applied to the resulting rules in order to provide a linguistic description of the intensity of the relationship captured by the rules. We have shown a practical example on a real data set. The main idea for the future is to apply the theory of generalized Aristotle's syllogisms and the theory of generalized Aristotle's square of opposition, which give to infer a new possible information from the result that was found before. Alternatively, we may use the syllogisms to prune the generated rules from rules that are potentially not so useful as those that cannot be derived from the others. Moreover, we would like to extend our approach to data with missing values by applying fuzzy partial logics [5].

Acknowledgement

The paper has been supported by the project "LQ1602 IT4Innovations excellence in science".

References

[1] Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proc. 20th Int. Conf. on Very Large Databases, pp. 487–499. AAAI Press, Chile (1994)

[2] Anderson, E.: The species problem in iris. Annals of the Missouri Botanical Garden **23**(3), 457–509 (1936). URL http://www.jstor.org/stable/2394164

[3] Burda, M.: Linguistic fuzzy logic in R. In: Proc. IEEE International Conference On Fuzzy Systems (FUZZ-IEEE), pp. 1–7. IEEE, Istanbul (2015)

[4] Burda, M., Rusnok, P., Štěpnička, M.: Mining linguistic associations for emergent flood prediction adjustment. Advances in Fuzzy Systems **2013** (2013)

[5] Běhounek, L., Novák, V.: Towards fuzzy partial logic. In Proceedings of the IEEE 45th International Symposium on Multiple-Valued Logics (ISMVL 2015) pp. 139–144 (2015)

[6] Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics **7**(2), 179–188 (1936). DOI 10.1111/j.1469-1809.1936.tb02137.x. URL http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

[7] Geng, L., Hamilton, H.J.: Interestingness measures for data mining: A survey. ACM Computing Surveys **38**(3), 9 (2006)

[8] Glöckner, I.: Fuzzy Quantifiers: A Computational Theory. Springer, Berlin (2006)

[9] Hájek, P.: The question of a general concept of the GUHA method. Kybernetika **4**, 505–515 (1968)

[10] Hájek, P., Havránek, T.: Mechanizing hypothesis formation: Mathematical foundations for a general theory. Springer-Verlag, Berlin/Heidelberg/New York (1978)

[11] Lakoff, G.: Linguistics and natural logic. Synthese **22**, 151–271 (1970)

[12] Murinová, P., Novák, V.: A formal theory of generalized intermediate syllogisms. Fuzzy Sets and Systems **186**, 47–80 (2013)

[13] Murinová, P., Novák, V.: Analysis of generalized square of opposition with intermediate quantifiers. Fuzzy Sets and Systems **242**, 89–113 (2014)

[14] Murinová, P., Novák, V.: The structure of generalized intermediate syllogisms. Fuzzy Sets and Systems **247**, 18–37 (2014)

[15] Murinová, P., Novák, V.: Generalized conjunctive syllogisms with more premisses in fuzzy naturral logic. Proocedings of FLINS (2016)

[16] Murinová, P., Novák, V.: Syllogisms and 5-square of opposition with intermediate quantifiers. Logica Universalis (2016)

[17] Novák, V.: On fuzzy type theory. Fuzzy Sets and Systems **149**, 235–273 (2005)

[18] Novák, V.: A comprehensive theory of trichotomous evaluative linguistic expressions. Fuzzy Sets and Systems **159**(22), 2939–2969 (2008)

[19] Novák, V.: A formal theory of intermediate quantifiers. Fuzzy Sets and Systems **159**(10), 1229–1246 (2008)

[20] Novák, V.: Linguistic characterization of time series. Fuzzy Sets and Systems **285**, 52–72 (2015)

[21] Novák, V., Lehmke, S.: Logical structure of fuzzy IF-THEN rules. Fuzzy Sets and Systems **157**, 2003–2029 (2006)

[22] Novák, V., Murinová, P.: On general properties of intermediate quantifiers. In IPMU 2014, Part II, Communications in Computer and Informationa Science, Springer, Berlin pp. 355–364 (2014)

[23] Novák, V., Perfilieva, I., Dvořák, A., Chen, G., Wei, Q., Yan, P.: Mining pure linguistic associations from numerical data. Int. Journal of Approximate Reasoning **48**, 4–22 (2008)

[24] Novák, V., Perfilieva, I., Pavliska, V.: The use of higher order f-transform in time series analysis. World Congress IFSA 2011 and AFSS 2011 pp. 2211–2216 (2011)

[25] Novák, V., Štěpnička, M., Kupka, J.: Linguistic descriptions: their structure and applications. in: H. Larsen, M. Martin-Bautista, M. Vila, T.Andreasen, H. Christiansen (Eds.), Flexible Query Answering Systems, Springer, Berlin, Heidelberg **39**, 209–220 (2013)

[26] Novák, V., Štěpnička, M., Perfilieva, I., Vavříčková, L., Dvořák, A.: Analysis of seasonal time series using fuzzy approach. Int. J. Gen. Sys. **39**, 305–328 (2010)

[27] Perfilieva, I.: Fuzzy transform: theory and applications. Fuzzy sets and systems **157**, 993–1023 (2006)

[28] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2016). URL https://www.R-project.org/

[29] Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right objective measure for association analysis. Information Systems **29**(4), 293–313 (2004)

[30] Štěpnička, M., Burda, M., Štěpničková, L.: Fuzzy rule base ensemble generated from data by linguistic associations mining. Fuzzy Sets and Systems **285**, 140 – 161 (2016). Special Issue on Linguistic Description of Time Series

[31] Westerståhl, D.: Aristotelian syllogisms and generalized quantifiers. Studia Logica: An International Journal for Symbolic Logic **48**, 577–585 (1989)

[32] Westerståhl, D.: Quantifiers in formal and natural languages. In: D. Gabbay, F. Guenthner (eds.) Handbook of Philosophical Logic vol. IV, pp. 1–131. D. Reidel, Dordrecht (1989)

Experimental Comparison of Three On-Line Fuzzy Classification Systems

Tomoharu NAKASHIMA

Osaka Prefecture University
Gakuen-cho 1-1, Naka-ku, Sakai, Osaka 599-8531
Japan
tomoharu.nakashima@kis.osakafu-u.ac.jp

Abstract

This paper compares the performance of the on-line version of fuzzy classifiers in a dynamic environment. The classification problems that are considered in this paper assume that the true classification boundaries changes over time. In order to tackle this difficult situation, on-line learning frameworks are employed in the fuzzy classifier. Three on-line learning methods as well as a famous delta learning method are investigated for their classification performance.

Keywords: Machine learning, Fuzzy if-then rules, Classification, On-line learning.

1 Introduction

Information explosion in recent years has brought the need for those information systems that can deal with a massive volume of data within a reasonable computa-tional time. Not only the data are growing, but also the property of the data chang-es over time. Even the source of the data might change itself over time, making the previous data rather obsolete.

On-line learning is one of the promising methods for the circumstances de-scribed above. On-line learning methods modify the models such as classifiers and regression functions based on a few training patterns. They can be applied for learning problems with an intractably large number of training patterns. If the number of training patterns is too large to handle at once, only a small subset of the training patterns is used to train a learning model and incrementally update it using another small subset of the training patterns. On-line learning techniques are also useful in dynamic problem where the true input-output relation or the true classification boundaries changes over time. In this case, available training pat-terns at each time can be used to update the learning model.

There are several machine learning approaches for on-line learning including confidence-weighted learning [1] and passive-aggressive learning [2]. These ap-proaches are proposed for linear classifiers. The weights of a linear classifier are adaptively updated once a training pattern are newly available. The weight update for both passive-aggressive and confidence-weighted learning is conducted so that the amount of weight change is minimized while the loss function of the new training patterns is zero or probabilistically small.

Fuzzy rule-based systems have been shown their good performance in classifi-cation tasks [3]. In the case of classification, a fuzzy classifier is constructed from a given set of training patterns. The fuzzy classifier consists of a set of fuzzy if-then rules. In this paper, fuzzy if-then rules with antecedent fuzzy sets and a con-sequent real value are considered for the fuzzy classifier. The antecedent fuzzy sets are assumed to be given a priori by using human knowledge. The consequent real values are obtained by using some learning technique from given training patterns.

Although the fuzzy systems have been shown to be effective for classification, the classification problems have been assumed to be static. That is, true classifica-tion boundaries did not change but only remained still. In this paper, two on-line learning algorithms are introduced to fuzzy rule-based classification systems in order to handle dynamic classification where the true classification boundaries changes over time. As on-line learning algorithms, the paper focuses passive-aggressive learning and confidence-weighted learning.

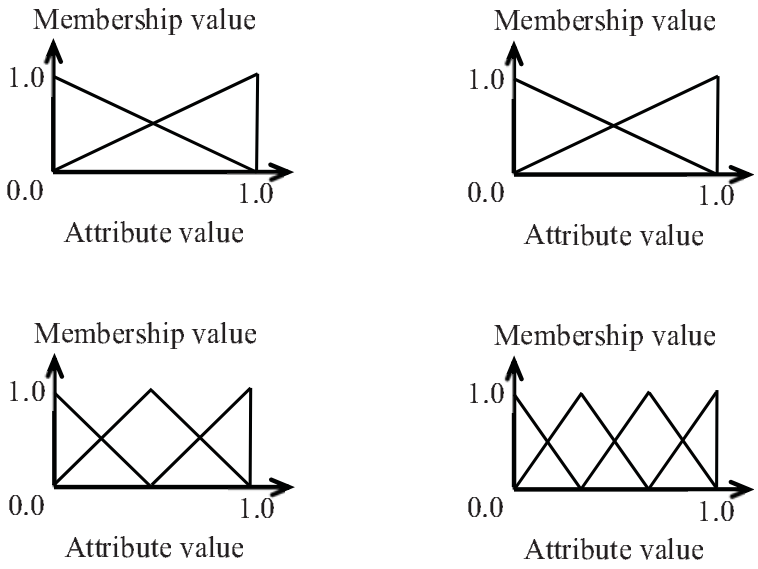


Figure 1: Triangular-type membership functions.

First, it is shown that a fuzzy classification system can be viewed as a linear classification system in a high-dimensionally mapped space. Next, the formulation of the on-line learning algorithms is provided for fuzzy classification systems. Then, a series of computational experiments are conducted in order to investigate the classification performance of on-line fuzzy classification systems. In the com-putational experiments, two on-line classification problems are used where a two-dimensional training pattern is randomly generated with a rotating classification boundary at each time step

1.1 Fuzzy Classifier

A fuzzy classifier consists of fuzzy if-then rules. Let us assume that the classification problem has the dimensionality of n. It is also assumed that there are only two classes in the classification problem in a unit space without losing generality. The following type of fuzzy if-then rule is used for the fuzzy classifier in this paper:

$$\begin{aligned} \text{Rule } R_j: \quad & \text{If } x_1 \text{ is } A_{j1} \text{ and } x_2 \text{ is } A_{j2} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \\ & \text{then } y = b_j, \quad j = 1, 2, \ldots, N, \end{aligned} \tag{1}$$

where R_j is the label of the j-th fuzzy if-then rule, $A_j = (A_{j1}, A_{j2}, \ldots, A_{jn})$ is a set of antecedent fuzzy sets in R_j, b_j is the consequent real value, and N is the total number of fuzzy if-then rules in the fuzzy classifier. The number of fuzzy if-then rules is determined once the fuzzy partition for each attribute is determined by a human user. For example, if three fuzzy sets are used for each attribute in a four-dimensional classification problem, then the total number of generated fuzzy if-then rules is $3^4 = 81$. For the antecedent fuzzy sets, any type of membership functions can be used while this paper considers only triangular-type fuzzy membership functions. Figure 1 shows the triangular-type membership functions with various fuzzy partitions (from two to five fuzzy sets for an attribute). The membership functions are homogeneously distributed. That is, the width of the fuzzy sets are equal except the rightmost and leftmost ones, which are half-size of the other membership functions.

Let $\vec{x} = (x_1, x_2, \ldots, x_n)$ be an unknown pattern for a fuzzy classifier to be classified. The classification rule of the unknown pattern is given as follows:

$$\vec{x} = \begin{cases} \text{Class 1, if } \sum_{j=1}^{N} b_j \cdot \mu_j(\vec{x}) > 0, \\ \text{Class 2, otherwise,} \end{cases} \tag{2}$$

where $\mu_j(\vec{x})$ is the compatibility of the j-th fuzzy if-then rule with the unknown pattern, which is calculated using multiplication of membership values as follows:

$$\mu_j(\vec{x}) = \mu_{j1}(x_1) \cdot \mu_{j2}(x_2) \cdot \ldots \cdot \mu_{jn}(x_n), \tag{3}$$

where $\mu_{ji}(x_i)$, $i = 1, 2, \ldots, n$ is the membership value for the i-th attribute x_i.

2 On-Line Learning Techniques

On-line learning techniques in this paper incrementally update the parameters of classifier models. The parameter update occurs every time a new training pattern becomes available. It is assumed that only a single pattern becomes available at a time. Let us define $\vec{x}_t$ as the training pattern that is available at Time t. The class of the training pattern $\vec{x}_t$ is defined as c_t ($c_t \in \{1, 2\}$). It is assumed that only one training pattern becomes available at each time step. This section introduce a linear classifier as the conventional method as well as two on-line learning methods that are applicable to the linear classifier.

2.0.1 Linear Classifier

The class of an n-dimensional pattern $\vec{x} = (x_1, x_2, \ldots, x_n)$ is determined by the following rule:

$$\vec{x} = \begin{cases} \text{Class 1, if } \sum_{i=1}^{n} w_i \cdot x_i + w_0 > 0, \\ \text{Class 2, otherwise,} \end{cases} \tag{4}$$

where $\vec{w} = (w_1, w_2, \ldots, w_n)$ is a weight vector. The weight vector is determined so that the classification error of the linear classifier is minimized in terms of a given set of training patterns. Typically, a least mean squared method is used. However, this is not exactly applicable in this paper as there is no training pattern available for training the linear classifier model. It is also possible that the classification boundary changes over time. There are implicit requirements in the least mean square method that all training patterns are available for training and the classification boundary stays still and never changes.

The assumption in this paper is that not all training patterns are available before training classifier models but gradually made available over time. Thus the classi-fier model requires to perform on-line learning where the model adapts itself every time a new training pattern becomes available.

2.1 Passive-Aggressive Learning

The training of the linear classifier described in Subsection 2.0.1 is generally formulated as an optimization problem where the weight vector $\vec{w} = (w_0, w_1, \ldots, w_n)$ is determined in a way the error between the target class of training patterns and the classification results by the linear classifier is minimized. This formulation is possible if all the training patterns are available at a time. However, this paper considers a different situation where not all training patterns are available but only a limited number of training patterns is available at each time step. Crammer [3] proposed an on-line learning called passive-aggressive learning in order to tackle this special situation of training linear classifiers. The passive-aggressive learning uses the following hinge function for calculating the error between the target class and the classification result:

$$l_{\text{hinge}}(\vec{x}, c, \vec{w}) = \max\{0, 1 - c \times \vec{w} \cdot \vec{x}\}, \tag{5}$$

where $\vec{x}$ and c are a training pattern and its target class, respectively. This hinge loss function gives the degree of errors in terms of the current weight vector $\vec{w}$. In most of the cases, there exist

multiple weight vectors that makes the value of the loss function. The weight vector is modified so that the amount of the update is minimized. Thus the optimization problem at time t is formulated as follows:

$$\vec{w}^{t+1} = \arg\min_{\vec{w}} \frac{1}{2}||\vec{w} - \vec{w}^t||^2 \tag{6}$$

$$\text{subject to } l_{\text{hinge}}(\vec{x}_t, c_t, \vec{w}) = 0 \tag{7}$$

where $\vec{w}^{t+1}$ is the updated weighted vector which will be used at the next time step, and $\vec{w}^t$, $\vec{x}^t$, and c are the current weight vector, the available training pattern and the target class at time t, respectively.

2.2 Confidence-Weighted Learning

The passive-aggressive learning in Subsection 2.1 has shown its powerful on-line performance in classification and is already used in real-world systems. There is, however, a disadvantage in this method that the learning technique is too sensitive to noisy data. In order to overcome this problem, confidence-weighted learning technique was proposed by Dredze and Crammer [1]. This learning technique assumes that the weights in the linear function are not real values, but follows Gaussian probability distributions. In this way, each weight has two parameters: the mean value and the width of a Gaussian function. The mean value shows the probable value of the corresponding weight, and the width shows the confidence in the mean value. When the confidence in the mean is low, the width of the Gaussian function is large while the width is small when the confidence is high. In the update of the Gaussian functions, the mean value will not largely modified when the width is small as the confidence in the mean value is high. On the other hand, when the width of the Gaussian function is large, the mean value will be moved largely because it is not sure if the current mean value is appropriate or not.

Now that the weight vector $\vec{w}$ follows the Gaussian function, it is written as

$$\vec{w} \sim N(\vec{\mu}, \sum), \tag{8}$$

where $\vec{\mu}$ and $\sum$ are the mean vectors and the co-variance matrix, respectively. When a training pattern $\vec{x}_t$ becomes available with its target class c_t for training the linear function at time t, the update of the weight values are performed so that the amount of the modification should be minimized subject to the restriction that the probability of correct classification for the new training is higher than η. This is written as an optimization problem that is formulated as follows:

$$(\vec{\mu}^{t+1}, \sum^{t+1}) = \arg\min_{\vec{\mu}, \sum} D_{KL}(N(\mu, \sum)||N(\vec{\mu}^t, \sum^t)), \tag{9}$$

$$\text{subject to } \text{Prob}_{\vec{w}\ N(\vec{\mu}, \sum^t)}\{c_t \cdot (\vec{w} \cdot \vec{x}_t) \geq 0\} \geq \eta, \tag{10}$$

where D_{KL} is the Kalback-Leibler divergence between the two Gaussian distributions.

2.3 Adaptive Regularization of Weight Vectors

Adaptive regularization of weight vectors (AROW) [4] is an improved version of the confidence weighted learning. This method has the following features:

- It decreases the variance for each vector only gradually.
- It correctly classifies the currently available training pattern as in the confidence weighted learning and passive-aggressive learning.
- It minimizes the variation in the confidence distributions by the update.

Mathematical formulation of the adaptive regularization of weight vectors are written as follows:

$$(\vec{\mu}^{t+1}, \sum^{t+1}) = \arg\min_{\vec{\mu}, \sum} D_{KL}(N(\mu, \sum)||N(\vec{\mu}^t, \sum^t)) + \lambda_1 l(c_t, \vec{\mu} \cdot \vec{x}_t) + \lambda_2 \vec{x}_t^T \sum_t \vec{x}_t, \tag{11}$$

$$l(c_t, \vec{\mu} \cdot \vec{x}_t) = (\max 0, 1 - c_t(\vec{\mu} \cdot \vec{x}_t)^2. \tag{12}$$

As it can be seen from (11), each term in the formulation corresponds to the three features that were mentioned above.

2.4 Applying the On-Line Learning Techniques to Fuzzy Classifiers

The on-line learning techniques in Subsections 2.1 and 2.2 can be applied to fuzzy classifiers. Equations (2) and (4) are the classification rule for an input pattern by a fuzzy classifier and a linear classifier, respectively. They can be seen as the same if the membership values of fuzzy if-then rules is treated as an input vector in a mapped space. That is, the fuzzy classifier performing a linear classification in the mapped space by the fuzzy if-then rules. In the mapped space, $\vec{\mu(\cdot)}$ is the input vector and $\vec{b} = (b_1, b_2, \ldots, b_N)$ can be viewed as the weight vector in the linear classification. In this perspective, the fuzzy classifier can be updated in an on-line manner using either the passive-aggressive learning or the confidence-weighted learning.

If the passive-aggressive learning is applied to the fuzzy classifier, the vector of the consequent real values $\vec{b}$ is updated by solving the following optimization problem:

$$\vec{b}^{t+1} = \arg\min_{\vec{b}} \frac{1}{2}||\vec{b} - \vec{b}^t||^2 \tag{13}$$

$$\text{subject to } l_{\text{hinge}}(\vec{\mu}(\vec{x}_t), c_t, \vec{b}) = 0 \tag{14}$$

where

$$l_{\text{hinge}}(\vec{\mu}(\vec{x}_t), c_t, \vec{b}) = \max\left\{0, 1 - c_t \times \vec{b} \cdot \mu(\vec{x}_t)\right\} \tag{15}$$

.

In the case of the confidence-weighted learning, the consequent real values of fuzzy if-then rules are assumed to follow Gaussian functions with the mean vector $\vec{\mu}_b$ and the co-variance matrix $\sum_b$. Then the parameters are modified by solving the following optimization problem:

$$(\mu_b^{t+1}, \sum_b^{t+1}) = \arg \min_{\vec{\mu}_b, \sum_b} D_{KL}(N(\vec{\mu}_b, \sum_b)||N(\vec{\mu}_b^t, \sum_b^t)) \tag{16}$$

$$\text{subject to } \text{Prob}_{\vec{b}\ N(\vec{\mu}_b^t, \sum_b^t)}\left\{c_t \cdot (\vec{b} \cdot \mu(\vec{x}_t)) \geq 0\right\} \geq \eta. \tag{17}$$

3 Noisy and Dynamic Classification Problems

3.1 Generating noisy training patterns

Without the loss of generality, let us consider a two-class classification problem in this paper. The classification problem in the computational experiments in this paper is a noisy and dynamic problem. The noise is added in the generation process of training patterns. First, a two-dimensional vector is randomly generated in the unit space $[0, 1]^2$. Secondly, the class of the generated vector is decided from the current classification boundary. Then, noise is added to the generated vector. The modified vector is only employed as a training pattern of the classifiers if the modifid vector lies inside the unit two-dimensional space.

3.2 Dynamic Classification Boundary

The classification problems in the computational experiments of this paper is dynamic in the sense that the true classification boundary is not static but keep changing over time. The classification boundary rotates around some point. This kind of information is not used as a prior knowledge when a classifier is constructed. The only available information is the training pattern generated by the procedure described in Subsection 3.1. Only a single training pattern is generated at a time step. We consider a linear and a non-linear classification boundary in the computational experiments in this paper as follows:

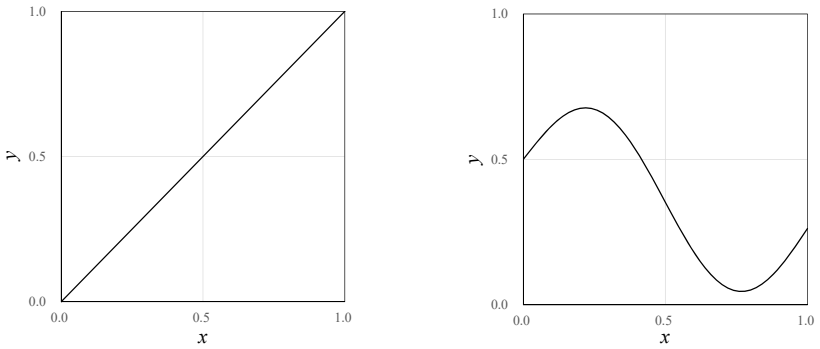
[Linear classification boundary]

$$y = x. \tag{18}$$

[Non-linear classification boundary]

$$y = \frac{1}{2}\sin(2\pi \cdot x \cdot 0.4) \cdot \cos(2\pi \cdot x \cdot 0.6) + \frac{1}{2}. \tag{19}$$

Note that the above equations are at time step 0 and will rotate around a prespecified point by the degree of one. The boundaries at time step 0 are shown in Fig. 2.



(a) Linear boundary (b) Non-linear boundary

Figure 2: Classification boundaries at time step 0. (a) the linear case and (b) the non-linear case.

4 Experiments

4.1 Experimental Settings

We consider two problems in the computational experiments of this paper. One is a static on-line problem where from a set of training patterns, randomly selected one is used to train the classification models at one time. The other is its dynamic version where the classification boundary changes over time.

4.2 Experimental Results

The computational experiments involve various on-line learning for fuzzy systems such as the confidence-weighted learning, passive-aggressive learning, delta learning, and the adaptive regularization of weight vectors. In this section we only show the results for the classification problem with the linear classification boundary in Fig. 2(a). First we show the experimental results for the static problem in Fig. 3.

From Fig. 3, it is shown that the AROW method quickly adapts the fuzzy classifier and keeps high classification rates throughout the experiments. Next, we show the experimental results for the dynamic problem in Fig. 4.

From Fig. 4, it is shown that the classification accuracy by the AROW is not stable at all but goes ups and downs as the classification boundaries changes over time. This shows the poor classification performance of AROW for dynamic problems.

5 Conclusions

This paper investigates the performance of AROW method for the on-line learning of fuzzy classifiers. Noticing that the fuzzy classifier performs a linear classification in a mapped space, we can re-formulate the on-line learning of the fuzzy classifiers as the optimization of consequent real vectors in the fuzzy if-then rules. AROW method is the target of the investigation. The experimental results show that AROW performs almost equally as the other on-line methods for static problem.

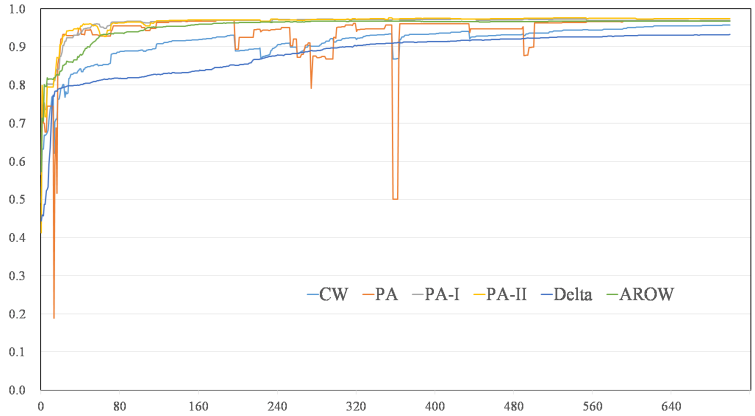


Figure 3: Experimental results for the static problem.

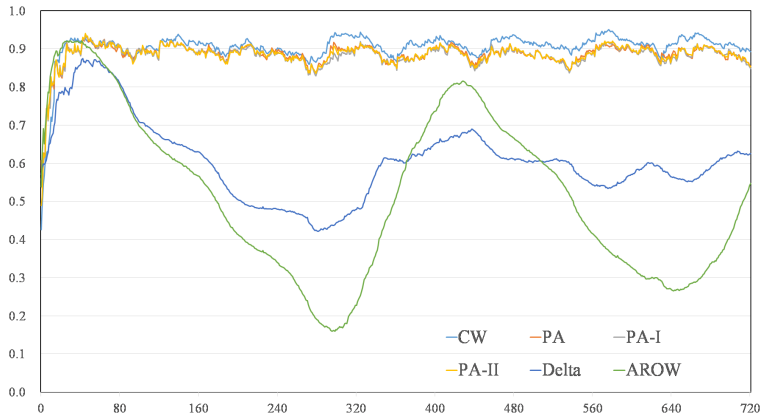


Figure 4: Experimental results for the dynamic problem.

However, for a dynamic problem it is not the case as AROW is too slow to adapt the weight vector to the changing environment.

References

[1] Dredze, M. and Crammer, K.: Confidence-Weighted Linear Classification. Proc. of the 25th international conference on Machine learning, 264–271 (2008)

[2] Crammer, K., Dekel, O., Keshet, J., S-Shwartz, S., and Singer Y.: Online Passive-Aggressive Algorithms. Journal of Machine Learning Research 7, 551–585 (2006)

[3] Ishibuchi, H., Nakashima, T., and Nii, M.: Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining, Springer (2004)

[4] Crammer, K., Kulesza, A., and Dredze, M.: Adaptive Regularization of Weight Vectors. Machine Learning 7 (2), 155–187 (2013)

Triangular Norms in Max-t fuzzy algebras

Zuzana NĚMCOVÁ

University of Hradec Králové
Rokitanského 62, Hradec Králové
Czech Republic
zuzana.nemcova@uhk.cz

Abstract

Many issues in the subject areas of optimization, geometry, control theory, machine scheduling, discrete event processes, manufacturing systems, telecommunication networks, and traffic control can be solved by using various types of extremal algebras, according to the character of the study. Extremal algebras can help to understand the development of the system dynamics. The attractiveness of these tools is grounded in the fact that the classes of the above mentioned nonlinear problems can be studied by applying a linear-algebraic approach. The paper gives an overview and basic characteristics of a triangular norms (briefly t-norms) used in the so called max-t fuzzy algebras to study steady states of discrete event systems. Among four main t-norms belong Łukasiewicz, Gödel, Product and Drastic norm. Also an overview of basic literature covering this domain is given in the text.

Keywords: T-norm, Extremal algebra, Max-t fuzzy algebra.

1 Extremal Algebras

Extremal algebras are used mainly for describing and studying systems working in discrete time. During the operation of such a system, a stable state can arise. The system is described by a transition matrix and the eigenvectors of such a matrix represent the stable states of the system.

With the use of extremal algebras, we are able to get various answers about system characteristics. Some examples of these questions follow.

- What is the maximum speed at which the system could run?
- What is the latest time to start so that the production meets the delivery dates?
- If all machines start working at time zero, at what time will each machine finish work on the project?
- Can the system repeatedly reach some sequence of states?
- How great a delay and to which events can be tolerated without prejudice to the calculated finishing time of the project?
- How to choose a sequence of nodes when travelling through some network, so that we visit as many nodes as possible before the given supplies are exhausted?

The term *extremal algebra* was used for the first time by [24]. This concept occurs under different names in the literature, for example, tropical algebra, max-algebra, or fuzzy algebra. It is a special algebraic structure equipped with two operations $\oplus$ and $\otimes$. The operation $\oplus$ is not addition as in usual algebra, but is one of the operations of extremum: maximum or minimum. The operation of $\otimes$ is a chosen group or semigroup operation so that these two operations can be commutative and associative. The distributive law holds, and there exist neutral elements with respect to both operations [26]. For example, if one chooses $\oplus = \max$ and $\otimes = \min$, one speaks of a max-min algebra; if one chooses $\oplus = \max$ and $\otimes = +$, one speaks of max-plus algebra.

The operations $\oplus$, $\otimes$ are then extended in a natural way to Cartesian products of vectors and matrices. The detailed mathematical investigation at a general level can be found in [3].

2 Max-t Fuzzy Algebras

The discrete time systems studied with the use of max-t fuzzy algebras can be sometimes intuitively called fuzzy systems. But this can be confusing in some ways because the whole study of the system in this way is not about studying "classical" fuzzy systems, which means constructing the function of membership, fuzzification, inference and defuzzification. The main focus lies in the description of the development of the system's states in time. Both the states described by vectors and the transition matrix have fuzzy numbers as entries. And this may be the reason why these systems are sometimes called fuzzy. Naturally, the terminology is continually developing. For the transition from one state to another, various operations for the conjunction of the sets are used. And these operations are also much used for studying "classical" fuzzy systems.

Fuzzy logic appeared as a necessary extension of Zadeh's theory of fuzzy sets [25]. The development of fuzzy logic over the last two decades has been driven by the applications in artificial intelligence and decision making processes.

The standard membership degrees of fuzzy sets (values from the unit interval), can be understood in a natural way as the truth degrees of an infinite valued logic, with $[0, 1]$ as its truth degree set.

According to [10], fuzzy algebra claims that each object, say x, of the universe, denoted by X, belongs to a fuzzy set A with respect to a characteristic real number $\mu_A(x)$ called the degree of membership. In other words, it is the truth degree of the formula $x \in A$.

With this understanding, the intersection of the fuzzy sets A and B, i.e., $A \cap B$, is $\mu_{A\cap B}(x)$, and this corresponds to the truth degree of the statement $x \in A$ AND $x \in B$. The word AND has the meaning of a conjunction connective. It can be claimed that the theory of crisp sets becomes presented in the language of classical logic. Already the first paper of Zadeh [25] offers different proposals for understanding the above mentioned operation of conjunction, AND, as taking the minimum (or as taking a usual, algebraic product). Similarly, for the operation of union $A \cup B$, Zadeh suggests taking the maximum of the membership degrees in A, B (or their algebraic sum). These relationships have been widely used in theoretical considerations about the foundations of fuzzy set theory. In particular, in the last decade, there has been initiated an enormous development in the field of fuzzy logic in the narrow sense. [7]

Max-t fuzzy algebra is defined over the interval $[0, 1]$, and, as the name suggests, has two binary operations: the operation of maximum (denoted by $\oplus$) and one of the t-operations, (denoted by $\otimes$) are used. These operations are extended to vectors and matrices in a formal way. According to the chosen t-operation, we can then talk, for example, about max-Łukasiewicz, max-min, max-prod or max-drast fuzzy algebra.

3 Triangular Norms

Triangular norms (briefly t-norms) were first introduced by Karl Menger in 1942 in his paper [18]. The main idea was to construct metric spaces where probability distributions (instead of numbers) describe the distance between two elements of the space. [12]

Schweizer and Sklar in [23] integrated t-norms into the context of probabilistic metric spaces. These t-norms are the operations used to interpret the conjunction in fuzzy logics and the intersection of fuzzy sets [13]. These functions have applications in many areas, such as decision making processes, statistics, game theory, information and data fusion, probability theory, and risk management. The t-norms, together with t-conorms, play a key role in fuzzy set theory. Interesting results with practical impact using t-norms in the calculations can be found for example in [4, 5, 15].

The t-norm is an operation of conjunction, denoted by $\triangle$, satisfying [11]:

$$\begin{aligned} 1\triangle x &= x \\ x\triangle y &= y\triangle x \\ x\triangle(y\triangle z) &= (x\triangle y)\triangle z \\ \text{If } w &\le x \text{ and } y \le z \text{ then } w\triangle y \le x\triangle z. \end{aligned}$$

In other words, the operation is associative, commutative, nondecreasing, and has 1 as a neutral element.

Remark 3.1. In fact, any operation that agrees with the truth table for two-valued conjunction is a candidate for a t-norm and that is why the definition of fuzzy logic operations may vary from application to application. [11]

Although there exist many families of t-norms (for example, the Aczel–Alsina, the Jane Doe1, Hamacher, Dombi, Yager, Einstein product, for an overview see [6, 12]), four main t-norms will be described in the following text: Łukasiewicz, Gödel, Product and Drastic.

How to choose a suitable t-norm is a good question. The answer is not obvious. Each t-norm offers a good approach and sees things from a different perspective. In general, the t-norms can not be linearly ordered. According to [10], some t-norms can be characterized as optimistic, with others as pessimistic. Others assign a case a high degree of membership when one or more parameters have extreme values.

To fulfill domain-context demands, various applications require various t-norms, and sometimes it is necessary to create a new one. In [20], the following three basic methods of forming a new t-norm are described in detail:

- construction of a new t-norm on the basis of some given t-norms with the use of a monotonic transformation,
- the use of the addition and multiplication of real numbers together with a function of one variable, called additive and multiplicative generators,
- forming a new t-norm from a family of given t-norms based on the concept of the "ordinal sum."

3.1 Łukasiewicz Norm

The main motivation for the development of logic was to create a logical system that is free of some of the paradoxes in two-valued Boolean logic. The revolutionary step was made by Jan Łukasiewicz with his introduction of three-valued logic, and later on also multi-valued logic [16, 17].

Both Łukasiewicz implication and negation were designed by Jan Łukasiewicz in 1930. A complete description of the derivation of Łukasiewicz conjunction can be also found in [7].

As mentioned above, the Łukasiewicz logic has the two following primitive connectives, implication and negation [7]:

$$x \to_L y = \min(1, 1 - x + y) \tag{1}$$

$$\neg x = 1 - x \tag{2}$$

From these primitive properties, further connectives are defined as syntactic abbreviations:

$$\varphi \,\&\, \psi = \neg(\varphi \to_L \neg\psi), \qquad \varphi \veebar \psi = \neg\varphi \to_L \psi \tag{3}$$

With the above defined properties, one gets a conjunction and a strong disjunction with truth degree functions that are usually called Łukasiewicz arithmetical conjunction and Łukasiewicz arithmetical disjunction:

$$x \,\&\, y = \max\{x + y - 1, 0\}, \qquad x \veebar y = \min\{u + v, 1\}. \tag{4}$$

Then, in the case of conjunction,

$$x \otimes_L y = \max\{x + y - 1, 0\}. \tag{5}$$

Łukasiewicz implication has applications in many fields. Together with the negation of Łukasiewicz implication, it was used by Mills in [19] to design sensors for small robots. Łukasiewicz implication is also used in [14] as a measure of the entropy of the fuzziness of messages, to develop fuzzy information theory. The use of the fuzzy conjunction in digital hardware implementation is formulated in [22].

The fuzzy extension of description logics using Łukasiewicz logic (the formalism for the representation of structured knowledge) is described in [1]. Description logic is frequently used in the design of ontologies, which have been successfully used as part of expert and multiagent systems, as a knowledge base in robotics, as well as the core element in the Semantic Web (which aims at converting the current Web into a "Web of data" by defining the meaning of information).

Łukasiewicz arithmetical conjunction can also be used in many types of situations in real life. The fact that the number 1 is subtracted and the maximum with zero is taken (in case that $x + y < 1$) leads to the observation that the count of the operation is the remainder, some part that is over the unit. Following this idea, the conjunction can be used to compute the amount of money that should be paid off for a phone bill, where x can be the price for the total SMSs, and y can represent the price for the calls in total. The number 1 here is replaced by the amount of the lump sum. There are other similar situations, for example, data backup on a server, the maximal capacity of a pond, overspending of project funds, or savings of partners with a common bill.

3.2 Gödel Norm

This is the simplest norm, in the literature known also as the "minimum norm" or the "strongest norm"; the conjunction is defined as the minimum of the entries: of the truth degrees of the constituents [6].

$$x \otimes_G y = min(x, y) \tag{6}$$

In contrast to Łukasiewicz logic (which might be considered as a logic of absolute or metric comparison), Gödel logic is a logic with a relative comparison.

3.3 Product Norm

The definition of the product norm follows, a complete description of this norm can be found in [9].

$$x \otimes_p y = x \cdot y \tag{7}$$

In [21] there can be found the application and comparison of the Gödel and Product norms in a classification problem—the ordering of some finite set of evaluations according to certain criteria. The product norm corresponds here to the stochastic or probability independence of elementary events (propositional letters), while the Gödel norm represents another kind of extreme situation: the logical dependence of propositional letters.

3.4 Drastic Norm

The drastic triangular norm (in the literature there can be also found the term "weakest norm" or "drastic product") is the basic example of a non-divisible t-norm on any partially ordered set, see [2]. The drastic triangular norm (the operation of $\otimes_d$) is defined as follows:

$$x \otimes_{\mathrm{d}} y = \begin{cases} \min(x, y) & \text{if } \max(x, y) = 1 \\ 0 & \text{if } \max(x, y) < 1 \end{cases} \tag{8}$$

The drastic norm is "drastic" in the sense that when there is some possibility of using, for example, two weighted paths, say path x and path y, they can be used if and only if one of them is equal to 1. Otherwise such a concatenated path can not be used. Moreover, if the walk consists of more than two paths, then at most one path can be weighted by a value less than 1. The interpretation of the drastic norm then reflects a situation when extreme demands are imposed on the reliability of a system: the walk is reliable if there is no more than one unsure segment.

The matrix powers in max-drast algebra behave differently than those in max-min algebra. However, the properties of the matrix periods in both cases show some similarity.

The differences between the four basic t-norms can be seen in Figure 2. Although the figure depicts the intersection of the fuzzy sets in the classical conception, the plots show the difference between the operations clearly. The contour plots for the basic t-norms can be seen in Figure 1 (the contour plot shows us the sets where the function in question has constant values). From the figure it can also be seen that the Gödel, Product and Łukasiewicz norms are continuous whereas the drastic norm is not. According to [20], the continuity of the operations on fuzzy sets is a highly desirable feature in many applications.

The norms can be also compared, for details see [12], since they are just functions. The comparison is done pointwise. If for two t-norms, say $\triangle_1$ and $\triangle_2$, it holds that

$$x \triangle_1 y \leq x \triangle_2 y$$

for all $x, y \in [0, 1]$, then $\triangle_1$ is weaker than $\triangle_2$ (or, equivalently, it can be said, $\triangle_2$ is stronger than $\triangle_1$), and thus one can write $\triangle_1 \leq \triangle_2$. This yields the following order of the four basic t-norms:

$$\text{Drastic} < \text{Łukasiewicz} < \text{Product} < \text{Gödel}.$$

It can also be observed from the definitions of the operations for the t-norms that by using the product or Łukasiewicz norm, new elements (values) are generated, whereas for the Gödel and drastic norm this is not true. The drastic norm is also characterized by the fact that small changes in the inputs produce big changes in the output. This results from the character of the operations.

The characteristic of idempotency can be formulated as follows: the value x is idempotent if $x \triangle x = x$. Note that the values 0 and 1 are idempotent elements for any t-norm. The t-norm is Archimedean (i.e., for any two elements a and b from the set there exists $n \in \mathbb{N}$ such that $n \cdot a > b$) if and only if its only idempotent elements are the trivial ones 0 and 1. [2]

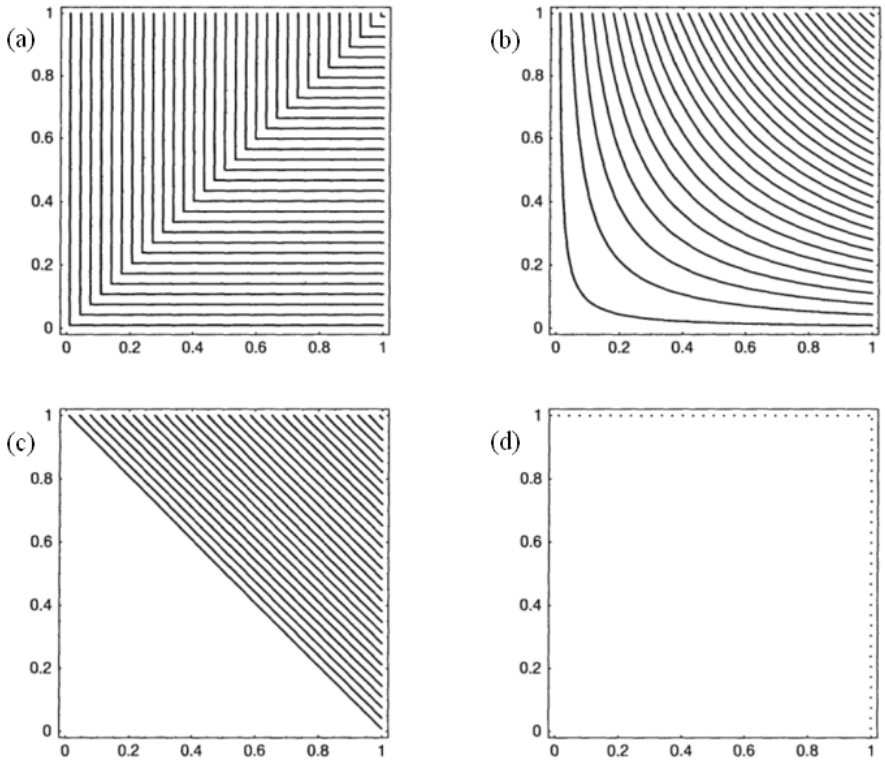


Figure 1: Contour plot of the t-norms: (a) Gödel, (b) Product, (c) Łukasiewicz, (d) Drastic. [12]

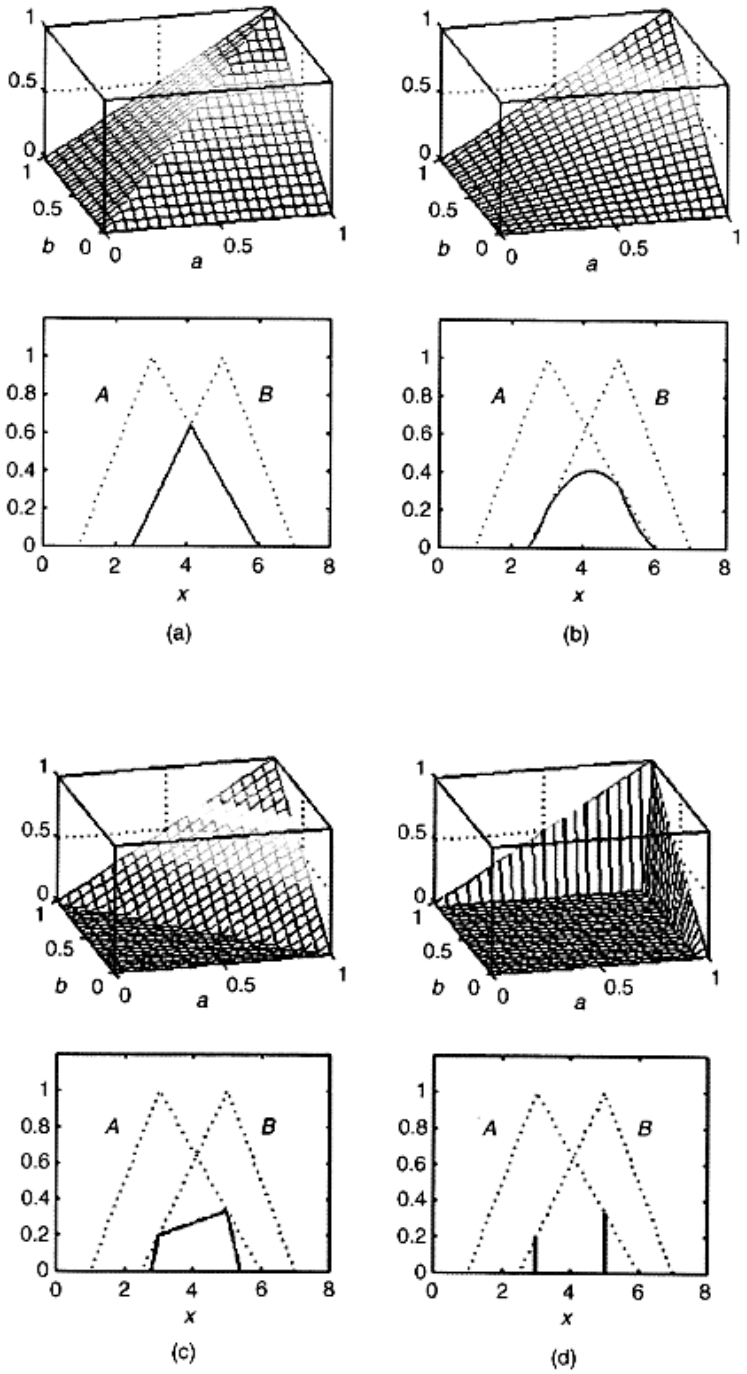


Figure 2: T-norms used in the realization of the intersection of fuzzy sets A and B: (a) Gödel, (b) Product, (c) Łukasiewicz, (d) Drastic. [20]

4 Triangular Conorms

A brief characterization of the *triangular conorms* (t -conorms for short) will be given next. A t-conorm is an operation of disjunction, the dual operation to that of a t-norm, on the unit interval,

denoted by ∇, and satisfying [11]

$$\begin{aligned} 0\nabla x &= x \\ x\nabla y &= y\nabla x \\ x\nabla(y\nabla z) &= (x\nabla y)\nabla z \\ \text{If } w &\leq x \text{ and } y \leq z \text{ then } w\nabla y \leq x\nabla z. \end{aligned}$$

A t-conorm can be generated from a t-norm by

$$x\nabla y = 1 - (x+1)\triangle(y+1) \tag{9}$$

Among the basic t-conorms there are [12]

$$\begin{aligned} &\text{Maximum:} && x \oplus_G y = \max(x,y) \\ &\text{Probabilistic sum:} && x \oplus_p y = x + y - x \cdot y \\ &\text{Łukasiewicz sum:} && x \oplus_L y = \min(x+y,1) \\ &\text{Drastic sum:} && x \oplus_{\mathrm{d}} y = \begin{cases} 1 & \text{if } \min(x,y) > 0 \\ \max(x,y) & \text{if } \min(x,y) = 0 \end{cases} \end{aligned}$$

It can be easily seen from the definition of a conorm that the difference between norms and conorms lies in their boundary conditions. The name, t-conorm, comes from the fact that in the unit interval, the value x acts in a similar way as its complement $1 - x$. [12]

Similarly to triangular norms, the conorms can be visualized, see Figure 4 for 3D and Figure 3 for contour plots.

Remark 4.1. Since both the operations for t-norms and t-conorms were introduced as operations with two arguments, they can be, of course, extended to n-ary operations in the usual way by induction.

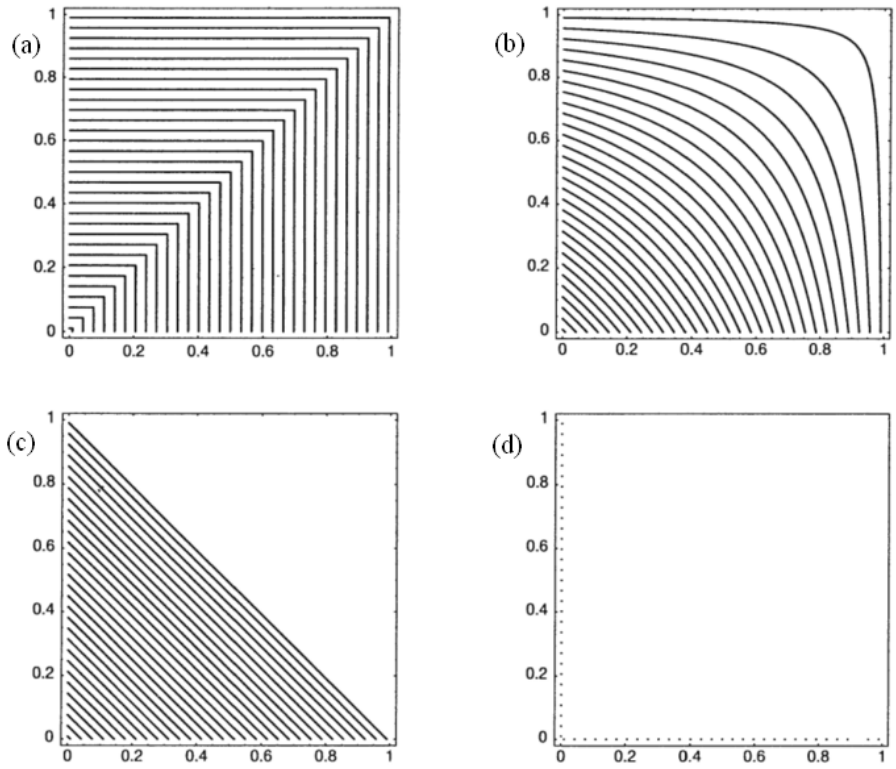


Figure 3: Contour plot of the t-conorms: (a) Maximum, (b) Product, (c) Łukasiewicz, (d) Drastic. [12]

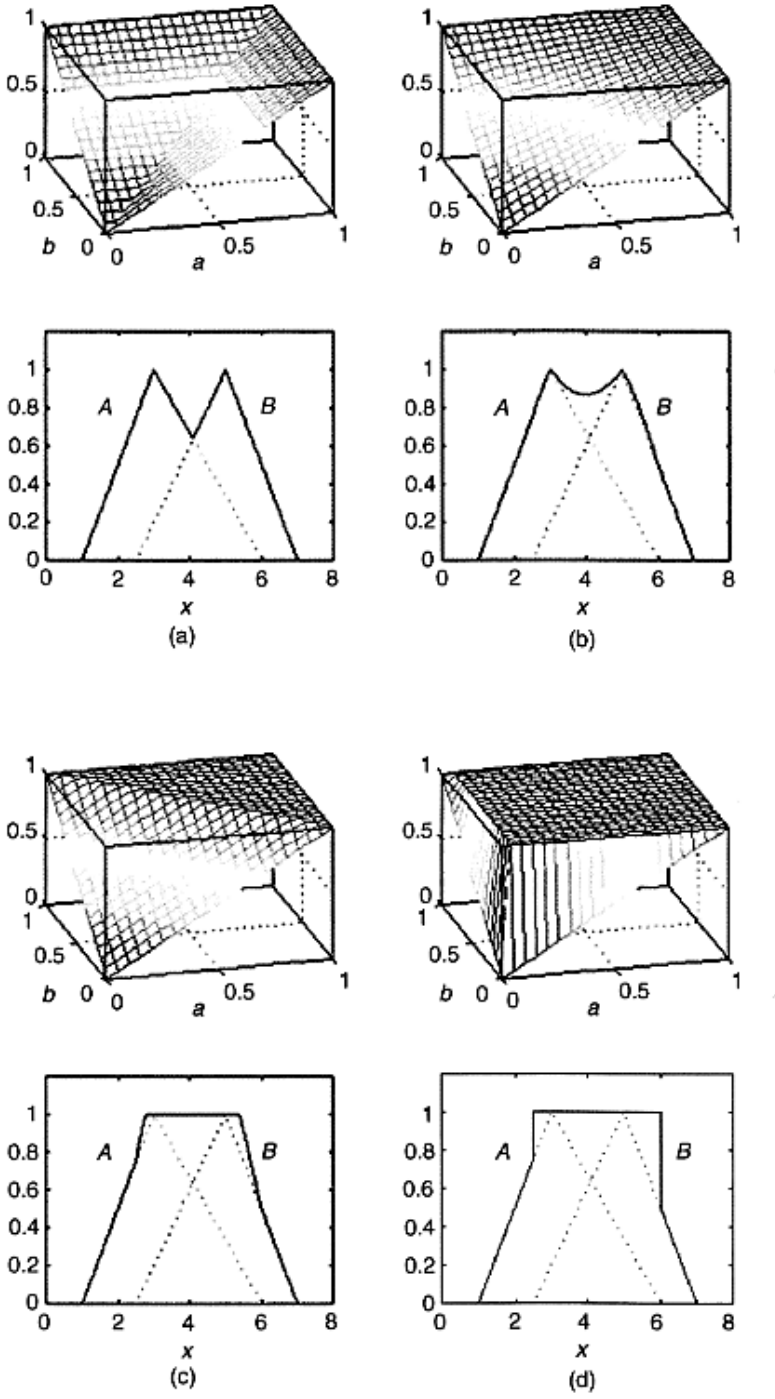


Figure 4: T-conorms used in the realization of the intersection of fuzzy sets A and B: ((a) Maximum, (b) Product, (c) Łukasiewicz, (d) Drastic. [20]

Acknowledgement
The financial support of the Excellence Project at FIM UHK is gratefully acknowledged.

References

[1] Bobillo, F. and Straccia, U. (2011). Reasoning with the finitely many-valued Łukasiewicz fuzzy description logic. *Information Sciences*, 181(4):758–778.

[2] Casasnovas, J. and Mayor, G. (2008). Discrete t-norms and operations on extended multisets. *Fuzzy Sets and Systems*, 159(10):1165–1177.

[3] Cuninghame-Green, R. A. (1979). *Minimax Algebra. Lecture Notes in Economics and Mathematical Systems* vol. 166. Springer-Verlag, Berlin.

[4] Chaira, T. (2015). Fuzzy mathematical morphology using triangular operators and its application to images. *Journal of Intelligent and Fuzzy Systems*, 28(5):2269–2277.

[5] Chen, X.-G. and Du, H. (2015). Interval-valued triangular fuzzy soft set and its method of dynamic decision making. *Xi Tong Gong Cheng Yu Dian Zi Ji Shu/Systems Engineering and Electronics*, 37(5):1111–1115.

[6] Gottwald, S. (2001). *A Treatise on Many-Valued Logics*, volume 3. Research Studies Press Baldock.

[7] Gottwald, S. (2005). Mathematical fuzzy logic as a tool for the treatment of vague information. *Information Sciences*, 172(1):41–71.

[8] Goverde, R. (2010). A delay propagation algorithm for large-scale railway traffic networks. *Transportation Research Part C: Emerging Technologies*, 18(3):269–287.

[9] Hájek, P., Godo, L., and Esteva, F. (1996). A complete many-valued logic with product-conjunction. *Archive for Mathematical Logic*, 35(3):191–208.

[10] Iliadis, L., Spartalis, S., and Tachos, S. (2008). Application of fuzzy t-norms towards a new artificial neural networks evaluation framework: A case from wood industry. *Information Sciences*, 178(20):3828–3839. Special Issue on Industrial Applications of Neural Networks 10th Engineering Applications of Neural Networks 2007.

[11] Jantzen, J. (2007). *Foundations of Fuzzy Control*. John Wiley & Sons.

[12] Klement, E. P., Mesiar, R., and Pap, E. (2000). *Triangular Norms*. Springer.

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

[14] Kosko, B. (1986). Fuzzy entropy and conditioning. *Information sciences*, 40(2):165–174.

[15] Liang, C., Zhao, S., and Zhang, J. (2014). Aggregation operators on triangular intuitionistic fuzzy numbers and its application to multi-criteria decision making problems. *Foundations of Computing and Decision Sciences*, 39(3):189–208.

[16] Łukasiewicz, J. (1930). Philosophische bemerkungen zu mehrwertigen systemen des aussagenkalk. *C. R. Soc. Sci. Lett. Varsovie*, 23:51–77.

[17] Łukasiewicz, J. (1970). *Studies in Logic and the Foundations of Mathematics*. North-Holland.

[18] Menger, K. (1942). Statistical metrics. *Proceedings of the National Academy of Sciences*, 28(12):535–537.

[19] Mills, J. W. (1993). Lukasiewicz' insect: The role of continuous-valued logic in a mobile robot's sensors, control, and locomotion. In *Proceedings of The Twenty-Third International Symposium on Multiple-Valued Logic, 1993*, pages 258–263. IEEE.

[20] Pedrycz, W. and Gomide, F. (2007). *Fuzzy Systems Engineering: Toward Human-centric Computing*. John Wiley & Sons.

[21] Perovic, A., Ognjanovic, Z., Raskovic, M., and Radojevic, D. (2011). Finitely additive probability measures on classical propositional formulas definable by Gödel's t-norm and product t-norm. *Fuzzy Sets and Systems*, 169(1):65–90.

[22] Rudas, I. J., Batyrshin, I. Z., Zavala, A. H., Nieto, O. C., Horváth, L., and Vargas, L. V. (2008). Generators of fuzzy operations for hardware implementation of fuzzy systems. In *MICAI 2008: Advances in Artificial Intelligence*, pages 710–719. Springer.

[23] Schweizer, B., Sklar, A., et al. (1960). Statistical metric spaces. *Pacific J. Math*, 10(19601):313–335.

[24] Vorobjov, N. N. (1967). Extremal Algebra of positive Matrices. *Datenverarbeitung und Kybernetik*, 3:37–71.

[25] Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8:338–353.

[26] Zimmermann, K. (2003). Disjunctive optimization, max-separable problems and extremal algebras. *Theoretical Computer Science*, 293(1):45–54.

Bivariate Fuzzy Transform Based on Tensor Product of Two Polynomial Spaces in Analysis of Correlation Functions

Linh NGUYEN and Michal HOLČAPEK

University of Ostrava
Institute for Research and Applications of Fuzzy Modelling, NSC IT4Innovations
30. dubna 22, 701 03 Ostrava 1, Czech Republic
{Linh.Nguyen, Michal.Holcapek}@osu.cz

Abstract

This paper introduces a new type of fuzzy transformation of higher degree applied to bivariate complex-valued functions, which is based on the tensor product of two polynomial spaces. We demonstrate that the new type of fuzzy transform is efficient in the study of random processes, namely, in the analysis of correlation functions of random processes after the application of higher degree fuzzy transform.

Keywords: Fuzzy transform, tensor product space, functional approximation, correlation function.

1 Introduction

The theory of fuzzy transform (F-transform) has been introduce by Perfilieva in the seminal paper [1] devoted to a new type of approximation of real-valued functions with a motivation in fuzzy modeling (Sugeno–Takagi models). The F-transform consists of two phases, namely, the direct and inverse F-transform. A generalization of the F-transform towards higher degrees (denoted as an F^m-transform for $m \in \mathbb{N}$) has been proposed in [2], where the original F-transform is of degree zero. The generalization uses the linear spaces of polynomials of degrees at most equal to m as its approximation spaces that are, in fact, the subspaces of weighted Hilbert spaces whose weighted functions together form a fuzzy partition of the real line or usually a closed subinterval. The direct F^m-transform transforms a given (real-valued or complex-valued) function into a vector of polynomials that are the orthogonal projections of the given function into the approximation spaces. The inverse F^m-transform provides an approximation of the original function made from its direct F^m-transform components. Among the most important advantages of F^m-transform belongs: the low computational cost, the very good compression ability, or the high efficiency in the noise reduction. Therefore, it is not surprising that the F^m-transform has been successfully applied in many fields of computer science as well as applied mathematics [3, 4, 5, 6, 7, 8, 9, 10].

The theory of fuzzy transform, particularly the F^m-transform, has been generalized to bivariate (multivariate) case in [11, 12, 13, 14]. The theoretical results on the bivariate F^m-transform gave rise efficient methods that have been successfully applied to image processing like the edge detection or the image reconstruction [11, 15].

Recently, we investigated the F^m-transform to analyze stationary random processes. The preliminary results can be found in [16]. Due to our analysis of the correlation function of a random process obtained by the inverse F^m-transform (an approximate random process), we realized that there is an interesting relationship between the correlation functions of the original random process and the approximate random process, which can be nicely described by a new type of bivariate higher degree fuzzy transform, where the standard approximation space of bivariate polynomials is replaced by the tensor product of two spaces of univariate polynomials. The aim of this paper is to introduced the "new" type of bivariate fuzzy transform including a computation method for the derivation of its components (bivariate polynomials) and demonstrate its usefulness in the investigation of the correlation function of random processes.

The paper is structured as follows. The next section contains a necessary background for the introduction of the bivariate F^m-transform based on the tensor product of two polynomial spaces. The third section provides the definition and basic properties. The fourth section is devoted to the analysis of the correlation function of a random process after the application of F^m-transform.

2 Preliminaries

2.1 Fuzzy partition

Fuzzy transform technique is based on a fuzzy partition of the real line (n-dimensional space) or a closed interval (n-dimensional cuboid). In this paper, we consider a generalized uniform fuzzy partition of the real line $\mathbb{R}$ and the plane $\mathbb{R}^2$. The both uniform fuzzy partitions are determined from generating functions defined as follows.

Definition 1. *A real-valued function $K : \mathbb{R} \to [0,1]$ is said to be a* generating function *if it is continuous, even, non-increasing on $[0,1)$ and vanishing outside of $(-1,1)$.*

The following example provides two generating functions which are frequently used in applications of the F-transform.

Example 1. *The functions $K^{tr}, K^{rc} : \mathbb{R} \to [0,1]$ defined by*

$$K^{tr}(t) = \max(1 - |t|, 0) \tag{1}$$

$$K^{rc}(t) = \begin{cases} \frac{1}{2}(1 + \cos(\pi t)), & -1 \le t \le 1; \\ 0, & \textit{otherwise,} \end{cases} \tag{2}$$

for any $t \in \mathbb{R}$, are called the triangle *and* raised cosine *generating functions, respectively.*

In [17], a generalized uniform fuzzy partition of the real line is introduced as follows. We use $\mathbb{R}_+$ and $\mathbb{Z}$ to denote the set of positive real numbers and integers, respectively.

Definition 2. *Let K be a generating function, let $x_0 \in \mathbb{R}$, and let $h, k \in \mathbb{R}_+$. A family $\mathcal{A} = \{A_k \mid k \in \mathbb{Z}\}$ of fuzzy sets on $\mathbb{R}$ determined by*

$$A_k(x) = K\left(\frac{x - x_0 - kr}{h}\right) \tag{3}$$

is said to be a generalized uniform fuzzy partition of $\mathbb{R}$ determined by the quadruple (K, h, r, x_0) *if the following condition is fulfilled*

$$\sum_{k \in \mathbb{Z}} A_k(x) = 1, \quad x \in \mathbb{R}. \tag{4}$$

*The fuzzy set A_k is called the k-*th basic function *of the generalized uniform fuzzy partition $\mathcal{A}$. The parameters h, r and x_0 are called the* bandwidth, shift *and* central node*.*

Note that condition (4) is called the Ruspiny condition or the partition of unity. For the sake of simplicity, we omit the term "generalized" in the expression "generalized uniform fuzzy partition".

For a bivariate fuzzy transform, we extend the previous definition to a uniform fuzzy partition of the plane, where the bivariate generating function is defined as the product of two univariate generating functions. A more general definition of bivariate (multivariate) uniform fuzzy partition can be found in [18].

Definition 3. *Let K_1, K_2 be generating functions. A function $\mathcal{K} : \mathbb{R}^2 \to [0,1]$ defined by*

$$\mathcal{K}(x, y) = K_1(x) \cdot K_2(y), \quad (x, y) \in \mathbb{R}^2,$$

is called the bivariate generating function determined by K_1 and K_2.

A uniform fuzzy partition of the plane is defined in the similar way as the uniform fuzzy partition of the real line. For the sake of simplicity, we omit a reference to uniform generating functions that determine a bivariate generating function if we do not need to specify them or no confusion can appear.

Definition 4. *Let $\mathcal{K}$ be a bivariate generating function, let $(x_0, y_0) \in \mathbb{R}^2$, and let $(h_1, h_2), (r_1, r_2) \in \mathbb{R}^2_+$. A family $\mathcal{A} = \{A_{k_1,k_2} \mid k_1, k_2 \in \mathbb{Z}\}$ of fuzzy relations on $\mathbb{R}^2$ determined by*

$$A_{k_1,k_2}(x, y) = \mathcal{K}\left(\frac{x - x_0 - k_1 r_1}{h_1}, \frac{y - y_0 - k_2 r_2}{h_2}\right),$$

is said to be a uniform fuzzy partition of $\mathbb{R}^2$ *determined by the quadruple $(\mathcal{K}, (h_1, h_2), (r_1, r_2), (x_0, y_0))$ if the following condition is fulfilled:*

$$\sum_{k_1,k_2 \in \mathbb{Z}} A_{k_1,k_2}(x, y) = 1, \quad \text{for any } (x, y) \in \mathbb{R}^2.$$

The fuzzy relation A_{k_1,k_2} is called the (k_1, k_2)-th basic function *of the uniform fuzzy partition $\mathcal{A}$. The pairs (h_1, h_2), (r_1, r_2) and (x_0, y_0) are called the* bandwidth, shift *and* central node, *respectively.*

Obviously, the adjustment of a central node x_0 has no influence on the theoretical results concerning the fuzzy transform. Therefore, we restrict our investigation to the uniform fuzzy partitions, where $x_0 = 0$, and we deal with the uniform fuzzy partitions determined by triplets (K, h, r). The same restriction is also assumed for the uniform fuzzy partitions of the plane that are determined simply by triplets $(\mathcal{K}, (h_1, h_2), (r_1, r_2))$. Additionally, we use $A_k[h](x)$ and $A_{k_1,k_2}[h_1, h_2](x, y)$ to denote the basic functions of uniform fuzzy partitions of $\mathbb{R}$ and $\mathbb{R}^2$, respectively, in all the situations when the bandwidth is important and should be emphasized. A simple consequence of the definition of uniform fuzzy partition of the plane is the following statement.

Lemma 1. *Let $\mathcal{A} = \{A_{k_1,k_2} \mid k_1, k_2 \in \mathbb{Z}\}$ be a uniform fuzzy partition of $\mathbb{R}^2$ determined by the tuple $(\mathcal{K}, (h_1, h_2), (r_1, r_2))$, where $\mathcal{K}$ is a bivariate generating function determined by the generating functions K_1 and K_2. If $\mathcal{A}_X = \{A_{X,k_1} \mid k_1 \in \mathbb{Z}\}$ and $\mathcal{A}_Y = \{A_{Y,k_2} \mid k_2 \in \mathbb{Z}\}$ denote the families of fuzzy sets determined by triplets (K_1, h_1, r_1) and (K_2, h_2, r_2) as in (3), respectively, then $\mathcal{A}_X$ and $\mathcal{A}_Y$ are uniform fuzzy partitions of $\mathbb{R}$ and*

$$A_{k_1,k_2}(x, y) = A_{X,k_1}(x) \cdot A_{Y,k_2}(y), \quad x, y \in \mathbb{R},$$

holds for any $k_1, k_2 \in \mathbb{Z}$.

Proof. It is sufficient to show that $\mathcal{A}_X$ and $\mathcal{A}_Y$ are uniform fuzzy partitions of $\mathbb{R}$; the rest of the proof immediately follows from the definition of the basic function A_{k_1,k_2}. Let $x, y \in \mathbb{R}$ be arbitrary. Since $\mathcal{A}$ is a uniform fuzzy partition, we find that

$$1 = \sum_{k_1,k_2 \in \mathbb{Z}} A_{k_1,k_2}(x, y) = \sum_{k_1 \in \mathbb{Z}} A_{X,k_1}(x) \cdot \sum_{k_2 \in \mathbb{Z}} A_{Y,k_2}(y),$$

which implies $\sum_{k_1 \in \mathbb{Z}} A_{X,k_1}(x) = 1$ and $\sum_{k_2 \in \mathbb{Z}} A_{Y,k_2}(y) = 1$. □ □

2.2 Univariate fuzzy transform

Let $L^2_{loc}(\mathbb{R})$ be a set of all complex-valued functions that are square integrable on any closed subinterval of the real line. Below, we recall the definition of the direct and inverse F^m-transform with the components (polynomials) derived from the monomial basis [14, 19, 20]. We use $\mathbb{N}$ to denote the set of natural numbers with zero.

Definition 5. *Let $f \in L_{loc}(\mathbb{R})$, $m \in \mathbb{N}$, and let $\mathcal{A}$ be a uniform fuzzy partition of the real line determined by the triplet (K, h, r). The* direct fuzzy transform of degree m (direct F^m-transform) *of f with respect to $\mathcal{A}$ is a family*

$$\mathbf{F}^m_{\mathcal{A}}[f] = \{F^m_k[f] \mid k \in \mathbb{Z}\},$$

of polynomials of the shape

$$F_k^m[f](x) = C_{z,0} + C_{k,1}\left(\frac{x-x_k}{h}\right) + \cdots + C_{k,m}\left(\frac{x-x_k}{h}\right)^m \tag{5}$$

with the coefficients determined by

$$(C_{k,0}, C_{k,1}, \ldots, C_{k,m})^T = (\mathcal{Z}_m)^{-1} \cdot \mathcal{Y}_{m,k},$$

where $\mathcal{Z}_m = (Z_{ij})$ *is a* $(m+1)\times(m+1)$ *invertible matrix defined by*

$$Z_{ij} = \int_{-1}^{1} x^{i+j-2} K(x)\,dx, \quad i,j = 1,\ldots,m+1, \tag{6}$$

and $\mathcal{Y}_{m,k} = (Y_{k,1}, \ldots, Y_{k,m+1})^T$ *is a vector defined by*

$$Y_{k,\ell} = \int_{-1}^{1} f(hx+x_k)\cdot x^{\ell-1} K(x)\,dx, \quad \ell = 1,\ldots,m+1. \tag{7}$$

The polynomial $F_k^m[f]$ *is called the* k-th component *of the direct* F^m*-transform with respect to* $\mathcal{A}$*.*

The inverse F^m-transform is a linear like combination of the components of the direct F^m-transform and the values of basic functions.

Definition 6. *Let* $f \in L^2_{loc}(\mathbb{R})$*, let* $\mathcal{A} = \{A_k \mid k \in \mathbb{Z}\}$ *be a uniform fuzzy partition of the real line, and let* $\mathbf{F}^m_{\mathcal{A}}[f] = \{F_k^m[f] \mid k \in \mathbb{Z}\}$ *denote the direct* F^m*-transform of the function* f *with respect to* $\mathcal{A}$*. The* inverse fuzzy transform of degree m (inverse F^m-transform) *of* f *with respect to* $\mathcal{A}$ *is defined by*

$$\hat{f}^m_{\mathcal{A}}(x) = \sum_{k\in\mathbb{Z}} F_k^m[f](x)\cdot A_k(x), \quad x \in \mathbb{R}. \tag{8}$$

The inverse F^m-transform $\hat{f}^m_{\mathcal{A}}$ approximates the original function f. The quality of the approximation depends on the parameter's setting of the fuzzy partition $\mathcal{A}$. For details, we refer to [2, 19, 20].

3 Bivariate fuzzy transform based on tensor product of two polynomial spaces

3.1 Definitions

Let $\mathbb{P}_m$ and $\mathbb{P}_n$ $(m,n \in \mathbb{N})$ denote the linear spaces of all complex-valued polynomials $P(x)$ and $P(y)$ of degree at most equal to m and n, respectively. For any $m,n \in \mathbb{N}$, we use $\mathbb{P}_m \otimes \mathbb{P}_n$ denote the tensor product of $\mathbb{P}_m$ and $\mathbb{P}_n$. It is easy to see that $\mathbb{P}_m \otimes \mathbb{P}_n$ is a linear space of dimension $N = (m+1)(n+1)$. For any $\ell = 1,\ldots,N$, consider $P_\ell(x,y) = x^{q_\ell} y^{r_\ell}$, where $q_\ell, r_\ell \in \mathbb{N}$ such that $\ell - 1 = (n+1)\cdot q_\ell + r_\ell$ and $r_\ell \le n$. Obviously, q_l is the quotient and r_ℓ is the remainder when we divide $\ell - 1$ by $n+1$ in long division. Then, $\{P_1,\ldots,P_N\} = \{x^i y^j \mid i = 0,1,\ldots,m, j = 0,1,\ldots,n\}$ forms a basis of the tensor product $\mathbb{P}_m \otimes \mathbb{P}_n$.

Let $L^2_{loc}(\mathbb{R}^2)$ be a set of all complex-valued functions f defined on $\mathbb{R}^2$ that satisfy

$$\int_I |f(x,y)|^2 dxdy < \infty,$$

for any compact set $I \subset \mathbb{R}^n$. Let $\mathcal{A} = \{A_{k_1,k_2} \mid k_1,k_2 \in \mathbb{Z}\}$ be a uniform fuzzy partition of $\mathbb{R}^2$ determined by a triplet $(\mathcal{K},(h_1,h_2),(r_1,r_2))$. For any $k_1,k_2 \in \mathbb{Z}$, put $x_{k_1} = k_1 r_1$ and $y_{k_2} = k_2 r_2$. Let $L^2(A_{k_1,k_2})$ be the set of complex-valued functions f defined on the rectangle $U_{k_1,k_2} = [x_{k_1} - h_1, x_{k_1} + h_1] \times [y_{k_2} - h_2, y_{k_2} + h_2]$ such that

$$\int_{U_{k_1,k_2}} |f(x,y)|^2 dxdy < \infty.$$

Define on $L^2(A_{k_1,k_2})$ an operator $\langle\cdot,\cdot\rangle_{A_{k_1,k_2}} : L^2(A_{k_1,k_2}) \times L^2(A_{k_1,k_2}) \to \mathbb{C}$, where $\mathbb{C}$ denotes the set of all complex numbers, as follows:

$$\langle f, g\rangle_{A_{k_1,k_2}} = \int_{U_{k_1,k_2}} f(x,y)\overline{g(x,y)}A_{k_1,k_2}(x,y)dxdy \tag{9}$$

where $\overline{g(x,y)}$ denotes the complex conjugate of $g(x,y)$. The operator forms an inner product in $L^2(A_{k_1,k_2})$. Furthermore, $L^2(A_{k_1,k_2})$ equipped by the inner product $\langle\cdot,\cdot\rangle_{A_{k_1,k_2}}$ is a Hilbert space, where we identify functions that differs in a set of zero measure. We refer to the space $L^2(A_{k_1,k_2})$ as *the weighted Hilbert space with respect to the basic function* A_{k_1,k_2}. For $(h_1,h_2) = (1,1)$ and $(k_1,k_2) = (0,0)$, i.e., $A_{0,0}[1,1] = \mathcal{K}$, we use $L^2(\mathcal{K})$ and $\langle\cdot,\cdot\rangle_{\mathcal{K}}$ to denote the space $L^2(A_{0,0}[1,1])$ and its inner product, respectively. Let $f, g \in L^2(A_{k_1,k_2})$, and let L be a linear subspace of $L^2(A_{k_1,k_2})$. We say that f *is orthogonal to* g *in* $L^2(A_{k_1,k_2})$ and denote it by $f\perp_{A_{k_1,k_2}} g$, if $\langle f, g\rangle_{A_{k_1,k_2}} = 0$. Similarly we say that f *is orthogonal to* L *in* $L^2(A_{k_1,k_2})$ and denote it by $f\perp_{A_{k_1,k_2}} L$, if $f\perp_{A_{k_1,k_2}} g$ for any $g \in L$.

Since any function from $L^2_{loc}(\mathbb{R}^2)$ restricted to U_{k_1,k_2} belongs to $L^2(A_{k_1,k_2})$, the inner product $\langle f|_{U_{k_1,k_2}}, g|_{U_{k_1,k_2}}\rangle_{A_{k_1,k_2}}$ is well defined for any $f, g \in L^2_{loc}(\mathbb{R}^2)$, where $f|_{U_{k_1,k_2}}$ and $g|_{U_{k_1,k_2}}$ denote the respective restrictions to U_{k_1,k_2}. In the sequel, the arguments of $\langle\cdot,\cdot\rangle_{A_{k_1,k_2}}$ or $\|\cdot\|_{A_{k_1,k_2}}$ (the norm of the Hilbert space $L^2(A_{k_1,k_2})$) are always the functions that are restricted to U_{k_1,k_2}. For example, if P and Q are two polynomials and we use $\langle P, Q\rangle_{A_{k_1,k_2}}$, then P and Q are polynomials restricted to U_{k_1,k_2}. Notice that the computation of the inner product with respect to a basic function does not require the restriction of functions on a rectangle, the restriction is important only from the formal point of view and ensures that we deal within our framework of the weighted Hilbert spaces.

For any $k_1, k_2 \in \mathbb{Z}$, let us denote by $\mathbb{P}_{m,k_1} \otimes \mathbb{P}_{n,k_2}$ the set of bivariate polynomials from the tensor product $\mathbb{P}_m \otimes \mathbb{P}_n$ that are restricted to U_{k_1,k_2}. It is easy to see that $\mathbb{P}_{m,k_1} \otimes \mathbb{P}_{n,k_2}$ is a closed linear subspace of $L^2(A_{k_1,k_2})$. Similarly to the definition of F^m-transform, the bivariate fuzzy transform based on the tensor product of polynomial spaces has two phases. Their definitions are as follows.

Definition 7. *Let* $\mathcal{A} = \{A_{k_1,k_2} \mid k_1, k_2 \in \mathbb{Z}\}$ *be a uniform fuzzy partition of* $\mathbb{R}^2$*, and let* f *be a function from* $L^2_{loc}(\mathbb{R}^2)$. The direct fuzzy transform based on $\mathbb{P}_m \otimes \mathbb{P}_n$ (the direct F$^{m\otimes n}$-transform) *of* f *with respect to* $\mathcal{A}$ *is a family*

$$\mathrm{F}^{m\otimes n}_{\mathcal{A}}[f] = \left\{F^{m\otimes n}_{k_1,k_2}[f] \in \mathbb{P}_{m,k_1} \otimes \mathbb{P}_{n,k_2} \mid k_1, k_2 \in \mathbb{Z}\right\}, \tag{10}$$

where $(f - F^{m\otimes n}_{k_1,k_2}[f])\perp_{A_{k_1,k_2}} (\mathbb{P}_{m,k_1} \otimes \mathbb{P}_{n,k_2})$ *in* $L^2(A_{k_1,k_2})$*. The bivariate polynomial* $F^{m\otimes n}_{k_1,k_2}[f]$ *is called the* (k_1,k_2)*-th component of the direct* $F^{m\otimes n}$*-transform of* f *with respect to* $\mathcal{A}$*.*

Definition 8. *Let* $\mathcal{A} = \{A_{k_1,k_2} \mid k_1, k_2 \in \mathbb{Z}\}$ *be a uniform fuzzy partition of* $\mathbb{R}^2$*, let* $f \in L^2_{loc}(\mathbb{R}^2)$*, and let* $\mathrm{F}^{m\otimes n}_{\mathcal{A}}[f] = \left\{F^{m\otimes n}_{k_2,k_2}[f] \mid k_1, k_2 \in \mathbb{Z}\right\}$ *denote the direct* $F^{m\otimes n}$*-transform of* f *with respect to* $\mathcal{A}$*. The function*

$$\hat{f}^{m\otimes n}_{\mathcal{A}}(x,y) = \sum_{k_1,k_2\in\mathbb{Z}} F^{m\otimes n}_{k_1,k_2}[f](x,y)\cdot A_{k_1,k_2}(x,y), \quad (x,y) \in \mathbb{R}^2, \tag{11}$$

is called the inverse fuzzy transform based on tensor product $\mathbb{P}_m \otimes \mathbb{P}_n$ (the inverse F$^{m\otimes n}$-transform) *of* f *with respect to* $\mathcal{A}$*.*

3.2 Computation of the direct F$^{m\otimes n}$-transform components

In this subsection, we introduce the computation of the direct F$^{m\otimes n}$-transform components for which the (k_1,k_2)-th component is computed with respect to the basis

$$\left\{P_\ell\left(\frac{x - t_{k_1}}{h_1}, \frac{y - t_{k_2}}{h_2}\right) \mid \ell = 1, \ldots, N\right\}$$

of the approximation space $\mathbb{P}_{m,k_1} \otimes \mathbb{P}_{n,k_2}$, where $P_\ell = P_\ell(x,y) = x^{q_\ell}y^{r_\ell}$ $(\ell = 1, \ldots, N)$ denotes the ℓ-th element of the basis of $\mathbb{P}_m \otimes \mathbb{P}_n$ restricted to the square $[-1,1] \times [-1,1]$.

Let us define the following square matrix of the size $N \times N$ ($N = (m+1)(n+1)$):

$$\mathcal{Z}_{m\otimes n} = \begin{bmatrix} \langle P_1, P_1\rangle_{\mathcal{K}} & \langle P_1, P_2\rangle_{\mathcal{K}} & \dots & \langle P_1, P_N\rangle_{\mathcal{K}} \\ \langle P_2, P_1\rangle_{\mathcal{K}} & \langle P_2, P_2\rangle_{\mathcal{K}} & \dots & \langle P_2, P_N\rangle_{\mathcal{K}} \\ \dots & \dots & \dots & \dots \\ \langle P_N, P_1\rangle_{\mathcal{K}} & \langle P_N, P_2\rangle_{\mathcal{K}} & \dots & \langle P_N, P_N\rangle_{\mathcal{K}} \end{bmatrix}, \tag{12}$$

where $\langle\cdot,\cdot\rangle_{\mathcal{K}}$ is the inner product in $L^2(\mathcal{K})$. Recall that

$$\langle P_\ell, P_k\rangle_{\mathcal{K}} = \int_{-1}^{1}\int_{-1}^{1} x^{q_\ell+q_k} y^{r_\ell+r_k}\, dxdy \tag{13}$$

For any $\ell, k = 1, \dots, N$. Recall that $\mathcal{Z}_m$ denotes the $(m+1)\times(m+1)$ matrix whose elements are the inner products of polynomials from the monomial basis $\{1, \dots, x^m\}$ of the space $\mathbb{P}_m$ (see Definition 5). Similarly, define the $(n+1)\times(n+1)$ matrix $\mathcal{Z}_n$ that consists of the inner product of polynomials from the monomial basis $\{1, \dots, y^n\}$ of the space $\mathbb{P}_n$. The following lemma is essential in the computation of the direct $\mathrm{F}^{m\otimes n}$-transform components.

Lemma 2. *Let $\mathcal{Z}_{m\otimes n}$ be the matrix defined in (12). Then, $\mathcal{Z}_{m\otimes n} = \mathcal{Z}_m \otimes \mathcal{Z}_n$, where $\otimes$ means the tensor (Kronecker) product of matrices. Moreover, $\mathcal{Z}_{m\otimes n}$ is a real and invertible matrix.*

Proof. The first part of the statement is a simple consequence of Fubini's theorem. By Lemma 2.8 in [20] (see also Lemma 8 in [19]), the matrices $\mathcal{Z}_m$ and $\mathcal{Z}_n$ are real and invertible. It is well known that $(\mathcal{Z}_{m\otimes n})^{-1} = (\mathcal{Z}_m)^{-1} \otimes (\mathcal{Z}_n)^{-1}$. Hence, $\mathcal{Z}_{m\otimes n}$ is real and invertible. □

Using the properties of the inner product and the previous lemma, one can prove the following statement that provides a simple approach based on matrix calculus to the computation of the direct $\mathrm{F}^{m\otimes n}$-transform components.

Theorem 3. *Let $f \in L^2_{loc}(\mathbb{R})$, and let $\mathcal{A}$ be a uniform fuzzy partition of $\mathbb{R}^2$ determined by the tuple $(\mathcal{K}, (h_1, h_2), (r_1, r_2))$. Let $m, n \in \mathbb{Z}$ and $N = (m+1)(n+1)$. Then, the (k_1, k_2)-th component of the direct $F^{m\otimes n}$-transform of f with respect to $\mathcal{A}$ has the following form*

$$F^{m\otimes n}_{k_1,k_2}[f](x,y) = \sum_{\ell=1}^{N} C_{k_1,k_2,\ell} \cdot P_\ell\left(\frac{x-t_{k_1}}{h_1}, \frac{y-t_{k_2}}{h_2}\right) \tag{14}$$

determined by

$$(C_{k_1,k_2,1}, C_{k_1,k_2,2}, \dots, C_{k_1,k_2,N})^T = (\mathcal{Z}_{m\otimes n})^{-1} \cdot \mathcal{Y}^{k_1,k_2}_{m\otimes n} \tag{15}$$

where $\mathcal{Z}_{m\otimes n}$ is the matrix defined in (12), and

$$\mathcal{Y}^{k_1,k_2}_{m\otimes n} = (Y_{k_1,k_2,1}, Y_{k_1,k_2,2}, \dots, Y_{k_1,k_2,N})^T$$

is defined by

$$Y_{k_1,k_2,k} = \int_{-1}^{1}\int_{-1}^{1} f(uh_1 + t_{k_1}, vh_2 + t_{k_2}) \cdot P_k(u,v) \cdot \mathcal{K}(u,v) du dv, \tag{16}$$

for $k = 1, \dots, N$.

From (15), one can see that also the $F^{m\otimes n}$-transform is a linear map. Moreover, if a bivariate function can be express as a product of two univariate functions, each component of the direct $\mathrm{F}^{m\otimes n}$-transform of the original function is the product of a component of the direct F^m-transform and a component of the direct F^n-transform applied to the univariate functions as the following statement shows.

Theorem 4. *Let f be a bivariate complex-valued function such that $f(x,y) = \mathscr{X}(x) \cdot \mathscr{Y}(y)$ where $\mathscr{X}, \mathscr{Y} \in L^2_{loc}(\mathbb{R})$. Let $\mathcal{A}$ be a uniform fuzzy partition of $\mathbb{R}^2$ determined by the tuple $(\mathcal{K}, (h_1, h_2), (r_1, r_2))$ with $\mathcal{K}(x,y) = K_1(x) \cdot K_2(y)$. Then, for any $k_1, k_2 \in \mathbb{Z}$, $(x,y) \in [t_{k_1} - h_1, t_{k_1} + h_1] \times [t_{k_2} - h_2, t_{k_2} + h_2]$, it holds that*

$$F^{m\otimes n}_{k_1,k_2}[f](x,y) = F^m_{k_1}[\mathscr{X}](x) \cdot F^n_{k_2}[\mathscr{Y}](y),$$

where $F^m_{k_1}[\mathscr{X}]$ and $F^n_{k_2}[\mathscr{Y}]$ are the k_1-th and k_2-th components of the direct F^m- and F^n-transforms of the function $\mathscr{X}$ and $\mathscr{Y}$ with respect to the uniform fuzzy partitions of the real line determined by the triplets (K_1, h_1, r_1) and (K_2, h_2, r_2), respectively.

Sketch of proof. It is a consequence of Lemma 2, Fubini's theorem and matrix calculus. □

The following corollary is a straightforward consequence of Theorem 4. It provides a simple way for the computation of the inverse $\mathrm{F}^{m\otimes n}$-transform of bivariate functions that can be separated into a product of two univariate functions.

Corollary 5. *Let the assumptions of Theorem 4 be satisfied. Then,*

$$\hat{f}^{m\otimes n}_{\mathcal{A}}(x,y) = \widehat{\mathscr{X}}^{m}_{\mathcal{A}_1}(x)\cdot\widehat{\mathscr{Y}}^{n}_{\mathcal{A}_2}(y),$$

where $\widehat{\mathscr{X}}^{m}_{\mathcal{A}_1}$ and $\widehat{\mathscr{Y}}^{n}_{\mathcal{A}_2}$ are the inverse F^m- and F^n-transforms of the function $\mathscr{X}$ and $\mathscr{Y}$ with respect to the uniform fuzzy partitions $\mathcal{A}_X$ and $\mathcal{A}_Y$ of the real line respectively determined by the triplets (K_1,h_1,r_1) and (K_2,h_2,r_2), respectively.

Example 2. *Let us consider two bivariate functions*

1. $f(x,y) = (x+y)\sin 3y - 3\cos x$,
2. $g(x,y) = 5\cos 2x\cdot\sin\left(3y+\frac{\pi}{3}\right)$.

In Figure 1, we display the inverse $F^{2\otimes 1}$-transform of f with respect to the bivariate uniform fuzzy partition determined by the triplet $(\mathcal{K},\mathbf{h},\mathbf{r})$, where $\mathcal{K}(x,y) = K^{tr}(x)\cdot K^{tr}(y)$ and $\mathbf{h}=\mathbf{r}=(0.4,0.4)$. The result is restrict on the rectangle $[0,4]^2$. Additionally, by using the same fuzzy partition we compute the inverse $F^{0\otimes 2}$-transform of function g. However, instead of using Theorem 3 and, then, Definition 8 for the computation, we apply Corollary 5 because g is a variable separable function. As a result, we only need to compute the inverse F^0-transform and inverse F^2-transform of the functions $\mathscr{X}(x) = 5\cos 2x$ and $\mathscr{Y}(y) = \sin\left(3y+\frac{\pi}{3}\right)$, respectively, with respect to the triangular uniform fuzzy partition determined by the triplet $(K^{tr},0.4,0.4)$. The result is then obtain by the application of Corollary 5, which is depicted in Figure 2.

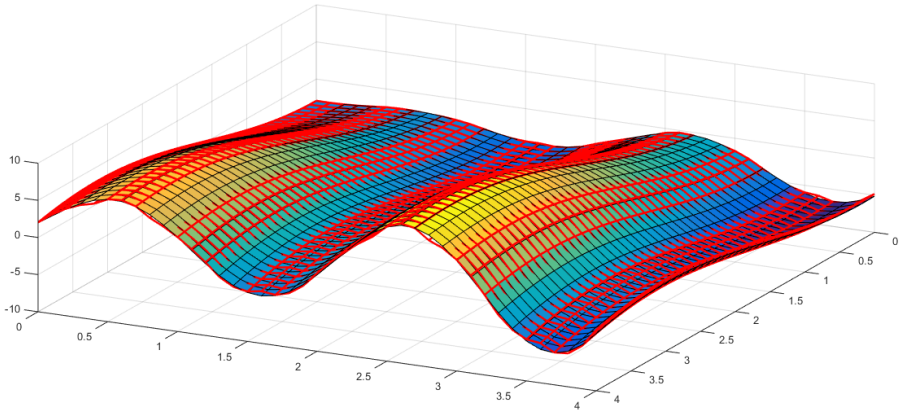


Figure 1: Inverse $\mathrm{F}^{2\otimes 1}$-transform of f

4 $\mathrm{F}^{m\otimes n}$-transform in the analysis of correlation functions

Let $\xi(t)$, $t\in\mathbb{R}$, be a complex-valued random process, and assume that $\Gamma(t,s)$ is its correlation function, i.e.,

$$\Gamma(t,s) = \mathbf{E}\left(\xi(t)\cdot\overline{\xi(s)}\right), \quad (t,s)\in\mathbb{R}^2.$$

We know that the theory of correlation functions is one of the fundamental tools in the investigation of the behavior of random processes (see, e.g., [21]). Recently, we provided in [16] an analysis of random processes using the fuzzy transform of higher degree. This analysis is a continuation of the research introduced in [22, 23]. In what follows, we show that the $\mathrm{F}^{m\otimes n}$-transform reveals

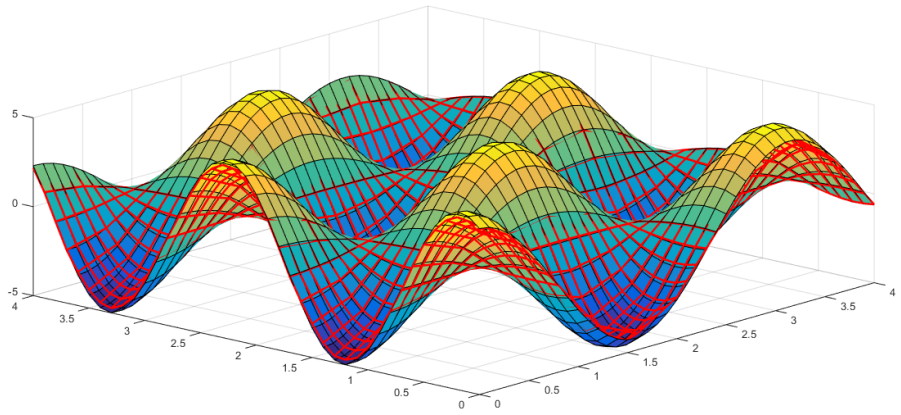


Figure 2: Inverse $\mathrm{F}^{0\otimes 2}$-transform of g

the relationship between the correlation function of a random processes $\xi(t)$ and the correlation function of the inverse F^m-transform of $\xi(t)$.

Let $\mathcal{A}_X$ be a uniform fuzzy partition of the real line determined by the triplet (K,h,r). It follows from [16] that the k_1-th component of the direct F^m-transform of $\xi(t)$ with respect to $\mathcal{A}_X$ is determined as follows:

$$F^m_{k_1}[\xi](t)=\sum_{i,p=0}^{m}\left(\frac{t-t_{k_1}}{h}\right)^i\cdot T_{i+1p+1}\cdot\int_{-1}^{1}\xi(h\tau-t_{k_1})\cdot\tau^pK(\tau)d\tau$$

where $(T_{ij})_{i,j=\overline{1,m+1}}=(\mathcal{Z}_m)^{-1}$. Let $F^{m\otimes m}_{k_1,k_2}[\Gamma]$ be the (k_1,k_2)-th component of the direct $\mathrm{F}^{m\otimes m}$-transform of Γ with respect to the uniform fuzzy partition of $\mathbb{R}^2$ determined by the tuple $(\mathcal{K},(h,h),(r,r))$ with $\mathcal{K}(x,y)=K(x)\cdot K(y)$. As a result, for any $k_1,k_2\in\mathbb{Z}$, $t\in[t_{k_1}-h,t_{k_1}+h]$ and $t'\in[t_{k_2}-h,t_{k_2}+h]$, we find that

$$\begin{aligned}
&\mathbf{Cor}\left(F^m_{k_1}[\xi](t),F^m_{k_2}[\xi](t')\right)=\mathbf{E}\left(F^m_{k_1}[\xi](t)\cdot\overline{F^m_{k_2}[\xi](t')}\right)\\
=&\sum_{i,j,p,\iota=0}^{m}\left(\frac{t-t_{k_1}}{h}\right)^i\cdot\left(\frac{t'-t_{k_2}}{h}\right)^j\cdot\\
&\qquad T_{i+1p+1}\cdot T_{j+1\iota+1}\cdot\mathbf{E}\left(\int_{-1}^{1}\xi(h\tau+t_{k_1})\cdot\tau^pK(\tau)d\tau\cdot\int_{-1}^{1}\overline{\xi(hs+t_{k_2})}\cdot s^\iota K(s)ds\right)\\
=&\sum_{i,j,p,\iota=0}^{m}\left(\frac{t-t_{k_1}}{h}\right)^i\cdot\left(\frac{t'-t_{k_2}}{h}\right)^j\cdot\\
&\qquad V_{(m+1)i+j+1(m+1)p+i+1}\cdot\int_{-1}^{1}\int_{-1}^{1}\Gamma\left(h\tau+t_{k_1},hs+t_{k_2}\right)\cdot\tau^ps^\iota K(\tau)K(s)d\tau ds\\
=&\sum_{\ell,\kappa=1}^{(m+1)^2}\left(\frac{t-t_{k_1}}{h}\right)^{q_\ell}\cdot\left(\frac{t'-t_{k_2}}{h}\right)^{r_\ell}\cdot V_{\ell\kappa}\cdot\int_{-1}^{1}\int_{-1}^{1}\Gamma\left(h\tau+t_{k_1},hs+t_{k_2}\right)\cdot\tau^{q_\kappa}s^{r_\kappa}K(\tau)K(s)d\tau ds\\
=&F^{m\otimes m}_{k_1,k_2}[\Gamma](t,t'),
\end{aligned}$$

where $(V_{\ell\kappa})_{\ell,\kappa=\overline{1,(m+1)^2}}=(\mathcal{Z}_{m\otimes m})^{-1}=(\mathcal{Z}_m)^{-1}\otimes(\mathcal{Z}_m)^{-1}$, and q_ℓ and r_ℓ are the quotient and remainder, respectively, of the division of $\ell-1$ by $m+1$. Finally, from this result, i.e.,

$$\mathbf{Cor}\left(F^m_{k_1}[\xi](t),F^m_{k_2}[\xi](t')\right)=F^{m\otimes m}_{k_1,k_2}[\Gamma](t,t'),$$

it is easy to prove that

$$\mathbf{Cor}\left(\hat{\xi}^m_{\mathcal{A}}(t),\hat{\xi}^m_{\mathcal{A}}(t')\right)=\widehat{\Gamma}^{m\otimes m}_{\mathcal{A}}(t,t'),\quad t,t'\in\mathbb{R}.$$

By the previous analysis, one can see that the $\mathrm{F}^{m\otimes n}$-transform can be used in the investigation of the correlation function of the inverse F^m-transform of a random process $\xi(t)$, namely, instead of the derivation of correlation function from the random process $F^m[\xi](t)$ one can directly apply the $\mathrm{F}^{m\otimes n}$-transform on the correlation function of $\xi(t)$. Furthermore, the properties of the $\mathrm{F}^{m\otimes n}$-transform can help to study the interesting properties of correlation functions of random processes that are obtained by the higher degree fuzzy transform. This investigation is a subject of our future research.

5 Conclusions

In this paper, we introduced a new type of fuzzy transformation of bivariate complex-valued functions which is called the $\mathrm{F}^{m\otimes n}$-transform. This novel transformation is defined within the framework of the bivariate higher degree fuzzy transform, where the (bivariate) polynomial space (approximation space) is replaced by the tensor product of two (univariate) polynomial spaces. We proposed an approach based on matrix calculus to the computation of $\mathrm{F}^{m\otimes n}$-transform components and show that the $\mathrm{F}^{m\otimes n}$-transform becomes very simple for bivariate functions that can be separated into a product of two univariate functions. Finally, we demonstrated the advantage of the application of the $\mathrm{F}^{m\otimes n}$-transform in the investigation of correlation functions of random processes.

Acknowledgement
This work was supported by the project LQ1602 IT4Innovations excellence in science. The additional support was provided by the Czech Science Foundation through the project of No.16-09541S.

References

[1] I. Perfilieva. Fuzzy transforms: theory and applications. *Fuzzy Sets and Systems*, 157:993–1023, 2006.

[2] I. Perfilieva, M. Daňková, and B. Bede. Towards a higher degree F-transform. *Fuzzy Sets and Systems*, 180:3–19, 2011.

[3] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, and L. Vavříčková. Analysis of seasonal time series using fuzzy approach. *Int. Journal of General Systems*, 39:305–328, 2010.

[4] M. Štěpnička, A. Dvŏrák, and V. Pavliska. A linguistic approach to time series modeling with the help of F–transform. *Fuzzy Sets and Systems*, 180:164–184, 2011.

[5] V. Novák, I. Perfilieva, and V. Krejnovich. F-transform in the analysis of periodic signals. In M. Inuiguchi, Y. Kusunoki, and M. Seki, editors, *Proc. 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, pages 150–158. Osaka University, Osaka, Japan, 2012.

[6] V. Novák, V. Pavliska, M. Štěpnička, and L. Štěpničková. Time series trend extraction and its linguistic evaluation using F-transform and fuzzy natural logic. In *Studies in Fuzziness and Soft Computing*, pages 429–442. Springer Verlag, 2014.

[7] V. Novák, I. Perfilieva, M. Holčapek, and V. Kreinovich. Filtering out high frequencies in time series using F-transform. *Information Sciences*, 274:192–209, 2014.

[8] V. Novák and I. Perfilieva. Forecasting Direction of Trend of a Group of Analogous Time Series Using F-Transform and Fuzzy Natural Logic. *The International Journal of Computational Intelligence Systems*, 8:15–28, 2015.

[9] Khastan Alireza, Irina Perfilieva, and Zahra Alijani. A new fuzzy approximation method to cauchy problems by fuzzy transform. *Fuzzy Sets and Systems*, 288:75–95, 2016.

[10] Irina Perfilieva, Petra Števuliáková, and Radek Valášek. F-transform-based shooting method for nonlinear boundary value problems. *Soft Computing*, page 34933502, 2017.

[11] P. Hodáková. *Fuzzy (F–)transform of functions of two variables and its applications in image processing, (PhD thesis)*. University of Ostrava, 2015.

[12] M. Holčapek. Discrete multivariate F-transform of higher degree. In *Proc. IEEE International Conference on Fuzzy Systems 2015*. Istanbul, Turkey, August 2-5 2015.

[13] M. Kokainis and S. Asmuss. Higher degree F-transforms based on B-splines of two variables. In *16th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2016*, pages 648–659. Eindhoven, The Netherlands, June, 20.–24. 2016.

[14] Linh Nguyen, Michal Holčapek, and Vilém Novák. Multivariate fuzzy transform of complex-valued functions determined by monomial basis. *Soft Computing*, pages 3641–3658, 2017.

[15] I. Perfilieva, P. Hodáková, and P. Hurtík. Differentiation by the F-transform and application to edge detection. *Fuzzy Sets and Systems*, 288:96–114, 2016.

[16] Linh Nguyen and Michal Holčapek. Higher degree fuzzy transform: Application to stationary processes and noise reduction. In *Proceeding of EUSFLAT 2017, Warsaw, Poland*. 2017.

[17] Michal Holčapek, Irina Perfilieva, Vilém Novák, and Vladik Kreinovich. Necessary and sufficient conditions for generalized uniform fuzzy partitions. *Fuzzy sets syst.*, 277:97–121, 2015.

[18] Michal Holčapek and Tomáš Tichý. Discrete multivariate F-transform of higher degree. In *Proc. of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2015.

[19] M. Holčapek, Linh Nguyen, and T. Tichý. Polynomial alias higher degree fuzzy transform of complex-valued functions. *Fuzzy Sets and Systems*, page submitted, 2016.

[20] M. Holčapek and Linh Nguyen. Trend-cycle estimation using fuzzy transform of higher degree. *Iranian Journal of fuzzy systems*, page In press, 2017.

[21] A.M. Yaglom. *An Introduction to The Theory of Stationary Random Functions. Revised English ed. Translated and edited by Richard A. Silverman.* Englewood Cliffs, NJ: Prentice-Hall, Inc. XIII, 1962.

[22] M. Holčapek, V. Novák, and I. Perfilieva. Analysis of stationary processes using fuzzy transform. In *Proc. European Society for Fuzzy Logic and Technology*, pages 714–721. Istanbul, Turkey, 2013.

[23] M. Holčapek, V. Novák, and I. Perfilieva. Noise reduction in time series using F-transform. In *Proc. IEEE International Conference on Fuzzy Systems 2013*. Hyderabad, India, 2013.

Computational Complexity of Coalitional Stable Assignments in Group Activity Selection Games

Natsuki NIWA[1], Souta SUZUKI[2], Takashi TSUBOTA[3] and Shao-Chin SUNG[4]

[1,3,4] *Aoyama Gakuin University,*
5-10-1 Fuchinobe, Chuo, Sagamihara, Kanagawa, 252-5258, Japan
[2] *Mitsui Knowledge Industry Co., Ltd.,*
2-5-1 Atago, Minato-ku, Tokyo, 105-6215, Japan
c5617162@aoyama.jp[1], suzuki-sota@mki.co.jp[2], d5615006@aoyama.jp[3], son@ise.aoyama.ac.jp[4]

Abstract

Group Activity Selection (GAS) is a model of coalition formation games introduced by Darmann *et al.* [2], where an outcome of such a game is an assignment which associates each player to one of given activities. It is assumed that each player cares not only which activity she or he is assigned, but also the number of participants are assigned to the same activity. An outcome is *coalitional stable* if there exists no single-player or group of players who has incentive to change the outcome. In this paper, we are concerned with the computational complexity of problems related to coalitional stable assignments in GAS. We investigate the computational complexity of the existence related to coalitional stable assignments. Moreover, we show that contractually coalitional stable assignments can be constructed in polynomial time of the number of players and activities.

Keywords: Group Activity Selection, Coalitional Stabilities, Computational Complexity.

1 Introduction

GAS occurs in situations where an organizer wants to arrange players for one or some group activities. Each player will participates in at most one activity, and cares not only the group activity he or she participates in, but also a number of participants in the group activity. Moreover, it is assume that each player has a preference over the all possible alternatives of (activity, group size), or a possibility of "doing nothing".

An outcomes is an assignment of players to activities. An assignment naturally partitions a set of players into one or some groups, where each group is a set of players whom assigned to same non-void activity, or a singleton group of player whom assigned to void activity. The goal is to find an outcome which satisfies a stability concept. Darmann *et al.* [2] figured that GAS is a generalized model of *anonymous hedonic games*, which is the coalition formation games where each player is concerned with the size of the coalition which he or she belongs to. By the analogies from such games, they introduced some stability concepts for special cases of GAS.

In this paper, we are concerned with the computational complexity of problems related to coalitional stable, which includes *core stable*, *strictly core (SC) stable*, *contractually core (CC) stable*, and *contractual strictly core (CSC) stable*, assignments in GAS. Darmann [1] claimed that checking whether a core stable assignment exists is NP-complete. First, we propose the proof of checking whether a core stable or a SC stable assignment exists is NP-complete. Secondly, we show that CC stable or CSC stable assignments always exists and propose the polynomial time algorithms to construct such stable assignments.

2 Preliminaries

Let $\mathcal{N} = \{1, 2, \ldots, n\}$ be a set of players and let $\mathcal{A} = \mathcal{M} \cup \mathcal{V}$ be a set of activities, where $\mathcal{M} = \{a_1, a_2, \ldots, a_m\}$ and $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$. Each $v_i \in \mathcal{V}$ is *void activity*, that means "doing nothing".

An *alternative* (a, k) is a pair of an activity $a \in \mathcal{A}$ and the number $k \in [1, n]$ of participants in a, where $[p, q] = \{r \in \mathbf{Z} \mid p \leq r \wedge r \leq q\}$ for any integers p and q. We denote by $X = \mathcal{A} \times [1, n]$ the collection of all possible alternatives. Each player $i \in \mathcal{N}$ is endowed with a *preference* $\succeq_i$ (i.e., a complete transitive binary) over X. By $\succ_i$ and $\sim_i$, we denote the corresponding strict preference and indifference with respect to $\succeq_i$. An *instance* $(\mathcal{N}, \mathcal{A}, \succeq)$ of GAS is a tuple of a set $\mathcal{N}$ of players, a set $\mathcal{A}$ of activities, and a profile $\succeq = (\succeq_1, \succeq_2, \ldots, \succeq_n)$ of players' preferences, such that $(v_i, 1) \succeq_i (v, k)$ for each player $i \in \mathcal{N}$ and for each alternative $(v, k) \in \mathcal{V} \times [1, n]$.

An outcome for an instance $(\mathcal{N}, \mathcal{A}, \succeq)$ is an *assignment* $\mu : \mathcal{N} \to \mathcal{A}$. To describe a group size of each activity at μ, we define $G_\mu(a) = \{i \in \mathcal{N} \mid \mu(i) = a\}$.

Here, we introduce some subdomain of preferences. We say that a preference profile $\succeq$ is *a-gasp* if each player $i \in \mathcal{N}$ has $\{x \in X \mid x \sim_i (v_i, 1)\} = \{(v_i, 1)\}$ and $x \sim_i y$ for each $x, y \in \{x' \in X \mid x' \succ_i (v_i, 1)\}$. We say that a preference profile $\succeq$ is *strict* if each player $i \in \mathcal{N}$ has either $x \succ_i y$ or $y \succ_i x$ for any $x, y \in X$ with $x \neq y$. For each player $i \in \mathcal{N}$, we say that the preference $\succeq_i$ is *increasing* with respect to $a \in \mathcal{M}$ if $(a, k) \succeq_i (a, k-1)$ for each $k \in [2, n]$. We say that the preference profile $\succeq$ is increasing, if for each player $i \in \mathcal{N}$, his or her preference $\succeq_i$ is increasing with respect to each activity $a \in \mathcal{M}$.

The description size of a profile $\succeq$ is $\mathcal{O}(n^3 m^2)$. We set ranking over X for all players $i \in \mathcal{N}$,

$$\textsc{rank}_{\succeq_i}(a, k) = \left|\left\{(a', k') \in X \mid (a, k) \succeq_i (a', k')\right\}\right|$$

From profile $\succeq$, rankings of all players can be obtained in $\mathcal{O}(n^3 m^2)$ time. Each query "$(a, k) \succeq_i (b, \ell)$" to player i's preference can be answered in constant time, because $(a, k) \succeq_i (b, \ell)$ if and only if $\textsc{rank}_{\succeq_i}(a, k) \geq \textsc{rank}_{\succeq_i}(b, \ell)$. We assume that rankings are computed as a preprocess.

2.1 On Coalitional Stabilities

An assignment μ is *individually rational (IR)* if $(v_i, 1) \succeq_i (\mu(i), |G_\mu(\mu(i))|)$ for each player $i \in \mathcal{N}$. By the analogies from hedonic games, Darmann *et al.* has defined some stability concepts for special cases of GAS [2] [1]. We define some coalitional stabilities for GAS by the analogies from hedonic games [6].

An assignment μ is *core stable* if there exists no *core deviation*, which is a pair $(\mathcal{T}, a)$ of a player set $\mathcal{T} \in \{\mathcal{T}' \subseteq \mathcal{N} \mid \mathcal{T}' \neq \emptyset\}$ and an activity $a \in \mathcal{A}$ with $G_\mu(a) \subseteq \mathcal{T}$ such that

- $(a, |\mathcal{T}|) \succ_i (\mu(i), |G_\mu(\mu(i))|)$ for each $i \in \mathcal{T}$.

An assignment μ is *strictly core (SC) stable* if there exists no *SC deviation*, which is a pair $(\mathcal{T}, a)$ of a player set $\mathcal{T} \in \{\mathcal{T}' \subseteq \mathcal{N} \mid \mathcal{T}' \neq \emptyset\}$ and an activity $a \in \mathcal{A}$ with $G_\mu(a) \subseteq \mathcal{T}$ such that

- $(a, |\mathcal{T}|) \succeq_i (\mu(i), |G_\mu(\mu(i))|)$ for each $i \in \mathcal{T}$, and
- $(a, |\mathcal{T}|) \succ_i (\mu(i), |G_\mu(\mu(i))|)$ for some $i \in \mathcal{T}$.

An assignment μ is *contractually core (CC) stable* if there exists no *CC deviation*, which is a pair $(\mathcal{T}, a)$ of a player set $\mathcal{T} \in \{\mathcal{T}' \subseteq \mathcal{N} \mid \mathcal{T}' \neq \emptyset\}$ and an activity $a \in \mathcal{A}$ with $G_\mu(a) \subseteq \mathcal{T}$ such that

- $(a, |\mathcal{T}|) \succ_i (\mu(i), |G_\mu(\mu(i))|)$ for each $i \in \mathcal{T}$, and
- $(\mu(j), |G_\mu(\mu(j)) \setminus \mathcal{T}|) \succeq_j (\mu(j), |G_\mu(\mu(j))|)$ for each $j \in \mathcal{N} \setminus \mathcal{T}$.

An assignment μ is *contractual strictly core (CSC) stable* if there exists no *CSC deviation*, which is a pair $(\mathcal{T}, a)$ of a player set $\mathcal{T} \in \{\mathcal{T}' \subseteq \mathcal{N} \mid \mathcal{T}' \neq \emptyset\}$ and an activity $a \in \mathcal{A}$ with $G_\mu(a) \subseteq \mathcal{T}$ such that

- $(a, |\mathcal{T}|) \succeq_i (\mu(i), |G_\mu(\mu(i))|)$ for each $i \in \mathcal{T}$,
- $(a, |\mathcal{T}|) \succ_i (\mu(i), |G_\mu(\mu(i))|)$ for some $i \in \mathcal{T}$, and
- $(\mu(j), |G_\mu(\mu(j)) \setminus \mathcal{T}|) \succeq_j (\mu(j), |G_\mu(\mu(j))|)$ for each $j \in \mathcal{N} \setminus \mathcal{T}$.

3 Core Stability

Darmann [1] claimed that it is NP-complete to decide whether GAS admits a core stable assignment even when preference profile is strict and increasing and the hardness-proof proceeds by a reduction from Exact Cover By 3-Sets (X3C) [3]. However, the approach of reduction was not stated. We now define a problem that is NP-complete is the strong sense.

Exact Cover By 3-Sets (X3C)

Instance : A pair $(\mathcal{R}, S)$, where $\mathcal{R} = \{1, 2, \ldots, 3q\}$ and $S \subseteq \{\mathcal{S} \subseteq \mathcal{R} \mid |\mathcal{S}| = 3\}$.

Question : Is there $S' \subseteq S$, where $\bigcup_{\mathcal{S} \in S'} \mathcal{S} = \mathcal{R}$ and $|S'| = q$?

In the following, we define the existence of a core stable assignment as a decision problem.

Existence of a core stable assignment (GC)

Instance : Instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GAS.

Question : Is there an core stable assignment in $(\mathcal{N}, \mathcal{A}, \succeq)$?

First, we guess how he construct an instance of GC from an instance of X3C by his previous research [2] which is also described reduction from X3C to GC. Let $(\mathcal{R}, S)$ an instance of X3C, an instance of GC as follows. Let $\mathcal{N} = \mathcal{R}$ and $\mathcal{A} = \mathcal{M} \cup \mathcal{V}$, where $\mathcal{M} = \{a_{\mathcal{S}} \mid \mathcal{S} \in S\}$ and $\mathcal{V} = \{v_1, v_2, \ldots, v_{3q}\}$. For each player $i \in \mathcal{N}$, let preference $\succeq_i$ such that $\{x \in \mathcal{A} \times [1, 3q] \mid x \succ_i (v_i, 1)\} = \{(a_{\mathcal{S}}, k) \in \mathcal{M} \times [1, 3q] \mid i \in \mathcal{S} \wedge k \geq 3\}$ and let $\succeq = (\succeq_1, \succeq_2, \ldots, \succeq_{3q})$.

Theorem 1. *There is an instance of X3C, where a "yes" answer to X3C and "no" answer to GC.*

Proof. We show that there is no core stable assignment. We set instance $(\mathcal{R}, S)$ of X3C as follows, $\mathcal{R} = \{1, 2, \ldots, 6\}$ and $S = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_5\}$, where $\mathcal{S}_1 = \{1, 2, 3\}, \mathcal{S}_2 = \{4, 5, 6\}, \mathcal{S}_3 = \{1, 3, 4\}, \mathcal{S}_4 = \{1, 2, 5\}, \mathcal{S}_5 = \{2, 3, 6\}$. It is obvious that the answer to X3C is "yes". We set instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GC from $(\mathcal{R}, S)$. For each player $i \in \mathcal{N}$, preference $\succeq_i$ satisfies as follows.

$$
\begin{aligned}
(a_{\mathcal{S}_3}, 3) \succ_1 (a_{\mathcal{S}_4}, 3) \succ_1 (a_{\mathcal{S}_1}, 3) \succ_1 (v_1, 1) \\
(a_{\mathcal{S}_4}, 3) \succ_2 (a_{\mathcal{S}_5}, 3) \succ_2 (a_{\mathcal{S}_1}, 3) \succ_2 (v_2, 1) \\
(a_{\mathcal{S}_5}, 3) \succ_3 (a_{\mathcal{S}_3}, 3) \succ_3 (a_{\mathcal{S}_1}, 3) \succ_3 (v_3, 1) \\
(a_{\mathcal{S}_3}, 3) \succ_4 (a_{\mathcal{S}_2}, 3) \succ_4 (v_4, 1) \\
(a_{\mathcal{S}_4}, 3) \succ_5 (a_{\mathcal{S}_2}, 3) \succ_5 (v_5, 1) \\
(a_{\mathcal{S}_5}, 3) \succ_6 (a_{\mathcal{S}_2}, 3) \succ_6 (v_6, 1)
\end{aligned}
$$

There is a core deviation for all assignments. Hence, there is no core stable assignment. □

It is obvious that this reduction is not enough. Hence, we propose addition of gadget and restriction of preferences. We start explaining how we construct an instance of GAS from an instance of X3C. Given an instance $(\mathcal{R}, S)$ of X3C, an instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GC is as follows. Let $\mathcal{N} = \mathcal{R} \cup \{3q + 1, 3q + 2\}$, and $\mathcal{A} = \mathcal{M} \cup \mathcal{V}$, where $\mathcal{M} = \{a_{\mathcal{S}} \mid \mathcal{S} \in S\} \cup \{b, c, d\}$ and $\mathcal{V} = \{v_1, v_2, \ldots, v_{3q+2}\}$. The player preferences are as follow. Let $Y_i = \{x \in X \mid x \succ_i (v_i, 1)\}$ for each player $i \in \mathcal{R}$. For each player $i \in \mathcal{R}$, let $Y_i = Y_i^1 \cup Y_i^2 \cup Y_i^3$,

- where $x \succ_i y$, for each $s \in \{1, 2\}$, for each $x \in Y_i^s$, and for each $y \in Y_i^{s+1}$ and
- where $x \sim_i y$, for each $s \in [1, 3]$, and for each $x, y \in Y_i^s$.

For each player $i \in \mathcal{R}$, Let $Y_i^1 = \{(b, 3q + 1)\}$, $Y_i^2 = \{(a_{\mathcal{S}}, k) \in \mathcal{A}' \times \{3\} \mid i \in \mathcal{S}\}$ and $Y_i^3 = \{d\} \times \{4, 7, \ldots, 3q+1\}$. Let $Y_{3q+1} = \{(c, 2), (b, 3q+1)\}$ such that $(c, 2) \succ_{3q+1} (b, 3q+1)$. Let $Y_{3q+2} = (\{d\} \times \{4, 7, \ldots, 3q+1\}) \cup \{(c, 2)\}$ such that $(d, k) \succ_{3q+2} (c, 2)$ for each $(d, k) \in \{d\} \times \{4, 7, \ldots, 3q + 1\}$. Let $\succeq = (\succeq_1, \succeq_2, \ldots, \succeq_{3q+2})$.

Observation 2. *From an instance of X3C, an instance of GC is constructed in polynomial time of $|\mathcal{R}|$ and $|S|$*

It is obvious to be able to construct $\mathcal{A}$ and $\mathcal{N}$ in polynomial time. In addition, preference $\succeq_i$ of each player $i \in \mathcal{N}$ can be constructed in polynomial time by setting $\textsc{rank}_{\succeq_i}(a, k)$ for each $(a, k) \in X$ and constructing a binary relation based on that rank.

Lemma 1. *A "yes" answer to X3C implies a "yes" answer to GC.*

Proof. If An answer to X3C is "yes", there exists S'. We prove that an assignment μ', where for each $\mathcal{S} \in S'$, and for each $i \in S, \mu'(i) = a_{\mathcal{S}}$ and $\mu'(3q+1) = c, \mu'(3q+2) = c$, is core stable.

We show that there is not a core deviation for each activity. For each player $i \in \mathcal{R}$ and for each $(a_{\mathcal{S}}, k), (a'_{\mathcal{S}}, k') \in \{a_{\mathcal{S}} \mid \mathcal{S} \in S\} \times [1, 3q+2]$, $(a_{\mathcal{S}}, k) \succ_i (v_i, 1) \wedge (a'_{\mathcal{S}}, k') \succ_i (v_i, 1)$ implies $(a_{\mathcal{S}}, k) \sim (a'_{\mathcal{S}}, k')$. Hence, there is not a core deviation for each activity $a_{\mathcal{S}} \in \{a'_{\mathcal{S}} \mid \mathcal{S} \in S\}$. Only each player $i \in \mathcal{R}$ strictly prefer $(\mu'(i), |G_{\mu'}(\mu'(i))|)$ to alternatives $(b, 3q+1)$. However, $|\mathcal{R}| = 3q$, there is not a core deviation for the activity b. Each player $i \in \mathcal{N}$ satisfies $(\mu'(i), |G_{\mu'}(\mu'(i))|) \succeq_i x$ for each alternative $x \in \{c\} \times [1, 3q+2]$, so there is not a core deviation for the activity c. Only the player $3q+2$ strictly prefer $(\mu'(3q+2), |G_{\mu'}(\mu'(3q+2))|)$ to alternatives (d, k) such that $k \in \{4, 7, \ldots, 3q+1\}$. there is not a core deviation for the activity d. Hence, μ' is core stable. □

Lemma 2. *A "yes" answer to GC implies a "yes" answer to X3C.*

Proof. We show that any IR assignment μ is not core stable. We set $S_\mu = \{\mathcal{S} \in S \mid \exists i \in \mathcal{N}[\mu(i) = a_{\mathcal{S}}]\}$. For each player $i \in \mathcal{R}$ and for each $\mathcal{S} \in S$, if $i \in \mathcal{S}$, it satisfies $(a_{\mathcal{S}}, 3) \succ_i (v_i, 1)$, so $G_\mu(a_{\mathcal{S}}) = \mathcal{S}$ for each $\mathcal{S} \in S_\mu$. From assumption, there is no S', so $|S_\mu| \neq q$. In addition, $|S_\mu| < q$ from each player $i \in \mathcal{N}$ should be assigned at most one activity. Let $\mathcal{P} = \mathcal{N} \setminus \bigcup_{\mathcal{S} \in S_\mu} \mathcal{S}$, $\mathcal{P}$ satisfies $\mathcal{P} \neq \emptyset$. For any IR assignment μ,

- $\mu(3q+1) = v_{3q+1}, \mu(3q+2) = v_{3q+2}$ implies there is core deviation $(\{3q+1, 3q+2\}, c)$.
- $\mu(3q+1) = c, \mu(3q+2) = c$ implies there is core deviation $(\mathcal{P} \cup \{3q+2\}, d)$.
- $\mu(i) = d$ for each $i \in \mathcal{P} \cup \{3q+2\}$ implies there is core deviation $(\mathcal{R} \cup \{3q+1\}, b)$.
- $\mu(i) = b$ for each $i \in \mathcal{R} \cup \{3q+1\}$ implies there is core deviation $(\{3q+1, 3q+2\}, c)$.

Therefore, there is not core stable assignment. □

Theorem 3. *It is NP-complete to decide whether GAS admits a CS assignment.*

Proof. We can verify of μ is indeed core stable in polynomial time of n and m. Hence, this theorem follows from lemma 1 and lemma 2. □

Dramann [1] stated that it is NP-complete to decide whether GAS admits a core stable assignment even when preference profile is strict and increasing. So we focus preference profile is strict and increasing.

Given an instance $(\mathcal{R}, S)$ of X3C, we construct an instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GC which preference profile is strict and increasing as follows. The set of players $\mathcal{N}$ and the set of activities $\mathcal{N}$ is same as general case. The preference profile is as follow. Let $Y_i = \{x \in X \mid x \succ_i (v_i, 1)\}$ for each player $i \in \mathcal{N}$. For each player $i \in \mathcal{R}$, let $Y_i = Y_i^1 \cup Y_i^2 \cup Y_i^3$,

- where $x \succeq_i y$, for each $s \in [1, 2]$, for each $x \in Y_i^s$, and for each $x \in Y_i^{s+1}$.

For each player $i \in \mathcal{R}$, let $Y_i^1 = \{(b, 3q+1), (b, 3q+2)\}$, $Y_i^2 = \{(\alpha_{\mathcal{S}}, k) \in \mathcal{A}' \times [1, 3q+2] \mid i \in \mathcal{S} \wedge k \geq 3\}$ and $Y_i^3 = \{(d, \ell) \mid \ell \geq 2\}$. In addition, for each $\mathcal{S} \in S$, there is a player $i \in \mathcal{S}$ such that,

- $i \in \mathcal{S}'$ implies $(\alpha_{\mathcal{S}'}, 3) \succ_i (\alpha_{\mathcal{S}}, 3)$ for each $\mathcal{S}' \in S \setminus \{\mathcal{S}\}$.

We set $Y_{3q+1} = \{c, b\} \times [2, 3q+2]$ such that $(c, 2) \succ_{3q+1} (b, 3q+2)$ and $Y_{3q+2} = \{c, d\} \times [2, 3q+2]$ such that $(d, 2) \succ_{3q+2} (c, 3q+2)$. Let $\succeq = (\succeq_1, \succeq_2, \ldots, \succeq_{3q+2})$.

It is known that X3C remains NP-complete even if each $i \in \mathcal{R}$ occurs in at most three members of S [3]. We use this assumption in the proof of lemma 3, to prove there always exists instance above. Let $\mathcal{M}_S = \{a_{\mathcal{S}} \mid \mathcal{S} \in S\}$. Let a bipartite graph $G_{(\mathcal{N}, \mathcal{A}, \succeq)}$, where $V(G_{(\mathcal{N}, \mathcal{A}, \succeq)}) = \mathcal{M}_S \cup \mathcal{R}$ and $E(G_{(\mathcal{N}, \mathcal{A}, \succeq)}) = \{\{\alpha_{\mathcal{S}}, i\} \in \mathcal{M}_S \cup \mathcal{R} \mid \mathcal{S} \in S \wedge i \in \mathcal{S}\}$. We show there exists the complete matching of $G_{(\mathcal{N}, \mathcal{A}, \succeq)}$ from $\mathcal{M}_S$ to $\mathcal{R}$ from Hall's theorem.

Lemma 3. *For any X3C instance, there is an instance of GC which preference profile is strict and increasing.*

Proof. From $|\mathcal{S}| = 3$ for each $\mathcal{S} \in S$, each $a_{\mathcal{S}} \in \mathcal{M}_S$ connects three vertex of $\mathcal{R}$. In addition, from each $i \in \mathcal{R}$ occurs in at most three members of S, each vertex $i \in \mathcal{R}$ connects three vertex of $\mathcal{M}_S$. We assume there is not a complete matching. From Hall's theorem, there is $\mathcal{U} \subseteq \mathcal{M}_S$ such that $|\Gamma(\mathcal{U})| < |\mathcal{U}|$. There is a vertex $a_{\mathcal{S}} \in \mathcal{U}$ which connects at most two vertex of $\mathcal{R}$ or there is a vertex $i \in \mathcal{R}$ which connects greater than or equal to four vertex of $\mathcal{M}_S$. Therefore, from Hall's theorem, there exists a complete matching of $G_{(\mathcal{N},\mathcal{A},\succeq)}$ from $\mathcal{M}_S$ to $\mathcal{R}$. □

From the bipartite graph $G_{(\mathcal{N},\mathcal{A},\succeq)}$ contains a complete matching, maximum matching is a complete matching from $\mathcal{M}_S$ to $\mathcal{R}$. Find a maximum matching $\mathcal{M}_S$ to $\mathcal{R}$ by Hopcroft-Karp algorithm [4], whose running time is $\mathcal{O}(|E|\sqrt{|V|})$.

Observation 4. *From instance of X3C, an instance of GC which preference profile is strict and increasing is constructed in polynomial time of $|\mathcal{R}|$ and $|S|$.*

From $(\mathcal{R}, S)$, an instance $(\mathcal{N}, \mathcal{A}, \succeq)$ is constructed in polynomial time of n and m. It is obvious to be able to construct $\mathcal{A}$ and $\mathcal{N}$ in polynomial time. In addition, preference $\succeq_i$ of each player $i \in \mathcal{N}$ can be constructed in polynomial time by setting rank $\text{RANK}_{\succeq_i}(a, k)$ for each $(a, k) \in X$ based on complete matching of G from $\mathcal{M}_S$ to $\mathcal{R}$.

Theorem 5. *It is NP-complete to decide whether GAS admits a core stable assignment even when preference profile is strict and increasing.*

Proof. We can verify of μ is indeed core stable in polynomial time of n and m. Therefore we prove that an answer to X3C is "yes" if and only if an answer to GC is "yes".

First, we show a "yes" answer to X3C implies a "yes" answer to GC. If an answer to X3C is "yes", there exists S'. We show that an assignment μ', where $\mu'(i) = a_{\mathcal{S}}$ for each $\mathcal{S} \in S'$ and for each $i \in S$, and $\mu'(3q+1) = c, \mu'(3q+2) = c$, is core stable. For each alternative $(a_{\mathcal{S}}, k) \in \mathcal{M}_S \times [1, 3q+2]$, let $\mathcal{P}(a_{\mathcal{S}}, k) = \{i \in \mathcal{N} \mid (a_{\mathcal{S}}, k) \succ_i (v_i, 1)\}$. For each activity $a_{\mathcal{S}} \in \mathcal{M}_S$, $k < 3$ implies $P(a_{\mathcal{S}}, k) = \emptyset$ and $k \geq 3$ implies $P(a_{\mathcal{S}}, k) = \mathcal{S}$. From assumption, there is a player $i \in \mathcal{S}$ such that $(\mu'(i), |G_{\mu'}(\mu'(i))|) \succ_i (\alpha_{\mathcal{S}}, 3)$. Hence, there is no core deviation $(\mathcal{S}, a_{\mathcal{S}})$ for each activity $a_{\mathcal{S}} \in \mathcal{M}_S$. There is no core deviation related to the activity b, c and d as the proof of theorem 2. Hence, the assignment μ' is core stable.

Secondly, we show a "yes" answer to GC implies a "yes" answer to X3C. This proof is the same as the proof of theorem 2. For any IR assignment, there exists core deviation. □

4 Strictly Core Stability

In the following, define Existence of a SC stable assignment as a decision problem.

Existence of a SC stable assignment (GSC)

Instance : Instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GAS.

Question : Is there an assignment which is core stable in $(\mathcal{N}, \mathcal{A}, \succeq)$?

We start by first explaining how we construct an instance of GSC from an instance of X3C. Given an instance $(\mathcal{R}, S)$of X3C, we construct an instance $(\mathcal{N}, \mathcal{A}, \succeq)$of GSC as follows. Let $\mathcal{N} = \mathcal{R}$, and $\mathcal{A} = \mathcal{M} \cup \mathcal{V}$, where $\mathcal{M} = \{a_{\mathcal{S}} \mid \mathcal{S} \in S\}$ and $\mathcal{V} = \{v_1, v_2, \ldots, v_{3q}\}$. For each player $i \in \mathcal{N}$, let preference $\succeq_i$ such that, $\{x \in \mathcal{A} \times [1, 3q] \mid x \succ_i (v_i, 1)\} = \{(a_{\mathcal{S}}, k) \in \mathcal{M} \times [1, 3q] \mid i \in \mathcal{S} \wedge k \geq 3\}$ and $x \sim_i y$ for each alternative $x, y \in \{x \in \mathcal{A} \times [1, 3q] \mid x \succ_i (v_i, 1)\}$ and let $\succeq = (\succeq_1, \succeq_2, \ldots, \succeq_{3q})$.

Observation 6. *The above preference profile is a-GAS.*

Observation 7. *From instance of X3C, an instance of GSC is constructed in polynomial time of $|\mathcal{R}|$ and $|S|$.*

Observation 8. *We can verify an assignment is SC stable in polynomial time of $|\mathcal{R}|$ and $|S|$.*

We can also verify an assignment μ is indeed SC stable in polynomial time of n and m by scanning the whole profile for each $a \in \mathcal{A}$ and for each $k \in [1, n]$.

Theorem 9. *It is NP-complete to decide whether GAS admits SC assignment even when preference profile is a-GAS.*

Proof. We can verify of μ is indeed SC in polynomial time n and m, so we prove that an answer to X3C is "yes" if and only if an answer to GSC is "yes".

First, we show that a "yes" answer to X3C implies a "yes" answer to GSC. If an answer to X3C is "yes", there exists S'. Let an assignment μ', where $\mu'(i) = a_{\mathcal{S}}$, for each $\mathcal{S} \in S'$ and for each $i \in S$. From assumption, it is obvious that the assignment μ' is SC stable.

Secondly, we show that a "yes" answer to GSC implies a "yes" answer to X3C. We shows that any IR assignment μ is not core stable. We set $S_\mu = \{\mathcal{S} \in S \mid \exists i \in N[\mu(i) = a_{\mathcal{S}}]\}$. For each player $i \in \mathcal{R}$ and for each $\mathcal{S} \in S$, if $i \in \mathcal{S}$, it satisfies $(a_{\mathcal{S}}, 3) \succ_i (v_i, 1)$, so $G_\mu(a_{\mathcal{S}}) = \mathcal{S}$ for each $\mathcal{S} \in S_\mu$. By assumption, there is no S'. Hence $|S_\mu| \neq q$. In addition, from each player $i \in \mathcal{N}$ should be assigned at most one activity, S_μ satisfies $|S_\mu| < q$. Thus, let $\mathcal{P}$ is $\mathcal{N} \setminus \bigcup_{\mathcal{S} \in S_\mu} \mathcal{S}$, it satisfies $\mathcal{P} \neq \emptyset$. Each player $i \in \mathcal{P}$ satisfies $(a_{\mathcal{S}}, 3) \succ_i (\mu(i), |G_\mu(\mu(i))|)$ for $\mathcal{S} \in \{\mathcal{S}' \in S \mid i \in \mathcal{S}'\}$. Therefore, there exists SC deviation $(\mathcal{S}, a_{\mathcal{S}})$ for each $\mathcal{S} \in \{\mathcal{S}' \in S \mid \exists i \in \mathcal{P}[i \in \mathcal{S}']\}$ □

Darmann [1] pointed out that An assignment μ is core stable if and only if μ is SC stable when preference profile is strict. Hence, it is also NP-complete to decide whether GAS admits a SC stable assignment even when preference profile is increasing and strict.

5 Contractually Core Stability

In the following, we show that we can verify whether an assignment is CC stable in $\mathcal{O}(n^3 m)$ time, by scanning the whole profile. First, we define the extension knapsack algorithm.

Extension knapsack problem

Instance : Collection of Item Set $\mathcal{ISS} = \{IS_1, IS_2, \ldots, IS_p\}$, and Capacity W

- Each item $j \in IS_i$ has a weight w_j and a profit c_j, for each $IS_i \in \mathcal{ISS}$

Task : Find $S \subseteq \bigcup_{IS \in \mathcal{ISS}} IS$ such that $|S \cap IS_i| \leq 1$ for each $IS_i \in \mathcal{ISS}$, $\sum_{j \in S} w_j \leq W$ and $\sum_{j \in S} c_j$ is maximum.

Let $\ell = |\bigcup_{IS \in \mathcal{ISS}} IS|$. We finds optimum solution in $\mathcal{O}(\ell W)$ by dynamic programming algorithm [5].

INPUT : $\mathcal{ISS} = \{IS_1, IS_2, \ldots, IS_p\}$

OUTPUT : $S \subseteq \bigcup_{IS \in \mathcal{ISS}} IS$

Step 1. Set $x(0,0) := 0$ and $x(0,k) := 0$ for $k = 1, \ldots, W$

Step 2. For $i := 1$ to $|\mathcal{ISS}|$ do :

Step 2.1. For $k := 1$ to W do :

Step 2.1.1 $x(i,k) := x(i-1,k), interim := 0$

Step 2.1.2 Repeat the following steps for each item $j \in IS_i$

If $w_j \leq k \wedge x(j-1, k-w_j) + c_j > x(i,k)$ then,
$s(i,k,interim) := 0$, $s(i,k,j) := 1$, $interim := j$ and $x(i,k) := x(i-1, k-w_j) + c_j$

Step 3. Let $k := W$, Set $S := \emptyset$,

Step 3.1. For $i := |\mathcal{ISS}|$ down to 1 do :

For each item $j \in IS_i$, if $s(i,k,j) = 1$ then $S := S \cup \{j\}$, $k := k - w_j$.

Lemma 4. *The running time of the proposed algorithm is $\mathcal{O}(\ell W)$ time.*

Proof. Now we consider the running time of Step 2. Step 2 repeats $\mathcal{O}(|\mathcal{ISS}|)$ times. Step 2.1 repeats $\mathcal{O}(W)$ times and Step 2.1.2 repeats $\mathcal{O}(|IS_i|)$ times. Hence, the running time of the proposed algorithm is $\mathcal{O}(\ell W)$ time. □

Theorem 10. *The proposed algorithm finds an optimum solution in $\mathcal{O}(\ell W)$ time.*

Proof. The variable $x(j,k)$ denotes the minimum total weight of a subset $S \subseteq IS_1 \cup, \ldots \cup IS_i$ with $\sum_{j\in S} w_j \leq W$ and $\sum_{j\in S} c_j = k$. The algorithm correctly computes these values using the recursion formulas

$$x(i,k) = \begin{cases} x(i-1,k-w_i)+c_i & \text{if } w_j \leq k \wedge x(i-1,k-w_i)+c_j > x(i,k) \\ x(i-1,k) & \text{otherwise} \end{cases}$$

for $i = 1,\ldots,|\mathcal{ISS}|$ and $k = 0,\ldots,W$. For each $j \in IS_i$, the variables $s(i,k,j)$ indicate which of these two cases applies. So the algorithm enumerates all subsets $S \subseteq \bigcup_{IS\in\mathcal{ISS}} IS$ except those that are infeasible or those that are dominated by others: S is said to be dominated by S' if $\sum_{i\in S} c_i = \sum_{i\in S'} c_i$ and $\sum_{i\in S} w_i \leq \sum_{i\in S'} w_i$. In Step 3. the best feasible subset is chosen. □

For each $a \in \mathcal{A}$ and for each $k \in [1,n]$, let $\mathcal{T}_\mu(a,k) = \{i \in \mathcal{N} \mid (a,k) \succ_i (\mu(i), |G_\mu(\mu(i))|)\}$. For each $a \in \mathcal{A}$ and for each $k \in [1, |G_\mu(a)|-1]$, let

- $cont_\mu(k,a) = \{i \in G_\mu(a) \mid (a,|G_\mu(a)|) \succ_i (a, |G_\mu(a)|-k)\}$.

We show that we can verify whether an assignment is CC stable in $\mathcal{O}(n^3m)$.

INPUT : An instance $(\mathcal{N},\mathcal{A},\succeq)$ of GAS and an assignment μ.

OUTPUT : A CC deviation on μ if it exists. "no" otherwise.

Step 1. Set $\mathcal{T}_\mu(a,k)$ for each $a \in \mathcal{A}$ and for each $k \in [1,n]$.

Step 2. Set $cont_\mu(k,a)$ for each $a \in \mathcal{A}$ and for each $k \in [1,|G_\mu(a)|-1]$.

Step 3. Repeat the following steps for each $a \in \mathcal{A}$ and for each value $k \in [1,n]$.

Step 3.1. if $|\mathcal{T}_\mu(a,k)| \geq k \wedge G_\mu(a) \subseteq \mathcal{T}_\mu(a,k) \wedge |G_\mu(\alpha)| < k$, then go to Step 3.2.

Step 3.2. Repeat the following steps for each $b \in \{a' \in \mathcal{A}\backslash\{a\} \mid \exists i \in \mathcal{T}_\mu(a,k)[\mu(i) = a']\}$ and for each $s \in [1, |\mathcal{T}_\mu(a,k) \cap G_\mu(b)|]$.

Step 3.2.1 If there exists $\mathcal{P} = \{i \in G_\mu(b) \mid (a,k) \succ_i (\mu(i), |G_\mu(\mu(i))|)\}$ such that $|\mathcal{P}| = s \wedge cont_\mu(s,b) \subseteq \mathcal{P}$ then $dev(b) := dev(b) \cup \{P\}$.

Step 3.3. Find $\mathcal{T} \subseteq \mathcal{T}_\mu(a,k)$ such that $|\mathcal{T}| = k - |G_\mu(a)|$ by extension knapsack algorithm.

Step 3.3.1 If there exists, return $(\mathcal{T} \cup G_\mu(a), a)$.

Step 3.4. For each $b \in \mathcal{A} \setminus \{a\}$, $dev(b) := \emptyset$.

Step 4. return "no".

We apply extension knapsack problem in Step 3.3. In more details, we can set $dev(b)$ as a set of item IS for each activity $b \in \{a' \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}_\mu(a,k)[\mu(i) = b]\}$, and set each item's value and weight as the size of each element of $dev(b)$. we set $k - |G_\mu(a)|$ as the capacity W.

Lemma 5. *The running time of the proposed $\mathcal{O}(n^3m)$ time.*

Proof. Step 1 is $\mathcal{O}(n^2m)$ time and Step 2 is $\mathcal{O}(nm)$ time. Now we consider the running time of Step 3. Step 3 repeats nm times and From lemma 4, the running time of Step 3.3 is time $\mathcal{O}(n^2)$. Therefore, the total running time is $\mathcal{O}(n^3m)$ time. □

For CC deviation $(\mathcal{T},a)$, let $t_{a'} = |G_\mu(a') \cap \mathcal{T}|$ for each $a' \in \mathcal{A} \setminus \{a\}$. The correctness of the algorithm is described by showing that $cont_\mu(t_{a'},a') \subseteq \mathcal{T}$ for each $a' \in \{b \in \mathcal{A}\backslash\{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$ if and only if core deviation $(\mathcal{T},a)$ is CC deviation.

Theorem 11. *The proposed algorithm find a CC deviation in $\mathcal{O}(n^3m)$.*

Proof. First, we show that $cont_\mu(t_{a'},a) \subseteq \mathcal{T}$ for each $a' \in \{b \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$ implies core deviation $(\mathcal{T},a)$ is CC deviation. We assume that, for each $a' \in \{b \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$, $cont_\mu(t_{a'},a') \subseteq \mathcal{T}$ and a pair $(\mathcal{T},a)$ is not CC deviation. From the assumption, if $cont_\mu(t_{a'},a') \subseteq \mathcal{T}$ for each $a' \in \{b \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$, then any player $i \in \mathcal{N} \setminus \mathcal{T}$ does not satisfy $(\mu(i), |G_\mu(\mu(i)) \setminus \mathcal{T}|) \succeq_i (\mu(i), |G_\mu(\mu(i))|)$. Hence, the pair $(\mathcal{T},a)$ is CC deviation. This contradicts that $(\mathcal{T},a)$ is not CC deviation.

Next, we show that, if core deviation $(\mathcal{T}, a)$ is CC deviation, $cont_\mu(t_{a'}, a) \subseteq \mathcal{T}$ for each $a' \in \{b \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$. We assume that, for some $a' \in \{b \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$, $cont_\mu(t_{a'}, a') \not\subseteq \mathcal{T}$ and core deviation $(\mathcal{T}, a)$ is CC deviation. From the assumption, each player $i \in \mathcal{N} \setminus \mathcal{T}$ satisfies $(\mu(i), |G_\mu(\mu(i)) \setminus \mathcal{T}|) \succeq_i (\mu(i), |G_\mu(\mu(i))|)$. From the assumption, there exists $a' \in \{b \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$ such that $cont_\mu(t_{a'}, a') \not\subseteq \mathcal{T}$, so, there is player $i \in G_\mu(a) \setminus \mathcal{T}$ such that $(\mu(i), |G_\mu(\mu(i))|) \succ_i (\mu(i), |G_\mu(\mu(i)) \setminus \mathcal{T}|)$. This contradicts that $(\mathcal{T}, a)$ is CC deviation. □

A CC stable assignment can be obtained by the following algorithm.

INPUT : An instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GAS.

OUTPUT : An assignment μ.

Step 1. Set $\mu(i) = v_i$ for each $i \in \mathcal{N}$.

Step 2. Repeat the following steps until there exists no CC deviation from μ.

Step 2.1. Find an arbitrary CC deviation $(\mathcal{T}, a)$

Step 2.2. Set $\mu(i) = a$ for each $i \in \mathcal{T}$.

Step 3. Return μ

Lemma 6. *The running time of the proposed algorithm is $\mathcal{O}(n^5m^2)$ time.*

Proof. Step 2 repeats at most $n(nm-1)$ times. Step 2.1. is $\mathcal{O}(n^3m)$ time. Therefore, the running time of the proposed algorithm is $\mathcal{O}(n^5m^2)$ time. □

Theorem 12. *The proposed algorithm construct a CC stable assignment in $\mathcal{O}(n^5m^2)$ time.*

Proof. By implementing of a CC deviation $(\mathcal{T}, a)$, no one worse off and there exists some player in $\mathcal{T}$ such that strictly better off. Hence, above algorithm will be terminated in a finite number of steps (at most $n(nm-1)$ implementations). When the algorithm does terminate, it returns an assignments which no CC deviation exists, which an assignments is CC stable. □

6 Contractual Strictly Core Stability

In the following, we show that a CSC stable assignments of each instance can be constructed in $\mathcal{O}(n^6m^2)$ time. First, we show that we can also verify whether an assignment is CSC stable in $\mathcal{O}(n^4m)$ as the algorithm of CC stable. For each $a \in \mathcal{A}$ and for each $k \in [1, n]$, let $\mathcal{WT}_\mu(a, k) = \{i \in \mathcal{N} \mid (a, k) \succ_i (\mu(i), |G_\mu(\mu(i))|)\}$.

INPUT : An instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GAS and an assignment μ.

OUTPUT : A CSC deviation on μ if it exists. "no" otherwise.

Step 1. Set $\mathcal{WT}_\mu(a, k)$ for each $a \in \mathcal{A}$ and for each$k \in [1, n]$.

Step 2. Set $cont_\mu(k, a)$ for each $a \in \mathcal{A}$ and for each $k \in [1, |G_\mu(a)| - 1]$.

Step 3. Repeat the following steps for each $a \in \mathcal{A}$ and for each value $k \in [1, n]$.

Step 3.1. If $|\mathcal{WT}_\mu(a, k)| \geq k \wedge G_\mu(a) \subseteq \mathcal{WT}_\mu(a, k) \wedge |G_\mu(\alpha)| < k$, then go to Step 3.2.

Step 3.2. Repeat the following steps for each $b \in \{a' \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{WT}_\mu(a, k)[\mu(i) = a']\}$ and for each $s \in [1, |\mathcal{WT}_\mu(a, k) \cap G_\mu(b)|]$.

Step 3.2.1 If there exists $\mathcal{P} = \{i \in G_\mu(b) \mid (a, k) \succ_i (\mu(i), size_\mu(i))\}$ such that $|\mathcal{P}| = s \wedge cont_\mu(s, b) \subseteq \mathcal{P}$ then $dev(b) := dev(b) \cup \{P\}$.

Step 3.3. Find $\mathcal{T} \subseteq \mathcal{WT}_\mu(a, k)$ such that $|\mathcal{T}| = k - |G_\mu(a)|$ by extension knapsack algorithm.

Step 3.3.1 If there exists and it satisfies $(a, k) \succ_i (\mu(i), |G_\mu(\mu(i))|)$ for some player $i \in \mathcal{T} \cup G_\mu(a)$, return $(\mathcal{T} \cup G_\mu(a), a)$.

Step 3.4. For each $b \in \mathcal{A} \setminus \{a\}$, $dev(b) := \emptyset$.

Step 4. return "no".

Step 3.3 is applied by the extension of knapsack algorithm. However, there is a difference from the algorithm of CC stable. For CSC deviation $(\mathcal{S}, a)$, there exists a player $i \in \mathcal{S}$ such that $(a, |\mathcal{S}|) \succ_i (\mu(i), |G_\mu(mu(i))|)$. Hence, we put one item before applying the extension of knapsack and apply the extension of knapsack algorithm at most $n-1$ times more than the algorithm of CCS.

Lemma 7. *The running time of the proposed algorithm is $\mathcal{O}(n^4m)$ time.*

Proof. Step 1 of the algorithm is $\mathcal{O}(n^2m)$ time. Step 2 of the algorithm is $\mathcal{O}(nm)$ time. Now we consider the running time of Step 3. Step 3 repeats $\mathcal{O}(nm)$ times. In a straightforward way, Step 3.1 can be implemented with running time $\mathcal{O}(1)$. Step 3.2 can be implemented with running time $\mathcal{O}(n^2)$ and Step 3.3 can be implemented with running time $\mathcal{O}(n^3)$. Therefore, the total running time is $\mathcal{O}(n^4m)$ time. □

Theorem 13. *CSC deviation can be found in $\mathcal{O}(n^4m)$.*

Proof. For CSC deviation $(\mathcal{T}, a)$, let $t_{a'} = |G_\mu(a') \cap \mathcal{T}|$ for each $a' \in \mathcal{A} \setminus \{a\}$. The correctness of the algorithm is described by showing that $cont_\mu(t_{a'}, a') \subseteq \mathcal{T}$ for each $a' \in \{b \in \mathcal{A} \setminus \{a\} \mid \exists i \in \mathcal{T}[\mu(i) = b]\}$ if and only if $(\mathcal{T}, a)$ is CSC deviation. This proof is the same way as proof of theorem 11. □

A CSC stable assignment can be obtained by the following algorithm.

INPUT : An instance $(\mathcal{N}, \mathcal{A}, \succeq)$ of GAS.

OUTPUT : An assignment μ.

Step 1. Set $\mu(i) = v_i$ for each $i \in \mathcal{N}$.

Step 2. Repeat the following steps until there exists no CSC deviation from μ.

Step 2.1. Find an arbitrary CSC deviation $(\mathcal{T}, a)$

Step 2.2. Set $\mu(i) = a$ for each $i \in \mathcal{T}$.

Step 3. Return μ

Lemma 8. *The running time of the proposed algorithm is $\mathcal{O}(n^6m^2)$ time.*

Proof. Step 2 repeats at most $n(mn-1)$ times. The Step 2.1. is $\mathcal{O}(n^4m)$ time. Therefore, the running time of the proposed algorithm is $\mathcal{O}(n^5m^2)$ time. □

Theorem 14. *The proposed algorithm construct a CSC stable assignment in $\mathcal{O}(n^5m^2)$ time.*

Proof. By implementing of a CSC deviation $(\mathcal{T}, a)$, no one worse off and there exists some player in $\mathcal{T}$ such that strictly better off. Hence, above algorithm will be terminated in a finite number of steps (at most $n(nm-1)$ implementations). When the algorithm does terminate, it returns an assignments which no CSC deviation exists, which an assignments is CSC stable. □

For the *strict* preference, we show that the CSC stable assignments can be efficiently constructed. We show that a CSC stable assignments of each instance can be constructed in $\mathcal{O}(n^2m)$ time.

INPUT : An strict GAS instance $(\mathcal{N}, \mathcal{A}, \succeq)$

OUTPUT : An assignment μ

Step 1. Set $\mu(i) := v_i$, for all $i \in \mathcal{N}$ and $\mathcal{P} := \mathcal{N}, \mathcal{B} = \mathcal{A}$.

Step 2. Repeat the following steps until $\mathcal{P} = \emptyset$

Step 2.1. Select a player $i \in \mathcal{N}$

Step 2.2. Find $(a, k) \in M \times [1, |\mathcal{P}|]$ such that $(a, k) \succ_i (b, s)$ for each $(b, s) \in (\mathcal{B} \times [1, |\mathcal{P}|]) \setminus \{(a, k)\}$

Step 2.3. If $(a, k) \succ_i (v_i, 1)$, set $b := a$ and $\mathcal{S} \subseteq \mathcal{P}$ such that $|\mathcal{S}| = k \wedge i \in \mathcal{S}$. Otherwise set $b := v_i$ and $\mathcal{S} := \{i\}$

Step 2.4. For each $j \in \mathcal{S}$, $\mu(j) := b$. Set $\mathcal{P} := \mathcal{P} \setminus \mathcal{S}$ and $\mathcal{B} := \mathcal{B} \setminus \{b\}$

Step 3. Return μ.

For instances of GAS, the ranking calculation time $\mathcal{O}(n^3m^2)$ can not be included in the calculation time of the proposed algorithm. From this reason, we will discuss it supposing that it is an *oracle* preference profile. We define the oracle for query about a player's preference. Before we proceed the definition of the oracle, we define query about a player's preference as a decision problem.

Query about player's preference

Instance : GAS instance $(\mathcal{N}, \mathcal{A}, \succeq)$, player $i \in \mathcal{N}$, Alternative $x, y \in X$.

Question : Player i weakly prefer x than y?

An *oracle* problem query about player's preference is an abstract device which, for any instance, returns "yes" or "no". It is assumed that the oracle may return the answer in just one step.

Lemma 9. *The running time of the proposed algorithm is $\mathcal{O}(n^2m)$ time.*

Proof. Step 2 repeats $\mathcal{O}(n)$ times. Step 2.2 can be implemented with running time $\mathcal{O}(nm)$. Therefore, the total running time is $\mathcal{O}(n^2m)$ time. □

Theorem 15. *A CSC stable assignments of each instance with strict preference can be constructed in $O(n^2m)$ time.*

Proof. Let activities which players are assigned to at each repetition of Step 2 be $(b_1, b_2, \ldots, b_q)$, and let $\mathcal{P}_i = \{j \in \mathcal{N} \mid \mu(j) = b_i\}$. It satisfies that collect all $\mathcal{P}_i$ is a partition of $\mathcal{N}$. For each repetition $k \in [1, q]$, it always holds that

- There exists player $i \in \mathcal{P}_r$ such that $(b_r, |\mathcal{P}_r|) \succ_i (b_r, |\mathcal{P}_r| + s)$ for each $r \in [1, k-1]$ and for each $s \in [1, |\mathcal{N} \setminus \bigcup_{i=1}^{k} \mathcal{P}_i|]$.
- There exists player $i \in \mathcal{P}_k$ such that $(b_k, |\mathcal{P}_k|) \succ_i (b_r, |\mathcal{P}_r| - s)$ for each $r \in [k+1, q]$ and for each $s \in [1, |\mathcal{N} \setminus \bigcup_{i=1}^{k} \mathcal{P}_i|]$
- There exists player $j \in \mathcal{P}_k$ such that $(b_k, |\mathcal{P}_k|) \succ_j (b_k, |\mathcal{P}_k| - s)$ for each $s \in [1, |\mathcal{N} \setminus \bigcup_{i=1}^{k} \mathcal{P}_i|]$

Hence output μ is CSC stable assignment. □

Contractually strict core stability implies contractually core stability. Hence, CC stable assignments of each instance can be constructed in $\mathcal{O}(n^2m)$ time when preference profile is strict.

References

[1] A. Darmann, 'Group Activity Selection from Ordinal Preferences' *Algorithmic Decision Theory Volume 9346*, pp 35-51, (2015).

[2] A. Darmann, E. Elkind, J. Lang, S. Kurz, J. Schauer, G. Woeginger, 'Group Activity Selection Problem' *Internet and Network Economics Volume 7695*, pp 156-169, (2012).

[3] M.R. Garey, D.S. Johnson, 'Computers and Intractability; A Guide to the Theory of NP-Completeness' *W. H. Freeman & Co.*, (1990)

[4] J.E. Hopcroft, R.M. Karp, 'An $n^{5/2}$ algorithm for maximum matching in bipartite graphs.' *SIAM J. Comput 2*, pp 225-231, (1973)

[5] S. Martello, P. Toth, 'Knapsack Problems: Algorithms and Computer Implementations' *John Wiley & Sons, Inc* (1990).

[6] S.C. Sung, D. Dimitrov, 'On Myopic Stability Concepts for Hedonic Games' *Theory and Decision Volume 62*, pp 31-45, (2007).

Question Selection Methods for Adaptive Testing with Bayesian Networks

Martin PLAJNER[1] and Amal MAGAUINA[1] and Jiří VOMLEL[2]

[1] *Faculty of Nuclear Sciences and Physical Engineering,*
Czech Technical University, Prague
Trojanova 13, Prague, 120 00, Czech Republic
martin.plajner@fjfi.cvut.cz

[2] *Institute of Information Theory and Automation,*
Czech Academy of Sciences,
Pod vodárenskou věží 4, Prague 8, 182 08, Czech Republic

Abstract

The performance of Computerized Adaptive Testing systems, which are used for testing of human knowledge, relies heavily on methods selecting correct questions for tested students. In this article we propose three different methods selecting questions with Bayesian networks as students' models. We present the motivation to use these methods and their mathematical description. Two empirical datasets, paper tests of specific topics in mathematics and Czech language for foreigners, were collected for the purpose of methods' testing. All three methods were tested using simulated testing procedure and results are compared for individual methods. The comparison is done also with the sequential selection of questions to provide a relation to the classical way of testing. The proposed methods are behaving much better than the sequential selection which verifies the need to use a better selection method. Individually, our methods behave differently, i.e., select different questions but the success rate of model's predictions is very similar for all of them. This motivates further research in this topic to find an ordering between methods and to find the best method which would provide the best possible selections in computerized adaptive tests.

Keywords: Computerized Adaptive Testing, Question Selection, Bayesian Networks.

1 Introduction

In our research we focus on Computerized Adaptive Testing (CAT). In CAT there is not a single static version of a test distributed to many students but an individual test is dynamically created during the course of testing for each individual participant. The next question is selected with regard to student's previous answers. This leads to several benefits as a better student assessment, a better motivation, etc. [4, 8]

We employ Bayesian networks for our research in this domain. The most recent papers we have published consider the beneficial effect of monotonicity conditions while learning model parameters. In this paper we aim at the testing process itself while the network is already learned. We take a closer look at the question selection procedure. There are many options how to select the next question from a bank of possible questions. This selection process is crucial for a successful adaptive testing procedure because the order in which questions are selected affects the rate in which the model improve its estimations. There is so far no definite answer which objective function produces the best possible results. In this article we discuss several question selection functions and compare them on two real models.

The paper is organized as follows. First, we describe the concept of Computerized Adaptive Testing, our models, and the notation we use. A short overview of two empirical data sets is presented. Both sets contain results of a paper (written) test collected for the purpose of our

research. The first dataset is formed by results of high school tests of mathematical skills in the domain of functions; the second dataset has been collected from test results of foreign students of Czech language. In Section 4, we propose three different types of methods and to compare we also use linear selection process (questions are asked in the same order as they are ordered in the set of possible questions). All methods are tested on two available data sets. Results of experiments are presented in Section 5 of this paper where methods are compared and contrasted. The concluding section summarizes our results and points out possibilities for further improvements in this area.

2 Computerized Adaptive Testing

CAT is a concept of testing which is getting a large scientific attention for about two decades [9, 10, 12]. With CAT we build computer administered and computer controlled tests. The computer system is selecting questions for a student taking the test and evaluating his/her performance.

The process can be divided into two phases: model creation and testing. In the first phase the student model is created while in the second phase the model is used to actually test examinees. There are many different model types usable for adaptive testing as can be found, for example, in [1, 2, 3]. In this work we are working with Bayesian Networks. Regardless of the model the testing part follows the same scheme. With a prepared and calibrated model, CAT repeats following steps:

- The next question to be asked is selected.
- This question is asked and an answer is obtained.
- This answer is inserted into the model.
- The model (which provides estimates of the student's skills) is updated.
- Answers to all questions are estimated given the current estimates of student's skills. (optional)

This procedure is repeated until a termination criterion is reached. Criteria can be of various types, for example, a time restriction, a number of questions, or a confidence interval of the estimated variables (i.e., reliability of the test).

In this article we consider the first step of the testing procedure which is the question selection procedure.

3 Bayesian Network Models

We use Bayesian Networks (BNs) to model students. Details about BNs can be found in, for example, [6, 5]. We restrict ourselves to the following BN structure. Networks have two levels, variables in the parent's level are addressed as skill variables $S \in \mathcal{S}$ where $\mathcal{S}$ is the set of all skills. The children level contains question variables $X \in \mathcal{X}$ where $\mathcal{X}$ is the set of all questions.

- We will use symbol $\boldsymbol{X}$ to denote the multivariable $(X_1, \ldots, X_n)$ taking states $\boldsymbol{x} = (x_1, \ldots, x_n)$. The total number of question variables is n, the set of all indexes of question variables is $\boldsymbol{N} = \{1, \ldots, n\}$. Question variables are binary and they are observable.
- We will use symbol $\boldsymbol{S}$ to denote the multivariable $(S_1, \ldots, S_m)$ taking states $\boldsymbol{s} = (s_1, \ldots, s_m)$. The set of all indexes of skill variables is $\boldsymbol{M} = \{1, \ldots, m\}$. In this article we use only binary skill variables. The set of all possible state configurations of $\boldsymbol{S}$ is $Val(\boldsymbol{S})$. Skill variables are all unobservable.

3.1 Data and specific models

To test our theoretical methods we have collected empirical data. We obtained two different data sets which are described here.

First, we designed a paper test of mathematical knowledge of grammar school students. The test focuses on simple functions (mostly polynomial, trigonometric, and exponential/logarithmic).

Students were asked to solve various mathematical problems (reffered as questions) including graph drawing and reading, calculating points on the graph, root finding, describing function shapes and other function properties. Questions are open (the mathematical problem's solution has to be included) and results are stored as binary correct/wrong values. In total 281 participants took the test. For the purpose of this paper this data set is modeled by two different Bayesian network models. One of them is shown in Figure 1. It consists of 53 questions and 8 skill nodes. These skill nodes represent different student skills connected to questions. This models is further referred to as Mathematical knowledge test Model (MM).

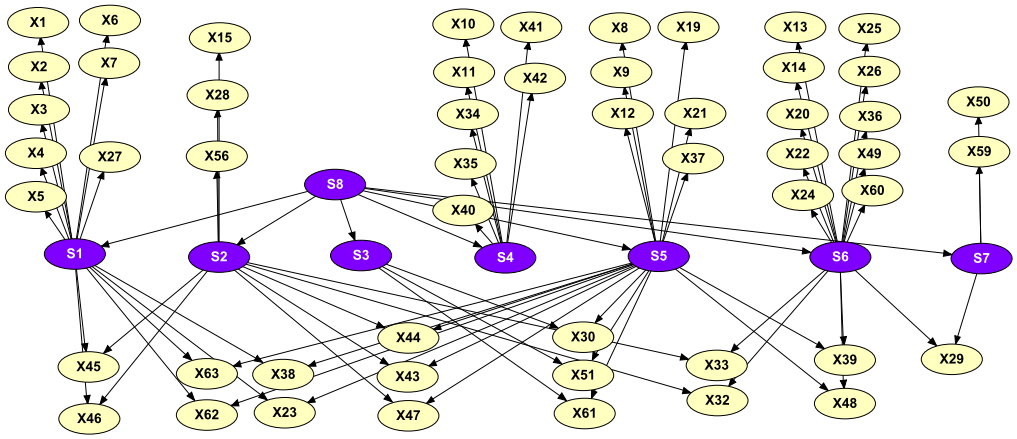


Figure 1: CAT MM structure

The second dataset was collected with a test of Czech language for non-native speaker students. This test contained multiple choice questions with four possible answers. One answer was correct. The test was assessed in a binary way where each question was either correct or incorrect. This test contains 30 questions and 143 students participated in the testing process. The model which was created by a domain expert is shown in Figure 2. Apart from 30 question nodes it has 11 skill nodes. Each skill, again, represents a specific ability a student should have to answer a connected question correctly. The skills include abilities related to morphology, vocabulary, conjugation, etc. This model is referred to as Czech language test Model (CM)[1].

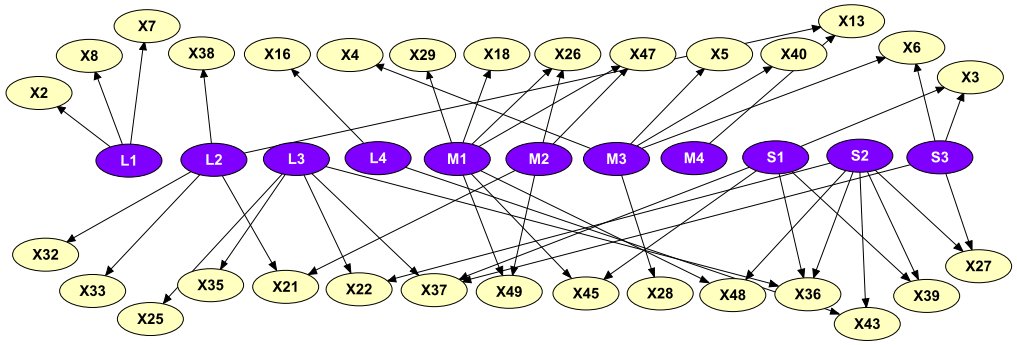


Figure 2: CAT CM structure

4 Question Selection Methods

The task of the question selection is repeated in every step of testing of an individual student. Its process is described in detail below.

[1]More detailed information can be found (in Czech) in the master thesis of Amal Magauina available at https://dspace.cvut.cz/

We define the question evidence e as:

$$e = \{X_{i_1} = x_{i_1}, \ldots, X_{i_n} = x_{i_n} | i_1, \ldots, i_n \in \boldsymbol{N}\}.$$

where $\{i_1, \ldots, i_n\} = \boldsymbol{I}$ are indexes of already answered questions. Remaining questions are unobserved (unanswered) $\hat{\mathcal{X}} = \{X_i | i \in \boldsymbol{N} \setminus \boldsymbol{I}\}$.

The goal is to select a question from $\hat{\mathcal{X}}$ to be asked next. The selection is dependent on a criterion function which may take different forms. Below, we describe three possible question selection methods. In this paper we also use, as a comparison method, a sequential selection. While using the sequential selection, the question we select is simply chosen in the same order as they are ordered in the question input list. This type of question selection is often used in non-adaptive tests where questions are always asked in the same sequence.

Three methods, we present further, are:

- Maximization of the Expected Entropy Reduction (also called Information Gain)
- Maximization of the Expected Skills Variance
- Maximization of the Expected Question Variance

The motivation for selecting these three possibilities is discussed for each criterion separately.

4.1 Maximization of the Expected Entropy Reduction

The purpose of an adaptive test is to provide the best possible information about a tested student. Each student is modeled by his skills. The criterion described in this section uses the Shannon entropy calculated over all skill values which we define in this section. It is a measure of the certainty of skills estimation. Because of that we want to select a question which provides the largest expected information gain if asked, i.e., a question which reduces uncertainty the most. This method is further referred to as Skills' Entropy.

We compute the cumulative Shannon entropy over all skill variables of S given the evidence e:

$$H(e) = \sum_{j \in \boldsymbol{M}} \sum_{s=0}^{1} -P(S_j = s|e) \cdot \log P(S_j = s|e) \ . \tag{1}$$

Assume we decide to ask a question $\hat{X} \in \hat{\mathcal{X}}$. After inserting the observed outcome the entropy over all skills changes. We can compute the value of new entropy for evidence extended by $\hat{X} = \hat{x}$ as:

$$H(e, \hat{X} = \hat{x}) = \sum_{j \in \boldsymbol{M}} \sum_{s=0}^{1} -P(S_j = s|e, \hat{X} = \hat{x}) \cdot \log P(S_j = s|e, \hat{X} = \hat{x}) \ . \tag{2}$$

This entropy $H(e, \hat{X} = \hat{x})$ is the sum of individual entropies over all skill nodes. Another option would be to compute the entropy of the joint probability distribution of all skill nodes. This would take into account correlations between these nodes. In our task we want to estimate marginal probabilities of all skill nodes. In the case of high correlations between two (or more) skills the latter criterion would assign them a lower significance in the model. This is the behavior we wanted to avoid. The first criterion assigns the same significance to all skill nodes which is a better solution. Moreover, the computational time required for the proposed method is lower.

Now, we can compute the expected entropy after answering question $\hat{X}$:

$$EH(\hat{X}, e) \quad = \quad \sum_{\hat{x}=0}^{1} P(\hat{X} = \hat{x}|e) \cdot H(e, \hat{X} = \hat{x}) \ . \tag{3}$$

Finally, we choose a question X^* that maximizes the information gain $IG(\hat{X}, e)$

$$X^* \quad = \quad \arg\max_{\hat{X} \in \hat{\mathcal{X}}} IG(\hat{X}, e) \ , \text{ where} \tag{4}$$

$$IG(\hat{X}, e) \quad = \quad H(e) - EH(\hat{X}, e) \ . \tag{5}$$

4.2 Maximization of the Expected Skills Variance

With this criterion we want to select a question which leads to the largest variance of state probabilities of skill variables. The rationale behind this selection is very similar to the one discussed in the previous method. The goal is to provide the most accurate estimation of student's skills and also to provide the best separation of students based on their skills. We measure the variance between skill's state probabilities (student having the skill). The variance is measured for two possible answers to one question, i.e., correct and incorrect. The criterion searches for a question which provides the largest variance in these two possibilities. This method is further referred to as Skills' Variance.

We consider unaswered question $\hat{X} \in \hat{\mathcal{X}}$ to be asked. First, we establish following notation:

$$p_0^j = P(S_j = 1|\hat{X} = 0, e) \ ,$$
$$p_1^j = P(S_j = 1|\hat{X} = 1, e) \ ,$$

where $S_j \in \mathcal{S}$. The symbol p_0^j stands for the probability of a student having the examined skill S_j even though the answer to the question $\hat{X}$ was incorrect. p_1^j is the case where the answer was correct and the student has the skill S_j. Naturally, the value of p_1^j should be larger than p_0^j. We compute the average value $\overline{p}^j$:

$$\overline{p}^j = P(\hat{X} = 0|e) \cdot p_0^j + P(\hat{X} = 1|e) \cdot p_1^j \ .$$

Then, the expected variance of states' probabilities of the skill S_j after answering the question $\hat{X}$ can be obtained using the following formula:

$$var_j(S_j|e, \hat{X}) = (\overline{p}^j - p_0^j)^2 \cdot P(\hat{X} = 0|e) + (\overline{p}^j - p_1^j)^2 \cdot P(\hat{X} = 1|e) \ . \tag{6}$$

This value has to be computed for each skill in the model. Afterwards, we compute the average of these values for the question $\hat{X}$:

$$var(\mathcal{S}|e, \hat{X}) = \frac{1}{m} \sum_{j \in \boldsymbol{M}} var_j(S_j|e, \hat{X}) \ . \tag{7}$$

We select a question which has the highest average value computed from (7):

$$X^* \quad = \quad \arg\max_{\hat{X} \in \hat{\mathcal{X}}} var(\mathcal{S}|e, \hat{X}) \ . \tag{8}$$

Maximization of (6) can be viewed as a generalization of the criterion of student separation described in our previous article [7]. The difference is that in this case we consider the probability of $S_j = 1$ after answering $\hat{X}$ instead of the most probable state of S_j.

4.3 Maximization of the Expected Question Variance

Previous two criteria aimed at skills directly. This third one aims at questions instead. From all unanswered questions we want to find a question with the highest expected variance of correct answer probabilities for all possible state combinations. This criterion is motivated as follows: if the question's correct answer probability varies a lot with changing skill states it means that this question is significantly affected when student skills shifts. It follows from the Bayes rule that this question also has a significant influence on the skills. This method is further referred to as Questions' Variance.

The expected variance of the question's $\hat{X}$ correct answer probability is computed given the following formula:

$$var(\hat{X}|e) = \sum_{\boldsymbol{s} \in Val(\boldsymbol{S})} (P(\hat{X} = 1|e) - P(\hat{X} = 1|\boldsymbol{s}, e))^2 \cdot P(\boldsymbol{s}|e) \ . \tag{9}$$

A question with the highest value of expected variance given by (9) is selected to be asked next. The use of this function for computations during testing is impractical because of its computational

complexity as it would take long time to select the next question. We propose an approximation of Formula (9). We compute the variance for a single skill node and then take into account their combined average instead of the full computation over all states' combinations.

We establish following notation:

$$r_0^j = P(\hat{X} = 1|S_j = 0) \ ,$$
$$r_1^j = P(\hat{X} = 1|S_j = 1) \ ,$$

where $S_j \in \mathcal{S}$, $\hat{X} \in \hat{\mathcal{X}}$. r_0^j stands for the probability, that the student answers correctly to the question even though he/she has no skill in question. r_1^j is the same situation while the student has all examined skills. Intuitively, the value r_1^j has to be larger than r_0^j.

With the average value

$$\overline{r}^j = P(S_j = 0|e) \cdot r_0^j + P(S_j = 1|e) \cdot r_1^j$$

we can compute the expected variance of correct answer probability for the question $\hat{X}$ using the next formula:

$$\begin{aligned} var_j(\hat{X}|e) &= (\overline{r}^j - r_0^j)^2 \cdot P(S_j = 0|e) + (\overline{r}^j - r_1^j)^2 \cdot P(S_j = 1|e) \ , \\ var(\hat{X}|e) &= \frac{1}{m} \sum_{j \in \boldsymbol{M}} var_j(\hat{X}|e) \ . \end{aligned} \tag{10}$$

A question X^* we select is maximizing this variance.

$$X^* \quad = \quad \arg\max_{\hat{X} \in \hat{\mathcal{X}}} var(\mathcal{S}|e, \hat{X}) \ . \tag{11}$$

The value $(\overline{r}^j - r_s^i)$ can be viewed as differential of $P(\hat{X} = 1|S_j = s)$ of skill variables S_j that have only two states $s \in \{0, 1\}$. Therefore, if P is the probability density function of the continuous skill variable S_j, we can view it as a finite equivalent of the probability $P(\hat{X} = 1|S_j = s)$ derivative with respect to s. It means that $var(\hat{X}|e)$ is similar to Fisher's information which is a commonly used criterion for IRT (Item Response Theory) [11] – another possible type of model for CAT. More detailed explanation of this criterion in the case of continuous skill variables can be found in [7].

5 Experiments

5.1 Experimental Setup

To evaluate models we have done experiments on both data sets with models MM and CM described above. We have used 10 fold cross-validation method for both data sets. Models were first learned using standard EM algorithm from learning data. Next, we performed a simulation of CAT test for every model and for every student using testing data.

During simulated testing we first estimated the skills of a student based on his/her answers. At the start of each step we compute marginal probability distributions for all skills $\boldsymbol{S}$. This happens before selecting a new question and updating the model with the new answer. We use evidence e obtained in previous steps which is at the start of testing empty. Then, based on estimated skills we predict answers to all questions, where we select the most probable state of each question $X \in \mathcal{X}$:

$$x^* = \arg\max_x P(X = x|\boldsymbol{S}) \ . \tag{12}$$

By comparing this value to the real answer x' of the question X we obtain a success rate of the response estimates for all questions $X \in \mathcal{X}$ of a test result t (particular student's result) in one step

$$\mathrm{SR}^t \quad = \quad \frac{\sum_{X \in \mathcal{X}} I(x^* = x')}{|\mathcal{X}|} \ , \text{ where} \tag{13}$$

$$I(expr) \quad = \quad \begin{cases} 1 & \text{if } expr \text{ is true} \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

The total success rate of one model in one step for all test data is defined as

$$\mathrm{SR} \quad = \quad \frac{\sum_{t=1}^{D} \mathrm{SR}^t}{D} \ , \tag{15}$$

where D is the dataset size.

5.2 Experimental Results

Average results of simulated tests for both data sets, i.e., both models described above, are displayed in graphs 3 and 4 for each model separately. Graphs show success rates SR in the first 30 steps. Step 0 is the state before asking any questions. At this point the prediction is based only on data itself. There is no evidence and the selection criterion adds no benefit. Therefore the SR is the same over all cases for a single model. For comparison we include also the sequential selection as described in Section 4.

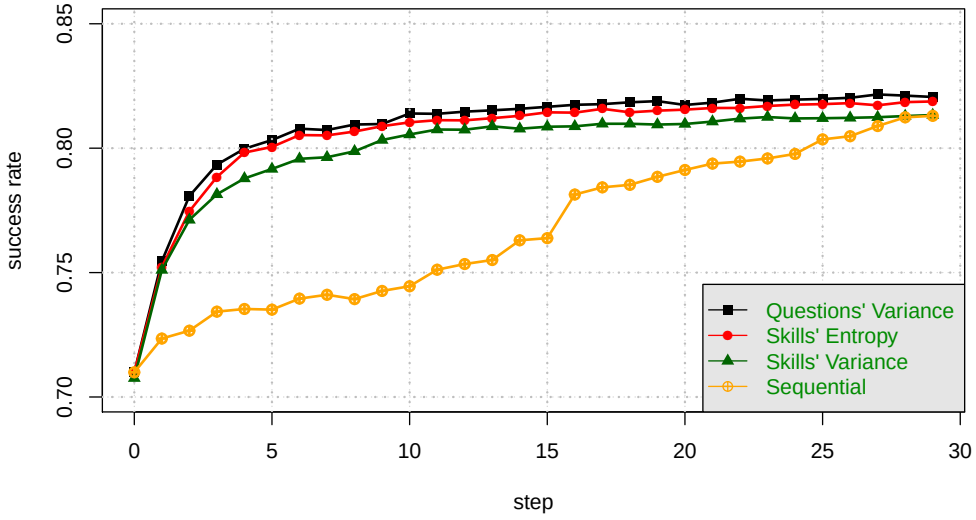


Figure 3: MM success rates for first 30 questions of simulated testing

As we can see in these graphs the worst performing method is the sequential selection. It is apparent that the rate in which this method improves its estimates is lower than the rate of the remaining methods. This is caused by the fact that it is selecting questions which are not most informative in the current test situation for the tested student.

The Skills' Variance method of the selection has the lowest performance (or same) from three proposed methods in both models. As explained below it is the only method which has statistically significantly worse results in one instance. Particular reasons for this behavior has to be explored further as there might be many possible causes.

Questions selected by individual methods are not the same even though the success rate of question estimates is very similar. The selected questions are displayed in the Tables 3 and 4. Numbers displayed correspond to the total number of selections of the particular question in the MM and the CM in the first five steps of simulated testing. Questions which were not selected at all are not included in the tables. By inspecting these tables we can easily see that there are differences in individual methods. Some questions were not selected at all by one method while the other two methods selected them in some cases only. For example, in the MM the question X42 was not selected by Skills' Entropy while Questions' Variance selected it in 112 cases in the first five steps. Nevertheless, we can see a trend of good questions which are selected very often and soon in the process of testing by all methods. For example, in the MM the question X43 was

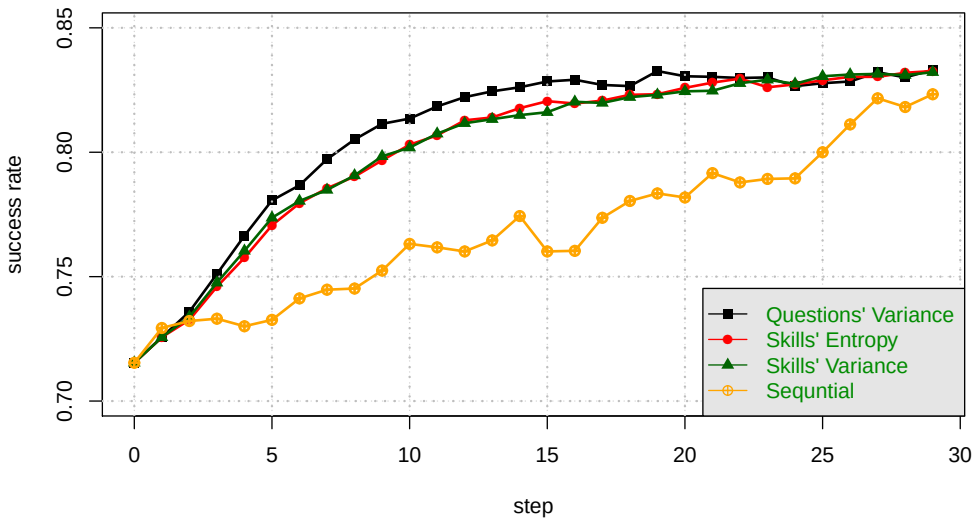


Figure 4: CM success rates for first 30 questions of simulated testing

selected for all students by Skills' Entropy and only for 7/10 of the dataset by the two remaining methods.

5.3 Wilcoxon tests

To confirm our conclusions described above we used the Wilcoxon signed-rank test. Tables 1 and 2 contain p-values obtained from Wilcoxon tests to compare the success rates of two criteria. An alternative hypothesis is that the overall success rate of the i-th criterion (row index) is greater than the overall success rate of the j-th criterion (column index).

Table 1: MM Wilcoxon tests p-values

	sequential	**Skills' Entropy**	**Skills' Variance**	**Questions' Variance**
sequential	-	1	1	1
Skills' Entropy	$1.17 \cdot 10^{-5}$	-	$1.62 \cdot 10^{-1}$	$9.39 \cdot 10^{-1}$
Skills' Variance	$2.20 \cdot 10^{-4}$	$8.40 \cdot 10^{-1}$	-	$9.99 \cdot 10^{-1}$
Questions' Variance	$2.04 \cdot 10^{-8}$	$6.15 \cdot 10^{-2}$	$9.22 \cdot 10^{-3}$	-

Table 2: CM Wilcoxon tests p-values

	sequential	**Skills' Entropy**	**Skills' Variance**	**Questions' Variance**
sequential	-	1	1	1
Skills' Entropy	$1.10 \cdot 10^{-4}$	-	$5.14 \cdot 10^{-1}$	$8.76 \cdot 10^{-1}$
Skills' Variance	$8.77 \cdot 10^{-5}$	$4.92 \cdot 10^{-1}$	-	$8.58 \cdot 10^{-1}$
Questions' Variance	$7.72 \cdot 10^{-6}$	$1.27 \cdot 10^{-1}$	$1.46 \cdot 10^{-1}$	-

As we can see in both cases, p-values of the sequential selection compared to all other criteria are much smaller than the borderline of $\alpha = 0.05$. This confirms the fact that the sequential selection provides the worst results. The table of the MM also shows that the success rate of Questions' Variance method is greater than the SR of Skills' Variance method (p-value $= 9.22 \cdot 10^{-3}$). This

shows there is statistically important improvement in success rates of the former over the latter method. All other pairs of different selection criteria show statistically insignificant difference within the selected confidence interval. Therefore, for the remaining pairs we can not establish any statistically sound order.

6 Conclusions and Future Work

This article considered different ways of selecting questions during the procedure of Computerized Adaptive Testing. We presented three different types of methods to select questions during CAT which were afterwards tested. For testing we used two data sets collected for this purpose.

The first important empirical observation is that the question selection method has a significant impact on the quality of predictions during the CAT procedure. In the comparisons all three proposed methods clearly outperformed the sequential selection. The motivation to study these methods is thus valid.

The next observation is that three proposed methods behave differently. In this case the difference in the quality of prediction is not large, it is statistically insignificant, but the methods are distinguishable since they select different questions. The first step in the future research is to provide generalizations of these methods to support multi-state skill variables. It seems that especially for Skills' Variance it may be very beneficial to test a model with skill nodes having more than two states. It is necessary to show if we can improve these methods and establish any ordering between them which would be valid generally over different models.

Acknowledgement
This work was supported by the Czech Science Foundation (project No. 16-12010S) and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS17/198/OHK4/3T/14.

Table 3: The frequency of questions X1-X61 selections during simulated testing using criterion (a) Skills' Entropy, (b) Skills' Variance, (c) Questions' Variance for the MM model. Questions which were never selected are not included.

	X3	X5	X7	X10	X11	X12	X13	X19	X25	X26	X27	X29	X30	X32	X33	X38	X39	X41	X42	X43	X44	X50	X51	X61
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	281	-	-	-	-
2	-	-	-	-	-	-	-	118	-	-	-	-	-	149	-	-	-	14	-	-	-	-	-	-
3	-	62	10	5	-	-	-	70	-	8	-	-	-	-	-	-	-	38	-	-	-	-	68	20
4	-	9	-	11	-	1	8	-	-	23	7	25	18	-	28	-	-	113	-	-	-	16	2	20
5	-	24	-	13	13	14	-	9	3	1	9	14	49	-	19	-	-	33	-	-	5	38	10	27

	X3	X5	X7	X10	X11	X12	X13	X19	X25	X26	X27	X29	X30	X32	X33	X38	X39	X41	X42	X43	X44	X50	X51	X61
1	-	-	-	-	-	-	-	-	-	-	-	-	-	56	-	-	-	-	-	197	28	-	-	-
2	-	-	-	-	-	-	-	49	-	-	-	-	-	105	-	-	-	108	-	10	9	-	-	-
3	10	53	-	-	-	-	-	48	-	-	-	-	-	14	38	-	-	35	9	15	-	-	45	14
4	4	6	5	7	-	1	-	16	-	13	-	-	-	-	11	3	-	80	35	13	9	49	18	11
5	9	8	2	1	-	-	8	32	-	24	1	14	9	-	13	11	-	32	20	2	7	40	20	28

	X3	X5	X7	X10	X11	X12	X13	X19	X25	X26	X27	X29	X30	X32	X33	X38	X39	X41	X42	X43	X44	X50	X51	X61
1	-	-	-	-	-	-	-	-	-	-	-	-	-	84	-	-	-	-	-	197	-	-	-	-
2	-	-	-	-	-	-	-	113	-	-	-	-	-	15	-	-	-	106	-	-	47	-	-	-
3	-	16	3	-	-	-	-	22	-	-	-	-	-	21	4	-	-	61	66	25	-	-	49	14
4	6	59	7	2	-	1	-	-	-	-	-	-	2	5	17	16	-	18	36	12	-	47	39	14
5	12	14	-	10	-	1	-	12	7	17	1	2	15	25	10	16	5	18	10	-	-	70	10	26

Table 4: The frequency of questions X2-X49 selections during simulated testing using criterion (a) Skills' Entropy, (b) Skills' Variance, (c) Xuestions' Variance for the CM model. Xuestions which were never selected are not included.

	X2	X3	X4	X6	X7	X13	X16	X18	X21	X22	X26	X28	X35	X36	X37	X39	X43	X45	X47	X48	X49
1	-	-	-	-	-	-	-	-	-	-	-	-	143	-	-	-	-	-	-	-	-
2	-	14	-	-	-	-	86	-	-	43	-	-	-	-	-	-	-	-	-	-	-
3	19	4	-	13	-	4	25	-	-	15	-	-	-	-	-	-	3	-	-	60	-
4	12	15	8	14	8	12	9	-	-	12	6	-	-	-	1	6	5	-	-	11	24
5	26	22	1	13	2	13	10	-	5	7	10	-	-	4	-	20	1	2	-	7	-

	X2	X3	X4	X6	X7	X13	X16	X18	X21	X22	X26	X28	X35	X36	X37	X39	X43	X45	X47	X48	X49
1	-	-	-	-	-	-	-	-	-	-	-	-	143	-	-	-	-	-	-	-	-
2	-	14	-	-	-	-	77	-	-	52	-	-	-	-	-	-	-	-	-	-	-
3	14	4	-	18	-	-	29	-	-	6	-	-	-	-	-	-	3	-	-	69	-
4	4	15	7	19	-	8	21	-	-	8	6	-	-	-	-	4	6	-	1	17	27
5	15	33	3	31	-	5	9	-	2	14	9	2	-	-	-	-	2	2	1	11	4

	X2	X3	X4	X6	X7	X13	X16	X18	X21	X22	X26	X28	X35	X36	X37	X39	X43	X45	X47	X48	X49
1	-	-	-	-	-	-	-	-	-	-	-	-	143	-	-	-	-	-	-	-	-
2	23	18	-	-	-	-	29	-	-	58	-	-	-	-	-	-	-	-	-	15	-
3	16	25	-	15	-	-	13	-	-	-	6	-	-	-	-	-	3	-	-	65	-
4	20	29	-	18	3	-	13	-	-	33	6	-	-	1	-	4	1	2	-	13	-
5	10	29	2	11	6	-	24	3	-	9	5	-	-	6	-	20	8	1	-	3	6

References

[1] R. G. Almond and R. J. Mislevy. Graphical Models and Computerized Adaptive Testing. *Applied Psychological Measurement*, 23(3):223–237, 1999.

[2] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 1999.

[3] M. J. Culbertson. *Graphical Models for Student Knowledge: Networks, Parameters, and Item Selection*. PhD thesis, University of Illinois at Urbana, 2014.

[4] K. C. Moe and M. F. Johnson. Participants' Reactions To Computerized Testing. *Journal of Educational Computing Research*, 4(1):79–86, jan 1988.

[5] T. D. Nielsen and F. V. Jensen. *Bayesian Networks and Decision Graphs (Information Science and Statistics)*. Springer, 2007.

[6] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., dec 1988.

[7] M. Plajner and J. Vomlel. Student Skill Models in Adaptive Testing. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*, pages 403–414. JMLR.org, 2016.

[8] S. Tonidandel, M. A. Quiñones, and A. A. Adams. Computer-adaptive testing: the impact of test characteristics on perceived performance and test takers' reactions. *The Journal of applied psychology*, 87(2):320–32, apr 2002.

[9] W. J. van der Linden and C. A. W. Glas. *Computerized Adaptive Testing: Theory and Practice*, volume 13. Kluwer Academic Publishers, 2000.

[10] W. J. van der Linden and C. A. W. Glas, editors. *Elements of Adaptive Testing*. Springer New York, NY, 2010.

[11] W. J. van der Linden and R. K. Hambleton. *Handbook of Modern Item Response Theory*. Springer New York, 2013.

[12] H. Wainer and N. J. Dorans. *Computerized Adaptive Testing: A Primer*. Routledge, 2015.

Strict and Strong Consistency in Pairwise Comparisons Matrix With Fuzzy Elements

Jaroslav RAMÍK

Faculty of Business Administration in Karviná
Silesian University in Opava
Czech Republic
ramik@opf.slu.cz

Abstract

This paper forms both theoretical and practical innovation basis for decision making process in micro and macro economics. The decision making problem considered here is to rank n alternatives from the best to the worst, using the information given by the decision maker(s) in the form of an $n \times n$ pairwise comparisons matrix. Here, we deal with pairwise comparisons matrices with fuzzy elements. Fuzzy elements of the pairwise comparisons matrix are applied whenever the decision maker is uncertain about the value of his/her evaluation of the relative importance of elements in question. We investigate pairwise comparisons matrices with elements from abelian linearly ordered group (alo-group) over a real interval which is a generalization of traditional multiplicative or additive approaches. The concept of reciprocity and consistency of pairwise comparisons matrices with fuzzy elements is well known. Here, we extend these concepts, namely to the strong as well as strict consistency of pairwise comparisons matrices with fuzzy elements (PCF matrices). We derive the necessary and sufficient conditions for strong/strict consistency and investigate their properties as well as some consequences to the problem of ranking the alternatives. Illustrating examples are presented and discussed.

Keywords: Multi-Criteria Optimization, Pairwise Comparisons Matrix, Fuzzy Elements, Strong Consistency, Strict Consistency, Alo-Group.

1 Introduction

A *decision making problem (DM problem)* which forms an application background in this paper can be formulated as follows:
Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of alternatives ($n > 1$). The DM aim is to rank the alternatives from the best to the worst (or, vice versa), using the information given by the decision maker in the form of an $n \times n$ pairwise comparison matrix.

Fuzzy sets being the elements of the pairwise comparison matrix can be applied whenever the decision maker (DM) is not sure about the preference degree of his/her evaluation of the pairs in question. Fuzzy elements may be taken also as aggregations of crisp pairwise comparisons of a group of decision makers in the group DM problem, see e.g. [21]. Decision makers acknowledge fuzzy pairwise preference data as imprecise knowledge about regular preference information. The preference matrix with fuzzy elements is then seen as a tool constraining an ill-known precise consistent comparison matrix. Inconsistencies, i.e. incompatibilities in comparison data are thus explicitly explained by the imprecise (or, inexact, vague etc.) nature of human-originated information. The former works that investigated the problem of finding a rank of the given alternatives based on some pairwise comparison matrices are e.g. [6] - [10] and [20]. In [20] some simple linear programming models for deriving the priority weights from various interval fuzzy preference relations are proposed. Leung and Cao [6] proposed a new definition of the pairwise comparison reciprocal matrix by setting deviation tolerances based on an idea of allowing inconsistent information. Mahmoudzadeh and Bafandeh [7] further discussed Leung and Cao's work and proposed a new method of fuzzy consistency test by direct fuzzification of QR (Quick Response) algorithm

which is one of the numerical methods for calculating eigenvalues of an arbitrary matrix. The recent results on pairwise comparison method are presented in [12]. Ramik and Korviny in [15] investigated inconsistency of pairwise comparison matrix with fuzzy elements based on geometric mean. They proposed a new inconsistency index which, however, does not measure satisfactorily inconsistency as well as uncertainty. In [16] and [17], the author presented a general approach for pairwise comparison matrices with fuzzy number elements based on alo-groups which unifies the previous approaches.

The recent paper is in some sense a continuation of [16]. In comparison to [16], we extend our approach from fuzzy number entries (where the core is a singleton) to fuzzy intervals (where the core is an interval). These elements are the matrix entries of pairwise comparison matrices (PCF matrices). We newly redefine the concepts of reciprocity and consistency and, moreover, introduce the strong reciprocity and strong consistency of PCF matrices. Then we derive necessary and sufficient conditions for strong reciprocity and strong consistency, which can be easily checked. Then we investigate some consequences of the new concepts of strong reciprocity and strong consistency to the problem of ranking the alternatives. Moreover, we also solve the problem of measuring the inconsistency of PCF matrices by defining corresponding indexes. We present several numerical examples in order to illustrate the proposed concepts and derived properties.

2 Preliminaries

Here, it will be useful to understand fuzzy sets as special nested families of subsets of a set, see [14].

A *fuzzy subset* of a nonempty set X (or a *fuzzy set* on X) is a family $\{A_\alpha\}_{\alpha\in[0,1]}$ of subsets of X such that $A_0 = X, A_\beta \subset A_\alpha$ whenever $0 \le \alpha \le \beta \le 1$, and $A_\beta = \cap_{0\le\alpha<\beta} A_\alpha$ whenever $0 < \beta \le 1$. The *membership function of* A is the function μ_A from X into the unit interval $[0,1]$ defined by $\mu_A(x) = \sup\{\alpha \mid x \in A_\alpha\}$. Given $\alpha \in]0,1]$, the set $[A]_\alpha = \{x \in X \mid \mu_A(x) \ge \alpha\}$ is called the *α-cut of fuzzy set A*.

Remark 1. Usually, an opposite approach is adopted: A fuzzy set A is defined by the membership function μ_A from X into the unit interval $[0,1]$ and then the family $\{A_\alpha\}_{\alpha\in[0,1]}$ of subsets of X is defined by the concept of α-cut $[A]_\alpha$ of fuzzy set A. Our approach, however, seems to be more logical, as the concept of fuzzy set is based on the concept of *set*, and not on the concept of *function*. It is more appropriate and elegant.

Remark 2. Let A be a subset of a set X and let $\{A_\alpha\}_{\alpha\in[0,1]}$ be the family of subsets of X defined by $A_0 = X$ and $A_\alpha = A$ for each positive $\alpha \in [0,1]$. It can easily be seen that this family is a fuzzy set on X and that its membership function is equal to the characteristic function of A; we call it the *crisp fuzzy set* on X.

If X is a nonempty subset of the n-dimensional Euclidean space $\mathbb{R}^n$, then a fuzzy set A in X is called *closed, bounded, compact* or *convex* if the α-cut $[A]_\alpha$ is a closed, bounded, compact or convex subset of X for every $\alpha \in]0,1]$, respectively.

We say that a fuzzy subset A of $\mathbb{R}^* = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ is a *fuzzy interval* whenever A is normal and its membership function μ_A satisfies the following condition: there exist $a, b, c, d \in \mathbb{R}^*$, $-\infty \le a \le b \le c \le d \le +\infty$, such that

$$\begin{array}{rl} \mu_A(t) = 0 & \text{if } t < a \text{ or } t > d, \\ \mu_A & \text{is strictly increasing and continuous on the interval } [a,b], \\ \mu_A(t) = 1 & \text{if } b \le t \le c, \\ \mu_A & \text{is strictly decreasing and continuous on the interval } [c,d]. \end{array} \tag{1}$$

A fuzzy interval A is *bounded* if $[a,d]$ is a compact interval. Moreover, for $\alpha = 0$ we define the *zero-cut* of A as $[A]_0 = [a,d]$. A bounded fuzzy interval A is the *triangular fuzzy number* if $b = c$.

An *abelian group* is a set, G, together with an operation $\odot$ (read: operation *odot*) that combines any two elements $a, b \in G$ to form another element in G denoted by $a \odot b$, see [1]. The symbol $\odot$ is a general placeholder for a concretely given operation. $(G, \odot)$ satisfies the following requirements known as the *abelian group axioms*, particularly: *commutativity*, *associativity*, there exists an *identity element* $e \in G$ and for each element $a \in G$ there exists an element $a^{(-1)} \in G$ called the *inverse element to* a.

The *inverse operation* $\div$ to $\odot$ is defined for all $a, b \in G$ as follows

$$a \div b = a \odot b^{(-1)}. \tag{2}$$

An ordered triple $(G, \odot, \leq)$ is said to be *abelian linearly ordered group*, *alo-group* for short, if $(G, \odot)$ is a group, $\leq$ is a linear order on G, and for all $a, b, c \in G$

$$a \leq b \text{ implies } a \odot c \leq b \odot c. \tag{3}$$

If $\mathcal{G} = (G, \odot, \leq)$ is an alo-group, then G is naturally equipped with the order topology induced by $\leq$ and $G \times G$ is equipped with the related product topology. We say that $\mathcal{G}$ is a *continuous alo-group* if $\odot$ is continuous on $G \times G$.

By definition, an alo-group $\mathcal{G}$ is a lattice ordered group. Hence, there exists $\max\{a, b\}$, for each pair $(a, b) \in G \times G$. Nevertheless, a nontrivial alo-group $\mathcal{G} = (G, \odot, \leq)$ has neither the greatest element nor the least element.

Because of the associative property, the operation $\odot$ can be extended by induction to n-ary operation.

$\mathcal{G} = (G, \odot, \leq)$ is *divisible* if for each positive integer n and each $a \in G$ there exists the (n)-th root of a denoted by $a^{(1/n)}$, i.e. $\left(a^{(1/n)}\right)^{(n)} = a$.

Let $\mathcal{G} = (G, \odot, \leq)$ be an alo-group. Then the function $\|.\| : G \to G$ defined for each $a \in G$ by

$$\|a\| = \max\{a, a^{(-1)}\} \tag{4}$$

is called a $\mathcal{G}$-*norm*.

The operation $d : G \times G \to G$ defined by $d(a, b) = \|a \div b\|$ for all $a, b \in G$ is called a $\mathcal{G}$-*distance*. Next, we present the well known examples of alo-groups, see also [2], [3], or, [17].

Example 1. *Additive alo-group* $\mathcal{R} = (]-\infty, +\infty[, +, \leq)$, see [17], is a continuous alo-group with: $e = 0,\ a^{(-1)} = -a,\ a^{(n)} = n.a$.

Example 2. *Multiplicative alo-group* $\mathcal{R}^+ = (]0, +\infty[, \bullet, \leq)$, see [17], is a continuous alo-group with:
$e = 1,\ a^{(-1)} = a^{-1} = 1/a,\ a^{(n)} = a^n$. Here, by $\bullet$ we denote the usual operation of multiplication.

Example 3. *Fuzzy additive alo-group* $\mathcal{R}_a$=$(]-\infty, +\infty[, +_f, \leq)$, see [17], is a continuous alo-group with:

$$a +_f b = a + b - 0.5,\ e = 0.5,\ a^{(-1)} = 1 - a, a^{(n)} = n.a - \frac{n-1}{2}.$$

Example 4. *Fuzzy multiplicative alo-group* $\mathbf{]0, 1[_m}$=$(]0, 1[, \bullet_f, \leq)$, see [2], is a continuous alo-group with:

$$a \bullet_f b = \frac{ab}{ab + (1-a)(1-b)}, e = 0.5, a^{(-1)} = 1 - a.$$

3 PCF matrices

In this paper we shall investigate $n \times n$ pairwise comparison matrices with elements being bounded fuzzy intervals of the alo-group $\mathcal{G}$ over an interval G of the real line $\mathbb{R}$. We call them shortly *PCF matrices*. Moreover, we assume that all diagonal elements of these PCF matrices are crisp (in the sense of Remark 2), particularly they are equal to the identity element of $\mathcal{G}$.

The general approach based on alo-group is useful as it unifies various approaches known from the literature. This fact has been demonstrated on the previous 4 examples, where the well known cases of alo-groups are shown. Particularly, all concepts and properties which will be presented bellow can be easily applied to any of those alo-groups. In practice, the type of alo-group will depend on the particular DM problem. In some DM problems, it is more appropriate to apply e.g. additive alo-group (see Example 1), in other cases, the multiplicative alo-group (Example 2), or fuzzy alo-groups (Examples 3 and 4) are more natural from an interpretation point of view.

Notice that elements of PCF matrices may be crisp and/or fuzzy numbers, and/or fuzzy intervals, and/or fuzzy intervals with bell-shaped membership functions, triangular fuzzy numbers, trapezoidal fuzzy numbers etc. Such fuzzy elements may be either evaluated by individual decision makers, or, they may be made up of crisp pairwise evaluations of decision makers in a group DM problem, see e.g. [19].

3.1 Reciprocity and strong reciprocity of PCF matrices

Now, we shall define reciprocity properties for PCF matrices. Reciprocity of a PC matrix is a natural property defining the evaluation of couples of alternatives in the reverse order. First, we define reciprocity for PCF matrices, then we define the concept of the strong reciprocity and investigate some relationships between the two properties. We derive necessary and sufficient conditions for a PCF matrix to satisfy strong reciprocity condition. Our approach will cover the classical definitions of reciprocity presented e.g. in [13], and [17].

Definition 1. Let $C = \{\tilde{c}_{ij}\}$ be an $n \times n$ PCF matrix, $\alpha \in [0,1]$. C is said to be α-$\odot$-*reciprocal*, if the following condition holds:
For every $i, j \in \{1, 2, ..., n\}$ there exist $c_{ij} \in [\tilde{c}_{ij}]_\alpha$ and $c_{ji} \in [\tilde{c}_{ji}]_\alpha$ such that

$$c_{ij} \odot \ c_{ji} = e, \tag{5}$$

or, equivalently,

$$c_{ji} = c_{ij}^{(-1)}. \tag{6}$$

$C = \{\tilde{c}_{ij}\}$ is said to be $\odot$-*reciprocal*, if C is α-$\odot$-reciprocal for all $\alpha \in [0,1]$.

Remark 3. If $C = \{\tilde{c}_{ij}\}$ is a PCF matrix with crisp elements, then $\tilde{c}_{ij} = c_{ij}, c_{ij} \in G$ for all i and j, and condition (5) coincides with the classical definition of reciprocity for crisp PCF matrices: A crisp PCF matrix $C = \{c_{ij}\}$ is $\odot$-*reciprocal* if for all i and j: $c_{ji} = c_{ij}^{(-1)}$.

Particularly, $C = \{c_{ij}\}$ is additive-reciprocal if $c_{ji} = -c_{ij}$ for all i and j; $C = \{c_{ij}\}$ is multiplicative-reciprocal if $c_{ji} = \frac{1}{c_{ij}}$ for all i and j.

Remark 4. Clearly, if $C = \{\tilde{c}_{ij}\}$ is an α-$\odot$-reciprocal PCF matrix, $\alpha, \beta \in [0,1]$, $\beta \leq \alpha$, then $C = \{\tilde{c}_{ij}\}$ is β-$\odot$-reciprocal.

Notation. Let $C = \{\tilde{c}_{ij}\}$ be a PCF matrix. As each element of C is a fuzzy interval, we have that each α-cut of C is a closed interval, we denote

$$[c_{ij}^L(\alpha), c_{ij}^R(\alpha)] = [\tilde{c}_{ij}]_\alpha.$$

Here, $c_{ij}^L(\alpha)$ is the left end point of this interval, $c_{ij}^R(\alpha)$ is the right end point of the interval.

In the following proposition two necessary and sufficient conditions for reciprocity of PCF matrices are derived. From now on, the proofs of all propositions and theorems are left to the kind reader.

Proposition 1. Let $C = \{\tilde{c}_{ij}\}$ be a PCF matrix, $0 \leq \alpha \leq 1$. The following three conditions are equivalent.

(i) C is α-$\odot$-reciprocal.

(ii) For all $i, j \in \{1, 2, ..., n\}$ it holds

$$[c_{ij}^L(\alpha), c_{ij}^R(\alpha)] \cap [(c_{ji}^R(\alpha))^{(-1)}, (c_{ji}^L(\alpha))^{(-1)}] \neq \emptyset. \tag{7}$$

(iii) For all $i, j \in \{1, 2, ..., n\}$ it holds

$$c_{ij}^L(\alpha) \leq c_{ji}^L(\alpha))^{(-1)}, \tag{8}$$

$$c_{ij}^R(\alpha) \geq c_{ji}^R(\alpha))^{(-1)}. \tag{9}$$

Now, consider the following optimization problem:

(P1)

$$\alpha \longrightarrow \max; \tag{10}$$

subject to

$$c_{ij}^L(\alpha) \leq c_{ji}^L(\alpha))^{(-1)}, \tag{11}$$

$$c_{ij}^R(\alpha) \geq c_{ji}^R(\alpha))^{(-1)}. \tag{12}$$

$$0 \leq \alpha \leq 1. \tag{13}$$

Remark 5. Generally, problem (P1) is a nonlinear one-dimensional optimization problem. In particular cases, e.g. if the membership functions of the elements of the PCF matrix are piece-wise linear, then the optimal solution can be easily calculated. The optimal solution α^* of (P1) gives the maximum level α such that a PCF matrix $C = \{\tilde{c}_{ij}\}$ is α- $\odot$-reciprocal. Moreover, $C = \{\tilde{c}_{ij}\}$ is an $\odot$-reciprocal PCF matrix, if and only if $\alpha^* = 1$ is the optimal solution of (P1).

For a PCF matrix with non-crisp elements the reciprocity condition investigated in this section is a rather weak condition which has nearly no impact on the shapes of membership functions of its elements. That is why we shall ask stronger requirements to define the concept of strong reciprocity.

Definition 2. Let $C = \{\tilde{c}_{ij}\}$ be an $n \times n$ PCF matrix, $\alpha \in [0, 1]$. C is said to be *strong α-$\odot$-reciprocal*, if the following condition holds:
For every $i, j \in \{1, 2, ..., n\}$ and for every $c_{ij} \in [\tilde{c}_{ij}]_\alpha$ there exists $c_{ji} \in [\tilde{c}_{ji}]_\alpha$ such that

$$c_{ij} \odot \ c_{ji} = e. \tag{14}$$

$C = \{\tilde{c}_{ij}\}$ is said to be *strong $\odot$-reciprocal*, if C is strong α-$\odot$-reciprocal for all $\alpha \in [0, 1]$.

Remark 6. Evidently, by Definition 2 and Definition 3, every strong α-$\odot$-reciprocal PCF matrix is α-$\odot$-reciprocal, but not vice-versa, see Example 5.

Proposition 2. Let $C = \{\tilde{c}_{ij}\}$ be an $n \times n$ PCF matrix, $\alpha \in [0, 1]$. Then the following three conditions are equivalent.

(i)

$$C \text{ is strong } \alpha\text{- } \odot \text{-reciprocal.}$$

(ii) For all $i, j \in \{1, 2, ..., n\}$,

$$c_{ij}^L(\alpha) \odot c_{ji}^R(\alpha) = e \text{ and } c_{ij}^R(\alpha) \odot c_{ji}^L(\alpha) = e. \tag{15}$$

(iii) For all $i, j \in \{1, 2, ..., n\}$,

$$[c_{ji}^L(\alpha), c_{ji}^R(\alpha)] = [(c_{ij}^R(\alpha))^{(-1)}, (c_{ij}^L(\alpha))^{(-1)}]. \tag{16}$$

Remark 7. When evaluating fuzzy elements of a PCF matrix $C = \{\tilde{c}_{ij}\}$, only one of the membership functions of elements $\tilde{c}_{ij}$ and $\tilde{c}_{ji}$, $i \neq j$, should be constructed, the other should naturally satisfy condition (15), or (16). Then the PCF matrix becomes strong $\odot$-reciprocal.

Remark 8. Notice that Remark 4 is not true for strong reciprocity, namely, if $C = \{\tilde{c}_{ij}\}$ is a strong α-$\odot$-reciprocal PCF matrix, $\alpha, \beta \in [0, 1], \beta < \alpha$, then $C = \{\tilde{c}_{ij}\}$ need not be strong β-$\odot$-reciprocal. It is, however, β-$\odot$-reciprocal.

Example 5. Consider the additive alo-group $\mathcal{R} = (\mathbb{R}, \odot, \leq)$ with $\odot = +$, see Example 1. Let PCF matrices $C = \{\tilde{c}_{ij}\}$, and $D = \{\tilde{d}_{ij}\}$ be given as follows:

$$C = \begin{bmatrix} 0 & (1;2;4) & (4;5;7) \\ (-4;-2;-1) & 0 & (3;4;4) \\ (-7;-6;-4) & (-4;-4;-3) & 0 \end{bmatrix},$$

$$D = \begin{bmatrix} 0 & (1;2;5) & (4;5;8) \\ (-5;-2;-1) & 0 & (3;4;5) \\ (-9;-5;-4) & (-5;-4;-3) & 0 \end{bmatrix}.$$

Here, C and D are 3×3 PCF matrices, particularly, PCF matrices with triangular fuzzy number elements and the usual piece-wise linear membership functions.
Solving (P1) with C, we obtain the optimal solution $\alpha^* = \frac{3}{4}$, hence, by Proposition 1, C is α-+-reciprocal for every $\alpha \leq 0.75$. This is, by condition (iii) in Proposition 2, not strong α-+-reciprocal

for $\alpha \leq 0.75$.
Solving (P1) with D, we obtain the optimal solution $\alpha^{**} = 1$, hence, by Proposition 1, D is α-+-reciprocal for every $\alpha \in [0,1]$, i.e. D is +-reciprocal. However, for each $\alpha \in]0,1[$ condition (iii) in Proposition 2 is not satisfied for the fuzzy element $(4;5;8)$ and $(-9;-5;-4)$ in the matrix D, hence, D is not strong α-+-reciprocal. If we change element $(-9;-5;-4)$ for $(-8;-5;-4)$, we obtain a new matrix which is strong +-reciprocal.

3.2 Consistency of PCF matrices

Rationality and compatibility of a decision making process can be achieved by the consistency property of PCF matrices.

Let $\mathcal{G} = (G, \odot, \leq)$ be as earlier a divisible and continuous alo-group, $C = \{\tilde{c}_{ij}\}$ be a crisp PCF matrix, where $c_{ij} \in G \subset \mathbb{R}$ for all $i, j \in \{1, 2, ..., n\}$. The following definition is known, see e.g. [3].

Definition 3. A crisp PCF matrix $C = \{c_{ij}\}$ is $\odot$-*consistent* if for all $i, j, k \in \{1, 2, ..., n\}$

$$c_{ik} = c_{ij} \odot c_{jk}. \tag{17}$$

The following equivalent condition for consistency of PCF matrices is popular e.g. in AHP, see [7], or [3].

Proposition 3. A crisp PC matrix $C = \{c_{ij}\}$ is $\odot$-consistent if and only if there exists a vector $w = (w_1, w_2, ..., w_n)$, $w_i \in G$ such that

$$w_i \div w_j = c_{ij} \text{ for all } i, j \in \{1, 2, ..., n\}. \tag{18}$$

Now, we extend Definition 4 to non-crisp PCF matrices as follows, see also [17].

Definition 4. Let $\alpha \in [0,1]$. A PCF matrix $C = \{\tilde{c}_{ij}\}$ is said to be α-$\odot$-*consistent*, if the following condition holds:
For every $i, j, k \in \{1, 2, ..., n\}$, there exists $c_{ik} \in [\tilde{c}_{ik}]_\alpha$, $c_{ij} \in [\tilde{c}_{ij}]_\alpha$ and $c_{jk} \in [\tilde{c}_{jk}]_\alpha$ such that

$$c_{ik} = c_{ij} \odot c_{jk}. \tag{19}$$

The matrix C is said to be $\odot$-*consistent*, if C is α-$\odot$-consistent for all $\alpha \in [0,1]$. If for some $\alpha \in [0,1]$ the matrix C is not α-$\odot$-consistent, then C is called α-$\odot$-*inconsistent.*

Remark 9. If C is crisp, then Definition 5 is equivalent to Definition 4.

Remark 10. Let $\alpha, \beta \in [0,1], \alpha \geq \beta$. Evidently, if $C = \{\tilde{c}_{ij}\}$ is α-$\odot$-consistent, then it is β-$\odot$-consistent.

Remark 11. Let $\alpha \in [0,1]$. By Definition 2 and Definition 5, if $C = \{\tilde{c}_{ij}\}$ is α-$\odot$-consistent, then it is α-$\odot$-reciprocal. Consequently, if $C = \{\tilde{c}_{ij}\}$ is $\odot$-consistent, then it is $\odot$-reciprocal.

In order to extend Proposition 3 to the non-crisp case we define the notion of consistent vector with respect to a PCF matrix.

Definition 5. Let $\alpha \in [0,1]$, $C = \{\tilde{c}_{ij}\}$ be a PCF matrix. A vector $w = (w_1, w_2, ..., w_n)$, $w_i \in G$ for all $i \in \{1, 2, ..., n\}$, is an α-$\odot$-*consistent vector with respect to* C if for every $i, j \in \{1, 2, ..., n\}$ there exists $c_{ij} \in [\tilde{c}_{ij}]_\alpha$ such that

$$w_i \div w_j = c_{ij}. \tag{20}$$

The next proposition gives three necessary and sufficient conditions for a PCF matrix to be α-$\odot$-consistent.

Proposition 4. Let $\alpha \in [0,1]$, $C = \{\tilde{c}_{ij}\}$ be a PCF matrix. The following four conditions are equivalent.

(i) $C = \{\tilde{c}_{ij}\}$ is α-$\odot$-consistent.

(ii) There exists a vector $w = (w_1, w_2, ..., w_n)$ with $w_i \in G, i \in \{1, 2, ..., n\}$, such that for all $i, k \in \{1, 2, ..., n\}$ it holds

$$c_{ik}^L(\alpha) \leq w_i \div w_k \leq c_{ik}^R(\alpha). \tag{21}$$

(iii) For all $i, j, k \in \{1, 2, ..., n\}$, it holds

$$[c_{ik}^L(\alpha), c_{ik}^R(\alpha)] \cap [c_{ij}^L(\alpha) \odot c_{jk}^L(\alpha), c_{ij}^R(\alpha) \odot c_{jk}^R(\alpha)] \neq \emptyset. \tag{22}$$

(iv) For all $i, j, k \in \{1, 2, ..., n\}$, it holds

$$c_{ik}^L(\alpha) \leq c_{ij}^R(\alpha) \odot c_{jk}^R(\alpha), \tag{23}$$

$$c_{ik}^R(\alpha) \geq c_{ij}^L(\alpha) \odot c_{jk}^L(\alpha). \tag{24}$$

Remark 12. Property (iv) in Proposition 4 is useful for checking α-$\odot$-consistency of PCF matrices. For a given PCF matrix $C = \{\tilde{c}_{ij}\}$ it can be easily calculated whether inequalities (23) and (24) are satisfied or not.

3.3 Strong consistency of PCF matrices

In this section the consistency property of PCF matrices investigated in the previous section, particularly in Definition 5, will be strengthen. Similarly to α-$\odot$-strong reciprocity defined in Definition 3, we define strong α-$\odot$-consistent PCF matrices and derive their properties.

Definition 5. Let $\alpha \in [0, 1]$. A PCF matrix $C = \{\tilde{c}_{ij}\}$ is said to be *strong α-$\odot$-consistent*, if the following condition holds:
For every $i, j, k \in \{1, 2, ..., n\}$, and for every $c_{ij} \in [\tilde{c}_{ij}]_\alpha$, there exist $c_{ik} \in [\tilde{c}_{ik}]_\alpha$ and $c_{jk} \in [\tilde{c}_{jk}]_\alpha$ such that

$$c_{ik} = c_{ij} \odot c_{jk}. \tag{25}$$

The matrix C is said to be *strong $\odot$-consistent*, if C is strong α-$\odot$-consistent for all $\alpha \in [0, 1]$.

Remark 13. If C is crisp, then Definition 4 is equivalent to Definition 7.

Remark 14. Evidently, by Definitions 5 and 7, each strong α-$\odot$-consistent PCF matrix is α-$\odot$-consistent.

Remark 15. Notice that Remark 10 is not true for strong consistency. Here, the situation is similar to Remark 8 for strong reciprocity. Particularly, if $C = \{\tilde{c}_{ij}\}$ is a strong α-$\odot$-consistent PCF matrix, $\alpha, \beta \in [0, 1], \beta < \alpha$, then $C = \{\tilde{c}_{ij}\}$ need not be strong β-$\odot$-consistent. However, it must be β-$\odot$-consistent.

In the following proposition, we prove a parallel property to the property of Remark 11, namely, that a strong α-$\odot$-consistent PCF matrix is strong α-$\odot$-reciprocal.

Proposition 5. Let $C = \{\tilde{c}_{ij}\}$ be a PCF matrix, $\alpha \in [0, 1]$. If $C = \{\tilde{c}_{ij}\}$ is strong α-$\odot$-consistent, then $C = \{\tilde{c}_{ij}\}$ is strong α-$\odot$-reciprocal. Moreover, if $C = \{\tilde{c}_{ij}\}$ is strong $\odot$-consistent, then $C = \{\tilde{c}_{ij}\}$ is strong $\odot$-reciprocal.

In the following proposition we derive two necessary and sufficient conditions for strong α-$\odot$-consistency of a PCF matrix. This property may be useful for checking strong consistency of PCF matrices.

Proposition 6. Let $\alpha \in [0, 1]$, $C = \{\tilde{c}_{ij}\}$ be a PCF matrix. The following three conditions are equivalent.

(i) $C = \{\tilde{c}_{ij}\}$ is strong α-$\odot$-consistent.

(ii) For all $i, j, k \in \{1, 2, ..., n\}$, it holds

$$[c_{ik}^L(\alpha), c_{ik}^R(\alpha)] \cap [c_{ij}^R(\alpha) \odot c_{jk}^L(\alpha), c_{ij}^L(\alpha) \odot c_{jk}^R(\alpha)] \neq \emptyset. \tag{26}$$

(iii) For all $i, j, k \in \{1, 2, ..., n\}$, it holds

$$c_{ik}^L(\alpha) \leq c_{ij}^L(\alpha) \odot c_{jk}^R(\alpha), \tag{27}$$

$$c_{ik}^R(\alpha) \geq c_{ij}^R(\alpha) \odot c_{jk}^L(\alpha). \tag{28}$$

Remark 16. Property (iii) in Proposition 6 is useful for checking strong α-$\odot$-consistency of PCF matrices. For a given PCF matrix $C = \{\tilde{c}_{ij}\}$ it can be easily calculated whether inequalities (27) and (28) are satisfied or not.

3.4 Strict consistency of PCF matrices

In this section the consistency property of PCF matrices investigated in the previous section, particularly in Definition 7, will be even more strengthen. Similarly to α-$\odot$-strong consistency defined in Definition 3, we define strict α-$\odot$-consistent PCF matrices and derive their properties.

Definition 6. Let $\alpha \in [0,1]$. A PCF matrix $C = \{\tilde{c}_{ij}\}$ is said to be *strict α-$\odot$-consistent*, if the following condition holds:
For every $i,j,k \in \{1,2,...,n\}$, for every $c_{ik} \in [\tilde{c}_{ik}]_\alpha$and for every $c_{ij} \in [\tilde{c}_{ij}]_\alpha$, there exist $c_{jk} \in [\tilde{c}_{jk}]_\alpha$ such that

$$c_{ik} = c_{ij} \odot c_{jk}. \tag{29}$$

The matrix C is said to be *strict $\odot$-consistent*, if C is strict α-$\odot$-consistent for all $\alpha \in [0,1]$.

Remark 17. If C is crisp, then Definition 8 is equivalent to Definition 7.

Remark 18. Evidently, by Definitions 5, 7 and 8, each strict α-$\odot$-consistent PCF matrix is α-$\odot$-consistent.

Remark 19. Similarly to Remark 11, we obtain that a strict α-$\odot$-consistent PCF matrix is strong α-$\odot$-reciprocal.

In the following proposition we derive two necessary and sufficient conditions for strong α-$\odot$-consistency of a PCF matrix. This property may be useful for checking strong consistency of PCF matrices.

Proposition 7. Let $\alpha \in [0,1]$, $C = \{\tilde{c}_{ij}\}$ be a PCF matrix. The following two conditions are equivalent.

(i) $C = \{\tilde{c}_{ij}\}$ is strict α-$\odot$-consistent.

(ii) For all $i,j,k \in \{1,2,...,n\}$, it holds

$$c^L_{ik}(\alpha) \geq c^R_{ij}(\alpha) \odot c^L_{jk}(\alpha), \tag{30}$$

$$c^R_{ik}(\alpha) \leq c^L_{ij}(\alpha) \odot c^R_{jk}(\alpha). \tag{31}$$

Example 6. Consider the additive alo-group $\mathcal{R} = (\mathbb{R}, \odot, \leq)$ with $\odot = +$, see Example 1. Let PCF matrices $B = \{\tilde{b}_{ij}\}$ be given as follows:

$$B = \begin{bmatrix} 0 & (1;3;4) & (4;6;8) \\ (-4;-2;-1) & 0 & (2;4;5) \\ (-8;-6;-4) & (-5;-4;-2) & 0 \end{bmatrix},$$

or, equivalently, using Definition 1 by α-cuts for $\alpha \in [0;1]$, we obtain

$$B = \begin{bmatrix} [0;0] & [1+2\alpha;4-\alpha] & [4+2\alpha;8-2\alpha] \\ [-4+2\alpha;-1-\alpha] & [0;0] & [2+2\alpha;5-\alpha] \\ [-8+2\alpha;-4-2\alpha] & [-5+\alpha;-2-2\alpha] & [0;0] \end{bmatrix}.$$

Here, B is a 3×3 PCF matrix, particularly a PCF matrix with triangular fuzzy number elements, see (1), and the corresponding piece-wise linear membership functions.
Checking (31) and (32), we obtain that B is not strict α-$\odot$-consistent PCF matrix for any α, $0 \leq \alpha \leq 1$.
Checking inequalities (27) and (28), we obtain that B is strong α-$\odot$-consistent PCF matrix for all α, $0 \leq \alpha \leq \frac{2}{3}$. Therefore, for all α, $0 \leq \alpha \leq \frac{2}{3}$, B is strong α-$\odot$-consistent but not strict α-$\odot$-consistent PCF matrix.
Checking inequalities (23) and (24), we obtain that α-$\odot$-consistent PCF matrix for all α, $0 \leq \alpha \leq \frac{5}{6}$.
It is also clear that B is α-$\odot$-inconsistent PCF matrix for all α, $\frac{5}{6} < \alpha \leq 1$.

4 Priority vectors, inconsistency of PCF matrices

Definition of the priority vector for ranking the alternatives will be based on Proposition 4, (ii), particularly on the optimal solution of the following optimization problem:

(P2)

$$\alpha \longrightarrow \max; \tag{32}$$

subject to

$$c_{ij}^L(\alpha) \leq w_i \div w_j \leq c_{ij}^R(\alpha) \text{ for all } i,j \in \{1,2,...,n\}, \tag{33}$$

$$\bigodot_{k=1}^{n} w_k = e, \tag{34}$$

$$0 \leq \alpha \leq 1, w_k \in G, \text{ for all } k \in \{1,2,...,n\}. \tag{35}$$

If optimization problem (P2) has a feasible solution, i.e. system of constraints (34) - (36) has a solution, then (P2) has also an optimal solution. Let α^* and $w^* = (w_1^*,...,w_n^*)$ be an optimal solution of problem (P2). Then α^* is called the $\odot$-*consistency grade of* C, denoted by $g_\odot(C)$, i.e.

$$g_\odot(C) = \alpha^*. \tag{36}$$

Here, by Definition 7, $w^* = (w_1^*,...,w_n^*)$ is an α^*-$\odot$-consistent vector with respect to C called the $\odot$-*priority vector of* C.

If optimization problem (P2) has no feasible solution, then we define

$$g_\odot(C) = 0. \tag{37}$$

In that case, the priority vector is not defined at this moment, it will be defined later.

Generally, problem (P2) is a nonlinear optimization problem that can be efficiently solved e.g. by the dichotomy method, which is a sequence of optimization problems, see e.g. [11]. For instance, given $\alpha \in [0,1]$, $\odot = +$, problem (P2) can be solved as an LP problem (with variables $w_1,...,w_n$).

The proof of the following proposition can be found in [17].

Proposition 8. Let $C = \{\tilde{c}_{ij}\}$ be a PCF matrix, where all entries $\tilde{c}_{ij}$ are triangular fuzzy numbers. If $w^* = (w_1^*,...,w_n^*)$ is an optimal solution of (P2), i.e. $\odot$-priority vector of C, then w^* is unique.

Remark 20. The optimal solution α^* and $w^* = (w_1^*,...,w_n^*)$ of problem (P2) should be unique as decision makers usually ask for unique decision, or, a unique ranking of the alternatives in X. A sufficient condition for uniqueness of the priority vector $w^* = (w_1^*,...,w_n^*)$ is that all elements $\tilde{c}_{ij}$ of the PCF matrix C are bounded fuzzy intervals (i.e. fuzzy numbers) and, particularly, that the core of each $\tilde{c}_{ij}$,

$$Core(\tilde{c}_{ij}) = \{t \in G | \mu_{c_{ij}}(t) = 1\},$$

is a singleton, see Proposition 8. However, this is not the case of PCF matrices where the entries are fuzzy intervals (i.e. trapezoidal fuzzy numbers). Then the uniqueness is not secured and multiple solutions of (P2) can occur. In practical decision making problems such cases usually require reconsidering evaluations of some elements of the PCF matrix.

If, at least for one triple of elements $i,j,k \in \{1,2,...,n\}$ and some $\alpha \in [0,1]$, the condition (19) is not satisfied for any $c_{ij}^{'} \in [\tilde{c}_{ij}]_\alpha$, $c_{ik}^{'} \in [\tilde{c}_{ik}]_\alpha$ and any $c_{kj}^{'} \in [\tilde{c}_{kj}]_\alpha$, then the PCF matrix C is α-$\odot$-*inconsistent.* If for all $\alpha \in [0,1]$ the PCF matrix C is α-$\odot$-inconsistent, then we say that C is $\odot$-*inconsistent.* It is an important task to measure an intensity of $\odot$-inconsistency of the given PCF matrix. In some cases the PCF matrix can be "close" to some $\odot$-consistent matrix, in the other cases $\odot$-inconsistency can be strong, meaning that the PCF matrix can be "far" from any $\odot$-consistent matrix.

The inconsistency of C will be measured by the minimum of the distance of the "ratio" matrix $W = \{w_i \div w_j\}$ to the "left" matrix $C^L = \{c_{ij}^L(0)\}$ and/or "right" matrix $C^R = \{c_{ij}^R(0)\}$, as follows.

Let $w = (w_1, ..., w_n), w_i \in G,\ i, j \in \{1, ..., n\}, i < j$. Denote

$$\begin{aligned} d_{ij}(L, w) &= e \text{ if } c_{ij}^L(0) \le w_i \div w_j \le c_{ij}^R(0), \\ &= \|c_{ij}^L(0) \div (w_i \div w_j)\| \text{ otherwise,} \end{aligned} \tag{38}$$

$$\begin{aligned} d_{ij}(R, w) &= e \text{ if } c_{ij}^L(0) \le w_i \div w_j \le c_{ij}^R(0), \\ &= \|c_{ij}^R(0) \div (w_i \div w_j)\| \text{ otherwise,} \end{aligned} \tag{39}$$

$$I_{ij}(C, w) = \min\{d_{ij}(L, w), d_{ij}(R, w)\}, \tag{40}$$

where $\|...\|$ is the $\mathcal{G}$-norm from Section 2. We define

$$I_{\odot}(C, w) = \max\{I_{ij}(C, w) | 1 \le i < j \le n\}. \tag{41}$$

Consider the following optimization problem.

(P3)

$$I_{\odot}(C, w) \longrightarrow \min; \tag{42}$$

subject to

$$\bigodot_{k=1}^{n} w_k = e, \tag{43}$$

$$w_k \in G, \text{ for all } k \in \{1, 2, ..., n\}. \tag{44}$$

The $\odot$-*inconsistency index of* C, $I_{\odot}(C)$, is defined as

$$I_{\odot}(C) = \inf\{I_{\odot}(C, w) | w_k \text{ satisfies (44), (45)}\}. \tag{45}$$

Remark 21. If $w^* = (w_1^*, ..., w_n^*)$ is the optimal solution of (P3), then

$$I_{\odot}(C) = I_{\odot}(C, w^*).$$

Generally, the uniqueness of optimal solution of (P3) is not saved. Depending on the particular operation $\odot$, problem (P3) may have multiple optimal solutions which is an unfavorable fact from the point of view of the DM. In this case, the DM should reconsider some (fuzzy) evaluations of pairwise comparison matrix.

Now, we define a priority vector also in case of $g_{\odot}(C) = 0$, i.e. if no feasible solution of (P2) exists. In contrast to the case of $g_{\odot}(C) > 0$, this priority vector cannot become an α-$\odot$-consistency vector of C for some $\alpha > 0$. If $g_{\odot}(C) = 0$, then the optimal solution $w^* = (w_1^*, ..., w_n^*)$ of (P3) will be called the $\odot$-*priority vector of* C.

Remark 22. If $C = \{\tilde{c}_{ij}\}$ is a PCF matrix, then exactly one of the following two cases occurs:

- Problem (P2) has a feasible solution. Then consistency grade $g_{\odot}(C) = \alpha$, for some $\alpha,\ 0 \le \alpha \le 1$, and $I_{\odot}(C) = e$.
- Problem (P2) has no feasible solution. Then consistency grade $g_{\odot}(C) = 0$, C is $\odot$-inconsistent, and $I_{\odot}(C) > e$.

Remark 23. In particular, assume that C is $\odot$-consistent, then $g_{\odot}(C) = 1$ and by the properties of the distance function, (41) and (42), we obtain $I_{\odot}(C) = e$. However, if C is $\odot$-inconsistent, then $g_{\odot}(C) = 0$ and $I_{\odot}(C) > e$.

Example 7. Let $E = \{\tilde{e}_{ij}\}$ be a PCF matrix on the fuzzy multiplicative alo-group $]\mathbf{0}, \mathbf{1}[_{\mathbf{m}} = (]0, 1[, \bullet_f, \le)$, with:

$$a \bullet_f b = \frac{ab}{ab + (1-a)(1-b)}, e = 0.5, a^{(-1)} = 1 - a, \tag{46}$$

$$\|a\| = max\{a, 1-a\}.$$

Fuzzy multiplicative alo-group $]\mathbf{0},\mathbf{1}[_{\mathbf{m}}$ is divisible and continuous. For more details and properties, see Example 4, [4], and [17]. Let

$$E = \begin{bmatrix} 0.5 & (0.45;0.6;0.7) & (0.4;0.5;0.65) \\ (0.3;0.4;0.55) & 0.5 & (0.3;0.4;0.55) \\ (0.35;0.5;0.6) & (0.45;0.6;0.7) & 0.5 \end{bmatrix},$$

or, equivalently by α-cuts, $\alpha \in [0;1]$, we obtain

$$E = \begin{bmatrix} [0.5;0.5] & [0.45+0.15\alpha; 0.7-0.1\alpha] & [0.4+0.1\alpha; 0.65-0.15\alpha] \\ [0.3+0.1\alpha; 0.5-0.1\alpha] & [0.5;0.5] & [0.3+0.1\alpha; 0.55-0.15\alpha] \\ [0.35+0.15\alpha; 0.6-0.1\alpha] & [0.45+0.15\alpha; 0.7-0.1\alpha] & [0.5;0.5] \end{bmatrix}.$$

Here, E is a 3×3 PCF matrix, the elements of E are triangular fuzzy numbers with the piece/wise linear membership functions. By property (ii) in Proposition 2, E is strong α-$\odot$-reciprocal for all $\alpha \in [0,1]$, i.e. E is strong $\odot$-reciprocal. Solving problem (P2), the $\bullet_f$-priority vector w^* of E is $w^* = (0.533, 0.500, 0.467)$ and, the consistency grade $g_{\bullet_f}(E) = 0.556$. Hence, E is α-$\odot$-consistent for all $\alpha \leq 0.556$. Moreover, inconsistency index $I_{\bullet_f}(E) = 0.5$.

5 Conclusion

This paper deals with pairwise comparison matrices with fuzzy elements. Fuzzy elements of the pairwise comparison matrix are usually applied whenever the decision maker is not sure about the value of his/her evaluation of the relative importance of elements in question. In comparison with PC matrices investigated in the literature, here we investigate PC matrices with elements from abelian linearly ordered group (alo-group) over a real interval (called PCF matrices). We generalized the concept of reciprocity and consistency of pairwise comparison matrices with triangular fuzzy numbers to bounded fuzzy intervals of alo-groups (trapezoidal fuzzy numbers). We also defined the concept of priority vector which is an extension of the well known concept in crisp case and which is used for ranking the alternatives. Such an approach allows for extending the additive, multiplicative and also fuzzy approaches known from the literature. The concept of reciprocity and consistency of pairwise comparison matrices with triangular fuzzy numbers (PCF matrices) was already studied in the former author's works as well as in the other literature on this subject, see e.g. [8], [11], [17], [18], [20] and [12]. Here, we defined stronger concepts, namely the strong reciprocity and strong consistency of PCF matrices and investigated their properties as well as some consequences to the problem of ranking the alternatives. Particularly, we derived necessary and sufficient conditions for PCF matrices to be strong reciprocal and strong consistent. Strong reciprocity and/or strong consistency may be useful when constructing membership functions of fuzzy elements of PCF matrices. Moreover, we also solved the problem of measuring the inconsistency of PCF matrices by defining corresponding indexes. The first index, called the consistency grade, $g(C)$, is the maximal α of the α-cut, such that the corresponding PCF matrix is α-consistent. Moreover, the inconsistency index $I(C)$ of the PCF matrix C was defined for measuring the fuzziness of this matrix by the distance of the PCF matrix to the closest crisp consistent matrix. Consequently, a PCF matrix is either consistent, then $0 \leq g \leq 1$, and the inconsistency index I is equal to the identity element e of $\mathcal{G}$, or, the PCF matrix is inconsistent, then $g = 0$, and I is greater than the identity element e. Several numerical examples were presented to illustrate the concepts and properties.

Acknowledgement
This paper was supported by the Ministry of Education, Youth and Sports Czech Republic within the Institutional Support for Long-term Development of a Research Organization in 2017.

References

[1] Bourbaki, N. *Algebra II.* Springer Verlag, Heidelberg-New York-Berlin, 1990.

[2] Cavallo, B., D'Apuzzo, L. (2009). A general unified framework for pairwise comparison matrices in multicriteria methods. *International Journal of Intelligent Systems*, 24: 4, 377–398, 2009.

[3] Cavallo, B., D'Apuzzo, L. Characterizations of Consistent Pairwise Comparison Matrices over Abelian Linearly Ordered Groups. *International Journal of Intelligent Systems*, 25:1035–1059, 2010.

[4] Cavallo, B., D'Apuzzo, L. Deriving weights from a pairwise comparison matrix over an alo-group. *Soft Computing*, 16:353–366, 2012.

[5] Ishizaka, A., Nguyen, N.H. Calibrated fuzzy AHP for current bank account selection. *Expert Systems with Applications*, 40:3775–3783, 2013.

[6] Leung, L.C., Cao, D. On consistency and ranking of alternatives in fuzzy AHP. *European Journal of Operational Research*, 124:102–113, 2000.

[7] Mahmoudzadeh, M., Bafandeh, A.R. A new method for consistency test in fuzzy AHP. *Journal of Intelligent and Fuzzy Systems*, 25: 2 457–461, 2013.

[8] Mikhailov, L. (2003). Deriving priorities from fuzzy pairwise comparison judgments. *Fuzzy Sets and Systems*, 134:365–385, 2003.

[9] Mikhailov, L. A fuzzy approach to deriving priorities from interval pairwise comparison judgements. *European Journal of Operational Research*, 159:687–704, 2004.

[10] Mikhailov, L., Tsvetinov, P. Evaluation of services using a fuzzy analytic hierarchy process. *Applied Soft Computing*, 5:23–33, 2004.

[11] Ohnishi, S., Dubois, D. et al. A fuzzy constraint based approach to the AHP. In: Uncertainty and Intelligent Inf. Syst., World Sci., Singapore, 2008, 217–228, 2008.

[12] Kou, G., Ergu, D., Lin, C. S.; Chen, Y. Pairwise comparison matrix in multiple criteria decision making. *Technological and Economic Development of Economy*,22:5, 738-765, 2016.

[13] Ohnishi, S., Yamanoi, T., Imai, H. A weights representation for fuzzy constraint-based AHP. IPMU 2008 http://www.gimac.uma.es/ipmu08/proceedings/papers/036-OhnishiYamanoiImai.pdf

[14] Ramik, J., Vlach, M. *Generalized concavity in optimization and decision making.* Kluwer Academic Publishers, Boston-Dordrecht-London 2001.

[15] Ramik, J., Korviny, P. Inconsistency of pairwise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems*, 161:1604–1613, 2010.

[16] Ramik, J. Isomorphisms between fuzzy pairwise comparison matrices. *Fuzzy Optim. Decis. Making*, 14:199–209, 2014.

[17] Ramik, J. Pairwise comparison matrix with fuzzy elements on alo-group. *Information Sciences*, 297:236–253.

[18] Salo, A.A. (1996). On fuzzy ratio comparison in hierarchical decision models. *Fuzzy Sets and Systems* 84 21–32, 2015.

[19] Bilgic, T., Turksen, I.B. Measurement of membership functions: theoretical and empirical work. In: Dubois, D., Prade, H., Eds. *Fundamentals of fuzzy sets*, Kluwer Academic Publ., New York, 2000, 195–227.

[20] Xu, Z.S., Chen, J. Some models for deriving the priority weights from interval fuzzy preference relations. *European Journal of Operational Research*, 184:266–280, 2008.

[21] Zhang, H. Group decision making based on multiplicative consistent reciprocal preference relations. *Fuzzy Sets and Systems*, 282:31–46, 2016.

A machine learning method for incomplete and imbalanced medical data

Issam SALMAN[1] and Jiří VOMLEL[2]

[1]*Department of Software Engineering*
Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University in Prague
Trojanova 13, 12000, Prague, Czech Republic
issam.salman@fjfi.cvut.cz

[2]*Institute of Information Theory and Automation*
Academy of Science of the Czech Republic
Pod vodárenskou věží 4, 18208, Prague, Czech Republic
vomlel@utia.cas.cz

Abstract

Our research reported in this paper is twofold. In the first part of the paper we use standard statistical methods to analyze medical records of patients suffering myocardial infarction from the third world Syria and a developed country - the Czech Republic. One of our goals is to find whether there are statistically significant differences between the two countries. In the second part of the paper we present an idea how to deal with incomplete and imbalanced data for tree-augmented naive Bayesian (TAN). All results presented in this paper are based on a real data about 603 patients from a hospital in the Czech Republic and about 184 patients from two hospitals in Syria.

Keywords: Machine Learning, Data analysis, Bayesian networks, Missing data, Imbalanced data, Acute Myocardial Infarction.

1 Introduction

Acute myocardial infarction (AMI) is commonly known as a heart attack. A heart attack occurs when an artery leading to the heart becomes completely blocked and the heart doesn't get enough blood or oxygen. Without oxygen, cells in that area of the heart die. AMI is responsible for more than a half of deaths in most countries worldwide. Its treatment has a significant socioeconomic impact.

One of the main objectives of our research is to design, analyze, and verify a predictive model of hospital mortality based on clinical data about patients. A model that predicts well the mortality can be used, for example, for the evaluation of the medical care in different hospitals. The evaluation based on mere mortality would not be fair to hospitals that treat often complicated cases. It seems better to measure the quality of the health care using the difference between predicted and observed mortality.

A related work was published by [1]. The authors analyze the mortality data in U.S. hospitals using the logistic regression model. Other work was published by [2]. The authors compare different machine learning methods using a real medical data from a hospital.

2 Data

Our dataset contains data about 787 patients characterized by 24 variables. 603 patients of them are from the Czech Republic [2] and 184 are from Syria. The attributes are listed in the Table 1. Most of the attributes are real valued, four attributes are nominal. Only a subset of attributes was measured for the Syrian patients.

Most records contain missing values, i.e., for most patients only some attribute values are available. The thirty days mortality is recorded for all patients. In the Czech Republic the results of blood tests are reported in millimoles per liter of blood. In Syria some of the measurements are reported in milligrams per liter and some in millimoles per liter. We standartize all measurements to the millimoles per liter scale.

We will note $\mathbf{U} = \{X_1, X_2, \ldots, X_m\}$ for a discrete domain, where $X_i, i \in \{1, 2, \ldots, m\}$ is a discrete attribute and take on values from a finite set, denoted by $Val(X_i)$. We use capital letters such as X, Y, Z for attribute names, and lower-case letters such as x,y,z to denote specific values taken by those variables. Sets of variables are denoted by boldface capital letters such as $\mathbf{X}$,$\mathbf{Y}$,$\mathbf{Z}$ and assignments of values to the variables in these sets are denoted by boldface lowercase letters $\mathbf{x}$,$\mathbf{y}$,$\mathbf{z}$. A classified discrete domain is a discrete domain where one of the attributes is distinguished as class. We will use $\mathbf{U}_C = \{A_1, A_2, \ldots, A_n, C\}$ for a classified discrete domain. A dataset $D = \{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$ of instances of $\mathbf{U}_C$, where each $\mathbf{u}_i, i \in \{1, \ldots, N\}$ is a tuple of the form $(a_i^1, \ldots, a_i^n, c_i)$ where $a_i^1 \in Val(A_1), \ldots, a_i^n \in Val(A_n)$ and $c_i \in Val(C)$. Also we note that the class is always known, and a missing value in the dataset is denoted by NA.

Table 1: Attributes

Attribute	**Code**	**type**	**value range in data**	**Country**
Age	AGE	real	[23, 94]	SYR, CZ
Height	HT	real	[145, 205]	CZ
Weight	WT	real	[35, 150]	CZ
Body Mass Index	BMI	real	[16.65, 48.98]	CZ
Gender	SEX	nominal	{male, female}	SYR, CZ
Nationality	NAT	nominal	{Czech, Syrian}	SYR, CZ
STEMI Location	STEMI	nominal	{inferior, anterior, lateral}	SYR, CZ
Hospital	Hospital	nominal	{CZ, SYR1, SYR2}	SYR, CZ
Kalium	K	real	[2.25, 7.07]	CZ
Urea	UR	real	[1.6, 61]	SYR, CZ
Kreatinin	KREA	real	[17, 525]	SYR, CZ
Uric acid	KM	real	[97, 935]	SYR, CZ
Albumin	ALB	real	[16, 60]	SYR, CZ
HDL Cholesterol	HDLC	real	[0.38, 2.92]	SYR, CZ
Cholesterol	CH	real	[1.8, 9.9]	SYR, CZ
Triacylglycerol	TAG	real	[0.31, 11.9]	SYR, CZ
LDL Cholesterol	LDLC	real	[0.261, 7.79]	SYR, CZ
Glucose	GLU	real	[2.77, 25.7]	SYR, CZ
C-reactive protein	CRP	real	[0.3, 359]	SYR, CZ
Cystatin C	CYSC	real	[0.2, 5.22]	SYR, CZ
N-terminal prohormone of brain natriuretic peptide	NTBNP	real	[22.2, 35000]	CZ
Troponin	TRPT	real	[0, 25]	CZ
Glomerular filtration rate (based on MDRD)	GFMD	real	[0.13, 7.31]	CZ
Glomerular filtration rate (based on Cystatin C)	GFCD	real	[0.09, 7.17]	CZ

3 Preliminary Statistical Analysis

For a preliminary statistical analysis [3] we selected a subset of attributes that are highly correlated with the class [5] and present in both groups, namely, we considered these variables: age, nationality, gender, STEMI location, and the class mortality.

The STEMI location encoded by 1 denotes a STEMI.inf, 2 denotes a STEMI.ant, and 3 denotes a STEMI.lat. The nationality is encoded by a binary variable, where 0 means Czech and 1 means Syrian. The Gender is encoded by a binary variable where 0 denotes a man, while 1 stands for

a female. The mortality is also encoded as a binary variable, where 0 means that the patient survived 30 days, while 1 means that he/she did not.

Already from Figure 1, where the histogram of the age values is presented, we can see that from patients that didn't survive a high percentage are young patients from Syria.

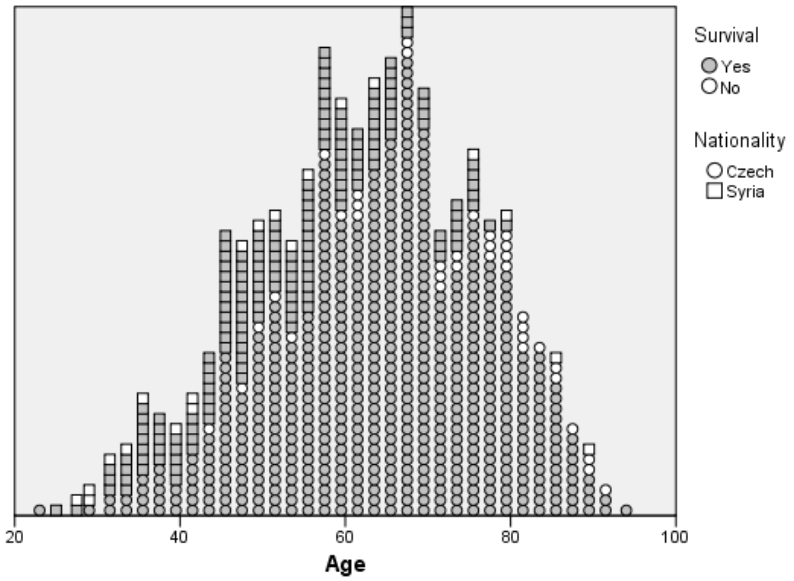


Figure 1: Histogram of the age values

Table 2: The Chi-Square Test of conditional independence

		gender	STEMI loc.	mortality	nationality
age	value	**.174**	-.010	.048	**-.381**
	sign.	**.0001**	.775	.181	**.0001**
gender	value		.022	.068	**.92**
	sign.		.53	.057	**.01**
STEMI loc.	value			-.026	-.036
	sign.			0.46	.312
mortality	value				**.089**
	sign.				**0.013**

The standard chi-square test of conditional independence between two variables reveals (see Table 2) that there is a significant dependence (at the level 0.05) between the mortality and nationality, the gender and nationality, also there are a significant dependencies between the gender and age, the mortality and gender – the patients from Syria have the lowest probability to survive, also they are younger and there is higher percentage of woman.

Finally, we learned the logistic regression model, that describes the relationship between the considered independent variables and the mortality as the dependent variable. We have got:

$$\begin{aligned} \text{logit } P(C=1|A=a) &= \beta_0 + \beta_1 a_1 + \ldots + \beta_4 a_4 \\ &= -0.034 + 0.001 \cdot a_1 + 0.027 \cdot a_2 - 0.007 \cdot a_3 + 0.065 \cdot a_4 \end{aligned}$$

where a_1: age, a_2: gender, a_3: STEMI loc, and a_4: nationality. Variables age and nationality appeared to be statistically significant for mortality prediction.

From the preliminary statistical analysis we can conclude that:

- In Syria the mortality from AIM is significantly higher than in the Czech Republic – 87.3% Syrian patients survive, while 94.7% patients from the Czech Republic survive.
- The age of patients in Syria is lower in average (the average difference is 13 years) and there is a higher prevalence of women among the patients with AIM in Syria than in the Czech Republic.
- The STEMI location is related to the mortality.

4 Machine Learning Methods

The preliminary statistical analysis studied mostly the pairwise relations only. Since the explanatory variables may combine their influence and the influence of a variable may be mediated by another variable it is worth of studying the relations of variables alltogether. Our data are incomplete and imbalanced. We will present an idea for dealing with that type of data using tree-augmented naive Bayesian (TAN).

4.1 Bayesian networks

A Bayesian network [6] is an annotated directed acyclic graph that encodes a mass probability distribution over a set of random variables $\mathbf{U}$. Formally, a Bayesian network for $\mathbf{U}$ is a pair $B = \langle G, \Theta \rangle$. The first component, G, is a directed acyclic graph whose vertices correspond to the random variables $\mathbf{U} = \{X_1, X_2, \ldots, X_m\}$, and whose edges represent direct dependencies between the variables. The graph G encodes independence assumptions: each variable X_i is independent of its non-descendants given its parents in G. The second component of the pair, namely Θ, represents the set of parameters that quantifies the network. It contains the parameter $\theta_{x_i|\Pi_{x_i}} = f(x_i|\Pi_{x_i})$ for each possible value x_i of X_i and Π_{x_i} of Π_{X_i}, where Π_{X_i} denotes the set of parents of X_i in G. Accordingly, a Bayesian network B defines a unique joint probability distribution over $\mathbf{U}$ given by:

$$f(X_1 = x_1, \ldots, X_m = x_m) \quad = \quad \prod_{i=1}^{m} f(X_i = x_i | \Pi_{X_i} = \Pi_{x_i}) \quad = \quad \prod_{i=1}^{m} \theta_{x_i|\Pi_{x_i}}$$

for each Π_{X_i} which is a parent of X_i.

4.2 Learning with Trees

A directed acyclic graph on $\{X_1, X_2, \ldots, X_n\}$ is a tree if Π_{X_i} contains exactly one parent for all X_i, except for one variable that has no parents (this variable is referred to as the root). A tree network can be described by identifying the parent of each variable [7]. A function $\pi : \{1, \ldots, n\} \to \{0, \ldots, n\}$ is said to define a tree over $X_1, X_2, \ldots, X_n$ if there is exactly one i such that $\pi(i) = 0$ (namely the root of the tree), and there is no sequence $i_1, \ldots, i_k$ such that $\pi(i_j) = i_{j+1}$ for $i \leq j < k$ and $\pi(i_k) = i_1$ (i.e., no cycles). Such a function defines a tree network where $\Pi_{X_i} = \{X_{\pi(i)}\}$ if $\pi(i) > 0$ and $\Pi X_i = \emptyset$ if $\pi(i) = 0$.

4.3 Learning Maximum Likelihood TAN

Let $\{A_1, A_2, \ldots, A_n\}$ be a set of attribute variables and C be the class variable. We say that B (Bayesian network) is a TAN model if $\Pi_C = \emptyset$ and there is a function that defines a tree over $\{A_1, A_2, \ldots, A_n\}$. The optimization problem consists on finding a tree defining function π over $\{A_1, A_2, \ldots, A_n\}$ such that the log likelihood is maximized [8] $LL(B_T|D) = \sum_{\mathbf{u} \in D} \log f(\mathbf{u})$. To learn the maximum likelihood TAN we should use the following equation to compute the parameters [8], $\theta_{a_i,\Pi_{a_i}} = \frac{N_{a_i,\Pi_{a_i}}(a_i,\Pi_{a_i})}{N_{\Pi_{a_i}}(\Pi_{a_i})}$ where $N_{a_i,\Pi_{a_i}}(a_i, \Pi_{a_i})$ stands for the number of times that attribute i has value a_i and its parents have values Π_{a_i} in the dataset. Similarly, $N_{\Pi_{a_i}}(\Pi_{a_i})$ is the number of times that the parents of attribute A_i have values Π_{a_i} in the dataset.

5 Learning TAN from incomplete data

Missing data are a very common problem which is important to consider in a many data mining applications, and machine learning or pattern recognition applications. Some variables may not be observable (i.e. hidden) even for training instances. Now more and more datasets are available, and most of them are incomplete. Therefore, we want to find a way to build a new model from an incomplete dataset. Normally, to learn the maximum likelihood TAN structure [8], we need a complete data, such that all

instances $\mathbf{u}_i, i \in \{1,\ldots,N\}$ from $\mathbf{U}_C$ are complete and don't have any missing value. In case the data are incomplete and there is an instance which has a missing value, we will not use the whole instance in TAN structure learning i.e. not use the other known values from that instance in TAN structure learning. Note that the class is always known, and a missing value in the dataset is denoted by NA. Our goal is to learn a tree-augmented naive Bayesian (TAN) from incomplete data. Some previous work by [13] propose maximizing conditional likelihood for BN parameter learning. They apply their method to MCAR (Missing Completely At Random) incomplete data by using available case analysis in order to find the best TAN classifier. In other work by [9] also deals with TAN classifiers and expectation-maximization (EM) principle for partially unlabeled data. In their work, only the variable corresponding to the class can have missing. Also, other work by [10] deals with TAN based on the EM principle, where they have proposed an adaptation of the learning process of Tree Augmented Naive Bayes classifier from incomplete data. In their work, any variable can have missing values in the dataset. The TAN algorithm can be adapted to learn from incomplete datasets, such that most available data will be used in TAN structure learning. The procedure is shown in Algorithm 1, where the Conditional Mutual Information "CMI" is defined as:

$$I(X,Y|Z) = \sum_{\mathbf{x,y,z}} f(\mathbf{x,y,z}) \log \frac{f(\mathbf{z})f(\mathbf{x,y,z})}{f(\mathbf{x,z})f(\mathbf{y,z})}$$

where the sum is only over $\mathbf{x,y,z}$ such that $f(\mathbf{x,z}) > 0$ and $f(\mathbf{y,z}) > 0$.

Algorithm 1 TAN For Incomplete Data

1: **procedure** CMI($A_i, A_j, C\}$) ▷ // Conditional Mutual Information
2: $\overline{D} = \{\overline{\mathbf{u}}_1,\ldots,\overline{\mathbf{u}}_N\}, \overline{\mathbf{u}}_m = (a_i, a_j, c), m \in \{1,\ldots,N\}$, such that $\mathbf{u}_m = (a_1,\ldots,a_n,c) \in D$
3: **Foreach** $\overline{\mathbf{u}}_m \in \overline{D}$
4: **If**$(a_i == NA | a_j == NA)$
5: Delete $\overline{\mathbf{u}}_m$ from $\overline{D}$
6: **endfor**
7: Compute $I_p = I(A_i, A_j|C)$ from $\overline{D}$
8: **return** I_p
9: **Endprocedure**
10: Read $D = \{\mathbf{u}_1,\ldots,\mathbf{u}_N\}, \mathbf{u}_m = (a_1,\ldots,a_n,c), m \in \{1,\ldots,N\}$
11: var:
12: n the number of attribute variables A;
13: $\mathbb{I}_p[n][n]$ the WeightMatrix;
14: UG the UndirectedGraph;
15: UT the UndirectedTree;
16: T the DirectedTree;
17: TAN the DirectedGraph;
18: **Foreach** $A_i, i \in \{1,\ldots,n-1\}$
19: **Foreach** $A_j, j \in \{2,\ldots,n\}$
20: $I_{p_ij} = CMI(A_i, A_j, C)$
21: $\mathbb{I}_p[i][j] = I_{p_ij}$
22: $\mathbb{I}_p[j][i] = I_{p_ij}$
23: **EndForeach**
24: **EndForeach**
25: G = ConstructUndirectedGraph($\mathbb{I}_p[i][j]$)
26: UT = MaximumWeightedSpanningTree(G);
27: T = MakeDirected(UT);
28: TAN = AddClass(T);

In Algorithm 1, on line 25 we build a complete undirected graph in which the vertices are the attributes $A_1,\ldots,A_n$. Annotate the weight of an edge connecting A_i to $A_j, i \neq j$ by $I_{p_ij} = I(A_i, A_j|C)$ One line 26 we build a subgraph from G, without any cycles and with the maximum possible total edge weight. On line 27 we transform the resulting undirected tree to a directed one by choosing a root variable and setting the direction of all edges to be outward from it. On line 28 we add the class C to the graph as a node and add edges from C to all other nodes in the graph

The idea behind Algorithm 1 is that we believe if we use more data then the estimates of conditional mutual information are more reliable.

6 Imbalanced Data

In case of imbalanced data the classifiers are more sensitive to detecting the majority class and less sensitive to the minority class. Thus, if we don't take care of the issue, the classification output will be biased, in many cases resulting in always predicting the majority class. Many methods have been proposed in the past few years to deal with imbalanced data. In our research the mortality rate of patients with myocardial infarction refers to the percentage of patients who have not survived more than 30 days, where the results are 89% of patients survive and 11% of patients do not survive, therefore the data are quite imbalanced. One of the most common and simplest strategies to handle imbalanced data is to under-sample the majority class [11, 12]. While different techniques have been proposed in the past, they did not bring any improvement with respect to simply selecting samples at random. So, for this analysis we propose the following steps:

- Let M be the number of samples for the majority class, and N be the number of samples for the minority class, and M be L times greater than N.
- Divide the instances which have majority class into L distinct clusters.
- Train L predictors, where each predictor is trained on only one of the distinct clusters, but on all of the data from the rare class. To be clear, the data from the minority class are used in the training of all L predictors.
- Use model averaging for the L learned predictors as your final predictor. i.e (in our case we will compute a conditional mutual information between each pair of attributes $(A_i, A_j), i, j \in 1, 2, \ldots, n, i \neq j$ given the class L times for each pair, in each time will use only one of the distinct clusters and all data from the minority class, then we will use the average of conditional mutual information for each pair to compute a weight matrix).

After integrating this step into the Algorithm 1, we will have a TAN algorithm which deals with an incomplete and imbalance data 2:

Algorithm 2 TAN for incomplete and imbalance data

1: var
2: M The number of samples for the majority class
3: N The number of samples for the minority class
4: D_T All instances of the majority class, $D_T \subset D$
5: D_F All instances of the minority class, $D_F \subset D$
6: integer division $L = M/N$
7: Divide D_T to L parts, $D_{T_k}, k \in \{1, \ldots, L\}$
8: **Foreach** D_{T_k}
9: $D_k = D_{T_k} \cup D_F$
10: **EndForeach**
11: Compute WeightMatrix $\mathbb{I}_{p_k}[n][n]$ foreach D_k
12: $\hat{\mathbb{I}}_p[n][n]$ = the average of $\mathbb{I}_{p_k}[n][n], k \in 1, \ldots, L$ ▷ // $\hat{\mathbb{I}}_p$ is the WeightMatrix which wwill be used in Algorithm 1
13: Continue from line 26 in Algorithm 1 using $\hat{\mathbb{I}}_p$

7 Results

For each data record classified by a classifier there are four possible classification results. Either the classifier got a positive example labeled as positive (in our data the positive example is the patient survived) or it made a mistake and marked it as negative. Conversely, a negative example may have been mislabeled as a positive one, or correctly marked as negative. Our results are summarized in Figure 2 using the ROC curves. We use the 10 fold cross validation as the model evaluation method. The ROC curve shows how the classifier can sacrifice the true positive rate (TP rate: number of positive examples, labeled as such over total positives) for the false positive rate(FP rate: number of negative examples, labeled as positive over total negatives) (1-specificity) by plotting the TP rate to the FP rate. In other words, it shows how many correct positive classifications can be gained as you allow for more and more false positives by changing the threshold.

In Figure 2 we compare our results with normal TAN ([8]) and SMOTE algorithm ([4]) for TAN. Algorithm 2 has achieved the highest area under the ROC curve (AUC) with 0.82. The results of Algorithm 1 (ROC = 0.77) is better than the normal TAN algorithm (ROC = 0.62). But SMOTE algorithm with TAN (ROC = 0.802) is better than Algorithm 1.

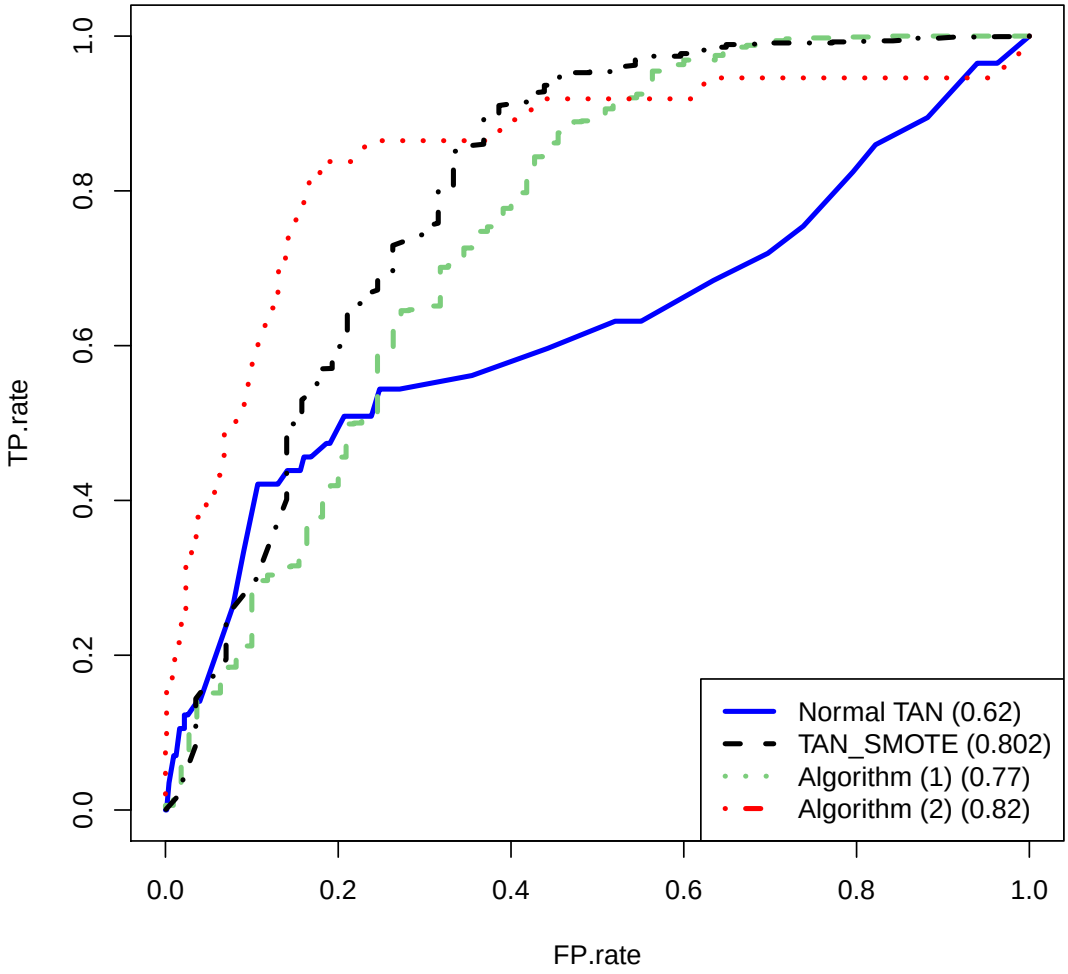


Figure 2: ROCs (TAN , TAN_SMOTI , Algorithm(1) , Algorithm(2))

8 Conclusions

First, we used medical data on patients with AIM for preliminary statistical analysis. We found a significant difference between Syrian patients and Czech patients. Second, Bayesian networks are a tool of choice for reasoning in uncertainty, with incomplete data. However, often, Bayesian network structural learning only deals with complete data. We have proposed here an adaptation of the learning process of the Tree Augmented Naive Bayes classifier from incomplete and imbalanced datasets. This methods have been successfully tested on our dataset. We have seen that our Algorithm 2 performed better than normal TAN and TAN-SOMTE.

Acknowledgement
This work was supported by the Czech Science Foundation through projects 16-12010S, and the student grant CTU SGS16/253/OHK3/3T/14.

References

[1] H. M. Krumholz, S.-L. T. Normand, D. H. Galusha, J. A. Mattera, A. S. Rich, Y. Wang and Y. Wang, *Risk-Adjustment Models for AMI and HF 30-Day Mortality, Methodology*, Harvard Medical School,

Department of Health Care Policy, (2007).

[2] J. Vomlel and H. Kružík and P. Tůma and J. Přeček, and M. Hutyra, *Machine Learning Methods for Mortality Prediction in Patients with ST Elevation Myocardial Infarction*, In the Proceedings of The Nineth Workshop on Uncertainty Processing WUPES'12, Czech Republic, 204–213, (2012).

[3] L. Wasserman. *All of Statistics*, Springer-Verlag New York, (2004).

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, *Synthetic Minority Over-sampling Technique*, Journal of Artificial Intelligence Research, Volume 11, Issue 16, 321–357, (2002).

[5] M. Hall and E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and H. Witten, *The WEKA Data Mining Software: an Update*, In 'ACM SIGKDD Exploration ACM SIGKDD Explorations', Volume 11, Issue 1. (2009), 10–18.

[6] F. V. Jensen, *An Introduction to Bayesian Networks*, Springer, (1996).

[7] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, (1988).

[8] N. Friedman, D. Geiger, and M. Goldszmidt, *Bayesian network classifiers*, Machine Learning Journal, Volume 29, Issue 2. (1997). 131–163.

[9] I. Cohen and F. Cozman and N. Sebe and M. C. Cirelo and T. S. Huang, *Semi-supervised learning of classifiers: theory, algorithms and their application to human-computer interaction*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 26, Issue 12, 1553–1568, (2004).

[10] O. C. H. Francois and P. Leray, *Learning the Tree Augmented Naive Bayes Classifier from incomplete datasets*,

Third European Workshop on Probabilistic Graphical Models, 91–98, (2006).

[11] R. Laza, R. Pavon, M. Reboiro-Jato and F. Fdez-Riverola R. Laza and et al, Evaluating the effect of unbalanced data in biomedical document classification, Journal of Integrative Bioinformatics, Volume 16, Issue 3, pp. 177, (2011).

[12] M. M. Rahman and D. N. Davis, Addressing the Class Imbalance Problem in Medical Datasets, International Journal of Machine Learning and Computing, volume 3, Issue 2, 224-228,(2013)

[13] R. Greiner and W. Zhou, Structural extension to logistic regression, Eighteenth Annual National Conference on Artificial Intelligence (AAI02), 167–173, (2002).

Standard Path Reduction based Algorithm for Optimal Paths in Multi-Objective Networks

Natsumi TAKAHASHI[1], Hisashi YAMAMOTO[2], Tomoaki AKIBA[3] and Shao-Chin SUNG[4]

[1,4] *Aoyama Gakuin University*
5-10-1 Fuchinobe, Chuo-ku, Sagamihara Kanagawa, 252-5258, Japan
[1]takahashi@ise.aoyama.ac.jp, [4]son@ise.aoyama.ac.jp

[2] *Tokyo Metropolitan University*
6-6 Asahigaoka, Hino Tokyo 191-0065, Japan
yamamoto@tmu.ac.jp

[3] *Chiba Institute of Technology*
2-17-1 Tsudanuma Narashino Chiba 275-0016, Japan
tomo@akiaki.net

Abstract

Many networks have been applied extensively in the real world, for example, scheduling of a production and distribution management system and search system for optimal routes in the Internet Services. These networks are formulated as a multi-objective network which has multi-criteria. In this study, we obtain optimal paths (Pareto solutions) for such a multi-objective network. Extended Dijkstra's algorithm is effective in obtaining Pareto solutions of multi-objective network. However, this algorithm takes large memory area to obtain optimal paths when criteria of network are increase. Therefore, our previous study proposed the algorithm that standard path reduced search space from that of extended Dijkstra's algorithm. This study considers several standard paths and evaluates the efficiency of standard path reduction. We conduct experiments to compare computing times of the proposed algorithms which differ in the number of standard paths.

Keywords: Multi-objective Network, Dijkstra's Algorithm, Optimal Path Problem, Pareto Solutions.

1 Introduction

Many network systems have been applied extensively in the real world, for example, scheduling of a production and distribution management system and search systems for optimal routes on Internet services. Many papers have deal with networks with two specific nodes. When network has two specific nodes (start node and terminal node), many paths between two specific nodes are formed by combinations of edges. Each edge has criteria, for example, distances, costs, and the time required. We obtain the particular path which minimizes the values of criteria. Such a path is called optimal path in this study.

Dijkstra [5] and the Bellman-Ford algorithms [4, 6] are well known to be effective in obtaining the optimal path for one-objective networks. In this paper, we consider multi-objective networks, where each edge has two and more criteria. For multi-objective networks, several studies have considered one criterion as an objective function and the other criteria as the constraints [3, 7]. However, a few studies have been conducted on multi-objective networks for obtaining Pareto solutions. Akiba *et al.* [1, 2] proposed an extended Dijkstra's algorithm that obtained all Pareto solutions in two and three objective networks by applying the algorithm for an optimal path with a constraint condition. In multi-objective networks, the search for paths uses a large memory area and takes much computing time in the extended Dijkstra's algorithm. For three-objective network, Akiba *et al.* [2] made their algorithm more efficient than the extended Dijkstra's algorithm.

However, their algorithms could not obtain all Pareto solutions. Thus, an efficient and exact algorithm that can obtain all Pareto solutions is required.

Our previous study considered two or more objective network. We extended the idea of Takahashi *et al.* [8] to network with more than three objective functions. That study proposed the reduction idea of search space by using standard paths. In our study, "Standard path reduction" means reducing search space by standard path. More standard paths tend to reduce more search space. But, it needs computing time to obtain standard paths. Therefore, this study evaluates efficiency of standard path reduction. We conduct numerical experiments and compare computing times of the proposed algorithms which differ in the number of standard paths.

2 Definition of Problem

2.1 Network Systems

This section defines some notations to state our problem. Let $G = (V, E)$ be a given network. V denotes a set of nodes. Node $s \in V$ is the start node, and $t \in V$ is the terminal node in a network. E denotes a set of directed edges $E = \{e_1, e_2, \ldots, e_n\}$, where n is the number of edges, and e_j is the j-th edge for $j = 1, 2, \ldots, n$. Connection of node s to node t needs some edges. The set of such edges is called a path from node s to node t. M denotes the number of objective functions. Each edge $e_j \in E$ has a non-negative cost vector $\mathbf{c}_j = (c_{1j}, c_{2j}, \ldots, c_{Mj})$, where its i-th component c_{ij} is the cost associated with the i-th objective. Also, we suppose costs $c_{1j}, c_{2j}, \ldots, c_{Mj}$ in $\mathbf{c}_j$ occur if edge e_j is included in a path. For example, for each simple path P, the cost of P is defined as follows.

$$\sum_{e_j \in E_P} \mathbf{c}_j = \left(\sum_{e_j \in E_P} c_{1j}, \sum_{e_j \in E_P} c_{2j}, \ldots, \sum_{e_j \in E_P} c_{Mj} \right),$$

where E_P is the set of edges appearing in P.

2.2 Optimal Path Search Problem of Multi-Objective Network

For $j = 1, 2, \ldots, n$, let $x_j \in 0, 1$ be a binary variable, where $x_j = 1$ if the path includes edge e_j, and $x_j = 0$ if not. Each edge in a path can not be gone through at the same time. So, the following equations hold. For $v \in V$,

$$\sum_{e_j \in E: start(e_j) = v} x_j - \sum_{e_j \in E: end(e_j) = v} x_j = \begin{cases} 1 & \text{if } v = s, \\ -1 & \text{if } v = t, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

where $start(e_j)$ and $end(e_j)$ mean the start and end node of an edge e_j, respectively.

By using above variables, we define n-dimensional vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Any paths from node s to node t can be expressed by vector $\mathbf{x}$, whose components satisfy equation (1). Then, X denotes a set of paths from node s to node t. Next, we define the following function. For $i = 1, 2, \ldots, M$,

$$g_i(\mathbf{x}) = \sum_{e_j \in E} c_{ij} x_j. \tag{2}$$

$g_i(\mathbf{x})$ is i-th total cost to construct path $\mathbf{x}$. Next, we describe the definition of Pareto solutions considered in this study.

Definition of Pareto solutions

We consider two paths $\mathbf{x}, \mathbf{x}' \in X$. $\mathbf{x}$ dominates $\mathbf{x}'$, when $g_i(\mathbf{x}) \leq g_i(\mathbf{x}')$ for $i = 1, 2, \ldots, M$ and at least one strict inequality holds. Then $\mathbf{x}$ is called as Pareto solution, if any other paths do not dominate $\mathbf{x}$.

Therefore, multi-objective optimal path problem in this study can be expressed as follows.

Definition of optimal path problem

For $i = 1, 2, \ldots, M$

$$g_i(\mathbf{x}) = \sum_{e_j \in E} c_{ij} x_j \to \min$$
$$s.t. \mathbf{x} \in X. \tag{3}$$

3 Extended Dijkstra's Algorithm

As mentioned in section 1, the extended Dijkstra's algorithms for a constrained optimal path problem were proposed by Aneja *et al.* [3]. Akiba *et al.* [1, 2] extended Dijkstra's algorithm for Pareto solutions in two and three objective network. Our algorithm is based on this algorithm. This section describes the extended Dijkstra's algorithm. To explain procedure of extended Dijkstra's algorithm, we define following notations. For $v \in V$,

$\mathbf{x}_v$: n-dimensional vector, that means a path from start node s to node v.
W_v: set of adjacent nodes to node v.

And for $v \in V$ and $i = 1, 2, \ldots, M$,

ℓ_{iv}: total cost associated with the i-th objective when a target goes through the path from node s to node v, that is $\ell_{iv} \equiv g_i(\mathbf{x}_v)$.

Next, we define "label" $(v, \ell_{1v}, \ell_{2v}, \ldots, \ell_{Mv})$ as the combination of node v and M total costs $(\ell_{1v}, \ell_{2v}, \ldots, \ell_{Mv})$. Let L_v be a set of labels for node v. Suppose that there are two labels, $(v, \ell_{1v}, \ell_{2v}, \ldots, \ell_{Mv})$ and $(v, \ell'_{1v}, \ell'_{2v}, \ldots, \ell'_{Mv})$. For $i = 1, 2, \ldots, M$, if $\ell_{iv} \geq \ell'_{iv}$ are satisfied and at least one strict inequality holds, then the label $(v, \ell'_{1v}, \ell'_{2v}, \ldots, \ell'_{Mv})$ dominates the label $(v, \ell_{1v}, \ell_{2v}, \ldots, \ell_{Mv})$.

Figure 1 illustrates how to calculate labels of nodes. In this example, $M = 3$. There are two paths from start node s to node 1. One path goes through edge e_1. This path generates the label (1, 15, 10, 30) for node 1. The other path uses edges e_2 and e_5. The label by this path is calculated as (1, 10 + 5, 20 + 15, 10 + 10) = (1, 15, 35, 20). Compared with these two labels, no labels are dominated. Therefore {(1, 15, 10, 30), (1, 15, 35, 20)} is assigned to the set L_1. For node 2, we can obtain two labels, (2, 10, 20, 10) and (2, 25, 20, 45). Compared with these two labels, (2, 25, 20, 45) is dominated by (2, 10, 20, 10). Then L_2 = {(2, 10, 20, 10)}. These labels of node 1 and node 2 generate the labels of terminal node t. In the similar way, L_t = {(t, 25, 30, 35), (t, 20, 40, 20)}. This is the procedure for calculating labels by extended Dijkstra's algorithm. Details of the extended Dijkstra's algorithm are described as follows.

Extended Dijkstra's algorithm

STEP 1 : (Initialize) Set $L_v \leftarrow (s, 0, 0, \ldots, 0), W_v \leftarrow \phi, v \leftarrow s$.
STEP 2 : Obtain the set W_v of adjacent nodes to node v.
STEP 3 : (Obtain the path to adjacent nodes)

STEP 3-1 : Select node $\omega (\in W_v)$.

STEP 3-2 : Select $(v, \ell_{1v}, \ell_{2v}, \ldots, \ell_{Mv}) \in L_v$.

STEP 3-3 : Calculate $(\omega, \ell_{1v} + c_{1j}, \ell_{2v} + c_{2j}, \ldots, \ell_{Mv} + c_{Mj})$ for edge e_j when $start(e_j)$ is v and $end(e_j)$ is ω, and $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega}) \leftarrow (\omega, \ell_{1v} + c_{1j}, \ell_{2v} + c_{2j}, \ldots, \ell_{Mv} + c_{Mj})$.

STEP 3-4 : Compare $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega})$ with all components of L_ω.
If label $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega})$ is Pareto solution, $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega})$ is memorized to a set of L_ω.

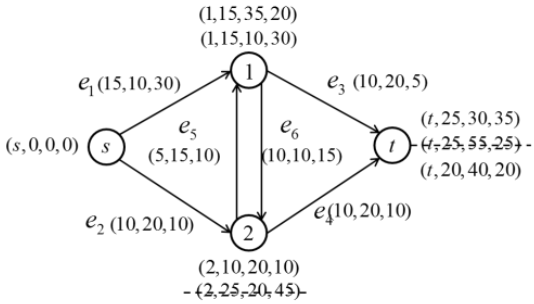


Figure 1: Image of label Calculation

STEP 3-5 : Go to STEP 3-6 if all $(v, \ell_{1v}, \ell_{2v}, \ldots, \ell_{Mv})$ are selected in STEP 3-2. Go to STEP 3-2 otherwise.

STEP 3-6 : Go to STEP 4 if all nodes in W_v (that is, adjacent nodes to v) are selected in STEP 3-1. Go to STEP 3-1 otherwise.

STEP 4 : (Obtain the path from known path to adjacent nodes and calculate the labels)

STEP 4-1 : Select node $\omega \in W_v$.

STEP 4-2 : Select label $(\sigma, \ell_{1\sigma}, \ell_{2\sigma}, \ldots, \ell_{M\sigma}) \in L_\sigma$ for node σ which is one of all nodes connecting with node ω.

STEP 4-3 : Calculate $(\omega, \ell_{1\sigma}+c_{1j}, \ell_{2\sigma}+c_{2j}, \ldots, \ell_{M\sigma}+c_{Mj})$ for edge e_j when $start(e_j)$ is σ and $end(e_j)$ is ω, and set $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega}) \leftarrow (\omega, \ell_{1\sigma}+c_{1j}, \ell_{2\sigma}+c_{2j}, \ldots, \ell_{M\sigma}+c_{Mj})$.

STEP 4-4 : Compare $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega})$ with all elements of L_ω.
If label $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega})$ is Pareto solution, $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega})$ is memorized to a set of L_ω.

STEP 4-5 : If all $(\sigma, \ell_{1\sigma}, \ell_{2\sigma}, \ldots, \ell_{M\sigma})$ are selected in STEP 4-2, go to STEP 4-6. Else go to STEP 4-2.

STEP 4-6 : If all nodes in W_v are selected in STEP 4-1, go to STEP 5. Else go to STEP4-1.

STEP 5 : (Select next node or output Pareto solutions)

STEP 5-1 : If $v = t$ and all nodes $v(\in V)$ are selected in STEP 2, go to STEP 5-2. Else, go to STEP 2 after selecting $v'(\in W_v)$ and setting $v \leftarrow v'$.

STEP 5-2 : Output all Pareto solutions L_v and the algorithm finishes.

Next, we show the procedure to obtain Pareto solutions, which is used at STEP 3-4 and STEP 4-4 in the above algorithm.

Searching procedure for Pareto solutions

STEP 1 : Receive L_ω and $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega})$.
STEP 2 : Select $(\omega, \ell'_{1\omega}, \ell'_{2\omega}, \ldots, \ell'_{M\omega}) \in L_\omega$.
STEP 3 : If all of inequalities $\ell^*_{1\omega} \geq \ell'_{1\omega}, \ell^*_{2\omega} \geq \ell'_{2\omega}, \ldots, \ell^*_{M\omega} \geq \ell'_{M\omega}$ are satisfied, go to STEP 5. Else, $L_\omega \leftarrow L_\omega \cup \{(\omega, \ell^*_1, \ell^*_2, \ldots, \ell^*_{M\omega})\}$ and go to STEP 4.
STEP 4 : If all of inequalities $\ell^*_{1\omega} \leq \ell'_{1\omega}, \ell^*_{2\omega} \leq \ell'_{2\omega}, \ldots, \ell^*_{M\omega} \leq \ell'_{M\omega}$ are satisfied, $L_\omega \leftarrow \{L_\omega \setminus (\omega, \ell'_{1\omega}, \ell'_{2\omega}, \ldots, \ell'_{M\omega})\}$.
STEP 5 : If all $(\omega, \ell'_{1\omega}, \ell'_{2\omega}, \ldots, \ell'_{M\omega})$ are selected in STEP 2, return L_ω. Else go to STEP 2.

4 Proposed Algorithm

As mentioned in section 1, extended Dijkstra's algorithm needs large memory area and much computing time to obtain Pareto solutions. Akiba *et al.* [2] improved extended Dijkstra's algorithm in order to reduce search space. However, this improved algorithm by Akiba *et al.* could not obtain

all Pareto solution. In this paper, based on the idea of standard path reduction by Takahashi *et al.* [8], we extend their idea to more than three-objective networks. In addition, our algorithm can obtain all Pareto solutions.

This section explains the property which is used to reduce search space. If some path $\mathbf{x}$ from start node to terminal node is obtained, then path $\mathbf{x}$ has the following property.

Property 1. The Space does not include Pareto solution, which satisfies all inequalities $g_i \geq g_i(\mathbf{x})$.

Property 1 is derived from the definition of Pareto solutions. Our study uses this property and obtains some path to reduce search space.The path is defined as standard path in our study. We consider standard path in the following section.

4.1 Standard Path Reduction

To obtain standard path, we consider the following one-objective optimal problem with a constraint condition. For $i = 1, 2, \ldots, M$,

$$\begin{aligned} &\sum_{i=1}^{M} w_i g_i(\mathbf{x}) \to min \\ s.t. \mathbf{x} \in X, &\sum_{i=1}^{M} w_i = 1, w_i > 0. \end{aligned} \tag{4}$$

For the path which satisfies this problem, following property is discussed.

Property 2. Let $\mathbf{px}$ be a path that minimizes expression (4) for $\mathbf{x} \in X$, $i = 1, 2, \ldots, M$. That is, $\mathbf{px} \equiv \arg\min_{\mathbf{x}\in X} \sum_{i=1}^{M} w_i g_i(\mathbf{x})$. Then $\mathbf{px}$ is Pareto solution.

To reduce larger search space, we propose the algorithm which uses several standard paths. We vary coefficients w_i of expression (4) and generate several standard paths. For $k = 0, 1, 2, \ldots, K$, let $h(k)$ be the weight variable, where $h(k) = 1$ if $k = 0$, and $h(k) > 1$ if $k \neq 0$. We define standard paths $\mathbf{px}_{k,u}$ generated by expression (4) as follows.

Property 3. For k, u, let $\mathbf{px}_{k,u}$ be the standard path $\mathbf{px}_{k,u} \equiv \arg\min_{\mathbf{x}\in X} \sum_{i=1}^{M} w_i g_i(\mathbf{x})$,

where

$$\mathbf{px}_{k,u} : w_u = \frac{h(k)}{h(k) + M - 1}, w_i = \frac{1}{h(k) + M - 1} (i = 1, 2, \ldots, M, i \neq u). \tag{5}$$

Figure 2 illustrates the search space of standard paths reduction in three-dimensional space with g_1, g_2 and g_3 as axes. For $i = 1, 2, 3$, g_i corresponds to $g_i(\mathbf{x})$. The standard paths $\mathbf{px}_{k,u}$ are intersections of solutions and the planes which differ in the coefficient of Property 3. In addition, these standard paths are Pareto Solutions. For each of standard paths $\mathbf{px}_{k,u}$, there is no Pareto solution in the space which satisfies all inequalities $g_i \geq g_i(\mathbf{px}_{k,u})(i = 1, 2, 3)$. Then this space is excluded from search space, and path search becomes efficient.

For each $k(= 1, 2, \ldots, K)$, M standard paths are derived from expression (5). When $k = 0$ (that is, $h(k) = 1$), the only one standard path $\mathbf{px}$ is obtained because $w_u = w_i$. We propose the algorithm that $KM + 1$ standard paths, $\mathbf{px}_{k,u}$, reduce search space. Each standard path is obtained by Dijkstra's algorithm before conducting the extended Dijkstra's algorithm. Then these standard paths reduce search space of the extended Dijkstra's algorithm.

To obtain above standard paths, our proposed algorithm adds the following steps which reduce search space for Pareto solutions.

Proposed Algorithm
(Add the STEP 0 to the extended Dijkstra's algorithm)

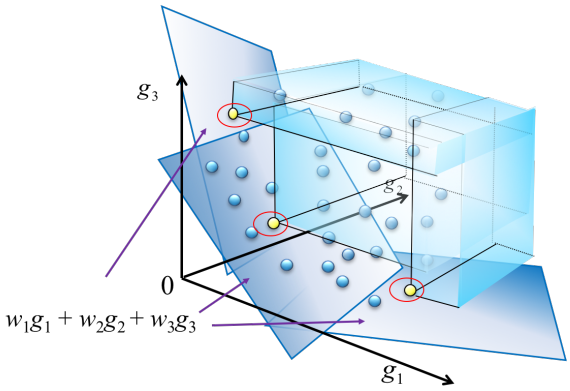


Figure 2: Image of reduced space by standard paths

STEP 0 :(Obtain the optimal path for $\mathbf{px}$ and $\mathbf{px}_{k,u}$)

STEP 0-1 : Obtain $\mathbf{px}$ by using the Dijkstra's algorithm.

STEP 0-2 : Obtain vector $(g_1(\mathbf{px}), g_2(\mathbf{px}), \ldots, g_M(\mathbf{px}))$ by the path $\mathbf{px}$ from STEP 0-1.

STEP 0-3 : $k \leftarrow 1$.

STEP 0-4 : $u \leftarrow 1$.

STEP 0-5 : Obtain $\mathbf{px}_{k,u}$ by using the Dijkstra's algorithm.

STEP 0-6 : Obtain vector $(g_1(\mathbf{px}_{k,u}), g_2(\mathbf{px}_{k,u}), \ldots, g_M(\mathbf{px}_{k,u}))$ by the path $\mathbf{px}_{k,u}$ from STEP 0-5.

STEP 0-7 : If $u = M$, go to STEP 0-8. Else, $u \leftarrow u + 1$ and go to STEP 0-5.

STEP 0-8 : If $k = K$, go to STEP 1. Else, $k \leftarrow k + 1$ and go to STEP 0-4.

(Substitute following STEPs for STEP 3-3 and STEP 4-3 of extended Dijkstra's algorithm)

STEP 3-3 : Calculate $(\omega, \ell_{1v}+c_{1j}, \ell_{2v}+c_{2j}, \ldots, \ell_{Mv}+c_{Mj})$ for edge e_j when $start(e_j)$ is v and $end(e_j)$ is ω, and set $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega}) \leftarrow (\omega, \ell_{1v}+c_{1j}, \ell_{2v}+c_{2j}, \ldots, \ell_{Mv}+c_{Mj})$.
If one of following $M+1$ conditions is satisfied, go to STEP 3-5;
(a) $\ell^*_{1\omega} \geq g_1(\mathbf{px}), \ell^*_{2\omega} \geq g_2(\mathbf{px}), \ldots, \ell^*_{M\omega} \geq g_M(\mathbf{px})$,
(b)For one of u, $\ell^*_{1\omega} \geq g_1(\mathbf{px}_{k,u}), \ell^*_{2\omega} \geq g_2(\mathbf{px}_{k,u}), \ldots, \ell^*_{M\omega} \geq g_M(\mathbf{px}_{k,u})$.
Else go to STEP 3-4.

STEP 4-3 : Calculate $(\omega, \ell_{1\sigma}+c_{1j}, \ell_{2\sigma}+c_{2j}, \ldots, \ell_{M\sigma}+c_{Mj})$ for edge e_j when $start(e_j)$ is σ and $end(e_j)$ is ω, and set $(\omega, \ell^*_{1\omega}, \ell^*_{2\omega}, \ldots, \ell^*_{M\omega}) \leftarrow (\omega, \ell_{1\sigma}+c_{1j}, \ell_{2\sigma}+c_{2j}, \ldots, \ell_{M\sigma}+c_{Mj})$. If one of following $M+1$ conditions is satisfied, go to STEP 4-5;
(a) $\ell^*_{1\omega} \geq g_1(\mathbf{px}), \ell^*_{2\omega} \geq g_2(\mathbf{px}), \ldots, \ell^*_{M\omega} \geq g_M(\mathbf{px})$,
(b) For one of u, $\ell^*_{1\omega} \geq g_1(\mathbf{px}_{u,k}), \ell^*_{2\omega} \geq g_2(\mathbf{px}_{k,u}), \ldots, \ell^*_{M\omega} \geq g_M(\mathbf{px}_{k,u})$.
Else go to STEP 4-4.

5 Numerical Experiments for Evaluation

Our proposed algorithm can reduce search space than the extended Dijkstra's algorithm. In case of more standard paths, reduced space more expands. But, standard path reduction needs the additional procedures to obtain the standard paths which minimize $\sum_{i=1}^{M} w_i g_i(\mathbf{x})$. So, we evaluate some algorithms which are different in the number of standard paths.

Numerical experiments were conducted by PC with Intel Core i7 CPU (2.4GHz), Memory 16.0 GB and MS-Windows 8.1 Update 64 bit. Used program was written in C programming language and compiled by Visual Studio 2010. We conducted numerical experiments in case of $M = 3$ and

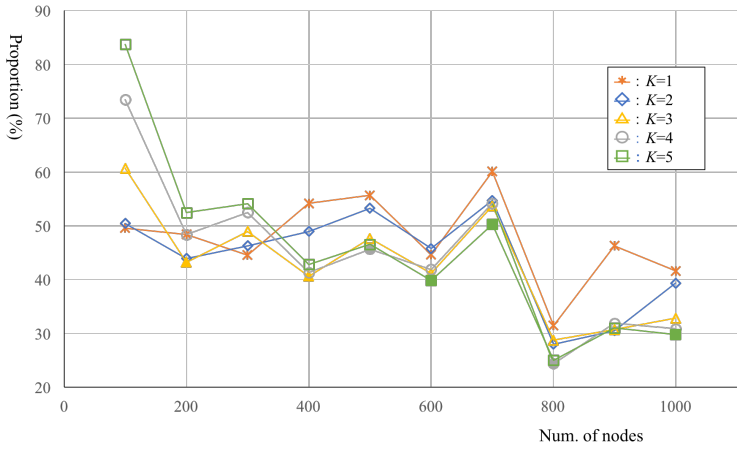


Figure 3: Comparison of computing time for $M = 3$

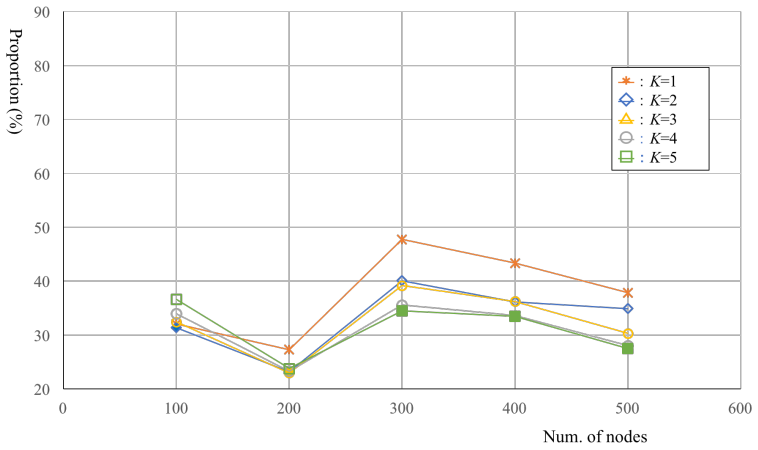


Figure 4: Comparison of computing time for $M = 4$

$M = 4$. For $j = 1, 2, \ldots, n$, components of cost vector $(c_{1j}, c_{2j}, c_{3j}(, c_{4j}))$ were generated by uniform random values from 1 to 100, from 1 to 1000, and from 1 to 10000. For each $k(= 0, 1, 2, \ldots, K)$, $h(k) = k \cdot 10 + 1$ ($h(k)$ is the parameter of expression(5)). Each experiment was conducted five times for different networks in cost vectors. The averages of these experiments are illustrated in following figures.

Following two figures show five cases which differ in the number of standard paths. K determines the number of $h(k)$. Figure 3 shows comparisons of computing times in $M = 3$. In this case, the number of nodes were from 100 to 1000. "Proportion" means the ratio of each algorithm's time to extended Dijkstra's algorithm's time. Figure 4 is the results of computing time in case that networks have from 100 to 500 nodes and $M = 4$. Compared with extended Dijkstra's algorithm, computing times of proposed algorithms are much reduced. Numerical experiments suggest using more standard paths is more efficient in cases that the number of nodes and objective functions are large. But the difference between $K = 4$ and $K = 5$ is small.

6 Conclusion

In this study, we considered the optimal path search problem of multi-objective network. Our study focused on the effective standard path reduction. Using these paths, we proposed the efficient algorithm for obtaining Pareto solutions. As the evaluation in $M = 3$ and $M = 4$, our proposed

algorithm was compared in terms of computing time. Our numerical experiments suggested that the number of standard path influenced effects of standard path reduction. Using more standard paths was more efficient in cases that the number of nodes and objective functions were large. However the difference between $K = 4$ and $K = 5$ tended to be small.

In this analysis, effective number of standard path is different depending on the number of nodes and the number of objective functions. In addition, suppose that weight variables $h(k)$ take different values, efficiency of standard path is probably influenced. As the future works, we need to consider effective relation among the number of standard paths, the number of nodes, and weight variables.

References

[1] T. Akiba, H. Yamamoto, T. Egashira and S. Iwakami. Fast Algorithm for the Pareto Solutions of a Two-objective Network. *Journal of the Society of Plant Engineers Japan* 24(2) 49-59, 2012.

[2] T. Akiba T, H. Yamamoto, D.Q. Rui and H. Nagatsuka. A Fast Calculation Method for the Partial Group of All Pareto Solutions at a Three-objective Network. *International Journal of Quality Technology and Quantitative Management* 9(3) 305-316, 2012.

[3] Y.P. Aneja, V. Aggarwal and K. P. K. Nair. Shortest chain subject to side constrains, *Networks* 13 295-302, 1983.

[4] R. Bellman. On a Routing Problem. *Quarterly of Applied Mathematics* 16(1), 87-90, 1958.

[5] E.W. Dijkstra. A note on two problems in connection with graphs. *Numerische Mathematik* 1, 269-271, 1959.

[6] L. R. Ford and D. R. Fulkerson. Flows in Networks. Princeton University Press, 1962.

[7] Y. Kambayashi, H. Yamachi, Y. Tsujimura and H. Yamamoto. Integrating Uncomfortable Intersection-Turns to Subjectively Optimal Route Selection Using Genetic Algorithm. *Proc. 5th IEEE Int. Conf. on Computational Cybernetics*, 203-208, 2007.

[8] N. Takahashi, H. Yamamoto, T. Akiba and X. Xiao. Algorithm for Obtaining Optimal Path in Three-objective Network with a Reduction of Search Space. *Innovation and Supply Chain Management* 9(3) 103-109, 2015.

Separable Hedonic Games and Contractually Stabilities

Masaya TANAKA[1] and Shao-Chin SUNG[2]

Aoyama Gakuin University,
5-10-1 Fuchinobe, Chuo, Sagamihara, Kanagawa,
252-5258, Japan
c5617157@aoyama.jp[1], son@ise.aoyama.ac.jp[2]

Abstract

We are concerned with the problem of constructing stable partitions for hedonic games with respect from *contractually stability concepts*, including contractually individual (CI), contractually core (CC), contractually strict core (CSC). In general, every instance of hedonic games has at least one stable partition when one of the contractually stability concepts is under consideration. For additive instances, it is known that a CI stable partition can be constructed in quadratic time in term of the number of players. In this paper, we consider a wider domain, namely *separable instances*. A instance of hedonic games is separable when any player classifies any other players as friends, enemies or others and the classification characterize her or his preference. However, it is not clear whether such a stable partition can be constructed efficiently with the characteristic of separable instance. We propose two quadratic time algorithms which respectively construct a CI stable partition and a CC stable partition for every separable instance. Moreover, we show that CSC stable partition can be obtained by one of the proposed algorithms when a subdomain of separable instance is under consideration.

Keywords: Hedonic Game, Separability, Contractually Stabilities.

1 Introduction

In situations of coalition formation, each player may only care about coalitions that she or he may belong. Such *hedonic aspect* of coalition formation games is pointed out by Dréze and Greenberg [6]. As a formal model of coalition formation with hedonic aspect, *hedonic coalition formation games* or simply *hedonic games* was introduced by Banerjee *et al.*[3]. Main study related to hedonic games includes (i) stability test of outcomes, (ii) existence of stable outcomes, (iii) construction of stable outcomes (e.g., see [4, 5, 8]).

In this paper, we are concerned with the problem of constructing *contractually individual (CI) stable*, *contractually core (CC) stable* or *contractually strict core (CSC) stable* partition for hedonic games. For the existence of such partitions, Ballester [2] pointed out that CI stable partitions exist, Sung and Dimitrov [8] pointed out that CSC stable partitions exist for all instances of hedonic games. In general, the size of each instance is exponential of the number of players, and in a straightforward way, a CI stable, CC stable or CSC stable partition can be constructed in polynomial time with respect to the size of instances.

Aziz *et al.* [1] claimed that a CI stable partition of each additive instance can be constructed in polynomial time in terms of the number of players; however their algorithm fails to produce CI stable partition for some additive instance. Recently, their claim is proved in a sophisticated way by Iwasaki and Sung [7], and Tanaka and Sung [9] proposed a more efficient algorithm to construct a CI stable partition with a simpler proof of correctness.

2 Preliminaries

2.1 Hedonic Games

Let $N = \{1, 2, ..., n\}$ be the set of n players. A *coalition* is a nonempty subset of N. For each player $i \in N$, we denote by $\mathcal{N}_i = \{X \subseteq N \mid i \in X\}$ the collection of all coalitions containing i. Each player $i \in N$ is endowed with a *preference* $\succeq_i$ (i.e., a complete transitive binary relation) over $\mathcal{N}_i$. An instance $\langle N, \succeq \rangle$ of *hedonic games* is a pair of the set $N = \{1, 2, ..., n\}$ of players and a profile $\succeq = (\succeq_1, \succeq_2, ..., \succeq_n)$ of player's preferences. In general, each player's preference can be described by a ranking of $\mathcal{O}(2^n)$ coalitions, and hence, the size of a preference profile is $\mathcal{O}(n2^n)$.

An *outcome* Π of $\langle N, \succeq \rangle$ is a partition of N, i.e., a collection of coalitions such that each player $i \in N$ appears in exactly one of its coalitions. For each partition Π of N, we denote by $\Pi(i)$ the coalition in Π containing player $i \in N$, i.e., $\{\Pi(i)\} = \Pi \cap \mathcal{N}_i$.

2.2 Separable Preferences

Let us introduce the domain of *separable preferences*. For each player $i \in N$, let

- $G_i = \{j \in N \setminus \{i\} \mid \{i, j\} \succ_i \{i\}\}$,
- $U_i = \{j \in N \setminus \{i\} \mid \{i, j\} \sim_i \{i\}\}$, and
- $B_i = \{j \in N \setminus \{i\} \mid \{i\} \succ_i \{i, j\}\}$.

Observe that (G_i, U_i, B_i) is a partition of $N \setminus \{i\}$. Then, a preference profile $\succeq = (\succeq_1, \succeq_2, ..., \succeq_n)$ is called *separable* if, for every $i, j \in N$ such that $i \neq j$ and for each $X \in \mathcal{N}^i \cap \mathcal{N}^j$,

- $j \in G_i$ if and only if $X \succ_i X \setminus \{j\}$,
- $j \in U_i$ if and only if $X \sim_i X \setminus \{j\}$, and
- $j \in B_i$ if and only if $X \setminus \{j\} \succ_i X$.

We notice that, for each $i \in N$, G_i, U_i and B_i characterize i'separable preference.

Let $\succeq_i$ be a separable preference, and let X and Y be coalition in $\mathcal{N}_i$. A sufficient condition of $X \succeq_i Y$ is $(Y \setminus X) \cap G_i = \emptyset$ and $(X \setminus Y) \cap B_i = \emptyset$. A sufficient condition of $X \succ_i Y$ is $X \neq Y$ and $Y \setminus X \subseteq B_i$ and $X \setminus Y \subseteq G_i$.

A *separable instance* of hedonic games is the one with an separable preference profile. The size of an separable instance is $\mathcal{O}(n2^n)$.

For some subdomain of preferences, each player's preference can be described in a more compact way. One of such subdomains is the domain of additive preferences. A preference profile $\succeq$ is *additive* if, for each $i \in N$, there exists a function $v_i : \mathcal{N}_i \to \mathbf{R}$ s.t. for all $X, Y \in \mathcal{N}_i$,

- $X \succeq_i Y$ if and only if $\sum_{j \in X} v_i(j) \geq \sum_{j \in Y} v_i(j)$.

An *additive instance* of hedonic games is the one with an additive preference profile. As in all previous results, we assume that each additive instance of hedonic games is described by n^2 real values $\{v_i(j)\}_{i,j \in N}$, and hence, the size of an additive instance is $\mathcal{O}(n^2)$.

2.3 Contractually Stabilities

Let us introduce contractually stabilities. Contractually stabilities are based on deviations which needs all player's agreement.

A partition Π is *contractually individual (CI) stable* with respect to $\langle N, \succeq \rangle$ if, for each $i \in N$ and each $X \in \Pi \cup \{\emptyset\}$, at least one of the following conditions holds.

- $\Pi(i) \succeq_i X \cup \{i\}$.
- $\Pi(i) \succ_j \Pi(i) \setminus \{i\}$ for some $j \in \Pi(i) \setminus \{i\}$.
- $X \succ_j X \cup \{i\}$ for some $j \in X$.

The first condition means player i has no incentive to leave $\Pi(i)$ and to join X.The second condition means some player $j \in \Pi(i)$ disagrees player i to leave $\Pi(i)$. The last condition means some player $j \in X$ disagrees player i to join X.

A partition Π is *contractually core (CC) stable* with respect to $\langle N, \succeq \rangle$ if, for any coalition $X \subseteq N$, at least one of the following conditions holds.

- $\Pi(i) \succeq_i X$ for some $i \in X$.
- $\Pi(i) \succ_i \Pi(i) \setminus X$ for some $i \in N \setminus X$.

The first condition means player i has no incentive to leave $\Pi(i)$ and to form X. The second condition means some player i disagrees the players of X to leave $\Pi(i)$.

A partition Π is *contractually strict core (CSC) stable* with respect to $\langle N, \succeq \rangle$ if, for any coalition $X \subseteq N$, at least one of the following conditions holds.

- $\Pi(i) \succ_i X$ for some $i \in X$.
- $\Pi(i) \succeq_i X$ for each $i \in X$.
- $\Pi(i) \succ_i \Pi(i) \setminus X$ for some $i \in N \setminus X$.

The first condition means player i disagrees to leave $\Pi(i)$ and to form X. The second condition means any player i in X has no incentive to leave $\Pi(i)$ and to form X. The third condition means some player i disagrees the players of X to leave $\Pi(i)$.

3 Existence of Contractually Stable Partitions

For any instance of hedonic games, contractually stable partition always exists.

Proposition 1. *A contractually stable partition always exists.* □

For the existence of contractually strict core (CSC) stable partitions, Sung and Dimitrov [8] pointed out that the answer is "*yes*" for all instances of hedonic games. Contractually strict core stability implies contractually core (CC) stability and contractually individual (CI) stability. Hence, there always exists a CC stable partition and a CI stable partition.

A CSC stable partition can be constructed by the following algorithm.

INPUT: An instance $\langle N, \succeq \rangle$ of hedonic games.

OUTPUT: A partition of N.

Step 1. Set Π be an arbitrary partition of N.

Step 2. Repeats the following until there exists no CSC deviation from Π.

Step 2-1. Find a CSC deviation X.

Step 2-2. Set $\Pi := \{Y \setminus X \mid Y \in \Pi \wedge Y \not\subseteq X\} \cup \{X\}$.

Step 3. Return Π.

Observe that, by implementing of a CSC deviation X, no one worse off and there exists some player in X such that strictly better off. Hence, above algorithm will be terminated in a finite number of steps (at most $n2^{n-1} - n$ implementations). When the algorithm does terminate, it returns a partition from which no CSC deviation exists, which means such a partition is CSC stable.

4 Efficient Construction of CI Stable Partitions

It is known that a contractually individual (CI) stable partition can be constructed for any additive instances in polynomial time [9].

Proposition 2. *A CI stable partition of an additive instance can be constructed in $\mathcal{O}(n^2)$ time.*

The following algorithm which can construct a CI stable partition for any additive instances in $\mathcal{O}(n^2)$ time.

Let $\langle N, \succeq \rangle$ be an additive instance of hedonic games. Consider the following directed graph $G_{\langle N,\succeq\rangle} = (V, E)$ induced by $\langle N, \succeq \rangle$.

- $V = N$ and $E = \{(i, j) \in V \times V \mid v_i(j) > 0\}$.

Let Γ be a collection of pairwise disjoint coalitions. By $U(\Gamma) = \bigcup_{X\in\Gamma} X$ we denote the union of all coalitions in Γ, and by $R(\Gamma) = N \setminus U(\Gamma)$ we denote the complement of $U(\Gamma)$. For each $i \in N$, let

- $\mathcal{A}_\Gamma(i) = \{X \in \Gamma \mid v_j(i) \geq 0 \text{ for each } j \in X\} \cup \{\emptyset\}$.

Input: An additive instance $\langle N, \succeq \rangle$ of hedonic games.

Output: A collection Γ of pairwise disjoint coalitions.

Step 1. Find the set S of players each of which is accessible (in $G_{\langle N,\succeq\rangle}$) from some player $j \in N$ whose strong connected component is not singleton.

Step 2. Set $\Gamma := \{S\}$ if S is nonempty, and otherwise set $\Gamma := \emptyset$.

Step 3. Repeat the following until $R(\Gamma) = \emptyset$.

Step 3-1. Select a player $i \in R(\Gamma)$ satisfying $v_i(j) \leq 0$ for each $j \in R(\Gamma)$.

Step 3-2. Find the best coalition $X \in \mathcal{A}_\Gamma(i)$ for player i, i.e., $X \succeq_i Y$ for each $Y \in \mathcal{A}_\Gamma(i)$.

Step 3-3. Update Γ by removing X from Γ (if $X \in \Gamma$), and by inserting $X \cup \{i\}$ into Γ, i.e., $\Gamma := (\Gamma \setminus \{X\}) \cup \{X \cup \{i\}\}$.

Step 4. Return Γ.

Now, we consider construction for a separable instance. Before we discuss how to construct, we define the problem of constructing a CI stable partition.

Construction of a CI stable partition

Instance: Separable hedonic game $\langle N, \succeq \rangle$, G_i, U_i, B_i for each $i \in N$.

Find: Construct a CI partition with respect to $\langle N, \succeq \rangle$.

By making little changes, the algorithm can be applied for any separable instances. Let $\langle N, \succeq \rangle$ be a separable instance. Firstly, we again define the directed graph $G_{\langle N,\succeq\rangle} = (V, E)$ induced by $\langle N, \succeq \rangle$.

- $V = N$ and $E = \{(i, j) \in V \times V \mid j \in G_i\}$.

Secondly, for each $i \in N$, let

- $\mathcal{A}_\Gamma(i) = \{X \in \Gamma \mid i \in G_j \cup U_j \text{ for each } j \in X\} \cup \{\emptyset\}$.

Lastly, we make a little change of Step 3-1. We select a player $i \in R(\Gamma)$ in Step 3-1 such that, for each $j \in R(\Gamma) \setminus \{i\}$, $j \notin G_i$. The changes about the algorithm is related to player's preference.

Before we consider the running time of the algorithm with above changes when the input is a separable instance. Now, we make a assumption for a preference. It is that we can directly query about the preference for any player. Why can we assume that? By the size of a preference, we may not describe all of the preference for any player. Hence, we define the oracle for query about a player's preference to reflect the assumption. Before the definition of oracle, we define query about a player's preference as a decision problem.

Query about player's preference

Instance: Hedonic game $\langle N, \succeq \rangle$, player $i \in N$, coalition $X, Y \in \mathcal{N}_i$.

Question: Does player i weakly prefer X than Y ?

An *oracle* for the problem, query about player's preference, is an abstract device which, for any instance, returns 'yes' or 'no'. It is assumed that the oracle may return the answer in just one step.

Observation 1. *For any separable instance $\langle N, \succeq \rangle$, $G_{\langle N,\succeq\rangle}$ is computed in $\mathcal{O}(n^2)$ time.* □

We don't make a change of Step 1. Hence, the running time of Step 1 is the same as before making the changes.

Proposition 3. *The set S defined in Step 1 can computed in $\mathcal{O}(n^2)$ time.* □

Lemma 1. *The total running time of the algorithm with the changes is $\mathcal{O}(n^2)$.* □

Proof. As shown in Proposition 3, Step 1 can be implemented with running time $\mathcal{O}(n^2)$. Obviously, Step 2 and Step 4 can be done with $\mathcal{O}(n)$ time.

Now we consider the running time of Step 3. First of all, Step 3 repeats $|N \setminus S| < n$ times. By maintaining in-degrees and out-degrees of all vertices in $G_{\langle R(\Gamma),\succeq\rangle}$ in such a way that all in-degrees and out-degrees are updated in $\mathcal{O}(n)$ time whenever a player is excluded from $R(\Gamma)$, Step 3-1 can be implemented with running time $\mathcal{O}(n)$. By the definition of the oracle, Step 3-2 can be implemented with running time $\mathcal{O}(n)$. In a straightforward way, Step 3-3 can be implemented with running time $\mathcal{O}(n)$ as well. Therefore, the total running time is $\mathcal{O}(n^2)$. □

Theorem 1. *A CI stable partition of a separable instance can be constructed in $\mathcal{O}(n^2)$ time.* □

Proof. Let Π be the outcome of the algorithm with the changes. By the definition of the set S, for each $i \in S$, there exists $j \in S$ such that $i \neq j$ and $i \in G_j$. Therefore, by the definition of separable preference, for each $i \in S$, there exists $j \in \Pi(i) \setminus \{i\}$ such that $\Pi(i) \succ_j \Pi(i) \setminus \{i\}$.

Let $i \in N \setminus S$, and let X be the best coalition in $\mathcal{A}_\Gamma(i)$ for player i found in Step 3-2. From the construction, $\Pi(i) \sim_i X$ and $\Pi(i)$ is the best coalition in $\mathcal{A}_\Pi(i) \cup \{\Pi(i)\}$. Hence, either player i has no incentive to leave $\Pi(i)$ or player i is not allowed to join any coalition Y in $\Pi \setminus \{\Pi(i)\}$ (i.e., $Y \notin \mathcal{A}_\Pi(i)$) satisfying $Y \cup \{i\} \succ_i \Pi(i)$. Therefore, Π is a CIS partition. □

We can construct for any separable instance defined in Section 2.2. Let us introduce the domain of *weakly separable preferences*. A preference profile $\succeq = (\succeq_1, \succeq_2, ..., \succeq_n)$ is *weakly separable* if, for every $i, j \in N$ such that $i \neq j$ and for each $X \in \mathcal{N}^i \setminus \mathcal{N}^j$,

- $\{i, j\} \succeq_i \{i\}$ if and only if $X \cup \{j\} \succeq_i X$ and
- $\{i\} \succeq_i \{i, j\}$ if and only if $X \succeq_i X \cup \{j\}$.

The output of the proposed algorithm is not always CI stable if the input is a weakly separable instance. The proposed algorithm is correct whenever it is applied for our separable instances.

5 Efficient Construction of CC Stable Partitions

The algorithm of Section 3 can construct a contractually core (CC) stable partition for any instance of hedonic games. However, the algorithm is not efficient. In the following, we propose the efficient algorithm to construct a CC stable partition for any separable instances. Now, we define the problem of constructing a CC stable partition.

Construction of a CC stable partition

Instance: Separable hedonic game $\langle N, \succeq \rangle$, G_i, U_i, B_i for each $i \in N$.

Find: Construct a CC partition with respect to $\langle N, \succeq \rangle$.

In the following, we show that a CC stable partition of each separable instance can be constructed in $\mathcal{O}(n^2)$ time. The running time is linear with respect to the size n^2 of an additive instance. Hence, the construction algorithm is efficiently.

Let $\langle N, \succeq \rangle$ be a separable instance of hedonic games. Let Γ be a collection of pairwise disjoint coalitions. Let k be a natural number which means how many times Step 2 repeated. Let $leader(k)$ be a player which be selected k times repeats in Step 2. By $R(\Gamma) = N \setminus \bigcup_{X \in \Gamma} X$, we denote the set of players which be not included any coalition of Π.

Input: An instance of Construction of a CC stable partition.

Output: A collection Γ of pairwise disjoint coalitions.

Step 1. Set $\Gamma := \emptyset$, $k := 1$.

Step 2. Repeats the following until $R(\Gamma) = \emptyset$.

Step 2-1. Select a player $i \in R(\Gamma)$ and set $leader(k) := i$.

Step 2-2. Let $X := \{i\} \cup (G_i \cap R(\Gamma))$.

Step 2-3. Let $\Gamma := \Gamma \cup \{X\}$, $k := k + 1$.

Step 3. Return Γ.

Let Π be the outcome of the proposed algorithm. The following observations follows from the above algorithm.

Observation 2. *The outcome* Π *is a partition of* N. □

Observation 3. *From the construction, for each* $k \in [1, |\Pi|]$ *and any coalition* $X \subseteq \Pi(leader(k)) \setminus \{leader(k)\}$, $X \subseteq G_{leader(k)}$. □

Let $\mathcal{A}$ and $\mathcal{B}$ be a collection of coalitions which is defined by the following.

- $\mathcal{A} = \{X \in 2^N \setminus (\Pi \cup \{\emptyset\}) \mid \exists k \in [1, |\Pi|] \Big[\Pi(leader(k)) \cap X \neq \emptyset \text{ and } leader(k) \notin X\Big]\}$.
- $\mathcal{B} = 2^N \setminus (\Pi \cup \{\emptyset\} \cup \mathcal{A})$.

Clearly, $(\mathcal{A}, \mathcal{B})$ is a partition of $2^N \setminus (\Pi \cup \{\emptyset\})$.

Observation 4. *For each coalition* $X \in \mathcal{B}$, *there exists* $leader(k) \in X$ *such that* $k \in [1, |\Pi|]$. □

Lemma 2. *The running time of the proposed algorithm is* $\mathcal{O}(n^2)$ *time.*

Proof. In a straightforward way, Step 1 and Step 3 of the algorithm is $\mathcal{O}(1)$ time.

Now we consider the running time of Step 2. Step 2 repeats $|N|$ times. In a straightforward way, Step 2-1 can be implemented with running time $\mathcal{O}(1)$. Step 2-2 and Step 2-3 can be implemented with running time $\mathcal{O}(n)$. Therefore, the total running time is $\mathcal{O}(n^2)$ time. □

Theorem 2. *A CC stable partition of a separable instance can be constructed in* $\mathcal{O}(n^2)$ *time.* □

Proof. We show that each coalition is not CC deviation from Π for $\langle N, \succeq \rangle$. $(\Pi, \mathcal{A}, \mathcal{B})$ is a partition of all coalitions $(2^N \setminus \{\emptyset\})$. Unfortunately, each coalition in Π is not deviation. Now, we consider coalitions in $\mathcal{A}$ or $\mathcal{B}$.

Firstly, we consider the coalitions in $\mathcal{A}$. By the definition of $\mathcal{A}$, for each coalition $X \in \mathcal{A}$, there exists player $leader(k) \in N \setminus X$ such that $k \in [1, |\Pi|]$ and $\Pi(leader(k)) \cap X \neq \emptyset$. Therefore, by Observation 3 and the definition of separable preference, $\Pi(leader(k)) \succ_{leader(k)} \Pi(leader(k)) \setminus X$, thus $leader(k)$ disagrees deviation X. Hence, X is not a CC deviation.

Lastly, we consider the coalitions in $\mathcal{B}$. By Observation 4, for each coalition $X \in \mathcal{B}$, there exists $leader(k) \in X$ such that $k \in [1, |\Pi|]$ and $leader(\ell) \notin X$ for each $\ell \in [1, k)$. From the construction, for each $\ell \in (k, |\Pi|]$, $G_{leader(k)} \cap \Pi(leader(\ell)) = \emptyset$. Therefore, by the definition of separable preference, $\Pi(leader(k)) \succeq_{leader(k)} X$, thus $leader(k)$ has no incentive to form coalition X. Hence, X is not a CC deviation. Therefore, we can now conclude that Π is a CC stable partition. □

The problem of construction of a CC stable partition has G_i, U_i and B_i for each $i \in N$ as the instance. But, we don't need U_i and B_i for each $i \in N$ on the algorithm. Hence, we can construct a CC stable partition for any separable instances if the instance of the problem is $\langle N, \succeq \rangle$ and G_i for each $i \in N$. Also, the oracle for query about player's preference is not necessary on the algorithm.

The output of the proposed algorithm is not always CC stable if the input is a weakly separable instance. The proposed algorithm are correct whenever this is applied for separable instances.

6 Efficient Construction of CSC Stable Partitions

Before we precede to our result on contractually strict core (CSC) stanility, let us introduce the domain of *strictly separable preferences*. A preference profile $\succeq = (\succeq_1, \succeq_2, ..., \succeq_n)$ is *strictly separable* if $\succeq$ is separable and $U_i = \emptyset$ for each $i \in N$.

Now, we define the problem of constructing a CSC stable partition.

Construction of a CSC stable partition

Instance: Strictly separable hedonic game $\langle N, \succeq \rangle$, G_i, U_i, B_i for each $i \in N$.

Find: Construct a CSC partition with respect to $\langle N, \succeq \rangle$.

The outcome of the proposed algorithm in section 5 is a CSC stable partition when the input of the algorithm is a hedonic game instance satisfying strictly separable. Let Π be the outcome of the algorithm when the input is a strictly separable instance. Let $\mathcal{A}$ and $\mathcal{B}$ be collection of coalitions same as in section 5.

Theorem 3. *A CSC stable partition of a strictly separable instance can be constructed in $\mathcal{O}(n^2)$ time.* □

Proof. We show that each coalition is not CSC deviation from Π for $\langle N, \succeq \rangle$. $(\Pi, \mathcal{A}, \mathcal{B})$ is a partition of all coalitions $(2^N \setminus \{\emptyset\})$. Unfortunately, each coalition in Π is not deviation. Now, we consider coalitions in $\mathcal{A}$ or $\mathcal{B}$.

Firstly, we consider the coalitions in $\mathcal{A}$. Any coalition in $\mathcal{A}$ is not deviation from the same reason of CC stability.

Hence, we consider the coalitions in $\mathcal{B}$. By Observation 4, for each coalition $X \in \mathcal{B}$, there exists $leader(k) \in X$ such that $k \in [1, |\Pi|]$ and $leader(\ell) \notin X$ for each $\ell \in [1, k)$. From the construction, for each $\ell \in (k, |\Pi|]$, $\Pi(leader(\ell)) \subseteq B_{leader(k)}$. Therefore, by the definition of separable preference, $\Pi(leader(k)) \succ_{leader(k)} X$, thus $leader(k)$ disagrees to form coalition X. Hence, X is not a CSC deviation. Therefore, we can now conclude that Π is a CSC stable partition. □

The problem of construction of a CSC stable partition has G_i, U_i and B_i for each $i \in N$ as the instance. But, we don't need U_i and B_i for each $i \in N$ on the algorithm. Hence, we can construct a CSC stable partition for any strictly separable instances if the instance of the problem is $\langle N, \succeq \rangle$ and G_i for each $i \in N$.

The output of the proposed algorithm is not always CSC stable if the input is a separable instance. The proposed algorithm are correct whenever this is applied for strictly separable instances.

References

[1] Aziz, H. Brandt, F. Seedig, H.G. (2013). Computing Desirable Partitions in Additively Separable Hedonic Games. Artificial Intelligence 195, 316-334.

[2] Ballester, C. (2004). NP-Completeness in Hedonic Games. Games and Economic Behavior 49, 1-30.

[3] Banerjee, S. Konishi, H. Sönmez, T. (2001). Core in a Simple Coalition Formation Game. Social Choice and Welfare 18, 135-153.

[4] Bogomolnaia, A. Jackson, M.O. (2002). The Stability of Hedonic Coalition Structures. Games and Economic Behavior 38, 201-230.

[5] Dimitrov, D. Borm, P. Hendrickx, R. Sung S.-C. (2006). Simple Priorities and Core Stability in Hedonic Games. Social Choice and Welfare 26, 421-433.

[6] Dreze, J. Greenberg, J. (1980). Hedonic Coalitions: Optimality and Stability. Econometrica 48, 987-1003.

[7] Iwasaki, H. Sung S.-C. (2015). Computing Stable Partition in Additively Separable Hedonic Games. In Proceedings of the 18th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, 65-70.

[8] Sung, S.-C. Dimitrov, D. (2007). On Myopic Stability Concepts for Hedonic Games. Theory and Decision 62, 31-45.

[9] Tanaka, H. Sung S.-C. (2016). Linear Time Construction of CIS Partitions. In Proceedings of the 19th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, 91-96.

Identification of Causal Effect in semi-Markovian Model On Example of Economic Model

Lucie VACHOVA

University of Economics, Prague/Faculty of Management, Department of Exact Methods
Jarosovska 1117, Jindrichuv Hradec, 37701
Czech Republic
vachova@fm.vse.cz

Abstract

Existence of causal links among variables belongs to topics connected (not only) to the field of economics. Based on theoretical background the relations among variables can be identified. In the acyclic causal diagram with nodes representing variables and arrows representing potential existence of causal relationships, the causal effect is the impact of changes in related variables in response to an intervention. In case of atomic intervention, the value of exactly one variable is set to the fixed value. In a model with existence of unobserved variables, the question of causal effect identifiability appears. Provided that desired causal effect is not possible to be computed in a model, the alterations of model can be employed. In the paper some possible alterations of semi-Markovian model to make the causal effect identifiable (as suggested in [5]) are used. The semi-Markovian model is a model that contains unobserved variables and each unobserved variable is a root node with exactly two observed children. The applicability of given alterations is examined on example of simple economic model.

Keywords: Causal Effect, semi-Markovian Model, Identifiability.

1 Introduction and Basic Notions

The paper considers the simple economic application of causality modelling. In causal model the arrows represent causal relationship between corresponding variables. In some cases the causal effect (as the impact of changes in related variables in response to an intervention) can be identified.

First, in Section 2 some basic notions related to causal modelling are introduced. In the paper we will use the methodology based on Judea Pearl's approach to causality. In case that model contains unobserved variables and each of them is a root node with exactly two observed children, we call this model as semi-Markovian. In Section 2.1 Theorems for identifying causal effect of atomic intervention in semi-Markovian models are presented (as introduced by Tien and Pearl, see [4]). In case of atomic intervention, the value of exactly one variable is set to the fixed value. Additionaly some alterations of model are introduced to be done for situation that the causal effect is not possible to be identified. In Section 3 example is presented to show the effect of some such alteration on causal effect identifiability.

2 Causal Modeling

First let us introduce some basic notions regarding causal modeling. Identification of causal effect is in center of interest as can be seen for example in work of Galles and Pearl [1], Pearl [2], Spirtes *et al.* [3], etc. Usually graphs are employed to represent the causal relations. A *directed acyclic graph* (DAG) is a graph with all edges directed and containing no directed cycles.

A *causal structure* of a set of variables V is a directed acyclic graph (DAG) in which each node corresponds to a distinct element of V, and each link represents a direct functional relationship among the corresponding variables. A *causal model* is a pair $M = \langle S, \Theta_S \rangle$ (S is a causal structure, Θ_S is a set of parameters that are compatible with S). The parameters Θ_S assign a function $x_i = f_i(pa_i, u_i)$ to each $X_i \in V$ and a probability measure $P(u_i)$ to each U_i, where pa_i are

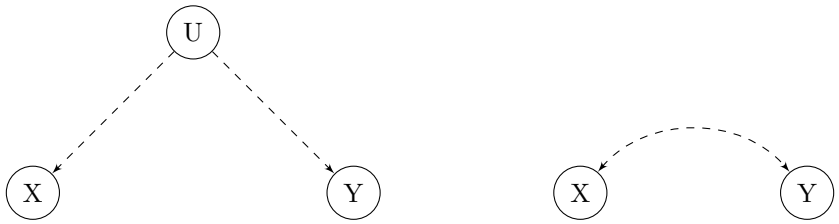


Figure 1: Confounding variable U in causal model

particular values of variables from PA_i, i.e., of the parents of X_i in S and where each U_i is a random disturbance (*unmeasured variable*) distributed according to $P(u_i)$, independently of all other U (u_i represents an instantiation of U_i) [2] .

An *intervention* (denoted $\hat{x}$) is setting the fixed value on the specific set of variables in the diagram. Given two disjoint sets of variables X and Y, a *causal effect* of X on Y is a function from X to the space of probability distributions on Y (denoted by $P(y \mid \hat{x})$). For each realization x of X, $P(y \mid \hat{x})$ gives the probability of $Y = y$ induced by deleting from the causal model all equations corresponding to variables in the set X and substituting $X = x$ in the remaining equations.

The causal effect of X on Y is *identifiable* from a graph G if the quantity $P(y \mid \hat{x})$ can be computed uniquely from any positive probability of the observed variables (see, e.g., [2]). In case that unmeasured variables are absent, all causal effects are identifiable. The joint probability function $P(v) = P(v_1, \ldots, v_n)$ is the product

$$P(v) \quad = \prod_i P(v_i|pa_i)$$

Considering all the variables as observable, the post-intervention distribution is expressed as follows (*truncated factorization* formula [2]):

$$P(v|\hat{x}) = \begin{cases} \prod_{i|V_i \notin X} P(v_i|pa_i) & \text{for all } v \text{ consistent with } X = x, \\ 0 & \text{for all } v \text{ inconsistent with } X = x \end{cases}$$

If some confounders are not observable, we have to deal with question of the causal effect identifiability. Whether the causal effect is identifiable or not depends on structure of diagram - namely, on position of variable under intervention, confounders and variables effected by intervention. Let us consider $V = V_1, \ldots, V_n$ as a set of observable variables and $U = U_1, \ldots, U_m$ as a set of unobservable variables. Then the joint probability function changes to

$$P(v) \quad = \sum_u \prod_{i|V_i \in V} P(v_i|pa_{v_i}) \prod_{i|U_i \in U} P(u_i|pa_{u_i})$$

where Pa_{v_i} and Pa_{u_i} denotes sets of parents of V_i and U_i respectively [2]. And the post-intervention distribution is then expressed as

$$P(v|\hat{x}) = \begin{cases} \sum_u \prod_{i|V_i \notin X} P(v_i|pa_{v_i}) \prod_i P(u_i|pa_{u_i}) & \text{if } v \text{ consistent with } x, \\ 0 & \text{if } v \text{ inconsistent with } x \end{cases}$$

2.1 Atomic Intervention in semi-Markovian Model

An *atomic intervention* is such an interventin when a single variable is forced to take on some fixed value ($X = x$). Given $V = V_1, \ldots, V_n$ is the set of observed variables, we denote $P(v|\hat{x})$ the causal effect of atomic intervention on X on all other observed variables (i.e., the causal effect of X on $V \backslash \{X\}$). In this section we will limit our interest to semi-Markovian models. We will understand *semi-Markovian model* as a model with unobserved variables, in which each unobserved variable is a root node with two observed children. In semi-Markovian model, the *bidirected edge* is often used to depict the confounding effect of unobserved variable. A *bidirected path* is a path that contains only bidirected edges (see, e.g., [4]). In Figure 1 the difference between using node and bidirected edge to indicate presence of confounding variable U in graph is shown.

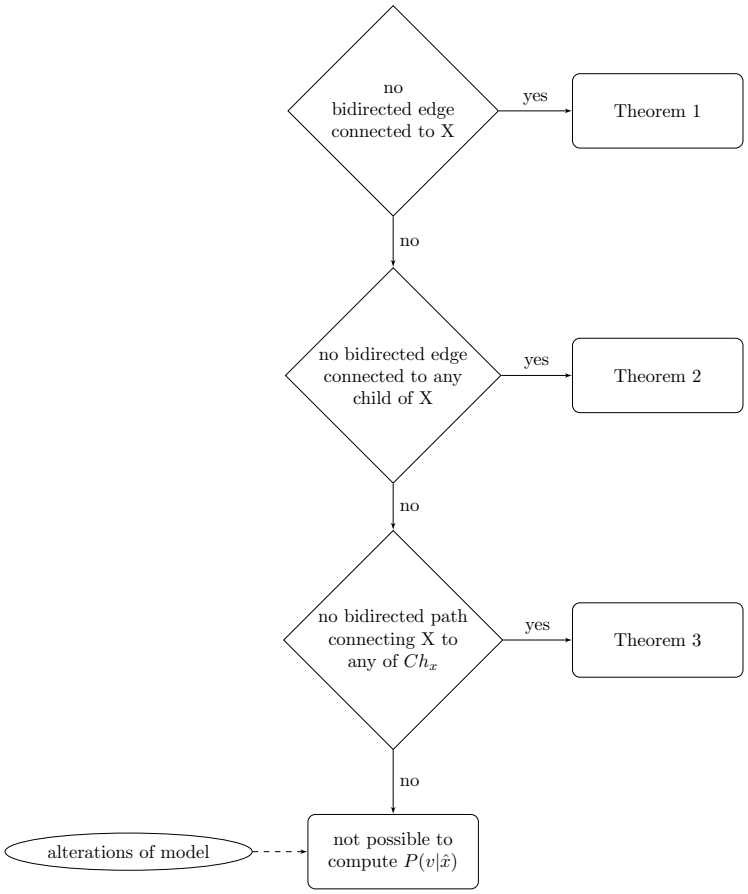


Figure 2: Identifiablity of Causal Effect $P(v|\hat{x})$

For a semi-Markovian model, the joint probability distribution $P(v)$ is expressed as

$$P(v) \quad = \sum_u \prod_i P(v_i|pa_i, u^i) \prod_i P(u_i)$$

where Pa_i and U^i are sets of measurable and unmeasurable parents of V_i respectively [4]. The post-intervention distribution is then written as

$$P(v|\hat{x}) = \begin{cases} \sum_u \prod_{i|V_i \notin X} P(v_i|pa_i, u^i) \prod_i P(u_i) & \text{if } v \text{ consistent with } x, \\ 0 & \text{if } v \text{ inconsistent with } x \end{cases}$$

In Figure 2 steps to be taken to decide on identifiability of $P(v|\hat{x})$ in semi-Markovian model are summarized. In the following Theorems three situations in which the causal effect $P(v|\hat{x})$ is identifiable are described, together with the formulas for its computing, as introduced by Tian and Pearl [4]. They differ according to structure of model, namely the position of bidirected edges. For more details and proofs see [4].

Theorem 1 (Tian-Pearl [4]) *If there is no bidirected edge connected to X, then $P(v|\hat{x})$ is identifiable and is given by*

$$P(v|\hat{x}) \quad = P(v|x, pa_x)P(pa_x)$$

Theorem 2 (Tian-Pearl [4]) *If there is no bidirected edge connected to any child of X, then $P(v|\hat{x})$ is identifiable and is given by*

$$P(v|\hat{x}) \quad = (\prod_{i|V_i \in Ch_x} P(v_i|pa_i)) \sum_x \frac{P(v)}{\prod_{i|V_i \in Ch_x} P(v_i|pa_i)}$$

where Ch_x depicts the set of childred of X.

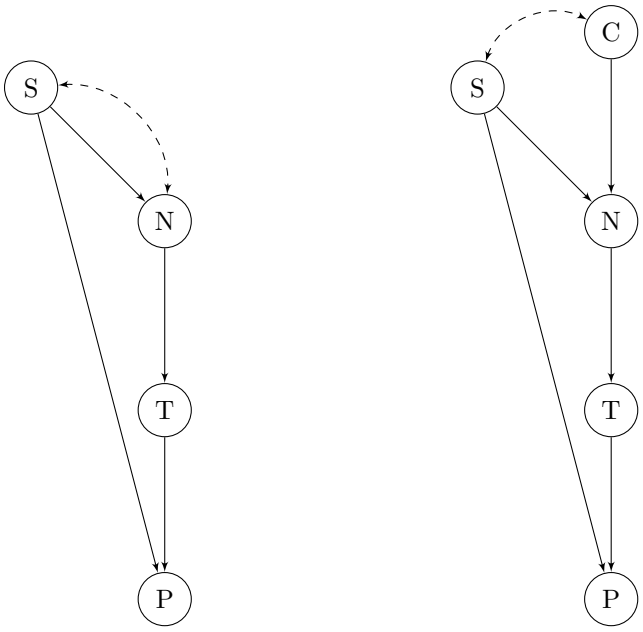


Figure 3: Alterations of causal model

Let us assign two variables to the same group if and only if these two variables are connected by a bidirected path. In this way the set of variables V can be partitioned into k disjoint groups $S_1, \ldots, S_k$. Depict N_j the set of U variables that are parents of variables in S_j. The sets $N_1, \ldots, N_k$ make a partition of U. Define

$$Q_j \quad = \sum_{n_j} \prod_{i|V_i \in S_j} P(v_i|pa_i, u^i) P(n_j), j = 1, \ldots, k.$$

Then $P(v)$ can be expressed as

$$P(v) \quad = \sum_u \prod_i P(v_i|pa_i, u^i) \prod_i P(u_i) \quad = \prod_{j=1}^{k} Q_j$$

Then S_j is a *c-component* of V in G, and Q_j is a *c-factor* corresponding to c-component S_j. Product $\prod_{j=1}^{k} Q_j$ is so called *Q-decomposition* of $P(v)$ [4].

Theorem 3 (Tian-Pearl [4]) *$P(v|\hat{x})$ is identifiable if and only if there is no bidirected path connecting X to any of ist children. When $P(v|\hat{x})$ is identifiable, it is given by*

$$P(v|\hat{x}) \quad = \frac{P(v)}{Q^X} \sum_x Q^X$$

where Q^X is the c-factor corresponding to the c-component S^X that contains X.

3 Modeling of Intervention

In case that the bidirected path connecting X to some child of X exists, the causal effect $P(v|\hat{x})$ is not computable using any above mentioned Theorems (1-3). In such a case the only possible way to identify the causal effect is to employ some alterations to model. Several types of alterations can be identified based on their position in the graph or according to form of alteration. As mentioned above, relations between variables are based on theoretical assumptions. So first it is necessary to verify that the model is structured in a correct way.

In the following text let us assume that in the model the relations were proven and the bidirected edges are correctly labeled. Thus we can focus strictly on position of alteration in model. In [5] some possible alterations of semi-Markovian model are suggested. Using simple economic model let us present their application.

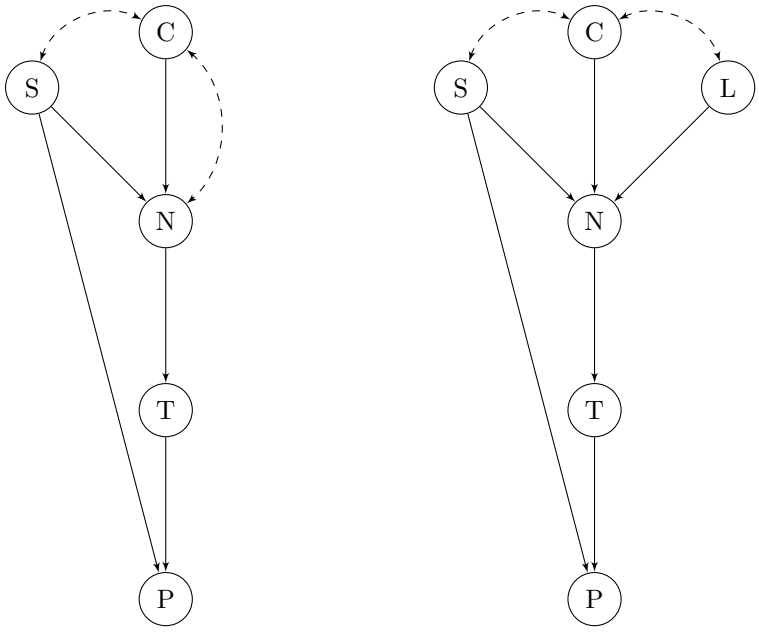


Figure 4: Alterations of causal model

Left part of Figure 3 presents causal relations among variables, namely selling price S, number of customers N, unit sales T and profit P. Moreover there is one unobservable variable represented by bidirected edge, that (in a graph) is a parent node of nodes S and N (we can understand this variable for example as crisis). For the purpose of computation let us depict this variable as U_1. We will consider selling price as a variable under intervention.

Given that bidirected path exists between S and its child (namely N), the causal effect $P(v|\hat{s})$ is not identifiable (see Theorem 3). Figure 2 then suggests to look for alteration of model. The goal is to introduce such an alteration that would allow to express $P(v|\hat{s})$ without using unobservable variable U_1. In order to compute causal effect one possible way is that the bidirected path connecting S to its child is to be broken directly by introducing new observable variable. Example of such situation is shown on right part of Figure 3 by introducing the node C (representing competitor price). By this step the bidirected path between S and N is interrupted and the causal effect is possible to be identified using the Theorem 2 as follows

$$P(v|\hat{s}) \quad = P(n|s,c) \cdot P(p|s,t) \sum_{s'} \frac{P(s',c,n,t,p)}{P(n|s',c) \cdot P(p|s',t)}$$

The desired causal effect is then identified as

$$P(v|\hat{s}) \quad = P(n|s,c) \cdot P(p|s,t) \cdot P(t|n) \cdot P(c)$$

and is expressed without necessity to include unobservable variable U_1.

In case we consider another unobservable variable to be included in model (such as on the left part of Figure 4 by adding bidirected edge between nodes C and N) we cause the indentifiability of causal effect $P(v|\hat{s})$ unenforceable, since again there is the bidirected path between nodes S and N. For the purpose of this example, let us call this variable reputation (U_2). In such a case we have to look for another suitable alteration to the model. By entering the new observable variable (such as customer loyalty L) we can again compute the causal effect $P(v|\hat{s})$ using Theorem 2. The situation is complicated by the fact that new variable is not allowed to become the child of S. On the right side of Figure 4, the new model is presented. And the causal effect is then computed without including U_2 as

$$P(v|\hat{s}) \quad = P(n|s,c,l) \cdot P(p|s,t) \sum_{s'} \frac{P(s',c,n,t,p,l)}{P(n|s',c,l) \cdot P(p|s',t)}$$

4 Conclusion and Future Research

In this paper we focused on identifiability of causal effect in semi-Markovian model and its use in a simple economic model. Our interest was limited only to atomic intervention and its effect to all other observable variables in the model ($P(v|\hat{x})$). Based on work of Tien and Pearl [4] and using some possible alterations of causal model suggested in [5] we presented simple example how to make the causal effect identifiable. There are several types of alterations that can be employed. The question then arrises about suitability of using different alterations and about the way how to decide on them.

Acknowledgement
The research was supported by GA ČR under grant no. 15-00215S.

References

[1] D. Galles. and J. Pearl. Testing identifiability of causal effects. *Uncertainty in Artificial Intelligence*, 11:185–195, 1995.

[2] J. Pearl, J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2009.

[3] P. Spirtes, C. Glymour, R. Scheines. *Causation, Prediction, and Search*. Springer-Verlag, New York, 1997.

[4] J. Tian, J. Pearl. On the Identification of Causal Effects. UCLA Cognitive Systems Laboratory, Technical Report (R-290-L), 2003.

[5] L. Vachova. On the Causal Effect Identification in semi-Markovian Models. In *CJS 2016: Proceedings of 19th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty*, Osaka, Japonsko, 2016.

A Note on Stepwise Solution Concepts for Two and Three Player Bargaining

Milan VLACH

Charles University
Department of Theoretical Computer Science and Mathematical Logic
Malostranské náměstí 25, 118 00 Praha 1, Czech Republic
milan.vlach@mff.cuni.cz

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation
Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic
vlach@utia.cas.cz

Abstract

We are concerned with stepwise solution concepts for two and three player cooperative bargaining problems whose Pareto sets are surfaces of a class of comprehensive closed convex sets bounded from above in two and three dimensional real linear spaces. After necessary preliminaries, we discuss the discrete Raiffa solution concept and its variants in two dimensions, and ordinal Shapley-Shubik solution concept in three dimensions. Then we propose an alternative stepwise solution concept in three dimensions that combines ideas from the Raiffa and Shapley-Shubik approaches.

Keywords: Cooperative bargaining, Stepwise solutions, Raiffa's solution, Ordinal solution.

1 Introduction

We continue our previous study [2, 16] of pure bargaining games. By a pure bargaining game we mean the question of solving conflicts of interests among a finite number of players in situations where the only possible outcomes are payoffs on which all the players will reach a unanimous agreement, or payoffs that each player will get in the event of a disagreement.

In terms of coalitional games this means that no coalitions except the grand coalition and the singleton coalitions are relevant. It follows that there is no difference between pure bargaining games involving two players only and two player coalitional games because no intermediate coalitions between the grand coalition and singleton coalitions exist in this case. If three or more players with conflicting interests are involved, then some intermediate coalitions of players may form and may act against the other players or coalitions. Therefore, the theory of pure bargaining games with more than two players is concerned with a special (rather narrow) class of coalitional games.

Since Nash's work [3, 4] in the early 1950's, there have been two different approaches to analyzing and solving bargaining games: a strategic (noncooperative) approach and an axiomatic (cooperative) approach. In the strategic approach the problem is formalized as a non-cooperative game in extensive form the solution of which is some of the Nash equilibriums, usually a subgame perfect equilibrium. Thus the strategic approach involves modeling the process of bargaining among players explicitly. In the axiomatic approach, one is abstracting away the details of the bargaining process. Players have the possibility of concluding mutually beneficial binding agreements, and the solution is defined by a list of reasonable conditions (axioms) that the solution is required to satisfy.

We are concerned with the axiomatic approach. The reader interested in the strategic approach or in the relationships between axiomatic and strategic approaches will find sufficient information, for example, in the book by Osborn and Rubinstein [5] and in a survey on the so-called Nash Program by Serrano [11].

Through the paper we use the following notation. For $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ from $\mathbb{R}^n$, we write $x < y$ and $x \leq y$ if, respectively, $x_i < y_i$ and $x_i \leq y_i$ for each i from $\{1, 2, \ldots, n\}$. The relations $>$ and $\geq$ between the elements of $\mathbb{R}^n$ are defined analogously. The sets $\{x \in \mathbb{R}^n : x \geq 0\}$ and $\{x \in \mathbb{R}^n : x > 0\}$ are denoted by $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$, respectively. If A is a subset of $\mathbb{R}^n$ and x is a point in $\mathbb{R}^n$, then we denote the sets $\{a + x : a \in A\}$ and $\{a - x : a \in A\}$ by $A + x$ and $A - x$, respectively. Similarly, if λ is a real number, we define λA as the set $\{\lambda a : a \in A\}$. We say that a point $a \in A$ is *Pareto optimal in* A if $x = a$ whenever x belongs to S and satisfies $x \geq a$.

Denoting by S the set of n-vectors $x = (x_1, x_2, \ldots, x_n)$ whose components represent the players payoffs that can be reached by unanimous agreements, we can represent a pure bargaining game by an ordered pair (S, d) where $d = (d_1, d_2, \ldots, d_n)$ is an element from $\mathbb{R}^n$ whose components represent the players payoffs in the case of disagreement. For obvious reason, we say that the set $S_d = \{x \in S : x \geq d\}$ is the individually rational part of (S, d).

As a rule, we are not interested only in one particular bargaining game (S, d) but in some fixed collection of such games. If $\mathcal{B}$ is such a fixed collection, then we say that $\mathcal{B}$ is a *bargaining problem* and define the notion of a *solution concept* for $\mathcal{B}$ as a function f from $\mathcal{B}$ to $\mathbb{R}^n$ such that, for each game (S, d) from $\mathcal{B}$, the value $f(S, d)$ of f at (S, d) belongs to $S \cup \{d\}$.

As mentioned previously, when there are only two players then there is no difference between bargaining games and coalitional games, while when more than two players are involved the bargaining games are very special cases of coalitional games. In the latter case, the differences in the geometry of two-dimensional spaces and of spaces of three or more dimensions may significantly influence the properties of problems and solution concepts.

First (Section 2) we consider the two player bargaining problems. Then (Section 3) we discuss the problems involving three players. In both cases we are interested in the stepwise solution concepts proposed by [1]. In particular, we will analyze a recently proposed solution concept [2, 16] which combines ideas of the discrete Raiffa solution concept [6, 7, 8] with the Shapley-Shubik ordinal solution concept [12, 9, 10].

2 Two players

The first mathematical models of pure bargaining games proposed and analyzed in the early 1950's by Nash [3, 4] and Raiffa [6, 7] are concerned with games of two players. For convenience and historical reason, we first recall Nash's axiomatic model.

2.1 The Nash concept

A particular solution concept for a bargaining problem can be defined in a number of different ways. For example, the Nash solution concept (let us denote it by N) for two player problem $\mathcal{B}$ where $\mathcal{B}$ is the set of games (S, d) in which S is a convex compact subset of $\mathbb{R}^2$ containing d, and at least one point x such that $x > d$, can be specified

- either explicitly by assigning to each bargaining game $(S, d) \in \mathcal{B}$ the maximizer of the product $(x_1 - d_1)(x_2 - d_2)$ over the individually rational part of (S, d) (see, Figure 1);
- or implicitly by requiring N to satisfy the following four conditions:
  1. Pareto optimality: $(N(S, d) + \mathbb{R}^2_+) \cap S = N(S, d)$ for each $(S, d) \in \mathcal{B}$.
  2. Symmetry: $N_1(S, d) = N_2(S, d)$ for each $(S, d) \in \mathcal{B}$ such that $(x_2, x_1) \in S$ whenever $(x_1, x_2) \in S$.
  3. Scale invariance: For each (S, d), if A is a positive affine transformation of $\mathbb{R}^2$ to itself, then $N(A(S), A(d)) = A(N(S, d))$.
  4. Independence of irrelevant alternatives: For every pair $(S, d), (T, d)$ of instances such that $S \subseteq T$, if $N(T, d)$ is in S, then $N(T, d) = N(S, d)$.

Note that neither of these two definitions specifies some explicit bargaining procedure.

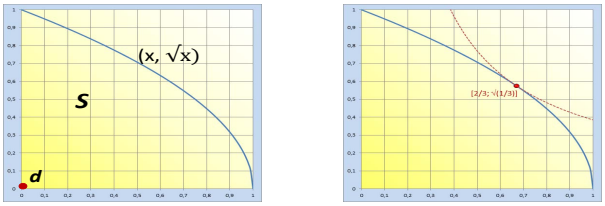


Figure 1: Nash's Solution

2.2 The Raiffa concept

Independently and almost at the same time as Nash, Raiffa [6, 7] proposed several solution concepts that differ from that of Nash. Two of them (one discrete, one continuous) arise as results of procedures in which the set S is kept unchanged while the disagreement point gradually changes. In the discrete case, which we are interested in, each new disagreement point is defined as the average of the players' most preferred points of S with respect to the old disagreement point.

To simplify description of the discrete Raiffa's solution concept, we reduce the domain of games considered by Nash. We assume that $\mathcal{B}$ consists of games (S, d) where d belongs to S, and where:

1. S is a nonempty, closed, and convex subset of $\mathbb{R}^2$.
2. S is bounded from above in the sense that there exist $a = (a_1, a_2)$ in $\mathbb{R}^2_{++}$ and a real number α such that $a_1x_1 + a_2x_2 \leq \alpha$ for all $x = (x_1, x_2)$ from S.
3. S is comprehensive in the sense that if x and z are points in S such that $x \leq z$, then the rectangle defined by x and z is also included in S.
4. Every point of "upper" (north-east) boundary of S is Pareto optimal in S.

The first two conditions guarantee that:

- The maximal payoff of each player is bounded.
- A weighted average of payoff vectors from S is also in S, which means that the payoffs can be based on the linear von Neumann-Morgenstern utilities, and the players can conduct lotteries.

The third condition can be interpreted as free disposability of utility in the sense the players can throw away, or donate, some of their gains from bargaining.

To define the discrete Raiffa solution concept for $\mathcal{B}$ formally, we first introduce for each (S, d) from $\mathcal{B}$ two mappings $y \mapsto u(S, y)$ and $y \mapsto m(S, y)$ from S to $\mathbb{R}^2$ as follows:

$$u(S, y) = (u_1(S, y), u_2(S, y)) \text{ and } m(S, y) = (m_1(S, y), m_2(S, y))$$

where

$$\begin{aligned} u_1(S, y) &= \max\{x_1 : (x_1, y_2) \in S\}, \\ u_2(S, y) &= \max\{x_2 : (y_1, x_2) \in S\}; \\ m_1(S, y) &= \frac{1}{2}(y_1 + u_1(S, y)) = y_1 + \frac{1}{2}(u_1(S, y) - y_1), \\ m_2(S, y) &= \frac{1}{2}(y_2 + u_2(S, y)) = y_2 + \frac{1}{2}(u_2(S, y) - y_2). \end{aligned}$$

It follows from the properties of S that these mappings are well defined[1]. Moreover, for each y in S, we have $m(S, y) \in S$ by convexity of S, and $u(S, y) \geq y$ by definition of u. The point $u(S, y)$,

[1] Sets $\{x_1 : (x_1, y_2) \in S\}$ and $\{x_2 : (y_1, x_2) \in S\}$ are nonempty for each $y \in S$, S is bounded from above, and S is closed.

which may (but not necessarily) belong to S, is usually called the ideal or utopia or bliss point for (S, y).

The *discrete Raiffa solution concept for* $\mathcal{B}$, let us denote it by R, is defined by setting, for each game (S, d) from $\mathcal{B}$, the value $R(S, d)$ as the limit of the sequence $\{x^k(S, d)\}$ of points generated as follows (see Figure 2):

$$x^0(S, d) = d, \tag{1}$$
$$x^{k+1}(S, d) = m(S, x^k(S, d)) \text{ for } k \geq 0. \tag{2}$$

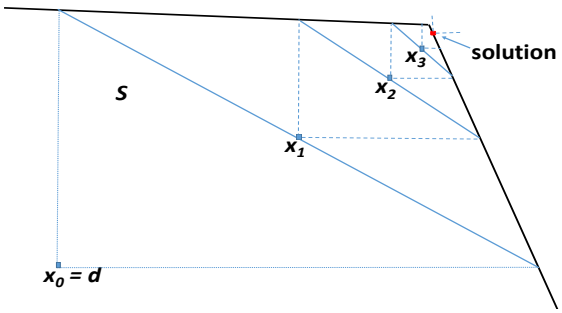


Figure 2: Discrete Raiffa Solution

It is easy to see that R is a well defined solution concept for $\mathcal{B}$. Indeed, the limit of $\{x^k(S, d)\}$ exists and belongs to S for every $(S, d) \in \mathcal{B}$ because: (i) S is closed, (ii) S is bounded from above, (iii) $x^k(S, d)$ belongs to S for each k, and (iv) $x^{k+1}(S, d) \geq x^k(S, d)$.

2.2.1 Properties

The discrete Raiffa solution concept has a number of desirable properties. Here we present some of them.

Property 1. *For each* $(S, d) \in \mathcal{B}$, $R(S, d)$ *is Pareto optimal in* S.

Property 2. *For each* $(S, d) \in \mathcal{B}$, $R(S, d)$ *belongs to the individually rational part of* (S, d).

Property 3. *For each* $(S, d) \in \mathcal{B}$, *if* d *is not Pareto optimal in* S, *then* $R(S, d) \neq d$.

Property 4. *For each* $(S, d) \in \mathcal{B}$, *if all players are symmetric in* (S, d), *then they are also symmetric in* $R(S, d)$.

Property 5. *For each* $(S, d) \in \mathcal{B}$, *if* A *is a positive affine transformation of* R^2 *to itself, then* $R(A(S), A(d)) = A(R(S, d))$.

Property 6. *For each* (S, d) *and* (T, d) *from* $\mathcal{B}$, *if* $S_d = T_d$, *then* $R(S, d) = R(T, d)$.

Let us verify that Property 1 is satisfied. The verification of the remaining properties is straightforward.

First we show that $u(S, \cdot)$ is continuous on S for each $(S, d) \in \mathcal{B}$. We know that the functions $u_1(S, \cdot)$ and $u_2(S, \cdot)$ are finite real valued functions on S. It follows from the convexity of S, that $u_1(S, \cdot)$ and $u_2(S, \cdot)$ are concave on S. Consequently, $u_1(S, \cdot)$ and $u_2(S, \cdot)$ are continuous in the interior of S. Now consider a boundary point x of S. According to the assumptions about games from $\mathcal{B}$, the point x is Pareto optimal in S. Therefore $u_1(S, x) = x_1$ and $u_2(S, x) = x_2$. Let $\{x^k\}$ be a sequence of point in S converging to x. It follows that $\liminf_{k\to\infty} u_1(S, x^k) \geq x_1$ because $u_1(S, x^k) \geq x_1^k$ for each k. Let y_1 denote the $\limsup_{k\to\infty} u_1(S, x^k)$, and assume that $y_1 > x_1$. Then the point (y_1, x_2) is the limit point of $(u_1(S, x^k), x_2^k)$. It follows that (y_1, x_2) is in S because S is closed. However this contradicts the Pareto optimality of x. Therefore $\lim_{k\to\infty} u_1(S, x^k) = x_1 = u_1(S, x^k)$. Similarly we obtain that $\lim_{k\to\infty} u_2(S, x^k) = x_2 = u_2(S, x^k)$.

Now we know that $u(S,\cdot)$ is a continuous not only in the interior of S but also on the boundary of S.

Next we observe that if x from S is such that $u(S,x) = x$, then x is Pareto optimal in S. Indeed: assume that $x \in S$ is not Pareto optimal in S. Then there exists a point $y \in S$ different from x such that $y_1 > x_1$ or $y_2 > x_2$. Since S is comprehensive, at least one of the points (y_1, x_2) and (x_1, y_2) belongs to S. Therefore $u_1(S,x) > x_1$ or $u_2(S,x) > x_2$.

To conclude the proof, it is sufficient to show that $u(S, R(S,d)) = R(S,d)$. Since $R(S,d)$ is the limit of the sequence $\{x^k(S,d)\}$ defined by (1) and (2), we have

$$\begin{aligned} R(S,d) &= \lim_{k\to\infty} x^{k+1}(S,d) = \lim_{k\to\infty} m(S, x^k(S,d)) = \lim_{k\to\infty} (x^k(S,d) + \frac{1}{2}(u(S,x^k(S,d)) - x^k(S,d)) \\ &= \lim_{k\to\infty} x^k(S,d) + \frac{1}{2}\lim_{k\to\infty} u(S,x^k(S,d)) - \frac{1}{2}\lim_{k\to\infty} x^k(S,d) \\ &= \frac{1}{2}\lim_{k\to\infty} x^k(S,d) + \frac{1}{2}\lim_{k\to\infty} u(S,x^k(S,d)) = \frac{1}{2}R(S,d) + \frac{1}{2}\lim_{k\to\infty} u(S,x^k(S,d)). \end{aligned}$$

By the continuity of $u(S,\cdot)$, we have

$$\lim_{k\to\infty} u(S,x^k(S,d)) = u(S, \lim_{k\to\infty} x^k(S,d)) = u(S,R(S,d)).$$

It follows that

$$R(S,d) = \frac{1}{2}R(S,d) + \frac{1}{2}u(S,R(S,d)) \quad \text{and} \quad R(S,d) = u(S,R(S,d)),$$

as required for completing the proof.

2.2.2 Remarks

Remark 1. Recently, Diskin et al. [1] have introduced a solution concept which is composed of two functions. One function specifies an interim agreement and the other specifies the terminal agreement. When specified for two player problems[2] this concept can be defined as follows:

A *stepwise solution concept* of a bargaining problem $\mathcal{B}$ is a pair (f,g) of functions from $\mathcal{B}$ into $\mathbb{R}^2$ such that, for each game (S,d) from $\mathcal{B}$,

- both $f(S,d)$ and $g(S,d)$ belong to S,
- and $f(S,d) = f(S, g(S,d))$.

Clearly, the discrete Raiffa solution concept can be viewed as a special stepwise solution concept (f,g) where the function f and g are defined by

$$f(S,d) = d + \frac{1}{2}(u(S,d) - d),$$

and $g(S,d)$ is the limit of the sequence $\{d^k(S,d)\}$ of points constructed inductively by

$$d^0(S,d) = d \text{ and } d^{k+1}(S,d) = f(S, d^k(S,d)).$$

Remark 2. Diskin et al. [1] extend Raiffa's approach by introducing a family $\{(f^p, g^p)\}_{0<p<1}$ of similar but different stepwise solution concepts where each member of the family is specified by a value of a real parameter p from the interval $(0,1)$ as follows: the function f^p is defined by

$$f^p(S,d) = d + \frac{p}{2}(u(S,d) - d),$$

and $g(S,d)$ is the limit of the sequence $\{d^k(S,d)\}$ of points constructed inductively by

$$d^0(S,d) = d \text{ and } d^{k+1}(S,d) = f^p(S, d^k(S,d)).$$

[2] Diskin et al. are concerned with problems in $\mathbb{R}^n$ for $n \geq 2$.

REMARK 3. Nash's and Raiffa's solution concepts (and many others) are invariant with respect to positive affine transformations (Property 5). Taking into account the fact that a large part of modern economic theory is based on ordinal preferences it is natural to ask whether these concepts are invariant with respect to every strictly increasing transformation of individual utilities. It is easy to construct examples to show that neither Nash's nor Raiffa's concept has this property.

However the situation is much worse in this respect. In the case of two player bargaining problems $\mathcal{B}$ considered in this section, all solution concepts satisfying this stronger requirement (let us call them ordinal) are quite uninteresting. To realize this, consider (S, d) from $\mathcal{B}$ where $d = 0$ and where

$$S = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}.$$

Let T be the transformation defined by

$$T(x_1, x_2) = \left(\frac{2x_1}{1 + x_1}, \frac{x_2}{2 - x_2}\right). \tag{3}$$

It can easily be verified that T preserves utility orderings of both players on the square

$$Q = \{(x_1, x_2) : 0 \leq x_1 \leq 1, \ 0 \leq x_2 \leq 1\},$$

and it maps the set S onto itself, see Figure 3. It follows that each solution concept f for $\mathcal{B}$ which

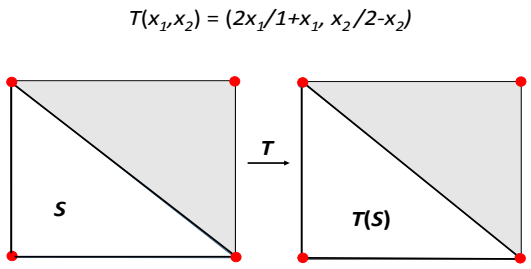


Figure 3: Fixed points of T

is invariant with respect to ordinal transformations must assign to game $(S, 0)$ a point in S which is also a fixed point (on S) of transformation (3). However, the only fixed points of T on S are the points

$$(0, 0), (0, 1), (1, 0), (1, 1).$$

The point $(1, 1)$ is uninteresting because it does not belong to S. The remaining points are also not appealing as results of bargaining: $(0, 0)$ is the point of disagreement, and points $(0, 1)$ $(1, 0)$, are maximally asymmetric (dictatorial).

Fortunately, the main argument of this proof of nonexistence of ordinal solution concepts cannot be extended to problems with three or more players. We will see in the next section that an interesting ordinal solution concept exists for problems with more than two players.

3 Three players

Analogously to the previous section we assume that bargaining problem $\mathcal{B}$ consists of games (S, d) with $d \in S$ and such that:

1. S is a nonempty, closed, and convex subset of $\mathbb{R}^3$.
2. S is bounded from above in the sense that there exist $a = (a_1, a_2, a_3)$ in $\mathbb{R}^3_{++}$ and a real number α such that $a_1x_1 + a_2x_2 + a_3x_3 \leq \alpha$ for all $x = (x_1, x_2, x_3)$ from S.
3. S is comprehensive in the sense that if x and z are points in S such that $x \leq z$, then the hyper-rectangle defined by x and z is also included in S.
4. Every point of "upper" boundary of S is Pareto optimal in S.

3.1 The Raiffa concept

The extension of the discrete Raiffa solution concept and of the family of concepts introduced by Diskin et al. to problems with three (or more) players is straightforward. First we define mappings u and m by

$$\begin{aligned} u_1(S,y) &= \max\{x_1 : (x_1,y_2,y_3) \in S\}, \\ u_2(S,y) &= \max\{x_2 : (y_1,x_2,y_3) \in S\}; \\ u_3(S,y) &= \max\{x_3 : (y_1,y_2,x_3) \in S\}; \\ m_1(S,y) &= \frac{1}{3}(y_1 + u_1(S,y)) = y_1 + \frac{1}{3}(u_1(S,y) - y_1), \\ m_2(S,y) &= \frac{1}{3}(y_2 + u_2(S,y)) = y_2 + \frac{1}{3}(u_2(S,y) - y_2), \\ m_3(S,y) &= \frac{1}{3}(y_3 + u_3(S,y)) = y_3 + \frac{1}{3}(u_2(S,y) - y_3), \end{aligned}$$

and then we use an analogue of (1) and (2); that is, we define $R(S,d)$ as the limit of the sequence $\{x^k(S,d)\}$ of points from $\mathbb{R}^3$ generated as follows

$$\begin{aligned} x^0(S,d) &= d, \\ x^{k+1}(S,d) &= m(S,x^k(S,d)) \text{ for } k \geq 0. \end{aligned}$$

3.2 The Shapley-Shubik concept

As mentioned previously, in the case of two player bargaining problems considered by Nash and Raiffa, all ordinal solution concepts are quite unappealing. This impossibility result, however, does not hold for problems with more than two players.

An ordinal solution for the three player problem has been constructed by Shapley and Shubik [13], see also [14]. Recently Safra and Samet [9], [10] have generalized the ordinal Shapley-Shubik concept for problems with more than three players and demonstrated that there is even a continuum of ordinal concepts that lead to efficient and symmetric outcomes.

The construction of such an ordinal solution concept for the three player problem is based on the fact that if Q is a Pareto surface in $\mathbb{R}^3$ and (a_1,a_2,a_3) is a point in $\mathbb{R}^3 \setminus Q$, then there is a unique point (b_1,b_2,b_3) such that the points

$$(a_1,b_2,b_3),(b_1,a_2,b_3),(b_1,b_2,a_3)$$

belong to Q, see Figure 4.

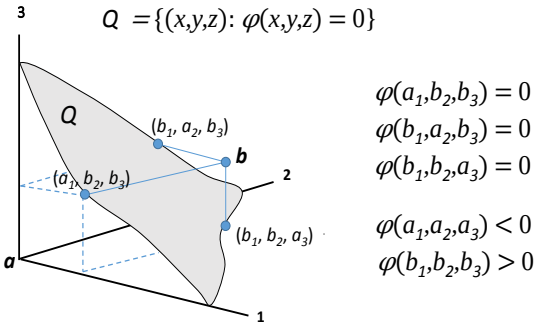


Figure 4: Shapley's construction

Using this fact, the outcome of the Shapley-Shubik concept for each instance (S,d) can be defined as the limit of the sequence $\{x^k\}$ of points obtained by setting $x^0 = d$ and generating

points x^k to be the unique points $x^k = (x_1^k, x_2^k, x_3^k)$ determined by the property that the points

$$(x_1^{k-1}, x_2^k, x_3^k), (x_1^k, x_2^{k-1}, x_3^k), (x_1^k, x_2^k, x_3^{k-1})$$

belong to Pareto surface of S. It can be seen that the outcomes of this procedure are individually rational, symmetric, and Pareto optimal, and that the concept can be formalized as stepwise one.

3.3 Alternatives

We propose to combine the Shapley-Shubik concept with the discrete Raiffa concept. The new interim point in the Raiffa solution procedure is obtained by averaging from three particular points on Pareto surface that are determined by the old interim point. We can see that in the Shapley-Shubik procedure each interim point determines three special points on the Pareto surface of S which are different from the points used in the Raiffa procedure, see Figure 5. Therefore it is of interest to apply averaging on the triple of points from the Shapley-Shubik procedure. In this way

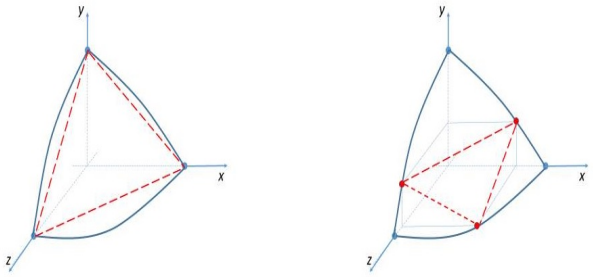


Figure 5: Alternative construction

we obtain the solution outcome to a game (S, d) from $\mathcal{B}$ as the limit of the sequence $\{y^k\}$ of points from S defined as follows.

Again we set $y^0 = d$. Let (x_1, x_2, x_3) be the point obtained from y^0 by one step of the Shapley-Shubik procedure. We construct the next interim point y^1 by the same averaging that is used in the Raiffa procedure, but now using the points[3]

$$(y_1^0, x_2, x_3),\ (x_1, y_2^0, x_3),\ (x_1, x_2, y_3^0)$$

instead of using the points

$$(m_1(S, y^0), y_2^0, y_3^0),\ (y_1^0, m_2(S, y^0), y_3^0),\ (y_1^0, y_2^0, m_3(S, y^0)).$$

Then we continue in the same way, that is, we construct y^{k+1} from y^k as follows. First we use the fact that there is a unique point (x_1, x_2, x_3) such that the points $(y_1^k, x_2, x_3), (x_1, y_2^k, x_3), (x_1, x_2, y_3^3)$ belong to S and set

$$y^{k+1} = \frac{1}{3}((y_1^k, x_2, x_3) + (x_1, y_2^k, x_3) + (x_1, x_2, y_3^k)).$$

Again the convexity of S guarantees that, for each k, the point y^k belongs to S and $y^{k+1} \geq y^k$. Because the set S is closed and bounded from above, we know that the sequence $\{y^k\}$ converges to a point in S. In fact, all properties analogous to those for two players are satisfied.

It is clear that the same results can be obtained for analogous combination of the Shapley-Shubik procedure with every member of the family $\{(f^p, g^p)\}$ proposed in [1].

Acknowledgement
This paper was written with the support of the Czech Science Foundation project No. P402/12/G097 DYME - Dynamic Models in Economics, and the Charles University project Progress Q48 - Computer Science.

[3]For a bargaining set in $\mathcal{B}$, they form a unique von Neumann-Morgenstern stable set, [17].

References

[1] Diskin, A., Koppel, M., Samet, D.: Generalized Raiffa solutions. Games and Economic Behavior 73, 452–458 (2011)

[2] Mihola, J., Valenčík, R., Vlach, M.: Game theoretic models of distribution systems. Proceedings of the 14th Czech-Japan Seminar on Data Analysis and Decision under Uncertainty, pp. 89–93, Charles University, Hejnice, (September 2011)

[3] Nash, J.: The Bargaining Problem. Econometrica 18, 155–162 (1950)

[4] Nash, J.: Two Person Cooperative Games. Econometrica 21, 128–140 (1953)

[5] Osborn, M.J., Rubinstein, A.: Bargaining and Markets. Academic Press, San Diego, California (1990)

[6] Raiffa. H.: Arbitration schemes for generalized two-person games. Report No. M720-1, R30, Engineering Research Institute, University of Michigan, Ann Arbor (June, 1950)

[7] Raiffa. H.: Arbitration Schemes for Generalized Two Person Games. Contributions to the Theory of Games II (H.W. Kuhn and A.W. Tucker eds.), 361–387, Princeton University Press, Princeton, New Jersey (1953)

[8] Luce, R.D., Raiffa, H.: Games and Decisions: Introduction and Critical Survey. Wiley, New York, (1957)

[9] Safra, Z., Samet, D.: An ordinal solution to bargaining problems with many players. Games and Economic Behavior 46, 129–142 (2004)

[10] Samet, D., Safra, Z.: A family of ordinal solutions to bargaining problems with many players. Games and Economic Behavior 50, 89–106 (2005)

[11] Serrano, R.: Fifty years of the Nash program. Investigaciones Economicas 29, 219–258 (2005)

[12] Shapley, L.S: Utility comparison and the theory of games. In: La Décision, edited by Guildband, G.Th., Editions du CNRS, Paris, 251–263 (1969). Reprinted in: The Shapley value (A.E. Roth ed.), Cambridge University Press, 307–319 (1988)

[13] Shapley, L.S., Shubik, M.: Game Theory in Economics – Chapter 4: Preferences and Utility. R-904/4-NSF, Published by The Rand Corporation (December, 1974)

[14] Shubik, M.: Game Theory in the Social Sciences. Cambridge, Mass. The MIT Press (1982, 1984)

[15] Thomson, W.: Cooperative models of bargaining. In: Aumann, R.J., Hart, S.(Eds.), Handbook of Game Theory, Chapter 35. North Holland, New York, pp. 1237–1248 (1994)

[16] Vlach, M.: Variant forms of the Raiffa solution to the bargaining problem. In *Proc. of the 33rd Int'l Conf. on Mahematical Methods in Economics*, pages 876–881, Cheb, Czech Republic, 2015.

[17] Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944)

Interval estimation for a fuzzy linear regression model

Jin Hee YOON[1], Hae Kyung KIM[2], Hye-Young JUNG[3], Woo-Joo LEE[4] and Seung Hoe CHOI[5]

[1] *School of Mathematics and Statistics, Sejong University*
209 Neungdong-ro Gwangjin-gu, Seoul 05006
South Korea
jin9135@sejong.ac.kr

[2,4] *Department of Mathematics, Yonsei University*
50 Yonsei-ro, Seodaemun-gu, Seoul 03722
South Korea
abc1217@yonsei.ac.kr;bobspace@yonsei.ac.kr

[3] *Department of Statistics, Seoul National University*
1 Gwanak-ro, Gwanak-gu, Seoul 08826
South Korea
hyjunglove@snu.ac.kr

[5] *School of Liberal Arts and Science, Korea Aerospace University, Koyang*
76 Hanggongdaehang-ro, Deogyang-gu, Koyang, Gyeonggi-do 10540
South Korea
shchoi@kau.ac.kr

Abstract

In this paper, a statistical inferences for linear regression models with vague data are studied. The least squares estimation for fuzzy parameters of regression models based on fuzzy-valued data is considered, and analogues of normal equations and fuzzy least squares estimators are derived. An examples is introduced to illustrate the results.

Keywords: Interval estimation, Fuzzy least squares estimation, Fuzzy regression model, Fuzzy data.

1 Introduction

In a real world, sometimes we face ambiguous or vague data in many situations, such as 'about 10', 'rather greater than 10', 'more or less between 5 and 10', etc. Moreover, linguistically expressed outcomes such as 'light', 'moderate', and 'heavy' to describe the degree or intensity of a certain event occur frequently in everyday life. The boundaries of these data or information are not well defined and often vague. In these cases, due to the vagueness of data, the traditional least squares method cannot be applied. Even if approximate estimates are computed based on vague data by some numerical methods, the resulting estimators do not enjoy any tractable finite optimality property, e.g., the best linear unbiased estimator, etc.

Linear models for regression analysis with vague observations have been used quite extensively in the past few years. One of the more widely used models is the one suggested by Tanaka *et. al* [16] . The model gives a crisp relationship between the dependent and independent variables based on the given data from a statistical viewpoint.

The constraints of vagueness to least squares approaches led to a new class of statistical regression problems, proposed by many authors such as [4, 5, 6, 9, 13, 14, 19, 22]. The fuzzy regression model, first proposed by Tanaka [16, 17], obtains parameters by solving linear programming problems. The basic notion of fuzzy variable and fuzzy sample that unites the randomness and fuzziness proposed in [10, 11, 12, 18]. A type of linear estimation theory with random fuzzy data was developed in [15] using a suitable notion of expectation and variance for fuzzy random variables. In

particular, Diamond [5] developed several models for least squares approaches for special types of parametric fuzzy data, so-called triangular fuzzy numbers, giving some criteria for when fuzzy data sets can be fitted to the models, and deriving analogues of the normal equations. The problems of statistical inference with vague data, i.e., estimation, testing hypotheses and analysis of variance with fuzzy data, are also considered by various authors [1, 2, 3, 7, 20, 21].

In this paper, we extend a linear regression model to include specific cases where observations are vague or even linguistic and consider some statistical inference of the model.

2 Preliminaries

Consider a classical univariate linear model. In matrix notation, the model can be given by

$$Y = X\boldsymbol{\beta} + \epsilon, \tag{1}$$

where $Y = (Y_1, \cdots, Y_n)^t$ is an $(n \times 1)$ vector of observable response variables, X is an $(n \times q)$ matrix of known constants x_{ij}, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_{p-1})^t$ denotes the $(p \times 1)$ vector of unknown parameters, and $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^t$ is an $(n \times 1)$ vector of unobservable random variables (errors) assumed to follow a distribution with d.f. F such that $E[\epsilon_n] = 0$ and $Var[\epsilon_n] = \sigma^2 I_n$, $\sigma^2 < \infty$. It is usual to take $x_{i1} \equiv 1$, $i = 1, \cdots, n$.

The usual method of estimation in this case is the least squares method, and the least squares estimator (LSE) is given by $\hat{\boldsymbol{\beta}}_n = (X_n^t X_n)^{-1} X_n^t Y$, where the design matrix X_n is supposed to have full rank. The LSE has the usual optimal properties under the Gaussian model, in which the error terms are assumed to be Gaussian. These are unbiased and asymptotically efficient. In fact, they have the minimum variance among unbiased estimators. The Gauss-Markov theorem presents a more general result in that it applies to the linear regression models with general error distributions. All that is assumed is that the error distribution has a mean of zero and a finite variance σ^2.

Throughout this work, the experimental data involved are assumed to be imprecise. Moreover, in order to obtain a generalization of the least squares method in relation to the situation of imprecise data, which reduces to the classical method when the model elements are crisp, we consider the following fuzzy linear regression model:

$$Y_i = B_0 \oplus B_1 x_{i1} \oplus \cdots \oplus B_{p-1} x_{i,p-1} \oplus \Phi_i, \qquad i = 1, \cdots, n \tag{2}$$

where Y_i is random fuzzy variable and x_{ij} $(j = 1, \cdots, p)$ are crisp variables, and B_j are unknown regression fuzzy parameters estimated on the basis of fuzzy and crisp observations Y_i and x_{ij}, respectively. Φ_i are assumed to be random fuzzy variables allowing negative spreads. In (2), the symbol $\oplus$ is the addition of fuzzy sets which is defined later.

Note that in model (2), two different kinds of uncertainty, vagueness and randomness, are considered simultaneously. In fact, randomness resulting from measurement errors and fuzziness resulting from system fuzziness are two different kinds of uncertainty that coexist in a regression analysis. In practice, precision of data is almost never attainable, and the most that one can expect is that observations are valid subject to random error. From this point of view, model (2) is different from the fuzzy regression models that are discussed in [5, 8] and other reports. Thus model (2) can be considered an extension of the standard linear regression model to the case where the observations of the explanatory and response variables are fuzzy numbers, since crisp values can be treated as degenerated fuzzy numbers.

Models dealing with imprecision or vagueness of data will be certain fuzzy sets of the space of real numbers. When taking the new model into account, we introduce some definitions in [5] regarding the fuzzy sets and the fuzzy numbers, as well as some basic results of fuzzy theory.

There are several metrics which can be defined on the more general class $\mathcal{F}_c(R^d)$. The distance between two fuzzy numbers is commonly based on the distance between their α-cuts. A useful type of metric can be defined via support functions. The support function of any compact convex set $A \subset R^d$ is defined as function $s_A : S^{d-1} \to R^1$ given by, for all $t \in S^{d-1}$

$$s_A(t) = \sup_{a \in A} < t, a >$$

where S^{d-1} is the $(d-1)$-dimensional unit sphere in R^d and $< \cdot, \cdot >$ denotes the scalar product on R^d. Note that for convex and compact $A \subset R^d$ the support function s_A is uniquely determined. A metric on $\mathcal{F}_c(R^d)$ is defined by the L_2-metric on the space of Lebesque integrable functions

$$\delta_2(A,B) = \left[d \cdot \int_0^1 \int_{S^{n-1}} |s_A(\alpha, r) - s_B(\alpha, r)|^2 \mu(dr) d\alpha \right]^{1/2}$$

for all $A, B \in \mathcal{F}_c(R^d)$. [14] proposed a metric between fuzzy numbers X and Y as the distance between the images of X and Y. For $X, Y \in \mathcal{F}_c(R^1)$, define

$$d_M^2(X,Y) = \int_0^1 (X_\alpha^- - Y_\alpha^-)^2 d\alpha + \int_0^1 (X_\alpha^+ - Y_\alpha^+)^2 d\alpha,$$

where X_α^-, X_α^+ are the lower and upper endpoints of X_α. Diamond [5] defined another metric in the set of all triangular fuzzy numbers as follows. Let $\mathcal{T}(R^1)$ denote the set of all triangular fuzzy numbers in R^1. For $X, Y \in \mathcal{T}(R^1)$, define

$$d^2(X,Y) = D_2^2(\text{supp}X, \text{supp}Y) + [m(X) - m(Y)]^2,$$

where supp$X$ denotes the compact interval of support of X, and $m(X)$ its mode. If $X = (x, \xi^l, \xi^r)_\triangle$, $Y = (y, \eta^l, \eta^r)_\triangle$, then

$$d_D^2(X,Y) = [y - \eta^l - (x - \xi^l)]^2 + [y + \eta^r - (x + \xi^r)]^2 + (y - x)^2. \tag{3}$$

3 Estimations for a fuzzy regression model

The subject of statistical analysis under (2) is concerned primarily with the problem of making inferences about the fuzzy parameters $B_0, B_1, \cdots, B_{p-1}$. We consider new multiple linear regression model :

$$Y_i = B_0 \oplus B_1 x_{i1} \oplus \cdots + B_{p-1} x_{i,p-1} \oplus \Phi_i, \qquad i = 1, \cdots, n, \tag{4}$$

when x_{ij} are the ith fixed input values of the jth variables ($j = 1, \cdots, p-1$) which gives rise to observations Y_i, which are triangular fuzzy numbers with vector representations $Y_i = (y_i - \eta_i^l, y_i, y_i + \eta_i^r)$ where y_i are the modes and η_i^l, η_i^r are the left and right spreads of Y_i, respectively. $B_0, B_1, \cdots, B_{p-1}$ are the unknown regression parameters which are triangular fuzzy numbers with vector representations $B_i = (b_i - \delta_i^l, b_i, b_i + \delta_i^r)$ with the mode b_i and the left and right spreads δ_i^l and δ_i^r, respectively. We assume that Φ_i are the crisp random vectors for expressing randomness which are expressed by $\Phi_i = (\epsilon_i^m - \epsilon_i^l, \epsilon_i^m, \epsilon_i^m + \epsilon_i^r)$ with crisp random variables ϵ_i^l, ϵ_i^m, ϵ_i^r.

Note that the fuzzy regression coefficient B_j may be interpreted as the increment in fuzzy Y corresponding to a unit increase in x_{ij} when all other variables are held constant. The component form of model (4) is given by the following crisp models:

$$\begin{aligned} \eta_i^l &= \delta_0^l + \delta_1^l x_{i1} + \cdots + \delta_{p-1}^l x_{i,p-1} + \epsilon_i^l, \\ y_i &= b_0 + b_1 x_{i1} + \cdots + b_{p-1} x_{i,p-1} + \epsilon_i^m, \\ \eta_i^r &= \delta_0^r + \delta_1^r x_{i1} + \cdots + \delta_{p-1}^r x_{i,p-1} + \epsilon_i^r, \end{aligned} \tag{5}$$

where $\eta_i^l \geq 0$, $\eta_i^r \geq 0$, and the error terms ϵ_i^l, ϵ_i^r in Φ_i are constrained to be $\eta_i^l \geq 0$, $\eta_i^r \geq 0$, a.s.

In the least squares approach, we seek estimators for $\boldsymbol{B} = (B_0, \cdots, B_{p-1})^t$ that minimizes the sum of squares of residuals of the n observed Y's from their predicted value $\hat{Y}$. Any vector $\hat{\boldsymbol{B}}_n = (\hat{B}_0, \cdots, \hat{B}_{p-1})^t$ which minimizes

$$Q(B_0, \cdots, B_{p-1}) = \sum_{i=1}^n d_D^2 \left(Y_i, \sum_{j=0}^{p-1} B_j x_{ij} \right)$$

is called a fuzzy least squares estimator of $\boldsymbol{B}$ based on fuzzy data $\{(x_{ij}, Y_i)\}$, $j = 0, 1, \cdots, p-1$; $i = 1, \cdots, n$, where $x_{i0} = 1$ and d_D is a metric defined in (3). Moreover, we have $Q(B_0, \cdots, B_{p-1})$

in as

$$\sum_{i=1}^{n}\left[\left(y_i-\sum_{j=0}^{p-1}b_jx_{ij}\right)^2+\left((y_i-\eta_i^l)-\sum_{j=0}^{p-1}(b_j-\delta_j^l)x_{ij}\right)^2+\left((y_i+\eta_i^r)-\sum_{j=0}^{p-1}(b_j+\delta_j^r)x_{ij}\right)^2\right],$$

where $x_{i0}=1$ for $i=1,\cdots,n$. The function Q is to be minimized with respect to $B_1,\cdots B_{p-1}$. The fuzzy least squares estimators of $B_1,\cdots B_{p-1}$ must satisfy

$$\frac{\partial Q}{\partial b_j}=0,\quad \frac{\partial Q}{\partial \delta_j^l}=0,\quad \frac{\partial Q}{\partial \delta_j^l}=0,\quad \text{for } j=0,1,\cdots,p-1.$$

In particular, for $p=2$, i.e., the simple fuzzy linear model, the normal equations becomes

$$\begin{cases}
3\sum y_i+\sum(\eta_i^r-\eta_i^l) &= 3nb_0+3b_1\sum x_i-n\delta_0^l+n\delta_0^r-\delta_1^l\sum x_i+\delta_1^r\sum x_i,\\
3\sum x_iy_i+\sum(\eta_i^r-\eta_i^l)x_i &= 3b_0\sum x_i+3b_1\sum x_i^2-\delta_0^l\sum x_i+\delta_0^r\sum x_i-\delta_1^l\sum x_i^2+\delta_1^r\sum x_i^2,\\
\sum y_i-\sum\eta_i^l &= nb_0+b_1\sum x_i-n\delta_0^l-\delta_1^l\sum x_i,\\
\sum y_i+\sum\eta_i^r &= nb_0+b_1\sum x_i+n\delta_0^r+\delta_1^r\sum x_i,\\
\sum x_iy_i-\sum x_i\eta_i^l &= b_0\sum x_i+b_1\sum x_i^2-\delta_0^l\sum x_i-\delta_1^l\sum x_i^2,\\
\sum x_iy_i+\sum x_i\eta_i^r &= b_0\sum x_i+b_1\sum x_i^2+\delta_0^r\sum x_i+\delta_1^r\sum x_i^2.
\end{cases}$$

It is more convenient to deal with the objective function of quadratic forms if they are expressed in vector notation. Let $\boldsymbol{y}=(y_1,\cdots,y_n)^t$, $\boldsymbol{\eta}^l=(\eta_1^l,\cdots,\eta_n^l)^t$ and $\boldsymbol{\eta}^r=(\eta_1^r,\cdots,\eta_n^r)^t$. Moreover, $\boldsymbol{b}=(b_0,b_1,\cdots,b_{p-1})^t$, $\boldsymbol{\delta}^l=(\delta_0^l,\delta_1^l,\cdots,\delta_{p-1}^l)^t$ and $\boldsymbol{\delta}^r=(\delta_0^r,\delta_1^r,\cdots,\delta_{p-1}^r)^t$. Then the vector terms of $Q(\boldsymbol{B})=Q(\boldsymbol{b},\boldsymbol{\delta}^l,\boldsymbol{\delta}^r)$ are given by

$$\begin{aligned}3(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b})^t(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b})-2(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b})^t(\boldsymbol{\eta}^l-\boldsymbol{X}\boldsymbol{\delta}^l)+(\boldsymbol{\eta}^l-\boldsymbol{X}\boldsymbol{\delta}^l)^t(\boldsymbol{\eta}^l-\boldsymbol{X}\boldsymbol{\delta}^l)\\ +2(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b})^t(\boldsymbol{\eta}^r-\boldsymbol{X}\boldsymbol{\delta}^r)+(\boldsymbol{\eta}^r-\boldsymbol{X}\boldsymbol{\delta}^r)^t(\boldsymbol{\eta}^r-\boldsymbol{X}\boldsymbol{\delta}^r).\end{aligned}$$

To find the values of $\boldsymbol{b}$, $\boldsymbol{\delta}^l$ and $\boldsymbol{\delta}^r$ in R^p that minimize the function Q, we set to zero the partial derivatives with respect to $\boldsymbol{b}$, $\boldsymbol{\delta}^l$ and $\boldsymbol{\delta}^r$. For this, first we obtain

$$\begin{aligned}\frac{\partial Q}{\partial \boldsymbol{b}}&=-6(\boldsymbol{X}^t\boldsymbol{y}-\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{b})-2(\boldsymbol{X}^t\boldsymbol{\eta}^l-\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\delta}^l)+2(\boldsymbol{X}^t\boldsymbol{\eta}^r-\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\delta}^r),\\ \frac{\partial Q}{\partial \boldsymbol{\delta}^l}&=\ 2(\boldsymbol{X}^t\boldsymbol{y}-\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{b})-2(\boldsymbol{X}^t\boldsymbol{\eta}^l-\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\delta}^l),\\ \frac{\partial Q}{\partial \boldsymbol{\delta}^r}&=-2(\boldsymbol{X}^t\boldsymbol{y}-\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{b})-2(\boldsymbol{X}^t\boldsymbol{\eta}^r-\boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\delta}^r)\end{aligned}$$

and let $\hat{\boldsymbol{b}}_n$, $\hat{\boldsymbol{\delta}}_n^l$ and $\hat{\boldsymbol{\delta}}_n^r$ denote the solution of the equations for $\boldsymbol{b}$, $\boldsymbol{\delta}^l$ and $\boldsymbol{\delta}^r$ when the derivatives are set equal to zero. We get the fuzzy least squares normal equations as

$$\begin{cases}\boldsymbol{X}^t\boldsymbol{X}\hat{\boldsymbol{b}_n}=\boldsymbol{X}^t\boldsymbol{y},\\ \boldsymbol{X}^t\boldsymbol{X}\hat{\boldsymbol{\delta}_n^l}=\boldsymbol{X}^t\boldsymbol{\eta}^l,\\ \boldsymbol{X}^t\boldsymbol{X}\hat{\boldsymbol{\delta}_n^r}=\boldsymbol{X}^t\boldsymbol{\eta}^r.\end{cases}\tag{6}$$

If $\operatorname{rank}(\boldsymbol{X})=p$, then the solution of (6) is given by the linear estimators

$$\begin{aligned}\hat{\boldsymbol{b}}_n&=(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y},\\ \hat{\boldsymbol{\delta}}_n^l&=(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{\delta}^r,\\ \hat{\boldsymbol{\delta}}_n^r&=(\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{\delta}^r.\end{aligned}\tag{7}$$

Note that if Y_i and B_j are crisp, i.e., $\eta_i^r=\eta_i^l=0$ for all i and $\delta_j^r=\delta_j^l=0$ for all j, then the normal equations (6) and its solution (7) are reduced to the ordinary results

$$\boldsymbol{X}^t\boldsymbol{X}\hat{\boldsymbol{b}}_n=\boldsymbol{X}^t\boldsymbol{y}$$

and

$$\hat{\boldsymbol{b}}_n = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

respectively.

NOTE. From the standard linear theory, the resulting least squares estimators in (7) enjoy many tractable finite sample optimality properties, e.g., the minimum variance linear unbiased estimators, etc. However, we need to restrict the coverage of the estimators $\hat{\boldsymbol{\delta}}_n^l$ and $\hat{\boldsymbol{\delta}}_n^r$ to some interval or region of $\hat{\boldsymbol{\delta}}_n^l \geq 0$ and $\hat{\boldsymbol{\delta}}_n^r \geq 0$ so that $\hat{\boldsymbol{b}}_n - \hat{\boldsymbol{\delta}}_n^l \leq \hat{\boldsymbol{b}}_n$ and $\hat{\boldsymbol{b}}_n + \hat{\boldsymbol{\delta}}_n^r \geq \hat{\boldsymbol{b}}_n$.

4 Interval estimation

In model (4), we now assume that the crisp random variables ϵ_i^m, ϵ_i^l, ϵ_i^r in error term Φ_i follow some distribution such that $E[\epsilon_i^m] = E[\epsilon_i^l] = E[\epsilon_i^r] = 0$ and $Var[\epsilon_i^m] = \sigma_m^2$, $Var[\epsilon_i^l] = \sigma_l^2$ and $Var[\epsilon_i^r] = \sigma_r^2$. Under these assumptions, (7) is the standard least squares estimator of (6) and the well-known Gauss-Markov Theorem states that for every fixed p-vector $\mathbf{c} \in R^p$, $\mathbf{c}^t\hat{\boldsymbol{b}}_n$, $\mathbf{c}^t\hat{\boldsymbol{\delta}}_n^l$ and $\mathbf{c}^t\hat{\boldsymbol{\delta}}_n^r$ are the best linear unbiased estimator (BLUE) of $\mathbf{c}^t\boldsymbol{b}_n$, $\mathbf{c}^t\boldsymbol{\delta}_n^l$ and $\mathbf{c}^t\boldsymbol{\delta}_n^r$, respectively, in the sense that they have the minimum variance in the entire class of linear unbiased estimators of $\mathbf{c}^t\boldsymbol{b}_n$, $\mathbf{c}^t\boldsymbol{\delta}_n^l$ and $\mathbf{c}^t\boldsymbol{\delta}_n^r$, respectively. Moreover, under mild regularity conditions it follow that $\hat{\boldsymbol{b}}_n$, $\hat{\boldsymbol{\delta}}_n^l$ and $\hat{\boldsymbol{\delta}}_n^r$ are weakly consistent of $\boldsymbol{b}$, $\boldsymbol{\delta}^l$ and $\boldsymbol{\delta}^r$, respectively, in fact they are strongly consistent too. Although these constitute important optimal properties of the FLSE, it is of little practical application unless we have some idea of the corresponding distributions. If we consider the additional assumption that the random errors in (4) are normally distributed, it follows that the FLSE $\hat{\boldsymbol{b}}_n$, $\hat{\boldsymbol{\delta}}_n^l$ and $\hat{\boldsymbol{\delta}}_n^r$ coincides with the MLE and that

$$\begin{aligned}
&\hat{\boldsymbol{b}}_n \sim N_p\,(\boldsymbol{b},\ \sigma_m^2(\boldsymbol{X}^t\boldsymbol{X})^{-1}),\\
&\hat{\boldsymbol{\delta}}_n^l \sim N_{p-1}(\boldsymbol{\delta}^l,\ \sigma_l^2(\boldsymbol{X}^t\boldsymbol{X})^{-1}),\\
&\hat{\boldsymbol{\delta}}_n^r \sim N_{p-1}(\boldsymbol{\delta}^r,\ \sigma_r^2(\boldsymbol{X}^t\boldsymbol{X})^{-1}).
\end{aligned} \tag{8}$$

Also, the confidence intervals and hypothesis tests for the fuzzy regression parameters, and prediction intervals for a new observation, based on known values of the fuzzy regression variables, can be applied using standard theory.

The least squares method gives no estimators for the error variances σ_l^2, σ_m^2 and σ_r^2, but the estimators based on the residuals of least squares are

$$\hat{\sigma}_l^2 = \frac{1}{n-p}\sum_{i=1}^{n}(\eta_i^l - \hat{\eta}_i^l)^2, \quad \hat{\sigma}_m^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \quad \hat{\sigma}_r^2 = \frac{1}{n-p}\sum_{i=1}^{n}(\eta_i^r - \hat{\eta}_i^r)^2 \tag{9}$$

which are unbiased estimators of σ_l^2, σ_m^2 and σ_r^2, respectively.

As in the crisp case, it is helpful to be able to extend a point estimate for a fuzzy parameter to an interval estimate so that its accuracy can be assessed. Now, we will consider three types confidence intervals: confidence interval on individual fuzzy regression parameters, confidence intervals on the fuzzy mean response, prediction intervals on future fuzzy observations.

First, we will derive the confidence intervals on individual fuzzy regression coefficients B_i. From the marginal distributions of (8), we have for $j = 0, 1, \cdots, p-1$

$$\hat{b}_i - \hat{\delta}_i^l \sim N(b_i - \delta_i^l, (\sigma_m^2 + \sigma_l^2)c_{ii}), \quad \hat{b}_i \sim N(b_i, \sigma_m^2 c_{ii}), \quad \hat{b}_i + \hat{\delta}_i^r \sim N(b_i + \delta_i^r, (\sigma_m^2 + \sigma_r^2)c_{ii}),$$

where c_{ii} is the $(i+1)$th diagonal element of the $(\boldsymbol{X}^t\boldsymbol{X})^{-1}$ matrix. Moreover, $\hat{\sigma}_m^2$, $\hat{\sigma}_l^2$ and $\hat{\sigma}_r^2$ are estimates of σ_m^2, σ_l^2 and σ_r^2 in (9), respectively, based on a χ^2 variable with common $(n-p)$ degrees of freedom, i.e., $(n-p)\hat{\sigma}_m^2/\sigma_m^2 \sim \chi^2(n-p)$, $(n-p)\hat{\sigma}_l^2/\sigma_l^2 \sim \chi^2(n-p)$ and $(n-p)\hat{\sigma}_r^2/\sigma_r^2 \sim \chi^2(n-p)$. Note that $\hat{b}_i$ and $\hat{\sigma}_m^2$, $\hat{\delta}_i^l$ and $\hat{\sigma}_l^2$, and $\hat{\delta}_i^r$ and $\hat{\sigma}_r^2$ are independent, respectively.

Now we assume that $\sigma_m^2 = \sigma_l^2 = \sigma_r^2 (= \sigma^2)$, and if

$$T_i^{(l)} = \frac{(\hat{b}_i - \hat{\delta}_i^l) - (b_i - \delta_i^l)}{\sqrt{\hat{\sigma}_m^2 + \hat{\sigma}_l^2}\sqrt{c_{ii}}}, \qquad T_i^{(m)} = \frac{\hat{b}_i - b_i}{\hat{\sigma}_m\sqrt{c_{ii}}}, \qquad T_i^{(r)} = \frac{(\hat{b}_i + \hat{\delta}_i^r) - (b_i + \delta_i^r)}{\sqrt{\hat{\sigma}_m^2 + \hat{\sigma}_r^2}\sqrt{c_{ii}}}$$

for $i = 0, 1, \cdots, p-1$, then the distributions of $T_i^{(l)}$, $T_i^{(m)}$ and $T_i^{(r)}$ are Student's t with $2(n-p)$, $n-p$ and $2(n-p)$ degrees of freedoms, respectively. In the case of $T_i^{(m)}$, for a fixed $i \in \{0, 1, \cdots, p-1\}$, let $t_{\frac{\alpha}{2(3)}}(n-p)$ be defined, as usual, by

$$P\left[T_i^{(m)} < t_{\frac{\alpha}{2(3)}}(n-p)\right] = 1 - \frac{\alpha}{2(3)}.$$

Therefore, using Bonferroni's inequality, a $100(1-\alpha)\%$ simultaneous confidence intervals for the mode b_i, the left and right spreads δ_i^l and δ_i^r of the regression coefficient $B_i = (b_i - \delta_i^l, b_i, b_i + \delta_i^r)$, $i = 0, 1, \cdots, p-1$, is

$$\begin{aligned}
(\hat{b}_i - \hat{\delta}_i^l) - t_{\frac{\alpha}{2(3)}}(2(n-p))\sqrt{\hat{\sigma}_m^2 + \hat{\sigma}_l^2}\sqrt{c_{ii}} \;&\leq \delta_i^l \leq (\hat{b}_i - \hat{\delta}_i^l) + t_{\frac{\alpha}{2(3)}}(2(n-p))\sqrt{\hat{\sigma}_m^2 + \hat{\sigma}_l^2}\sqrt{c_{ii}}, \\
\hat{b}_i - t_{\frac{\alpha}{2(3)}}(n-p)\hat{\sigma}_m\sqrt{c_{ii}} \;&\leq b_i \;\leq \hat{b}_i + t_{\frac{\alpha}{2(3)}}(n-p)\hat{\sigma}_m\sqrt{c_{ii}}, \\
(\hat{b}_i + \hat{\delta}_i^r) - t_{\frac{\alpha}{2(3)}}(2(n-p))\sqrt{\hat{\sigma}_m^2 + \hat{\sigma}_r^2}\sqrt{c_{ii}} \;&\leq \delta_i^r \leq (\hat{b}_i + \hat{\delta}_i^r) + t_{\frac{\alpha}{2(3)}}(2(n-p))\sqrt{\hat{\sigma}_m^2 + \hat{\sigma}_r^2}\sqrt{c_{ii}}.
\end{aligned} \tag{10}$$

5 Conclusions

In this paper, we have presented some results concerning statistical inference in the presence of vague data. We provided a confidence interval for a linear regression model based on the fuzzy least squares estimator which elucidates some of the statistical procedures for regression parameters, especially the three types of confidence regions for the fuzzy regression parameters, the fuzzy mean response and the fuzzy future responses. Analysis of such models is likely to be more complicated than that of the classical linear models. It will also be important to consider the problem of testing hypotheses concerning regression analysis in the presence of vague data which is going to be dealt with in our next research.

References

[1] B. F. Arnold, Statistical tests optimally meeting certain fuzzy requirements on the power function and on the sample size, *Fuzzy Sets and Systems*,75: 365-372, 1995.

[2] M. R.Casals, M. A. Gil, and P. Gil, A note on the operativeness of Neyman-Pearson tests with fuzzy information, *Fuzzy Sets and Systems*, 30:215-220, 1989.

[3] M. R. Casals, and P. Gil, Bayesian sequential test for fuzzy parametric hypotheses from fuzzy information, *Inform. Sci.* 80:283-298, 1994.

[4] Celminš, A. Least squares model fitting to fuzzy vector data, *Fuzzy Sets and Systems*, 22:245-269, 1987.

[5] Diamond, P. Fuzzy Least Squares, *Inform. Sci.* 46:141-157, 1988.

[6] Diamond, P. and Körner, R. Extended fuzzy linear models and least squares estimates, *Computers Math. Applic.*, 33:15-32, 1997.

[7] P. Grzegorzewski, Testing statistical hypotheses with vague data, *Fuzzy Sets and Systems*, 112:501-510, 2000.

[8] C. Kao and C. Chyu, A fuzzy linear regression model with better explanatory power, *Fuzzy Sets and Systems*, 126:401-409, 2002.

[9] H. K. Kim, J. H. Yoon, and Y. Li, Asymptotic properties of least squares estimation with fuzzy observations, *Inform. Sci.*, 178:439-451, 2008.

[10] R. Kruse, The strong law of large numbers for fuzzy random variables, *Inform. Sci.*, 28:233-241, 1982.

[11] H. Kwakernaak, Fuzzy random variables, Part I: definitions and theorems, *Inform. Sci.*, 15:1-19, 1978.

[12] H. Kwakernaak, Fuzzy random variables, Part II: algorithms and examples for the discrete case, *Inform. Sci.*, 17:253-278, 1978.

[13] W-J. Lee, H-Y. Jung, J. H. Yoon and S. H. Choi, The statistical inferences of fuzzy regression based on bootstrap Techniques, *Soft computing*, 19:883-890, 2015.

[14] M. Ming, M. Friedman. and A. Kandel, General fuzzy least squares, *Fuzzy Sets and Systems*, 88:107-118, 1997.

[15] W. Näther, Linear statistical inference for random fuzzy data, *Statistics*, 29:221-240, 1997.

[16] H. Tanaka, S. Uejima, and K. Asai, Linear regression analysis with fuzzy model, *IEEETrans. Syst., Man Cybernet.*, 12:903-907, 1982.

[17] H. Tanaka, Fuzzy data analysis by possibilistic linear models, *Fuzzy Sets and Systems*, 24:363-375, 1987.

[18] M. L. Puri, and D. A. Ralescu, Fuzzy Random Variables, *Jour. Math. Anal. Appl.*, 114:409-422, 1986.

[19] A. Wünsche and W. Näther, Least-squares fuzzy regression with fuzzy random variables, *Fuzzy Sets and Systems* 130:43-50, 2002.

[20] H. C. Wu, Statistical hypotheses testing for fuzzy data, *Inform. Sci.*, 175:30-56, 2005.

[21] H. C. Wu, Analysis of variance for fuzzy data, *Inter. Jour. of Syst. Sci.*, 38(3):235-246, 2007.

[22] J. H. Yoon, S. H. Choi and P. Grzegorzewski, On Asymptotic Properties of the Multiple Fuzzy Least Squares Estimator, *Soft Methods for Data Science*, 525-532, 2016.

Author index

Title: **Proceedings of the 20th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty**
CZECH-JAPAN SEMINAR 2017, Pardubice, Czech Republic, September 17–20, 2017

Editors: Vilém Novák, Masahiro Inuiguchi, Martin Štěpnička
Publisher: University of Ostrava, Dvořákova 7, 70103 Ostrava
Edition: First, 2017
Cover design: Tomáš Fabík

Typesetting: by authors using LaTeX
completion by Martin Dyba