PROCEEDINGS

OF THE

24th Czech-Japan Seminar on Data Analysis and Decision Making

Telč, Czech Republic, September 9-12, 2024

Proceedings of the 24th Czech-Japan Seminar on Data Analysis and Decision Making *(CJS 2024)* Telč, Czech Republic

Václav Kratochvíl, Masahiro Inuiguchi, Ondřej Čepek, Yoshifumi Kusunoki
(editors)

September 9-12, 2024

Published by:
ÚTIA - Institute of Information Theory and Automation,
Czech Academy of Sciences
Pod Vodárenskou věží 4, 182 00 Praha 8, Czech Republic

Printed in Prague — October 2024

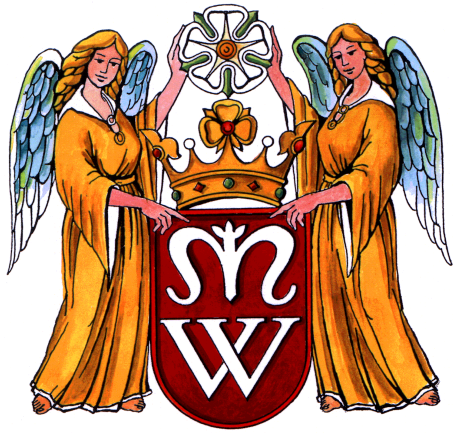
Organized by:
Institute of Information Theory and Automation, Czech Academy of Sciences

Sponsors:
Institute of Information Theory and Automation, Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University
Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Faculty of Management, Prague University of Economics and Business
Centre Telč, Institute of Theoretical and Applied Mechanics, Czech Academy of Sciences

Credits:
Editors: Václav Kratochvíl, Masahiro Inuiguchi, Ondřej Čepek, Yoshifumi Kusunoki
LaTeXeditor: Václav Kratochvíl
Cover design: Václav Kratochvíl
using LaTeX's 'confproc' package, version 0.8 by V. Verfaille

ISBN: 978-80-905688-0-8

24th CZECH-JAPAN SEMINAR ON DATA ANALYSIS AND DECISION MAKING



Telč coat of arms

Organized by:

Institute of Information Theory and Automation,
Czech Academy of Sciences

Telč

September 9-12, 2024

Programme and Conference Committee:
Radim Jiroušek - Chair, *Czech Academy of Sciences, Czechia*
Masahiro Inuiguchi - Chair, *Osaka University, Japan*
Ondřej Čepek, *Charles University, Czechia*
Yoshifumi Kusunoki, *Osaka Metropolitan University, Japan*
Shinji Mochida, *University of Marketing and Distribution Sciences, Japan*
Karel Mls, *University of Hradec Králové, Czechia*
Vilém Novák, *University of Ostrava, Czechia*
Irina Perfilieva, *University of Ostrava, Czechia*
Shao Chin Sung, *Aoyama Gakuin University, Japan*
Martin Štěpnička, *University of Ostrava, Czechia*
Milan Vlach, *Charles University, Czechia*

Organizing Committee:
Václav Kratochvíl - chair, *Czech Academy of Sciences, Prague*
Radim Jiroušek, *Czech Academy of Sciences, Prague*
Ondřej Čepek, *Charles University, Prague*

Foreword

The 24th edition of the seminar is now behind us. This year, the seminar took place in Telč, a beautiful Baroque town on the edge of the Bohemian-Moravian Highlands. We were delighted to welcome 26 participants, who, as the name of the seminar suggests, came from both the Czech Republic and Japan. All four days of the seminar were filled with engaging events and provided ample space for informal discussions on open topics related to mathematical decision-making and its associated theories.

Accommodation was provided by the Antoň Hotel, which offered lodging and excellent facilities with superb food, service, and a perfect setting for discussions. The seminar itself was held at the *Centre Telč*, part of the Institute of Theoretical and Applied Mechanics of the Czech Academy of Sciences, which generously offered its lecture room free of charge. We are particularly grateful to the center's director, Jakub Novotný, who not only arranged a guided tour for us but also allowed us a rare glimpse behind the scenes of this remarkable facility, including its unique wind tunnel, capable of simulating extreme weather conditions and their effects on the materials being studied.

In addition to the academic program, we had opportunities to refresh our minds in the open air. One afternoon, we took a hike to the nearby Roštejn Castle. Another day, we enjoyed Telč, a UNESCO World Heritage site recognized for its beauty and uniqueness.

The friendly atmosphere and the fantastic group of people who attended are reflected in the quality of this collection of papers, which were presented during the seminar. We hope you find them as inspiring and thought-provoking as we did.

We want to extend our heartfelt thanks to the seminar sponsors (listed above), whose support was essential to the event's success. Their contributions have allowed us to create a rich, stimulating environment for discussion and collaboration.

Enjoy your reading!

In Prague, September 5, 2024

Václav Kratochvíl, Radim Jiroušek

List of Content

Preliminary investigation of multimodal segmentation with shape metrics and key points for images of carotid atherosclerosis

Petr Čermák[1,3], Vladyslav Sherstobitov[2], David Školoudík[3], David Pakizer[3], and Jiří Kozel[3]

[1]Institute of Physics in Opava, Silesian University in Opava, Czechia
[2]Faculty of Applied Informatics, Tomas Bata University in Zlin, Czechia
[3]Faculty of Medicine, University of Ostrava, Czechia

Abstract

One of the greatest representatives of the diseases of civilization is stroke. Atherosclerosis, mainly atherosclerotic plaques cause this. The location and progression of atherosclerotic plaque in carotid arteries using image processing methods and artificial intelligence is the subject of this article. The determination of plaque location and border is provided by a residual learning neural network. Furthermore, multimodal segmentation and computation of object shape metrics along with key points are performed. This computation is performed for a stable and progressive plaque subset of the B-mode ultrasound echogenity images of the carotid artery. Each algorithm to compute shape metrics and key points is evaluated for its contribution to the classification between stable/progressive plaques. The contribution level is mapped into three groups. A neural network classification test is also performed.

Keywords: Carotid plaque, Computer analysis, Soft computing, Decision making, Image analysis, Shape metrics, Key points, Echogenity.

1 Introduction

The field of image data evaluation enables the analysis and prediction of civilization diseases. Stroke is one of the representatives of civilization's diseases and the main cause of morbidity and mortality [1]. These ischemic strokes are mainly caused by atherosclerosis, i.e., atherosclerotic plaques [2]. The atherosclerotic plaque in the carotid artery creates arterial stenosis, and as it grows, the risk of thrombosis or embolization of the brain

increases. Carotid plaque prone to rupture consists of a large lipid core or intraplaque hemorrhage (echolucent areas within the plaque on ultrasound) covered by a thin fibrous cap [3]. Various factors, including the accumulation of inflammatory cells and local hemodynamic factors, weaken the fibrotic cap, leading to plaque rupture and leaving behind ulceration or mural thrombus (Figure 1). These plaque features act as a thromboembolic source, allowing plaque components to be released into the blood and cause stroke [4].

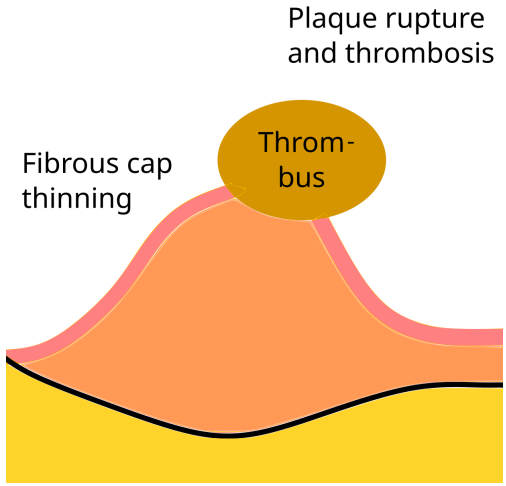


Figure 1: Visualization of unstable carotid plaque with a thin/fissured fibrous cap and mural thrombus.

To support decision making, it is necessary to characterize the properties of the plaque in terms of describing its echogenity (intensity), structural, textural, and shape properties. The structure will be essential as we will be looking for an area of possible instability/rupture of the plaque. With acquisition and evaluation are contacted problems namely

1. Different setup of the sonographic devices;
2. Perception and experience of the physician;
3. How to standardize acquisition and evaluation chain.

In terms of computed aided support for decision-making, the following main steps are taken

1. Localization ROI (Region of Interest);
2. Segmentation area of ROI;
3. Analyzing interesting segments in terms of echogenity, texture properties, and shape properties. Usually, we also analyze the relation between segments;

4. Make/calculate support decision, final decision made by a physician. In the process of making/calculating, we can also use other data about the patient(anamnesa, laboratory tests, etc.).

In the machine learning era, many methods are used [5], in this paper, the approaches of clustering - unsupervised learning, and at the same time neural network, i.e., supervised training, will be applied. Furthermore, the shape characteristics and key points will be investigated for part of an ANTIQUE database [8].

2 Detecting the position of the artery and determining its border

Detection of the position and identification of the artery border is a significant part of the calculation and analysis of carotid plaque. A deep learning network with ResNet50 residual learning is used to detect points at the artery border[6, 10]. The detected border points and point clouds were approximated by an ellipse due to the probe pressure on the carotid artery (Figure 2). The estimation and drawing of the ellipse was performed using the function cv2.fitEllipse(c) from the OpenCV library (Figure 2). Further developments could be made by switching the ellipsoid approximation to the spline approximation in the case of a sufficient number of correctly detected points on the carotid border. Image data selected from the ANTIQUE database [8].

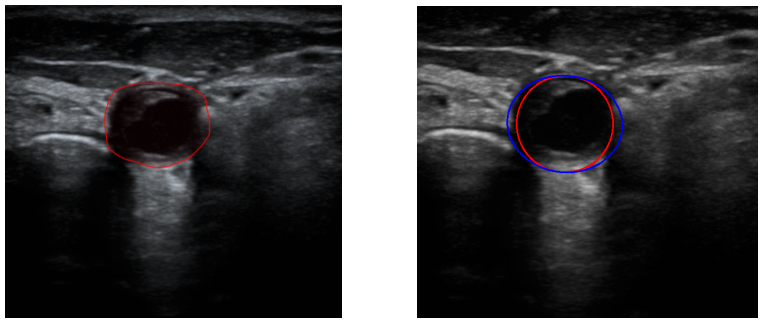


Figure 2: Example of labeling by a human physician and computer-aided SW[6]

3 Segmentation of the Carotide Artery

After detecting the position of the artery and determining its border, the segmenting of the carotid artery is done. Let us have an echogenity image from the B-Mode cut.

$$I_{i,j}, \tag{1}$$

where I is the echogenity image and i, j are the coordinates of the echogenity value stored.

$$SW_{i,j}, \tag{2}$$

where $SW_{i,j}$ is the sliding window as a submatrix of echogenity image I and its coordinates

$$\mathbf{Par} = \left\{ Par^{*}_{\frac{w}{2}+i,\frac{w}{2}+j} \right\}; i = 0, 1\cdot l, 2\cdot l, ..., n_x \cdot l; j = 0, 1\cdot l, 2\cdot l, ..., n_y \cdot l, \tag{3}$$

where **Par** is a set of parameter objects (numbers, matrices) calculated from sliding windows as input to the K-Mean algorithm. Indexes $\frac{w}{2}+i, \frac{w}{2}+j$ determine the center position of the sliding window in coordinates of the echogenity value in the image. The sliding window movement is denoted by l, C is there set of clusters and $K$is the preset number of clusters[14]

$$C = \{c_1, c_2, ..., c_K\}, \tag{4}$$

the objective function to minimize

$$J(C) = \sum_{k=1}^{K} J_k, \tag{5}$$

where $J_k$is the sum of squared error

$$J_k = \sum_{i=1}^{\|c_k\|} \left(\|\mathbf{x_{i,j}} - \mathbf{m}_k\| \right)^2. \tag{6}$$

As input of K-Means **Par** we can choose characteristic like DFT, DWL Coeficients. We can apply filtering of the coefficients using percentil of variance

$$Par^{*}_{i,j} \begin{cases} calc\ type = 1 & Filter\left(Par^{F}_{i,j}(k,l)\right) \\ calc\ type = 2 & \underset{s}{Filter}\left(Par^{W,s}_{i,j}(k,l)\right) \\ calc\ type = 3 & Par^{A}_{i,j} \\ calc\ type = 4 & Par^{D}_{i,j} \end{cases}. \tag{7}$$

DFT for sliding window as follows[14]

$$Par^{F}_{i,j}(k,l) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} SW_{i,j}(m,n)\, exp\left[-2\pi j \left(\frac{mk}{M} + \frac{nl}{N} \right) \right], \tag{8}$$

and DWT for given s [11]

$$Par^{W,s}_{i,j}(k,l) = \frac{1}{s} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} SW_{i,j}(m,n)\, \psi\left(\frac{m-k}{s} \right) \psi\left(\frac{n-l}{s} \right), \tag{9}$$

we also use average and variance

$$Par_{i,j}^{A} = \frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} SW_{i,j}(m,n)}{MN}, \tag{10}$$

,

$$Par_{i,j}^{D} = E\left[\left(Par_{i,j}^{A} - SW_{i,j}(m,n)\right)^{2}\right]. \tag{11}$$

.

As a clustering or classification tool, we can also use Neural Network

$$C = NN(\mathbf{Par}). \tag{12}$$

.

Those algorithms are implemented in software for the analysis of medical images [12] shown in Figure 3. The program was implemented in Python using several open-source libraries, namely OpenCV, pyWavelet, Numpy, Scipy, and Keras. During the analysis of the results, it was found that several iterations of segmentation are sometimes necessary to obtain meaningful and relevant results. It is required to change one or more parameters, such as the size of the sliding window, the number of steps the window goes through, or the number of clusters.

As an example of segmentation of the ultrasound image, we used sliding window techniques. The sliding window created a sub-matrix image for fast Fourier transform. We use these FFT coefficients to calculate the mean and variance. The filtering of SW the FFT coefficient is based on the percentile of the variance. K-means or artificial neural network classification is performed, and the clusters are colored (Figure 4).

$$C = \{c_1, c_2, ..., c_K\} \rightarrow \{RGB_1, RGB_2, ..., RGB_K\} \tag{13}$$

4 Shape metrics

We choose shape metrics based on the Euler characteristic as human perception [11, 14], curvature of the object border $K(t)$, and energy of the object border [14].

$$En = C - H \tag{14}$$

Euler characteristic is given by the difference in connected components and the number of holes[14].

Euler characteristic for Figure 3 is -1. The curvature function is defined [14]

$$K(t) = \frac{\theta(t)}{dt} = \frac{x'y'' - y'x''}{\left(x'^{2} + y'^{2}\right)^{\frac{3}{2}}}, \tag{15}$$

where $\theta(t)$ is definet follows

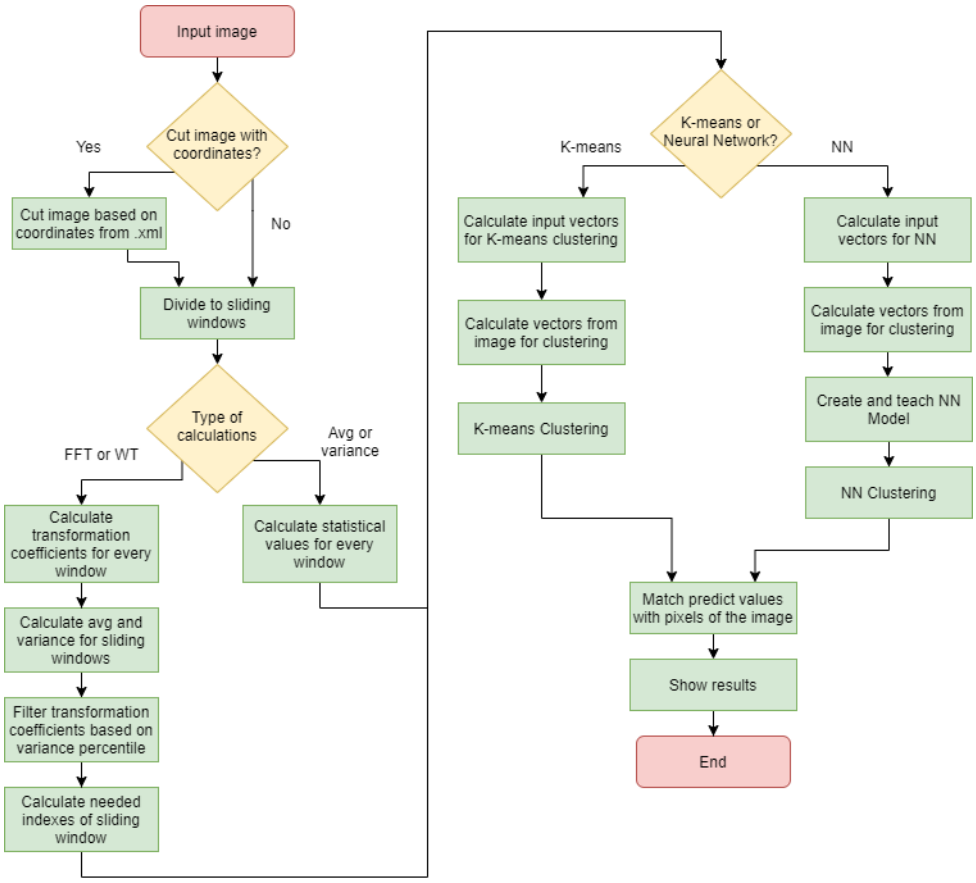


Figure 3: Algorithm of segmentation with sliding window [12]

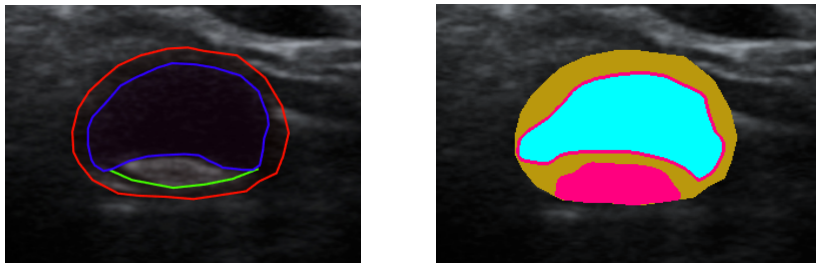


Figure 4: Labeling by physician, computer-segmentation[13, 20]

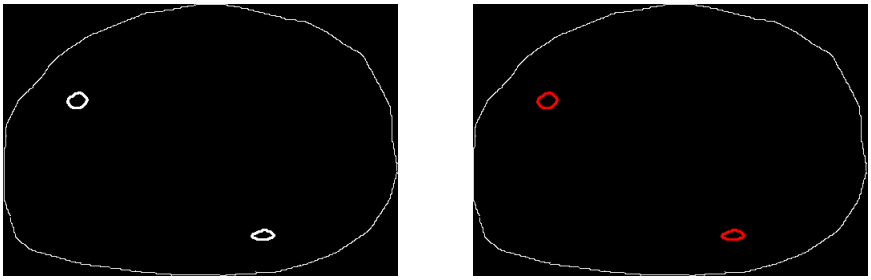


Figure 5: Finding holes

$$\theta(t) = arctan\frac{y(t) - y(t-w)}{x(t) - x(t-w)}, \tag{16}$$

finally border energy

$$Eb = \frac{1}{L}\int_0^L [K(u)]^2 \, du. \tag{17}$$

As a remark, $K(t)$ must be checked for gaps in curve approximation and using the neighboring connected algorithm. When we used bitmap images, there is a problem with discretization. Peeks in $K(t)$ must be filtered using a Gaussian filter. Three objects were extracted , and the border energy calculated, namely the carotid border, the carotid lumen, and the carotide segment from the progressive plaque in Figure 6.

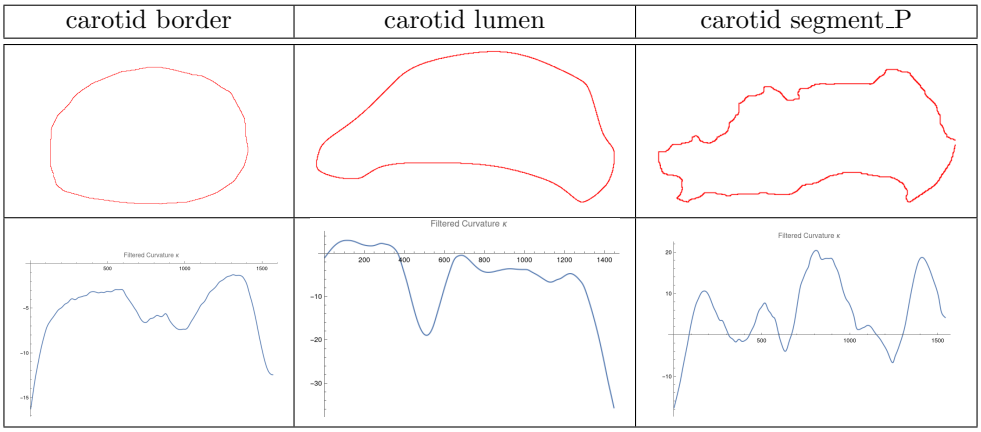


Figure 6: Segmented carotid borders and filtered curvature functions

For comparison, the functions of elipse, real and synthetic ulceration are also shown (Figure 7).

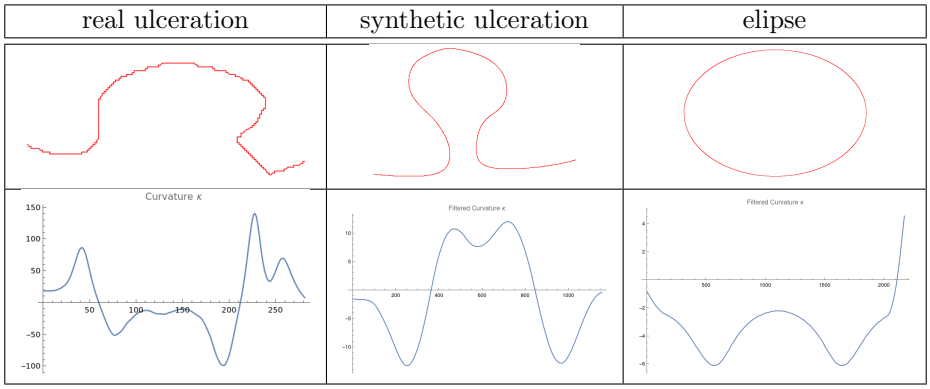


Figure 7: Borders of real ulceration, synthetic ulceration, and ideal elipse, their filtered curvature functions

It is clear that the circle has the lowest energy; the next elipse and carotid border. The highest value of border energy has the carotid segment of progressive plaque. The carotid lumen and ulceration plaque have near values. For comparison, we create a synthetic ulceration shape in the form of a mushroom. The energy of the curvature function could be used as a sign of complexity of the area, which could be relevant for the prediction of the rupture.

Object	Border energy (with Filtered $K(t)$)
elipse	15.2109
carotid border	35.255
carotid lumen	164.6
carotid segment_P	1610.21
carotid ulceration	118.903
synt. ulceration	70.6686

Table 1: Energy of different curvature functions filtered

The curvature function could also be used to predict possible rupture. Characteristic curvature is seen as synthetic ulceration and real ulceration. The curvature function can determine ROI with *rotation invariancy*. We can also use some type of curvature approximation, like Fourier coefficients, or classification using LSTM NN. However, the determination of how to create a representative template depends on the set of images.

5 Key Points

One of the techniques that allows for the identification of key points is OBR, which stands for Oriented FAST and Rotated BRIEF[19]. Compared to other widely used SIFT and

SURF algorithms, it has a distinct advantage in that it operates in real-time, is robust to changes in zoom, maintains accuracy even when the image is rotated, and is capable of analyzing fast amounts of data. To apply this method, a series of steps must be taken.

First, key points must be identified using oFAST, an enhanced version of the original FAST algorithm, which determines the direction of these points by ensuring their invariance. Thus, the initial step involves utilizing the FAST algorithm to detect potentially significant points and subsequently calculating a moment for each detected point based on the pixel intensity within a small region surrounding the point. This provides information regarding the primary orientation of the point, serving as a foundation for the *rotation invariant* local description of the points.

The second step in this process involves adapting the BRIEF descriptor to ensure rotational invariance. It involves creating a binary representation of the significant points by comparing the intensities of randomly placed pixel pairs around each point. This results in a concise and easily comparable descriptor. Subsequently, the orientation of this descriptor must be determined according to the predetermined orientation of the point that was established in the initial stage. This ensures that the descriptors remain comparable even if the images are subjected to rotation. In the following images, examples of the stable and progressive plaque and their processing are shown (Figure 8, 9).

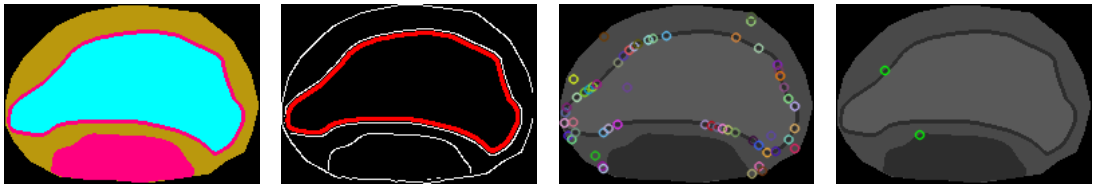


Figure 8: Dataset processing example, Stable plaque segmentation, detection hole, and SIFT, ORB methods[20]

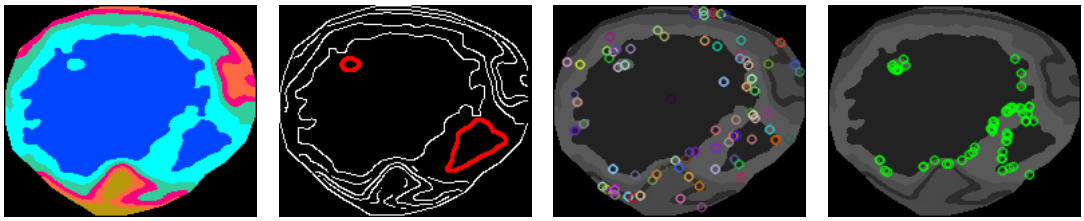


Figure 9: Dataset processing example, Progressive plaque segmentation, detection holes, and SIFT, ORB methods[20]

6 Shape metrics, Key points over stable and progressive plaque data subset

Let us calculate shape metrics and key points for

- Stable plaque data subset;
- Progressive plaque data subset;

on partial dataset from ANTIQUE study[8] results are presented in the Table 2 [20]. We classify the difference in characteristic average values according to stable and progressive plaques into three groups namely $\leq 10\%$, $10\% - 20\%$ and $\geq 20\%$. All other charectiristic which do not mentioned in this paper, we can found in [20].

Method/Algorithm	Category	Stable plaque subset	Progressive plaque subset
Perimeter	$\leq 10\%$	586	650
Circularity	$\leq 10\%$	0.429	0.442
Number of border points $K(p) > 0.5$	$\leq 10\%$	88	86.8
Centroid distance	$\leq 10\%$	9.31	9.57
Rectangularity	$\leq 10\%$	0.51	0.5
Elongation	$\leq 10\%$	1.151	1.2
FAST True	$10\% - 20\%$	72	88
Number of border points $K(p) = 0$	$10\% - 20\%$	315	353
Area	$10\% - 20\%$	66471	78050
Border energy (not filtered $K(t)$)	$\geq 20\%$	2556	3449
Euler characteristic	$\geq 20\%$	15.3	19.2
FAST False	$\geq 20\%$	562	670
SIFT	$\geq 20\%$	89	123
ORB	$\geq 20\%$	55	100

Table 2: Comparision of the results of Methods/Algorithms comparision with the stable plaque subset and the progressive plaque subset [20]

7 Conclusions

The research used a set of 151 segmented images of the carotid artery, some of which contained images of stable plaques, and others contained images of progressive plaques. First, there is the preprocessing of B-Mode carotid images to localize the position and carotide border. For shape descriptors, we need to provide segmentation of the carotid artery image. After pre-processing these images, 15 descriptors were used to help extract information on plaque structure. These were descriptors such as moments, perimeter,

circularity, number of border points, centroid distance, rectangularity, elongation, number of border points, area, border energy, Euler characteristic, two FAST versions, SIFT and ORB[20].

It was found that the complexity of the border, shape metrics, and key point computation have a greater influence on the ratio of parameter values between stable and progressing plaques. A threshold value of 20% was chosen. The ORB method had the highest difference of the average parameter values when comparing stable and progressing plaques. Through the use of deep learning algorithms, it was shown that the neural network tested can correctly classify 80% of medical images[20]. Although this is not a very high rate; this result can still be considered sufficiently valid for subsequent uses of the model in the laboratory.

Firstly, it is recommended to expand the database of medical images used for training and testing of the models. A total of 151 images were used in this research, which a relatively limited number for robust training of deep neural networks. Larger more data will ensure higher accuracy and reliability of the results[20]. Adding additional images to the dataset depends on the number of physicians, as labeling is a time-consuming process. It is also possible to use semi-automatic segmentation for gradually database extending, however, in questionable cases and images with artifacts, the cooperation of the physician is again required. The semi-automatic segmentation could be a other subtask, seems pretty important.

References

[1] Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 2006;367:1747–1757.

[2] KALVACH, Pavel. Mozkové ischemie a hemoragie. 3., přeprac. a dopl. vyd. Praha: Grada, 2010. ISBN 978-80-247-2765-3.

[3] Mathiesen EB, Bønaa KH, Joakimsen O. Echolucent plaques are associated with high risk of ischemic cerebrovascular events in carotid stenosis: the Tromsø study. Circulation. 2001;103(17):2171-2175. doi:10.1161/01.cir.103.17.2171

[4] Yuan J, Usman A, Das T, Patterson AJ, Gillard JH, Graves MJ. Imaging Carotid Atherosclerosis Plaque Ulceration: Comparison of Advanced Imaging Modalities and Recent Developments. AJNR Am J Neuroradiol. 2017;38(4):664-671. doi:10.3174/ajnr.A5026

[5] Cau R, Pisu F, Muscogiuri G, Mannelli L, Suri JS, Saba L. Applications of artificial intelligence-based models in vulnerable carotid plaque. Vessel Plus 2023;7:20. http://dx.doi.org/10.20517/2574-1209.2023.78.

[6] HRABAL, Patrik. Image Preprocessing for Detection of Atherosclerotic Plaques. Bakalářská práce. Olomouc: Moravská vysoká škola Olomouc, 2021.

[7] OPENCV. Structural Analysis and Shape Descriptors. Online. 2024, Sat Aug 31 2024. Available from: https://docs.opencv.org/3.4/d3/dc0/group__imgproc__shape.html#gaf259efaad93098103d6c27b9e4900ffa. [cit. 2024-09-01].

[8] ŠKOLOUDÍK, D., "Atherosclerotic plaque characteristics associated with a progression rate of the plaque in carotids and a risk of stroke.," (Feb. 2015). Clinical trial NCT02360137.

[9] HANSON, G., LIBBY, P. The immune response in atherosclerosis: a double-edged sword. Nat Rev Immunol 6, 508–519 (2006). https://doi.org/10.1038/nri1882

[10] HE, Kaiming, Xiangyu ZHANG, Shaoqing REN a Jian SUN. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs] [online]. 2015 [cit. 27.2.2021]. Available from: http://arxiv.org/abs/1512.03385

[11] SOJKA, Eduard, 2000. Digitální zpracování a analýza obrazů. Ostrava: VŠB-Technická univerzita. ISBN 80-7078-746-5.

[12] SHERSTOBITOV, Vladyslav. Texture-based segmentation with classification. Bakalářská práce. Olomouc: Moravská vysoká škola Olomouc, 2021.

[13] CVUT FBMI. Online. 20.08.2024. Available from: https://gdiag.fbmi.cvut.cz. [cit. 2024-08-20].

[14] HANG, Dengsheng. Fundamentals of image data mining: analysis, features, classification and retrieval. Second edition. Cham: Springer, [2021]. Texts in computer science. ISBN 978-3-030-69250-6.

[15] EHSAN, Shoaib, Adrian CLARK, Naveed REHMAN a Klaus MCDONALD-MAIER, 2015. Integral Images: Efficient Algorithms for Their Computation and Storage in Resource-Constrained Embedded Vision Systems. Sensors [online]. 15(7), 16804-16830 [cit. 2024-04-10]. ISSN 1424-8220. Available from: doi:10.3390/s150716804

[16] JOHRI, Amer M., Vijay NAMBI, Tasneem Z. NAQVI, Steven B. FEINSTEIN, Esther S.H. KIM, Margaret M. PARK, Harald BECHER a Henrik SILLESEN, 2020. Recommendations for the Assessment of Carotid Arterial Plaque by Ultrasound for the Characterization of Atherosclerosis and Evaluation of Cardiovascular Risk: From the American Society of Echocardiography. Journal of the American Society of Echocardiography [online]. 33(8), 917-933 [cit. 2024-08-14]. ISSN 08947317. Available from: doi:10.1016/j.echo.2020.04.021

[17] SABA, L., C. YUAN, T.S. HATSUKAMI, et al., 2018. Carotid Artery Wall Imaging: Perspective and Guidelines from the ASNR Vessel Wall Imaging Study Group and Expert Consensus Recommendations of the American Society of Neuroradiology. American Journalof Neuroradiology [online]. 2018-02-13, 39(2), 9-31 [cit. 2024-08-14]. ISSN 0195-6108. Dostupné z: doi:10.3174/ajnr.A5488

[18] SABA, Luca, Tobias SAAM, H Rolf JÄGER, et al., 2019. Imaging biomarkers of vulnerable carotid plaques for stroke risk prediction and their potential clinical implications. The Lancet Neurology [online]. 18(6), 559-572 [cit. 2024-08-14]. ISSN 14744422. Available from: doi:10.1016/S1474-4422(19)30035-3

[19] RUBLEE, Ethan, Vincent RABAUD, Kurt KONOLIGE a Gary BRADSKI, 2011. ORB: An efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision [online]. IEEE, s. 2564-2571 [cit. 2024-08-31]. ISBN 978-1-4577-1102-2. Available from: doi:10.1109/ICCV.2011.6126544

[20] SHERSTOBITOV, Vladyslav. Carotid Plaque Image Structure Analysis Using Descriptors and Deep Learning. Diplomová práce. Zlín: Univerzita Tomáše Bati ve Zlíně, Fakulta aplikované informatiky., 2024.

On cardinalities of different degrees of Belief functions conjunctive conflictness

Milan Daniel[1] and Václav Kratochvíl[2]

[1]Institute of Computer Sciences, Czech Academy of Sciences, Prague, Czech Republic
milan.daniel@cs.cas.cz
[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic
velorex@utia.cas.cz

Abstract

This paper examines mutual conflict behavior between belief function structures across different discernment frame sizes (Ω). Through experiments on Ω_2 to Ω_6, we observe that as frame size increases, non-conflicting pairs and higher-order hidden conflicts become exceedingly relatively rare despite of exponential grows of cardinalities of their classes. The super-exponential growth of possible belief structures complicates exhaustive analysis, leading us to employ random sampling. Our findings reveal that cardinality of class of first-degree hidden conflicts (HC_1) grows faster than cardinality of non-conflicts as frame size increases, highlighting the challenges and implications for applying belief function theory in complex decision-making scenarios.

1 Introduction

The theory of belief functions was developed to better express uncertainty in information, extending beyond traditional methods such as second-order probability, which represents the probability of a probability—or, more precisely, our confidence in the likelihood of a single phenomenon.

In real life, we constantly deal with uncertainty. Consider, for instance, the weather forecast many of us check on our smartphones. You may have noticed that different websites or applications often provide varying predictions. These discrepancies arise from the different models used and the data available to those models, resulting in different sources of information that are typically inconsistent with one another. When faced with this situation, how do we decide which prediction to trust? Instead of choosing just one,

belief functions allow us to combine all available sources of information using rules such as Dempster's rule or its non-normalized version, often referred to as the Conjunctive rule.

While combining conflicting information using these methods is possible, there are well-known examples where such combinations lead to paradoxical or meaningless results. A famous example is the combination of medical diagnoses, such as cancer and the flu, where conflicting evidence can produce highly counter-intuitive outcomes, such as assigning an unreasonably high degree of belief to an almost impossible event. This highlights the importance of quantifying the degree of inconsistency—often referred to as the magnitude of conflict—between belief functions.

The simplest definition of conflict comes directly from the Conjunctive rule, where the conflict is quantified by the probability mass assigned to the empty set by the combination rule. The critical difference between the Conjunctive rule and Dempster's rule lies in how they handle this mass: the Conjunctive rule retains it in the empty set, whereas Dempster's rule proportionally redistributes it among all non-empty sets in the resulting combination.

This definition of conflict is highly dependent on the structure of the belief functions involved. The existence of conflict depends on the structure, and its magnitude is influenced by the size of the assigned probability masses. In this article, we focus on the structures of belief functions and their impact on conflict. Specifically, we explore the likelihood that two random structures will generate a conflict, the nature of higher-level hidden conflicts, the numbers of conflicting and non-conflicting pairs, and the technical feasibility of examining these scenarios. These topics form the core of the following discussion.

2 Basic Notions

This section will recall some basic notations needed in this paper.

Assume a finite frame of discernment Ω with elements denoted usually by ω_i, i.e., $\{\omega_1, \omega_2, \ldots, \omega_n\}$ and their sets by capital letters. In the case of $|\Omega| = n$, we will highlight this fact using a subscript as Ω_n. $\mathcal{P}(\Omega) = \{X | X \subseteq \Omega\}$ is a *power-set* of Ω. $\mathcal{P}(\Omega)$ is often denoted also by 2^Ω, e.g., in Pichon et al. (2019).

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. We sometimes speak about m as of a mass function.

There are other equivalent representations of m: A *belief function (BF)* is a mapping $Bel : \mathcal{P}(\Omega) \longrightarrow [0, 1]$, $Bel(A) = \sum_{\emptyset \neq X \subseteq A} m(X)$. Because there is a unique correspondence between m and corresponding Bel we often speak about m as of a belief function.

Let m be a belief function defined on Ω and $A \subseteq \Omega$. If $m(A) > 0$ we say A is a *focal element* of m. The set of focal elements is denoted by $\mathcal{F}_m$ (or simply $\mathcal{F}$ for short), and we call it a *structure* of m. We say that a focal element $X \in \mathcal{F}$ is proper if $X \neq \Omega$. In the case of $m_{vac}(\Omega) = 1$ we speak about the *vacuous BF* (VBF) and about *a non-vacuous BF* otherwise. If all focal elements have a non-empty intersection, we speak about *consistent*

BF. If focal elements are nested, we speak about consonant BF.

The (non-normalized) conjunctive rule of combination $\circledcirc$, see e.g. Smets (2005), is defined by:

$$(m_1 \circledcirc m_2)(A) = \sum_{X \cap Y = A; X, Y \subseteq \Omega} m_1(X) m_2(Y)$$

for any $A \subseteq \Omega$. $\kappa = \sum_{X \cap Y = \emptyset; X, Y \subseteq \Omega} m_1(X) m_2(Y)$ is usually considered to represent a conflict of respective belief functions when $\kappa > 0$. By normalization of $m_{12} = m_1 \circledcirc m_2$ we obtain Dempster's rule $\oplus$, see Shafer (1976). To simplify formulas, we often use $\circledcirc_1^3 m = m \circledcirc m \circledcirc m$, and also $\circledcirc_1^k (m_1 \circledcirc m_2) = (m_1 \circledcirc m_2) \circledcirc \ldots \circledcirc (m_1 \circledcirc m_2)$, where $(m_1 \circledcirc m_2)$ is repeated k-times.

3 Hidden conflict

Let us assume conjunctively non-conflicting belief functions m_1 and m_2, i.e., $(m_1 \circledcirc m_2)(\emptyset) = m_{12}(\emptyset) = 0$. In the case that there exists $k \geq 1$ such that $(\circledcirc_1^{k+1} m_{12})(\emptyset) > 0$, then we say that there is a hidden conflict of degree k between m_1 and m_2. Note that k is the smallest with this property. We can formalize this in the following definition.

Definition 1 *Assume two BFs* m_1 *and* m_2 *such that for some* $k > 0$ $(\circledcirc_1^k (m_1 \circledcirc m_2))(\emptyset) = 0$. *If there further holds* $(\circledcirc_1^{k+1} (m_1 \circledcirc m_2))(\emptyset) > 0$ *we say that there is a* conflict of BFs m_1 *and* m_2 hidden in the k-th degree *(hidden conflict of k-th degree, abbreviated as* HC_k*). If there is already* $(\circledcirc_1^{k+1} (m_1 \circledcirc m_2))(\emptyset) = (m_1 \circledcirc m_2))(\emptyset) > 0$ *for* $k = 0$ *then there is a conflict of respective BFs which is not hidden or we can say that it is* conflict hidden in degree zero *(*HC_0*).*

Arnaud Martin called $(\circledcirc_1^k m)(\emptyset)$ *auto-conflict* of of k-th order of m in Osswald and Martin (2006). Thus conflict of m_1 and m_2 hidden in the k-th degree is auto-conflict of m_{12} hidden in k-th degree, specially positive a_{k+1} of combined $m_{12} = m_1 \circledcirc m_2$ hidden by zero $a_k(m_{12})$ (i.e., $a_{k+1}(m_{12}) > 0$ where $a_k(m_{12}) = 0$), see also our contribution in CJS'17 Daniel and Kratochvíl (2017).

Hidden conflict and its degrees are extensions of the classic Shafer's definition of conflict. It is not an alternative definition or approach but a more detailed classification of situations where $m(\emptyset) = 0$.

This technical definition defines all degrees of hidden conflict using repeated combinations[1]. However, the original observation of hidden conflict came from the analysis of conflict *Conf*, defined in Daniel (2014), and its comparison with conjunctive conflict in situations where $Conf(m_1, m_2) > 0$ while $(m_1 \circledcirc m_2)(\emptyset) = 0$. For better insight into

[1]Repeated combination is only a technical tool here: we are interested only in m_1, m_2, and m_{12}, not in bbms of any repeated combination either of them, with the exceptipon of bbm assigned to empty set by the non-normalized conjunctive rule $\circledcirc$, i.e., bbm which is normalized when Dempster's rule is applied. Regardles of this, considering two or more numerically same BFs does not mean anything about their depenency, we can assume that both / all of them come from indepenent sources.

this, we refer to the Introductory and Little Angel examples published in Daniel and Kratochvíl (2020), for brief presentation of these examples see Appendix 1.

From a large amount of results about hidden conflicts and their degrees, we recall the following principal theorem:

Theorem 1 *Hidden conflict of non-vacuous BFs on $\Omega_n, n > 1$ is always of a degree less or equal to $n-2$; i.e., the condition*

$$(\textcircled{\cap}_1^{n-1}(m_1 \textcircled{\cap} m_2))(\emptyset) = 0 \tag{1}$$

always means full non-conflictness of respective BFs, and no hidden conflict exists.

For more detail on the limitation of the degree of hidden conflicts in characteristic situations, see Daniel and Kratochvíl (2020) and for comparison with related Pichon's approach, see Daniel and Kratochvíl (2022).

Analogously to degrees of hidden conflicts, degrees of non-conflictness were defined in Daniel and Kratochvíl (2019). Analogously, to distinguishing internal conflict(s) of individual BFs from mutual conflict between them, also internal hidden conflicts are defined, and mutual hidden conflicts distinguished Daniel and Kratochvíl (2020), internal hidden conflict was presented for the first time in CJS 2017 in Pardubice by Daniel and Kratochvíl (2017) in fact. Considering this, a hidden conflict of two BFs is hidden internal conflict of their combination.

Preparing the actual presentation of hidden conflict we have the following observed:
1. Consistent BFs have no internal conflict nor hidden internal conflict, i.e., their auto-conflict any order is always equal to zero.
2. Non-consistent BFs always have some internal conflict, either hidden or non-hidden, i.e., there is always positive auto-conflict of some order.
3. BFs m_1 and m_2 with consistent $m_1 \textcircled{\cap} m_2$ are in no conflict nor hidden conflict of any degree. (Such m_is are always consistent itselves.)
4. BFs m_1 and m_2 with non-consistent $m_1 \textcircled{\cap} m_2$ are always in hidden conflict of some degree greater or equal to zero. (regardless whether m_1 and/or m_2 are/is consistent).

Lemma 1 *(i) Two BFs m_1 and m_2 are in a hidden conflict of a positive degree whenever $(m_1 \textcircled{\cap} m_2)(\emptyset) = 0$ and $m_1 \textcircled{\cap} m_2$ is not consistent.*
(ii) Specially, hidden conflict of the first degree appears whenever $(m_1 \textcircled{\cap} m_2)(\emptyset) = 0$ and $m_1 \textcircled{\cap} m_2$ has positive auto-conflict (of the second order: $a(m_1 \textcircled{\cap} m_2) = a_2(m_1 \textcircled{\cap} m_2) > 0$).

Proof. Proofs follow Theorem 5 from Daniel and Kratochvíl (2022) and the previous observations.

4 Conflict analysis

In our previous work Daniel and Kratochvíl (2022), we explored the concept of conjunctive conflict, specifically focusing on the amount of probabilistic mass that the non-normalized conjunctive rule assigns to the empty set. We found that the existence or absence of

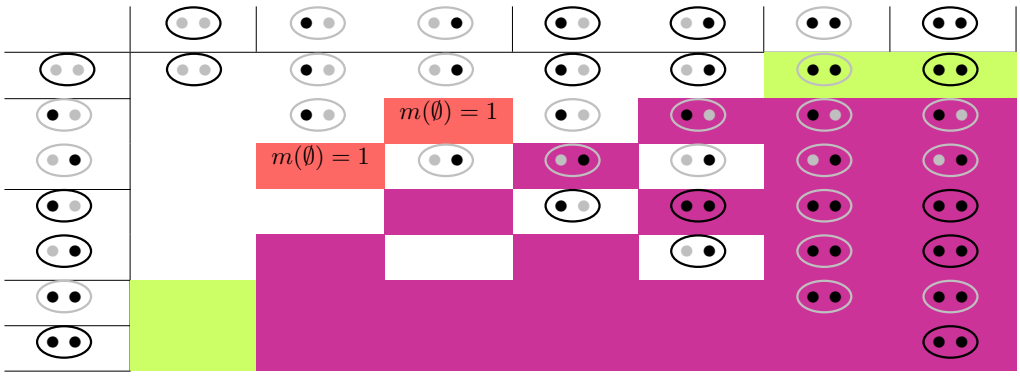


Table 1: All possible conjunctive combinations of belief structures and classes of non-conflictness on Ω_2: White – non-conflict, Green – hidden conflict, Red+Magenta – conflict. The table presents all possible combinations of belief function structures for Ω_2. The different structures are represented along the x and y axes. For Ω_2, there are two singletons and one set of cardinality 2, which represents the entire frame of discernment Ω_2. In the table, sets that are included in a particular structure are shaded black, while those not included are shaded grey.

conjunctive conflict is determined solely by the belief functions' structure. In contrast, the magnitude of the conflict depends on the individual probability masses. However, our primary interest lies in conflict's mere existence or non-existence rather than its magnitude in this study. Therefore, we concentrated on analysing the structures of belief functions.

Given the size of the frame of discernment Ω, we can enumerate the number of unique belief function structures. This enables us to calculate all possible combinations of these structures and determine how many are conflicting, hidden conflicting, or non-conflicting. In our previous study Daniel and Kratochvíl (2022), we performed this analysis for the three smallest frames, where $|\Omega| = 2, 3, 4$. The complete set of structure combinations for $|\Omega| = 2$ is illustrated in Table 1 and for $|\Omega_3| = 3$ by bitmap in Appendix 2, while the counts of conflicting and non-conflicting structures for different cardinalities are summarized in Table 2.

	NC	C i.e., HC_0	HC_1	HC_2	HC_3
Ω_2	17	28	4	–	–
Ω_3	649	14720	756	4	–
Ω_4	258.785	1.071.676.416	1.738.492	2.592	4

Table 2: Number of conflicting belief structure couples in different degrees hiddeness of conflictness/non-conflictness

We can immediately see that the most frequent cases are not hidden conflicts (C, i.e.,

hidden in degree zero HC_0). There are always four singular cases where hidden conflict of degree $n-1$ (HC_{n-1}) appears for Ω_n. Cardinalities of all other classes rapidly increase with the cardinality of the frame of discernment n.

For these small frames, it is interesting to note that the cardinality of HC_1, and thus the class of hidden conflicts in general, grows significantly faster than the cardinality of the non-conflict class (NC). There are significantly fewer hidden conflicts on Ω_2, but already more on Ω_3, and significantly more hidden conflicts than non-conflicts on Ω_4.

The cardinality of HC_2 is less than that of NC on Ω_3 and Ω_4. Nevertheless, it also grows quicker than the cardinality of NC: it is about 160 times less on Ω_3, while only about 100 times less on Ω_4.

5 Random sampling approach

As we attempt to extend our analysis to higher dimensions, we encounter a significant computational challenge due to the super-exponential growth in the number of structures. For example, with Ω_5, there are 31 possible focal elements, resulting in 2^{31} possible structures and 2^{62} combinations of these structures. The sheer magnitude of these numbers makes it infeasible to compute all possible combinations using current technology, and the problem only worsens with Ω_6 and beyond.

One potential solution is to employ random sampling. By selecting a sufficiently large random sample of structure combinations, we can estimate the distribution of different classes of conflicts, including different degrees of hidden conflict.

First, we validated our approach by performing random generation for Ω_2, Ω_3, and Ω_4 to ensure its accuracy. For these cases, we converted the results from Table 2 into percentages. Table 3 compares our randomly sampled results and the original results obtained from an exhaustive search of all possible combinations of the listed structures. The first three rows of Table 3 represent the exact results from the complete search, while the next three rows show the outcomes based on random sampling.

	NC	C	HC_1	HC_2	HC_3
Ω_2	34.7	57.1	8.2	–	–
Ω_3	4.02	91.26	4.69	0.025	–
Ω_4	0.024	99.814	0.162	0.00024	0.000000373
Ω_2 sampling	35.41	56.22	8.37	–	–
Ω_3 sampling	4.09	91.17	4.72	0.025	–
Ω_4 sampling	0.023	99.817	0.159	0.001	0.000

Table 3: Percentage representation of conflict levels across all possible combinations of Belief function structures for a given frame of discernment

The generation of random structures was carried out in two steps. First, we determined the number of focal elements for the generated structure. Then, we randomly selected the corresponding number of distinct subsets from Ω, forming the desired structure. The number of focal elements was generated such that the probability of selecting a given

number corresponded to the frequency distribution of focal elements among all possible structures on Ω. As is well-known from combinatorics, the most probable number of focal elements is approximately $2^{|\Omega|}/2$, which aligns with the highest number of possible combinations of subsets. Due to applying this two-step generation, we validated it as presented in Table 3.

Encouraged by our initial success, we proceeded to experiment with the Ω_5 frame. Given the vast number of possible combinations of the structures involved, we conducted eight separate experiments, each consisting of 100 million random combinations. We divided these into eight batches to assess the consistency of the intermediate results. The results from these individual experiments are remarkably consistent and are summarized in Table 4.

no.	NC	C	HC_1	HC_2	HC_3	HC_4
1	1	99999968	31	0	0	0
2	0	99999969	31	0	0	0
3	1	99999960	39	0	0	0
4	0	99999961	39	0	0	0
5	0	99999963	37	0	0	0
6	1	99999950	49	0	0	0
7	0	99999953	47	0	0	0
8	0	99999965	35	0	0	0

Table 4: Results of random sampling for Ω_5

For Ω_6, the results were even more compelling. Given the enormous number of possible structures and their combinations, we conducted 4 billion trials, organized into 160 sets of 25 million pairs each. Despite the extensive sampling, we did not find a single non-conflicting pair or a pair with a hidden conflict - summarized in Table 5. In other words, all pairs exhibited conflict. This suggests that the probability of encountering a pair/couple of non-conflicting belief functions for a larger frame of discernment is exceedingly close to zero, and the same holds for couples in hidden conflict of various degrees.

	NC	C	HC_1	HC_2	HC_3	HC_4	HC_5
Ω_6	0	4.000.000.000	0	0	0	0	0

Table 5: Results of random sampling for Ω_6

6 Summary and Results Analysis

Presenting the results of our experiments on Ω_5, we can confirm our observation that the cardinality of HC_1 grows much faster than NC, already reaching about a hundred times greater in this case. Unfortunately, the 100 million generated samples were insufficient to obtain a HC_i case for $i > 1$.

	NC	C	HC_1	HC_2	HC_3	HC_4	HC_5
Ω_2	17	28	4	–	–	–	–
Ω_3	649	14 720	756	4	–	–	–
Ω_4	258 785	1 071 676 416	1 738 492	2 592	4	–	–
Ω_5	1.725e10	4.611e18	1.771e12	(>> 3e3)	(>> 4)	4	–
Ω_6	?	8.507e37	?	?	?	?	4

Table 6: Numbers of conflicting belief structure couples in different degrees of conflictness/non-conflictness. There are precise number for Ω_2 – Ω_4, and estimation for entire space for Ω_5 and Ω_6

Since we know that the cardinality of HC_{n-1} is equal to 4 and that all HC_i values are greater for $i > n-1$, we have marked 0^+ in the cases where no samples were generated, although they could theoretically exist. Similarly, for Ω_6, even 25 million generated pairs were not enough to encounter anything other than a conflict that is not hidden.

Analyzing our results, it is clear that all classes of conflict/non-conflict increase with n. The largest class is always $C \sim HC_0$: the class of pairs with a classic unhidden conjunctive conflict. The cardinality of HC_i decreases from a maximum at $i = 0$ down to 4 for HC_{n-1}. The second-largest class, HC_1, is greater than NC (the class of non-conflicting pairs) for $n \geq 3$ and grows faster than NC as n increases.

As we are interesting only in the belief structures, we have sizes of no conflicts nor hidden conflicts here. Nevertheless, we should notice one exception which is structural: i.e., *full conflict* where all focal elements of m_1 have empty intersection with all focal elements of m_2, there appears $m_{12}(\emptyset) = 0$, the case where the conjunctive rule gives no other information and Dempster's rule is not applicable. There are two such cases on Ω_2, see red fields in Table 1, 36 on Ω_3, and 1 154 on Ω_4. This class also grows with the size of frame, but we can see that its cardinality incomparably less with the class of all conflicts HC_0 and also less than cardinalities of NC, HC_1 and on Ω_4 less than HC_2. I.e., cardinality of FC is less than cardinalities all classes investigated here, with the exception of HC_{n-1} which is always equal to 4 in any frame.

	NC	C	HC_1	HC_2	HC_3	HC_4	HC_5
Ω_2	34.7	57.1	8.2	–	–	–	–
Ω_3	4.02	91.26	4.69	0.025	–	–	–
Ω_4	0.024	99.814	0.162	0.00024	3.73e10	–	–
Ω_5	3.75e−9	100^-	3.85e−7	0^+	0^+	8.67e−15	–
Ω_6	0^+	100^-	0^+	0^+	0^+	0^+	4.70e−36

Table 7: Actual (for Ω_2 – Ω_4) and estimated (Ω_5 – Ω_6) percentages of conflict types among all possible combinations of structures for a given frame of discernment size

As the actual cardinalities of these rapidly growing classes are difficult to conceptualize, it is more intuitive to compare the percentages of belief structure pairs in particular conflict classes (see Table 7) or in direct comparison of the classes: for comparison of HC_i

classes with class of non-conflicting pairs and of class fully conflicting pairs with the other classes see Table 8. We can see rapid increase relative comparison of HC_1/N with the size of the frame, while decrease of any comparison of FC, especially with the class of all conflicting pairs C.

Since we did not observe any HC_i situations for $i > 1$ within our random generation for Ω_5 and Ω_6, but we know such situations exist, we have marked the corresponding fields in Table 7 for Ω_5 and Ω_6 with 0^+.

	HC_1/NC	HC_2/NC	HC_3/NC	FC / NC	FC / C	FC/HC_1	FC/HC_2
Ω_2	0.235	–	–	1.176e-1	7.143e-2	5.000e-1	–
Ω_3	1.166	0.006	–	5.547e-2	2.446e-3	4.762e-2	9.000e0
Ω_4	6.718	0.010	1.545e-5	4.459e-3	1.077e-6	6.638e-4	4.452e-1
Ω_5	102.666	(?)	(?)	(?)	(?)	(?)	(?)

Table 8: Increasing of relative frequencies of HC_i/NC with size of frame (columns 1–3). Relative frequencies of full conflict in comparison with conflicting/non-conflicting classes decreasing with size of frame (columns 4–7)

7 Conclusion

In this study, we investigated the behavior of conflict of couples of belief function structures, particularly focusing on the probability and distribution of conflicts across discernment frames Ω of different sizes. Through extensive experiments on Ω_2 to Ω_6, we observed that the cardinality of **all** conflict classes, particularly non-hidden conflicts (C), increases significantly with the size of Ω. Notably, our experiments on Ω_5 and Ω_6 revealed that despite the growing of cardinalities of all degrees of hidden conflicts, that even non-conflicting pairs and pairs of small positive higher-order degree of hidden conflicts ($i = 1, 2$) are exceedingly rare; confirming the hypothesis that the probability of encountering such pairs is almost negligible as the frame size grows and completely general[2] belief functions are considered.

The results underscore the computational challenges posed by the super-exponential growth in the number of possible belief structures and their combinations, making exhaustive searches infeasible for larger frames. Our use of random sampling provided valuable insights. However, the limitations of this method became apparent when no higher-order hidden conflicts were observed in larger frames despite their theoretical existence.

Furthermore, our analysis highlighted the disproportionate growth of certain conflict classes, particularly HC_1, which quickly surpasses non-conflicting pairs as Ω increases.

These findings suggest that as the frame of discernment expands, the likelihood of encountering meaningful, non-conflicting belief combinations diminishes, raising important questions about the practical implications of belief function theory in large-scale applications.

[2]It is, of course, different if some restricted class of belief function is considered: either from the point of view of their structure or the limitation of the number or size of focal elements.

In conclusion, while belief functions offer a robust framework for managing uncertainty, our findings indicate that the prevalence of conflict, particularly in larger frames, necessitates careful consideration in practical applications. Future work could explore alternative methods for managing or mitigating conflicts in belief structures, especially as the scale of analysis increases. Additionally, further research into the theoretical underpinnings of conflict distribution may yield new insights that can enhance the utility of belief function theory in complex decision-making scenarios.

References

M. Daniel. Conflict between Belief Functions: A New Measure Based on Their Non-conflicting Parts. In F. Cuzzolin, editor, *BELIEF 2014*, volume 8764 of *LNCS*, pages 321–330, Cham, 2014. Springer. doi: 10.1007/978-3-319-11191-9_35.

M. Daniel and V. Kratochvíl. Hidden Auto-Conflict in the Theory of Belief Functions. In *Proceedings of the 20th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty (CJS'17)*, pages 34–45, 2017. URL `http://hdl.handle.net/11104/0276743`.

M. Daniel and V. Kratochvíl. Belief Functions and Degrees of Non-conflictness. In G. Kern-Isberner and Z. Ognjanović, editors, *ECSQARU 2019*, volume 11726 of *LNCS*, pages 125–136, Cham, 2019. Springer. doi: 10.1007/978-3-030-29765-7_11.

M. Daniel and V. Kratochvíl. Classes of Conflictness / Non-Conflictness of Belief Functions. In *Proceedings of the 12th Workshop on Uncertainty Processing (WUPES'22)* , pages 97–110. MatfyzPress, 2022.

M. Daniel and V. Kratochvíl. Hidden conflicts of belief functions. *International Journal of Computational Intelligence Systems*, 14:438–452, 2020. ISSN 1875-6883. doi: https://doi.org/10.2991/ijcis.d.201008.001. URL `https://doi.org/10.2991/ijcis.d.201008.001`.

C. Osswald and A. Martin. Understanding the large family of Dempster-Shafer theory's fusion operators — a decision-based measure. In *9th International Conference on Information Fusion*. IEEE, 2006. doi: 10.1109/ICIF.2006.301631.

F. Pichon, A.-L. Jousselme, and N. B. Abdallah. Several shades of conflict. *Fuzzy Sets and Systems*, 366:63–84, 2019. doi: 10.1016/j.fss.2019.01.014.

G. Shafer. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.

P. Smets. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005. doi: 10.1016/j.ijar.2004.05.003.

Appendix 1: Hidden Conflict Examples

In accordance to the body of this study, only belief structures are important for existence and degree of conjunctive (hidden) conflict. Hence for our examples graphical presentations in Figures 1 and 2, are more important than any specific numeric belief masses.

In our examples, we would like to illustrate how a hidden conflict is revealed. Note that because of the commutativity of $\circledcirc$, we can rewrite $\left(\textcircled{\cap}_1^3(m_1 \circledcirc m_2)\right)$ into $\left(\textcircled{\cap}_1^3(m_1) \circledcirc \textcircled{\cap}_1^3(m_2)\right)$, etc. Once a positive mass is assigned to the empty set, it cannot be removed by $\circledcirc$. Let us highlight the first occurrence of a positive mass on an empty set to clarify the examples.

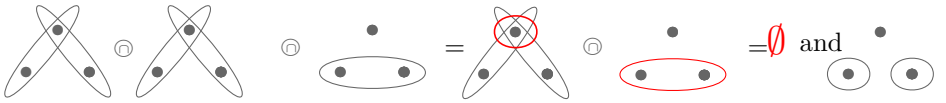


Figure 1: Arising of a hidden conflict: focal elements of m', m', m''; $m' \circledcirc m'$, m''; and of $m' \circledcirc m' \circledcirc m''$. Where ω_1 is the top one element, ω_2 and ω_3 clock-wise numbered.

Introductory Example. Let us assume Bel', Bel'' on Ω_3 where $\mathcal{F}' = \{\{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}\}$ and $\mathcal{F}'' = \{\{\omega_2, \omega_3\}\}$ (see Fig. 1). Then $(m' \circledcirc m'')(\emptyset) = 0$. But $(m' \circledcirc m' \circledcirc m'')(\emptyset) > 0$ (as highlighted in Figure 1, where conflicting focal elements are drawn in red), which implies $\textcircled{\cap}_1^2(m' \circledcirc m'')(\emptyset) > 0$ as well. Thus, there is a conflict hidden in the 1st degree. For detail, see Daniel and Kratochvíl (2020); and for an example of numeric bmms also Table 9. In comparison with $(m' \circledcirc m' \circledcirc m'')(\emptyset) = (m' \circledcirc m'' \circledcirc\ m' \circledcirc m'')(\emptyset) = 0.48$, the conflict based on non-conflicting parts of belief functions (see Daniel (2014)) $Conf(m', m'') = 0.40$, see Daniel and Kratochvíl (2020).

X	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_1,\omega_2\}$	$\{\omega_1,\omega_3\}$	$\{\omega_2,\omega_3\}$	$\{\omega_1,\omega_2,\omega_3\}$	$\emptyset$
$m'(X)$	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
$(m' \circledcirc m'' \circledcirc m' \circledcirc m'')(X)$	0.00	0.36	0.16	0.00	0.00	0.00	0.00	0.48

Table 9: Belief masses in the Introductory Example.

Little Angel Example. Let us have two BFs Bel_1 and Bel_2 on $\Omega_5 = \{\omega_1, \omega_2, ..., \omega_5\}$: $\mathcal{F}_1 = \{A, B, C\} = \{\{\omega_1, \omega_2, \omega_5\}, \{\omega_1, \omega_2, \omega_3, \omega_4\}, \{\omega_1, \omega_3, \omega_4, \omega_5\}\}$, $\mathcal{F}_2 = \{D\} = \{\{\omega_2, \omega_3, \omega_4, \omega_5\}\}$, i.e., $|\mathcal{F}_1| = 3$ while $|\mathcal{F}_2| = 1$. Respective structures can be seen in Fig. 2 where sets of focal elements of individual BFs m_1 ($3\times$) and m_2 ($1\times$) are depicted in its first row (ω_1 is on the top with ω_is clock-wise enumerated). Again, there is $(m_1 \circledcirc m_2)(\emptyset) = 0$ (there is no empty intersection of any $X \in \mathcal{F}_1$ with $Y \in \mathcal{F}_2$). Moreover, $\textcircled{\cap}_1^2(m_1 \circledcirc m_2)(\emptyset) = 0$ in this example. Finally, $(\textcircled{\cap}_1^3(m_1 \circledcirc m_2))(\emptyset) > 0$. Thus there is a hidden conflict of the 2-nd degree. Following the second line of Figure 2, the empty set emerges as the intersection of focal elements drawn by red color, i.e., it appears already in $m_1 \circledcirc m_1 \circledcirc m_1 \circledcirc m_2$.

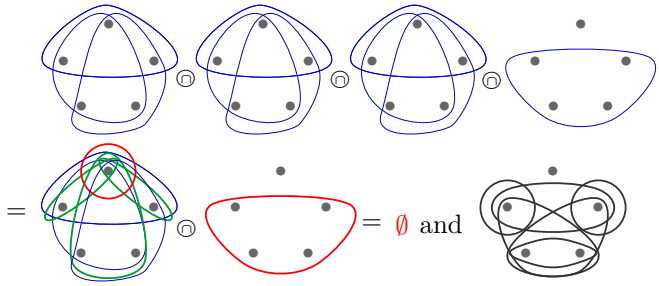


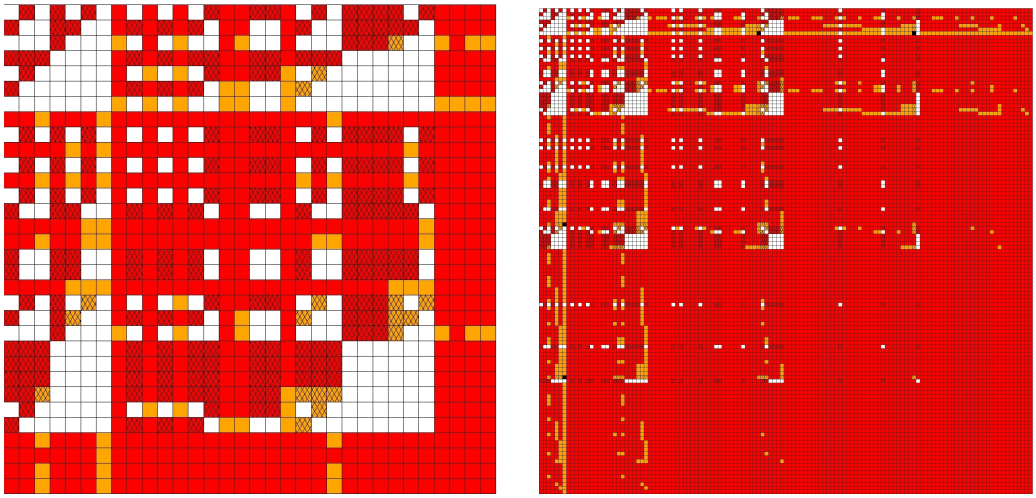
Figure 2: Little Angel example: focal elements of m_1, m_1, m_1, m_2; $\textcircled{\cap}_1^3 m_1, m_2$; $(\textcircled{\cap}_1^3 m_1) \textcircled{\cap} m_2$

Consistency of both m_1 and m_2 is underlined in Daniel and Kratochvíl (2020), nevertheless we already know, that non-consistency of their combination $m_1 \textcircled{\cap} m_2$ is more important.

For numeric values, see the original instance of Little Angel example in Daniel and Kratochvíl (2020), where is $(\textcircled{\cap}_1^3(m_1 \textcircled{\cap} m_1))(\emptyset) = 0.108 > 0$ and $Conf(m_1, m_2) = 0.1 > 0$.

Appendix 2: Bitmaps of Couples of Ω_3 Belief Structures

Analogously to Table 1 in Section 4 for Ω_2, we can present situation for Ω_3 by a bitmap presented in Figure 3.



(a) Zoom of left upper part (32×32)

(b) Full 127×127 bitmap

Figure 3: Hiddeness degree bitmap of conjunctive conflict of belief structures on Ω_3: White – full non-conflict (degree 3), Black – HC_2, Orange – HC_1, Red – conflict – HC_0 (degree 0).

Environmental Vulnerability and Uncertainty Processing

Milan Daniel[1] and Jan Geletič[1]

[1]Institute of Computer Sciences, Czech Academy of Sciences, Czechia
milan.daniel;geletic@cs.cas.cz

This contribution combines older Hájek-Valdés approach [3, 4] to uncertainty processing in MYCIN-like expert systems and analysis of Cadiag-2 fuzzy expert system [1] with development of City Simulation Software [2, 5] covering environmental, population, transportation and energetic aspects of smart city urban modelling and planning.

Uncertainty processing in MYCIN-like systems is based on operations corresponding to group operations enabling us to show isomorphicity of uncertainty processing in quite different classic expert systems MYCIN and PROSPECTOR. Analogously, some of norms and conorms used in fuzzy expert systems may be considered to be cones of corresponding group operations. Based on this, we can compare MYCIN-like systems also with fuzzy systems as Cadiag-2 and its variants are.

Finally, similar aspects appear when computing global environmental KPIs from local ones in City Simulation Software. After appropriate transformation even KPIs covering different environmental KPIs as different thermal KPIs, air quality and other KPIs can be aggregated.

References

[1] M. Daniel, P. Hájek, and H. P. Ngyuen. CADIAG-2 and MYCIN-like Systems.. *Artificial Intelligence in Medicine* **9** (3), 241–259 (1997)

[2] M. Daniel, R. Dostál, S. Kozhevnikov, A. Matysková, K. Moudrá, A. Maia Pereira, and O. Přibyl. City Simulation Software: Perspective of Mobility Modelling. *Smart City Symposium Prague 2021. IEEE Proceedings.* (2021)

[3] P. Hájek, T. Havránek, and R. Jiroušek. *Uncertain Information Processing in Expert Systems.* CRC Press, Boca Raton, Florida (1992)

[4] P. Hájek and J. J. Valdés. Generalized algebraic foundations of uncertainty processing in rule-based expert systems (dempsteroids). *Computers and Artificial Intelligence* **10** (1), 29–42 (1991)

[5] M. Svítek, O. Přibyl, J. Vorel, et al. *City simulation software for modeling, planning, and strategic assessment of territorial city units.* CTU & ICS CAS, 2021. 68 s. Technical Report. (2021)

On approximation of functions by lattice integral transforms

Michal Holčapek
IRAFM, NSC IT4Innovations
University Ostrava
michal.holcapek@osu.cz

Viec Bui Quoc
Deparment of Mathematics
Can Tho University
bqviec@ctu.edu.vn

Abstract

The aim of this paper is to analyze the approximation capabilities of lattice integral transforms, especially, using their appropriate composition. In particular, our objective is to introduce and analyze the concept of the inverse integral kernel function, specifically a Q–inverse for an integral kernel function K on X. Additionally, we demonstrate that a smooth variant of the original function can be approximated from both below and above using the suitable composition of lattice integral transforms. Ultimately, we assess the approximation quality of the original function using the modulus of continuity and the value obtained from the integration of the square of the integral kernel function Q, which specifies the inverse kernel function to K.

1 Introduction

Integral transforms are mathematical tools that generate a new function $g(y)$ by integrating the product of an existing function $f(x)$ and an integral kernel function $K(x, y)$ over specified limits. The integral kernel function serves as a bridge between the domains of $f(x)$ and $g(y)$. Among the most widely used integral transforms are the Fourier and Laplace transforms, applicable to both real and complex-valued functions. These transforms are highly effective in addressing practical problems in various scientific and engineering fields, such as solving (partial) differential equations, signal and image processing, and spectral analysis of stochastic processes (see, e.g. [2, 13, 15]).

When handling imprecise or vague data using fuzzy set theory, we typically work with vectors or functions whose values are part of a suitable algebra of truth values, such as a residuated lattice or its specific subclasses like BL-algebra, MV-algebra, or IMTL-algebra. In the paper [12], Perfilieva proposed the so-called direct (lower and upper) lattice fuzzy transforms, which transform residuum lattice-valued functions into component vectors according to the fuzzy partition of the function domain. The component vectors carry a certain sort of aggregated information

from the original functions, which may be further processed. In addition, the original functions can be reconstructed from below and above from the component vectors using the so-called inverse lattice fuzzy transforms, which were used in signal and image processing (e.g., pepper-and-salt noise suppression or image compression/decompression).

In the paper [8], Holčapek and Quoc generalized lattice fuzzy transforms and introduced lattice integral transforms, which follow the standard integral transform scheme. More precisely, for a given complete residuated lattice L with the multiplication operation $\otimes$ and the residuum operation $\rightarrow$, for a fuzzy measure space $\langle X, \mathcal{F}, \mu \rangle$, an integral kernel function $K : X \times Y \rightarrow L$, and $\star \in \{\otimes, \rightarrow\}$, the $\star$-lattice integral transform is defined as the map $F^{\star}_{(K,\mu)} : \mathcal{F}(X) \rightarrow \mathcal{F}(Y)$ given by

$$F^{\star}_{(K,\mu)}(f)(y) = \int_X^{\otimes} K(x,y) \star f(x)\, d\mu, \tag{1}$$

where $\int_X^{\otimes}$ is the Sugeno-like integral introduced in [3, 4] (see also [5]). Obviously, a lattice integral transform maps residuated lattice-valued functions over a domain X to residuated lattice-valued functions over a domain Y, in a similar way to the standard integral transforms such as Fourier or Hilbert transforms for both real or complex-valued functions. It should be noted that, in addition to the lattice integral transforms presented in [8], two other types of lattice integral transforms were introduced in [9]. A reconstruction of original functions and certain applications in decision-making and image processing were presented in [6, 7, 10].

One of the primary objectives for introducing lattice integral transforms was to enhance the approximation capabilities of the lower and upper fuzzy transforms. The first analysis of the approximation properties of lattice integral transforms was presented in [10]. In the work, the authors introduced the relationship between the integral kernel function and its inverse using another integral kernel function on one function domain and demonstrated that the reconstructions of the original functions are greater than or equal to (or less than or equal to) their lattice integral transforms with respect to the new integral kernel function. It is worth mentioning that this new integral kernel function represents, roughly speaking, the unit formed by the composition of the integral kernel function and its inverse. In the context of standard integral transforms, this unit is the Dirac delta function.

The purpose of this paper is to deepen the analysis of the approximation capabilities of lattice integral transforms started in [10]. In particular, we aim to introduce and analyze the concept of the inverse integral kernel function, specifically a Q–inverse for an integral kernel function K on X. This Q–inverse is crucial for reconstructing the original function through the composition of lattice integral transforms. Next, we show the conditions under which the following inequalities hold (pointwise):

$$F^{\rightarrow}_{(K^{-1},\mu_Y)} \circ F^{\otimes}_{(K,\mu_X)} \geq F^{\otimes}_{(Q,\mu_X)} \text{ and } F^{\otimes}_{(K^{-1},\mu_Y)} \circ F^{\rightarrow}_{(K,\mu_X)} \leq F^{\rightarrow}_{(Q,\mu_X)},$$

where K^{-1} is the Q–inverse of K. Finally, we introduce a modulus of continuity for lattice-valued functions, denoted as $\omega(f)$, and demonstrate that if both lattice

integral transforms preserve constant functions and K^{-1} is the Q–inverse of K then

$$F^{\rightarrow}_{(K^{-1},\mu_Y)} \circ F^{\otimes}_{(K,\mu_X)}(f)(x) \leftrightarrow f(x) \geq \omega(f),$$

and similarly for the opposite composition.

The paper is structured as follows. The next section provides a preliminary introduction to essential concepts such as residuated lattices, fuzzy sets, fuzzy relations, fuzzy measure spaces, and fuzzy integrals, which are utilized throughout the paper. The third section presents two types of lattice integral transforms and discusses their basic properties. The fourth section details the approximation capabilities of these transforms. Finally, the last section is the conclusion.

2 Preliminaries

Truth value algebras We assume that the algebra of truth values is a *complete residuated lattice*, i.e., an algebra $L = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ with four binary operations and two constants such that $\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice, where 0 is the least element and 1 is the greatest element of L, $\langle L, \otimes, 1\rangle$ is a commutative monoid (i.e., $\otimes$ is associative, commutative and the identity $a \otimes 1 = a$ holds for any $a \in L$) and the adjointness property is satisfied, i.e.,

$$a \leq b \rightarrow c \quad \text{iff} \quad a \otimes b \leq c \tag{2}$$

holds for each $a, b, c \in L$, where $\leq$ denotes the corresponding lattice ordering, i.e., $a \leq b$ if $a \wedge b = a$ for $a, b \in L$. The operations $\otimes$ and $\rightarrow$ are called the *multiplication* and *residuum*, respectively. For details, we refer to [1].

Example 1 *It is easy to prove that the algebra*

$$L_T = \langle [0,1], \min, \max, T, \rightarrow_T, 0, 1\rangle,$$

where T is a left continuous t-norm (see, e.g., [11]) and $a \rightarrow_T b = \bigvee\{c \in [0,1] \mid T(a,c) \leq b\}$ defines the residuum, is a complete residuated lattice.

A unary operation $N : L \rightarrow L$ is called a *generalized negation* (*negation* for short) on L if N is a non-increasing map such that $N(0) = 1$ and $N(1) = 0$.

Fuzzy sets Let L be a complete residuated lattice, and let X be a non-empty set. A function $A : X \rightarrow L$ is called a *fuzzy set on* X. The set of all fuzzy sets on X is denoted by $\mathcal{F}(X)$. A fuzzy set A on X is called *crisp* if $A(x) \in \{0, 1\}$ for any $x \in X$. The symbol $\emptyset$ denotes the empty fuzzy set on X, i.e., $\emptyset(x) = 0$ for any $x \in X$. The set of all crisp fuzzy sets on X (i.e., the power set of X) is denoted by $\mathcal{P}(X)$. A constant fuzzy set A on X (denoted as $\underline{a_X}$) satisfies $A(x) = a$ for any $x \in X$, where $a \in L$. The set $\mathrm{Core}(A) = \{x \mid x \in X \ \&\ A(x) = 1\}$ is called the the

core of a fuzzy set A, respectively. A fuzzy set A is called *normal* if $\mathrm{Core}(A) \neq \emptyset$. A fuzzy set K on the Cartesian product $X \times Y$ is called the fuzzy relation. We say that a fuzzy relation K is *normal* in the first argument if $\mathrm{Core}(K_x) \neq \emptyset$ for any $x \in X$, where $K_x(\cdot) = K(x, \cdot)$. Similarly, one can define the normality of K in the second component.

Fuzzy measure spaces Let X be a non-empty set. A subset $\mathcal{F}$ of $\mathcal{P}(X)$ is an *algebra of sets on* X provided that.

(A1) $X \in \mathcal{F}$,

(A2) if $A \in \mathcal{F}$, then $X \setminus A \in \mathcal{F}$,

(A3) if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$.

It is easy to see that if $\mathcal{F}$ is an algebra of sets, then the intersection of finite number of sets belongs to $\mathcal{F}$. A pair $(X, \mathcal{F})$ is called a *measurable space* (on X) if $\mathcal{F}$ is an algebra (σ-algebra) of sets on X. Let $(X, \mathcal{F})$ be a measurable space and $A \in \mathcal{F}(X)$. We say that A is $\mathcal{F}$*-measurable* if $A \in \mathcal{F}$. Obviously, the sets $\{\emptyset, X\}$ and $\mathcal{P}(X)$ are algebras of fuzzy sets on X.

A map $\mu : \mathcal{F} \to L$ is called a *fuzzy measure* on a measurable space $(X, \mathcal{F})$ if

(i) $\mu(\emptyset) = 0$ and $\mu(X) = 1$,

(ii) if $A, B \in \mathcal{F}$ such that $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

A triplet $(X, \mathcal{F}, \mu)$ is called a *fuzzy measure space* whenever $(X, \mathcal{F})$ is a measurable space and μ is a fuzzy measure on $(X, \mathcal{F})$. For details, we refer to [14]. Let μ be a fuzzy measure on $(X, \mathcal{F})$. We say that a map $\mu^{c,N} : \mathcal{F} \to L$ is N*-conjugate to* μ if $\mu^c(A) = N(\mu(X \setminus A))$ for any $A \in \mathcal{F}$, where $X \setminus A$ is the complement of A in X and N is a generalized negation on L (cf., [3]).

Example 2 *Let L_{T} be an algebra from Ex. 1, where T is a continuous t-norm. Let $X = \{x_1, \ldots, x_n\}$ be a finite non-empty set, and let $\mathcal{F}$ be an arbitrary algebra. A* relative *fuzzy measure μ^r on $(X, \mathcal{F})$ can be given as*

$$\mu^r(A) = \frac{|A|}{|X|}$$

for all $A \in \mathcal{F}$, where $|A|$ and $|X|$ denote the cardinality of A and X, respectively. Let $\varphi : L \to L$ be a monotonically non-decreasing map with $\varphi(0) = 0$ and $\varphi(1) = 1$. The relative measure μ^r can be generalized as a fuzzy measure μ^r_φ on $(X, \mathcal{F})$ given by $\mu^r_\varphi(A) = \varphi(\mu^r(A))$ for any $A \in \mathcal{F}$.

Multiplication-based fuzzy integral The integrated functions are fuzzy sets on X and are denoted by f, g etc. Let $(X, \mathcal{F}, \mu)$ be a fuzzy measure space, and let $f : X \to L$. The $\otimes$–*fuzzy integral of* f *on* X is given by

$$\int_X^{\otimes} f \, d\mu = \bigvee_{A \in \mathcal{F}} \mu(A) \otimes \left(\bigwedge_{x \in A} f(x) \right). \tag{3}$$

It should be noted that the previous definition of $\otimes$–fuzzy integral was proposed in [3] and coincides with the definition in [5] whenever $\otimes$ distributes over $\bigwedge$ in the algebra of truth values (e.g., an MV-algebra).

3 Lattice integral transforms

As mentioned in the introduction, the main parameter of lattice fuzzy transforms is a fuzzy partition of the function domain. It should be noted that a fuzzy partition is composed of normal fuzzy sets whose cores constitute a standard partition of the function domain. It can be demonstrated that each fuzzy partition defines a fuzzy relation with specific properties, and conversely, specific fuzzy relations can create a fuzzy partition. Building on the definition of a fuzzy partition, specifically the normality of fuzzy sets, we introduce the concept of an integral kernel function as follows.

Definition 1 *A fuzzy relation* $K : X \times Y \to L$ *that is normal in the second argument is called* integral kernel function *or simply* integral kernel.

The lattice integral transforms are constructed using the multiplication-based fuzzy integral, whose integrand is the transformed function multiplied by the integral kernel, where the multiplication $\star$ is one of the operations of $\otimes$ and $\to$. A lattice integral transform of fuzzy sets from $\mathcal{F}(X)$ to fuzzy sets from $\mathcal{F}(Y)$ is defined as follows.

Definition 2 *Let* $(X, \mathcal{F}, \mu)$ *be a fuzzy measure space, let* $K : X \times Y \to L$ *be an integral kernel, and let* $\star \in \{\otimes, \to\}$. *A map* $F^{\star}_{(K,\mu)} : \mathcal{F}(X) \to \mathcal{F}(Y)$ *defined by*

$$F^{\star}_{(K,\mu)}(f)(y) = \int_X^{\otimes} K(x, y) \star f(x) \, d\mu, \tag{4}$$

is called a $(K, \mu, \star)$–lattice integral transform.

In what follows, we assume that a fuzzy measure space $(X, \mathcal{F}, \mu)$ is given and $K : X \times Y \to L$ is an integral kernel. The following theorem provides a summary of the elementary properties of lattice integral transforms (see, [8]).

Theorem 1 *Let* $\star \in \{\otimes, \to\}$. *For any* $f, g \in \mathcal{F}(X)$ *and* $a \in L$, *we have*

(i) $F^{\star}_{(K,\mu)}(f) \leq F^{\star}_{(K,\mu)}(g)$ *if* $f \leq g$,

(ii) $F^{\star}_{(K,\mu)}(f \cap g) \leq F^{\star}_{(K,\mu)}(f) \wedge F^{\star}_{(K,\mu)}(g)$,

(iii) $F^{\star}_{(K,\mu)}(f) \vee F^{\star}_{(K,\mu)}(g) \leq F^{\star}_{(K,\mu)}(f \cup g)$,

(iv) $a \otimes F^{\star}_{(K,\mu)}(f) \leq F^{\star}_{(K,\mu)}(\underline{a}_X \otimes f)$,

(v) $F^{\star}_{(K,\mu)}(\underline{a}_X \to f) \leq a \to F^{\star}_{(K,\mu)}(f)$.

The following theorem shows conditions under which a constant function (constant fuzzy set) $\underline{a}_X$ on X is transformed into a constant function $\underline{a}_Y$ on Y, i.e., $F^{\star}_{(K,\mu)}(\underline{a}_X) = \underline{a}_Y$.

Theorem 2 *Let $a \in L$.*

(i) If for any $y \in Y$ there exists $A_y \in \mathcal{F}$ such that $A_y \subseteq \mathrm{Core}(K_y)$ and $\mu(A_y) = 1$, then $F^{\otimes}_{(K,\mu)}(\underline{a}_X) = \underline{a}_Y$.

(ii) If for any $y \in Y$ and for any $A \in \mathcal{F}$ with $A \subseteq X \setminus \mathrm{Core}(K_y)$ it holds that $\mu(A) \leq a$, then $F^{\rightarrow}_{(K,\mu)}(\underline{a}_X) = \underline{a}_Y$.

It is worth noting that the standard real-valued fuzzy transforms as well as lower and upper lattice fuzzy transforms preserve constant functions; therefore, it seems to be natural to assume that integral kernels and fuzzy measures as the parameters of the lattice integral transforms satisfy the conditions under which the constant functions are preserved. In addition, the preservation of constant functions proved to be an essential condition for the successful reconstruction of the original functions using lattice integral transforms, so we will discuss this property in more detail.

Definition 3 *Let $\langle X, \mathcal{F}, \mu \rangle$ be a fuzzy measure space, let K be an integral kernel, and let $a \in L$. The sufficient condition in Theorem 2(i) is denoted by (C1) and we say that* (K, μ) satisfies (C1). *The sufficient condition in Theorem 2(ii) is denoted by (C2) and we say that* (K, μ) satisfies (C2) for a, *where $a \in L$. If (K, μ) satisfies (C2) for any $a \in L$, then we say that* (K, μ) satisfies (C2).

The following theorem connects conditions (C1) and (C2) with fuzzy measures and N-conjugate fuzzy measures.

Theorem 3 *Let $\langle X, \mathcal{F}, \mu \rangle$ be a fuzzy measure space, let K be an integral kernel.*

(i) If (K, μ) satisfies (C1), then $(K, \mu^{c,N})$ satisfies (C2).

(ii) If (K, μ) satisfies (C2) and $\mathrm{Core}(K_y) \in \mathcal{F}$ for any $y \in Y$, then $(K, \mu^{c,N})$ satisfies (C1).

4 Approximation of functions using lattice integral transforms

This section focuses on the approximation capability of lattice integral transforms, particularly, on their composition which enable reconstructing original functions. For this purpose, we introduce the concept of the Q–inverse integral kernel and show that a smooth variant of the original function can be approximated from both below and above using a composition of lattice integral transforms. In addition, we estimate the quality of the approximation using the continuity modulus and the integral of the square of the kernel function Q.

4.1 Inverse integral kernel

Let $K : X \times Y \to L$ and $K' : Y \times X \to L$ be two integral kernels. First, we introduce a new type of integral kernel on X to express the relationship between K and K'.

Definition 4 *An integral kernel $Q : X \times X \to L$ is said to be* compatible with K *and K' or also* (K, K')–compatible *provided that*

$$Q(x, z) \otimes K'(y, z) \leq K(x, y) \tag{5}$$

holds for any $x, z \in X$ and $y \in Y$.

The notion of inverse of K is related to the integral kernel of Q on X, which allows to introduce a wider class of inverses that can be taken into account when approximating the original functions.

Definition 5 *Let $Q : X \times X \to L$ and $K : X \times Y \to L$ be integral kernels. An integral kernel $K' : Y \times X \to L$ is said to be an* Q–inverse of K *if*

$$Q(x, z) = \bigwedge_{y \in Y} K'(y, z) \to K(x, y), \quad x, z \in X. \tag{6}$$

From the previous definition, we can see that Q is uniquely determined from K and K' by (6). Interestingly, we can have different Q–inverses of K for the same Q as the following example shows, so the Q–inverse is not defined uniquely.

Example 3 *Let $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2\}$, and consider the Łukasiewicz algebra on $[0, 1]$. Assume the integral kernels expressed by matrices as follows:*

$$K = \begin{pmatrix} 1 & 0.8 \\ 0.9 & 1 \\ 0.7 & 1 \end{pmatrix} \quad K^T = \begin{pmatrix} 1 & 0.9 & 0.7 \\ 0.8 & 1 & 1 \end{pmatrix} \quad K' = \begin{pmatrix} 1 & 0.9 & 0.7 \\ 0.7 & 1 & 1 \end{pmatrix}.$$

We see that $K^T \neq K'$, particularly, $K^T > K'$. Introducing the matrix operation for a $p \times q$–matrix K and a $q \times p$–matrix L as

$$(K \to L)_{ik} = \min\{K_{jk} \to L_{ij} \mid j = 1, \ldots, q\} \tag{7}$$

for any $i, k = 1, \ldots, p$, we simply find that the integral kernel Q defined in (6) expressed by the matrix form is

$$Q = K \to K^T = K \to K' = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.9 & 1 & 1 \\ 0.7 & 0.8 & 1 \end{pmatrix}.$$

Hence, there are two different Q–inverses of K, namely, the transpose of K and the integral kernel K'.

In what follows, we use K^{-1} to denote an arbitrary Q–inverse of K. More precisely, if we use K^{-1}, then we mean one of the Q–inverses of K, including the maximal one.

4.2 Upper and lower approximation of functions

The following theorem gives a generalization of the inequality for upper lattice fuzzy transforms and in a sense shows the approximation from above of the original function using the composition of lattice integral transforms.

Theorem 4 *Let $F^{\otimes}_{(K,\mu_X)}$ be an lattice integral transform from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$, and let $F^{\to}_{(K^{-1},\mu_Y)}$ be a lattice integral transform from $\mathcal{F}(Y)$ to $\mathcal{F}(X)$. Then*

$$F^{\to}_{(K^{-1},\mu_Y)} \circ F^{\otimes}_{(K,\mu_X)} \geq F^{\otimes}_{(Q,\mu_X)}. \tag{8}$$

Unlike the original method, which approximates functions from above using a composition of lattice fuzzy transforms, the composition of lattice integral transforms approximates the original function in a different way. Specifically, the reconstructed function is above the lattice integral transform of the original function with respect to the integral kernel Q on X, which is determined by the kernel K and its inverse K^{-1}. The $(Q, \mu_X, \otimes)$–lattice integral transform on the right side of inequality (8) acts as a smoothing filter that suppresses high-frequency components present in the function. Therefore, the composition of lattice integral transforms generally does not approximate the original function from above but rather its smoothed version as produced by the transform. This property highlights the utility of lattice integral transforms as effective filters for high-frequency content or random noise.

Similarly, a generalization of the approximation from below of the original function by a composition of lattice integral transforms is given in the following theorem.

Theorem 5 *Let $F^{\to}_{(K,\mu_X)}$ be a lattice integral transform from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$, and let $F^{\otimes}_{(K^{-1},\mu_Y)}$ be a lattice integral transform $\mathcal{F}(Y)$ to $\mathcal{F}(X)$. Then*

$$F^{\otimes}_{(K^{-1},\mu_Y)} \circ F^{\to}_{(K,\mu_X)} \leq F^{\to}_{(Q,\mu_X)}. \tag{9}$$

Note that $(K, \mu, \to)$–lattice integral transform on X with respect to the integral kernel Q can be viewed as another smoothing filter that filters out the high

frequencies presented in the functions. Thus, the reverse composition of M–lattice integral transforms approximates from below the smoothed original function given by the $(K, \mu, \to)$–lattice integral transform on X with respect to Q. As a consequence of the previous theorems, we can derive another approximation of original functions using lattice integral transforms.

Corollary 1 *For any $f \in \mathcal{F}(X)$ and $y \in X$, it holds that*

(i) $F^{\to}_{(K^{-1},\mu_Y)} \circ F^{\otimes}_{(K,\mu_X)}(f)(y) \geq \int_X^{\otimes} f \otimes 1_{\mathrm{Core}(Q_y)}\, d\mu_X$,

(ii) $F^{\otimes}_{(K^{-1},\mu_Y)} \circ F^{\to}_{(K,\mu_X)}(f)(y) \leq \int_X^{\otimes} 1_{\mathrm{Core}(Q_y)} \to f\, d\mu_X$.

4.3 Estimation of approximation quality

In this part, we focus on assessing the quality of the approximation of the original function using lattice integral transforms. One method for evaluating this quality is by measuring the closeness between the original function and the reconstructed function based on the modulus of continuity. Let $\mathcal{E}(X)$ denote the set of all equivalences on X.

Definition 6 *The map $\omega : \mathcal{F}(X) \times \mathcal{E}(X) \to L$ given by*

$$\omega(f, E) = \bigwedge_{(x,y)\in E} f(x) \leftrightarrow f(y) \tag{10}$$

is called the modulus of continuity.

The modulus of continuity provides a degree of proximity of function values at points that are equivalent with respect to a chosen equivalence E.

Define $\nabla : L \to L$ as

$$\nabla(a) = \begin{cases} 0, & a = 0, \\ 1, & \text{otherwise}, \end{cases} \tag{11}$$

for any $a \in L$. The following theorem shows an estimate of the approximation quality for the $(K, \mu, \otimes)$–lattice integral transform.

Theorem 6 *Let (K, μ) satisfy (C1), and let $f \in \mathcal{F}(X)$ and $y \in Y$. Define an equivalence E_y on X as $(x, z) \in E_y$ if $x = z$ or $\nabla K_y(x) \otimes \nabla K_y(z) = \top$ for any $x, z \in X$. Then*

$$F^{\otimes}_{(K,\mu)}(f)(y) \leftrightarrow f(x) \geq \omega(f, E_y), \tag{12}$$

for any $x \in X$ such that $\nabla K_y(x) = 1$.

Similarly, the next theorem shows that the same estimate of approximation quality holds also for the $(K, \mu, \to)$–lattice integral transform.

Theorem 7 *Let (K, μ) satisfy (C2), and let $f \in \mathcal{F}(X)$ and $y \in Y$. Define the equivalence E_y on X as in Theorem 6. Then*

$$F^{\rightarrow}_{(K,\mu)}(f)(y) \leftrightarrow f(x) \geq \omega(f, E_y), \tag{13}$$

for any $x \in X$ such that $\nabla K_y(x) = 1$.

The following statement presents the estimation of the approximation quality of the reconstructed function.

Theorem 8 *Let K be an integral kernel, K^{-1} be an Q–inverse of K for a reflexive integral kernel Q, and let $f \in \mathcal{F}(X)$. Assume that (K, μ_X) satisfies (C1) and (K^{-1}, μ_Y) satisfies (C2) and define $\omega(f) = \bigwedge_{y \in Y} \omega(f, E_y)$, where E_y is the equivalence defined in Theorem 6. Then*

$$F^{\rightarrow}_{(K^{-1},\mu_Y)} \circ F^{\otimes}_{(K,\mu_X)}(f)(x) \leftrightarrow f(x) \geq \omega(f) \tag{14}$$

for any $x \in X$.

Note that a similar statement holds for the reverse composition of lattice integral transforms as well.

In the end of this subsection, we provide an estimate of the approximation quality of lattice integral transforms for very special functions which are extensional with respect to a fuzzy relation on the space X. We know that $K^{-1} = K^T$ is the Q–inverse of K, where Q is the fuzzy relation on X given by formula (6). Let $Y \subseteq X$. We say that a fuzzy relation Q on X is *Y–transitive* if

$$Q(x, y) \otimes Q(y, z) \leq Q(x, z) \tag{15}$$

holds for any $x, z \in X$ and $y \in Y$. Obviously, Q is transitive if $X = Y$. The following lemma shows the properties of Q, when the integral kernel K is determined by a similarity relation on X.

Lemma 1 *Let $Y \subseteq X$ be a non-empty set, let P be a similarity relation on X such that $K : X \times Y \to L$ given as $K(x, y) = P(x, y)$ for any $x \in X$ and $y \in Y$ is an integral kernel, and let $K^{-1} = K^T$ be a Q–inverse of K, i.e., Q is give by formula (6), Then Q is a reflexive and Y–transitive integral kernel on X such that $P \leq Q$. In addition, $P(x, y) = Q(x, y)$ holds for any $x \in X$ and $y \in Y$.*

Let $f : X \to L$ be a function, and let Q be a reflexive and Y–transitive fuzzy relation on X. We say that *f is extensional with respect to Q* if

$$f(x) \otimes Q(x, y) \leq f(y) \text{ and } Q(x, y) \otimes f(y) \leq f(x) \tag{16}$$

holds for any $x, y \in X$. The following theorems provide an estimate of the approximation of the extensional functions.

Theorem 9 *Let $Y \subseteq X$ be a non-empty set, let P be a similarity relation on X such that $K : X \times Y \to L$ given as $K(x, y) = P(x, y)$ for any $x \in X$ and $y \in Y$ is an integral kernel, and let $K^{-1} = K^T$ be a Q–inverse of K. If f is extensional with respect to Q and (K^{-1}, μ_Y) satisfies (C2), then*

$$F^{\rightarrow}_{(K^{-1},\mu_Y)} \circ F^{\otimes}_{(K,\mu_X)}(f)(x) \leftrightarrow f(x) \geq \int_X^{\otimes} Q^2(y, x)\, d\mu_X \tag{17}$$

for any $x \in X$.

Again, a similar statement holds for the reverse composition of lattice integral transforms as well.

We have shown that the composition of lattice integral transforms preserves constant functions under the satisfaction of conditions (C1) and (C2). The following corollary demonstrates that even the same conditions ensure preservation for the entire class of extensional functions with respect to Q, where the integral kernel Q is determined from a similarity relation P on X, as stated in the previous theorem.

Corollary 2 *Let the assumption of Theorem 9 be satisfied, and let (K, μ_X) satisfy (C1). Then*

$$F^{\rightarrow}_{(K^{-1},\mu_Y)} \circ F^{\otimes}_{(K,\mu_X)}(f) = f \tag{18}$$

for any extensional function f on X with respect to Q.

5 Conclusion

This paper analyzed the approximation capabilities of lattice integral transforms and their compositions that enable the reconstruction of original functions. The concept of the inverse integral kernel function, specifically a Q–inverse for an integral kernel function K on X, was introduced and some properties were examined. It was demonstrated that a smooth variant of the original function could be approximated from both below and above using suitable compositions of lattice integral transforms. The approximation quality of the original function was assessed using the continuity modulus and, for special functions, also the integral of the square of the kernel function Q, which serves as the inverse of K. These findings underscored the effectiveness of lattice integral transforms in achieving precise and reliable approximations, providing a robust framework for further applications in mathematical analysis and related fields.

References

[1] R. Bělohlávek. *Fuzzy Relational Systems: Foundations and Principles.* Kluwer Academic Publishers, New York, 2002.

[2] L. Debnath and D. Bhatta. *Integral Transforms and Their Applications*. CRC Press, New York, 1914.

[3] D. Dubois, H. Prade, and A. Rico. Residuated variants of sugeno integrals: Towards new weighting schemes for qualitative aggregation methods. *Information Sciences*, 329:765–781, 2016.

[4] A. Dvořák and M. Holčapek. **L**-fuzzy quantifiers of type $\langle 1 \rangle$ determined by fuzzy measures. *Fuzzy Sets and Systems*, 160(23):3425–3452, 2009.

[5] A. Dvořák and M. Holčapek. Fuzzy measures and integrals defined on algebras of fuzzy subsets over complete residuated lattices. *Information Sciences*, 185(1):205–229, 2012.

[6] M. Holčapek and V. B. Quoc. On an application of lattice-valued integral transform to multicriteria decision making. In *Soft Computing: Biomedical and Related Applications*, pages 137–149. Springer, 2021.

[7] M. Holčapek, V. B. Quoc, and P. Ferbas. On an application of lattice integral transforms in image processing. *Mathematics*, 10(21):4077, 2022.

[8] M. Holčapek and V. Bui. Integral transforms on spaces of complete residuated lattice valued functions. In *Proc. of IEEE World Congress on Computational Intelligence (WCCI) 2020*, pages 1–8. IEEE, 2020.

[9] M. Holčapek and V. Bui. On integral transforms for residuated lattice-valued functions. In *Proc. of Information Processing Management of Uncertainty (IPMU) 2020*, pages 1–14. IPMU, 2020.

[10] M. Holčapek and V. Quoc. Reconstruction of lattice-valued functions by integral transforms. In *International Workshop on Fuzzy Logic and Applications (WILF) 2021*, pages 1–8. WILF, 2021.

[11] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, volume 8 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, 2000.

[12] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy sets and systems*, 157(8):993–1023, 2006.

[13] F. C. Tenoudji. *Analog and Digital Signal Analysis: From Basics to Applications*. Springer, Switzerland, 2016.

[14] Z. Wang and G. J. Klir. *Fuzzy measure theory*. Plenum Press, New York, 1992.

[15] 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, 235 p. , 1962.

Pairwise Comparison-Based Decision Support in Multiple Criteria Problems with Inconsistent Preference Data*

Masahiro Inuiguchi
Department of Systems Innovation, Graduate School of Engineering Science,
Osaka University
inuiguti@sys.es.osaka-u.ac.jp

Abstract

The studies on the interval analytic hierarchy process (AHP) being conducted in our laboratory are systematically reviewed with their significance in this paper.

1 Introduction

The pairwise comparison matrix (PCM) given by the decision maker (DM) is often inconsistent. The analysis of the PCM has been allowed by the conventional AHP when the PCM is sufficiently consistent. The priority weight vector is estimated by minimizing a kind of total deviation. In the interval AHP, the inconsistency is assumed to come from the vagueness of human judgments. Therefore, the normalized interval priority weight vector is estimated to cover all preferential data given by the crisp PCM. We note that the condition of order preservation is never violated in the interval AHP because all components of the crisp PCM are treated as possible evaluations by the DM and covered by the estimated normalized interval priority weight vector. The conventional estimation method for a normalized interval priority weight vector from a given crisp PCM tends to derive a too-narrow interval priority weight vector. Therefore, several estimation methods for a normalized interval priority weight vector has been proposed and examined the advantages over the conventional estimation method by numerical experiments.

However, because the normalized interval weight vector corresponds to the consistent interval PCM is not always unique, the estimation problem of normalized interval priority weight vector under a given crisp/interval PCM can have non-unique solutions. Therefore, in our approach, we estimate a set of normalized interval priority weight vectors. It has been shown that the set of normalized

*Supported by JSPS KAKENHI Grant Number JP23K04272.

interval priority weight vectors is obtained as a line segment expressed by a parameter showing the sum of centers, if the estimation problem with the center vector normalization constraint has a unique solution. Because all normalized interval priority weight vectors in the set associate with the same conssistent interval PCM, their distances from the given crisp/interval PCM are the same. We regard each normalized interval priority weight vectors in the set as an interval priority weight vector expressing the DM's preference. As the evaluation of the DM is not yet consistent, s/he has multiple normalized interval priority weight vectors each of which includes many normalized crisp priority weight vectors.

For decision analysis using the set of normalized interval priority weight vectors, various approaches are conceivable. In the conventional AHP, the marginal utility values of alternatives for each criterion are estimated by a PCM. However, to avoid the rank reversal, we assume the marginal utility values of alternatives for each criterion have already been assessed properly. Therefore, we consider the set of normalized interval priority weight vectors only for criteria. Since a normalized interval priority weight vector includes the uncertainty and multiple normalized interval priority weight vectors are obtained for the DM, we face a decision analysis under double uncertainty.

One approach is to reduce the set of normalized interval priority weight vectors to a set of normalized crisp priority weight vectors included at least one normalized interval priority weight vector in the set. Then, we may apply one of decision rules developed for decision making under uncertainty. Namely, we may apply the maximin, maximax, Hurwicz and minimax regret rules for ranking the alternatives with marginal utility values and a set of normalized crisp priority weight vectors.

The other approach is to rank alternatives for each normalized interval priority weight vector in the set of solutions to the estimation problem by using one of the decision rules such as the maximin, maximax, Hurwicz and minimax regret rules. Then we obtain a set of weak orders of alternatives. As the set of normalized interval priority weight vectors treated in our study is a line segment, we may obtain a figure showing the transitions of interval total utility values of alternatives as functions of a parameter, and maps of weak orders of alternatives in the parameter space. The figure and maps visualize the DM's vague preference among alternatives. Observing those, we find alternative pairs such that the DM's preference between them is stable, alternative pairs such that the DM's preference between them is labile, and so on. Moreover, by the interaction with the DM, we may revise the set of weak orders to a more suitable one for the DM. Furthermore, by aggregating the weak orders in the set, we obtain an aggregated weak order. More rich analyses would be conceivable in the interval AHP.

The estimation methods of the normalized interval priority weight vector from a crisp PCM are compared also from the view of the accuracy of ranking alternatives. We have started the investigations of interval AHP but there are still many untouched topics. The author hopes that the interval AHP will be developed further as a useful tool for decision analysis and support even when the given PCM is not sufficiently consistent.

A tribute to Klir's research on entropy for belief functions

Radim Jiroušek **Václav Kratochvíl**
Institute of Information Theory and Automation,
Czech Academy of Sciences, Prague, Czech Republic
[radim/velorex]@utia.cas.cz

Abstract

This is a short paper to remind us that although Prof. George Klir passed away eight years ago, his ideas are still alive. He was one of the first who had the idea to extend information theory beyond probability theory and to see the potential of fuzzy sets and belief functions. Some of the problems he raised have not yet been satisfactorily solved. In this paper, we take a fresh look at the problem of what properties entropy should have if it is to be considered a measure of information. Based on these requirements, we consider 24 definitions of entropy for belief functions and study how well they satisfy the proposed conditions.

1 Introduction

We met Prof. Klir in the early 1990s when he was lecturing at the Institute of Information Theory and Automation of the Czech Academy of Sciences. He was talking about the open problem of defining entropy in the context of the theory of belief functions. At that time he had about five to seven definitions and was trying to answer several questions. He noticed that belief functions were burdened with different kinds of ambiguity. He recognized three types. The first was nonspecificity, which was related to the size of the focal elements; the larger the focal elements, the less specific the basic assignment. The second type of ambiguity measured internal conflict or dissonance in the belief function, and the third type was confusion. Thus, one question was which types of uncertainty were measured by the entropy functions under study. The main question, however, was what were the most important properties of the concept of entropy. Of course, the answer to this question depends on what we want to measure with this quantity. Is it a measure of uncertainty? If so, how should the measure deal with its different types? Or should entropy measure the information content of a random variable? Although much time has passed since then, no one has satisfactorily answered most of these questions. In this paper, however, we will show that if we restrict the problem

to finding an entropy that allows us to define some notions of information theory (in the case of this paper, the measure of mutual information) in the same way as Shannon did, we can deduce which properties are most important. We will also see how Prof. Klir's definitions meet these requirements.

2 Belief functions notation

We expect the reader to be familiar with the theory of belief functions (Shafer, 1976), so we will just briefly recall the main notions of this theory and introduce the notation used in this paper.

Thus, Ω denotes a finite *frame of discernment*. A *basic probability assignment* (BPA) is a function $m : 2^\Omega \to [0,1]$ such that $\sum_{a\subseteq\Omega} m(a) = 1$, $m(\emptyset) = 0$. We will also consider a two-dimensional case when $\Omega = \Omega_X \times \Omega_Y$. For such $\omega \in \Omega$, its projections (coordinates) will be denoted by $\omega^{\downarrow X}$ and $\omega^{\downarrow Y}$; i.e., $\omega = (\omega^{\downarrow X}, \omega^{\downarrow Y})$. The same symbol will also be used for pro projection of subsets $a \subseteq \Omega$, $a^{\downarrow X} = \{\omega^{\downarrow X} : \omega \in a\}$, and marginalization. $m^{\downarrow X} : 2^{\Omega_X} \to [0,1]$ is a marginal BPA defined defined on Ω_X by

$$m^{\downarrow X}(a) = \sum_{b\subseteq\Omega : b = a^{\downarrow X}} m(b)$$

for all $a \subseteq \Omega_X$. Analogous symbols will also be used for projections and marginalization concerning the other coordinate Y.

A subset $a \subseteq \Omega$ is said to be a *focal element* of m if $m(a) > 0$. A BPA with only one focal element is said to be *deterministic*, denoted ι_a, where $\iota_a(a) = 1$. Since ι_Ω represents total ignorance, it is called *vacuous*. *Bayesian* BPAs are those BPAs whose focal elements are only singletons.

We will also use the standard alternative representation of a BPA: the *belief function, plausibility function,* and *commonality function*.

$$Bel_m(a) = \sum_{b\subseteq a} m(b); \qquad Pl_m(a) = \sum_{b\subseteq\Omega : b\cap a\neq\emptyset} m(b); \qquad Q_m(a) = \sum_{b\subseteq\Omega : b\supseteq a} m(b).$$

A central concept in Dempster-Shafer's theory is Dempster's combination rule (Shafer, 1976), which combines information from two distinct sources: BPAs m_1 and m_2. The combined BPA $m_1 \oplus m_2$ is computed (for each subset $c \subseteq \Omega$) as follows:

$$(m_1 \oplus m_2)(c) = (1-K)^{-1} \sum_{a\subseteq\Omega} \sum_{b\subseteq\Omega : a\cap b = c} m_1(a)\cdot m_2(b),$$

where $K = \sum_{a\subseteq\Omega} \sum_{b\subseteq\Omega : a\cap b=\emptyset} m_1(a)\cdot m_2(b)$ is usually interpreted as the amount of conflict between m_1 and m_2 (if $K = 1$, then the combination is undefined). This rule is also used when creating a joint BPA for independent variables, i.e., for BPAs m_X and m_Y defined on Ω_X and Ω_Y, respectively. In this case, the focal elements of the combination are only those $c \subseteq \Omega$, for which $c = c^{\downarrow X} \times c^{\downarrow Y}$, and Dempster's rule simplifies to a simple product $(m_X \oplus m_Y)(c) = m_X(c^{\downarrow X}) \cdot m_Y(c^{\downarrow Y})$.

Recall that each BPA m is associated with a set of probability distributions π defined on Ω. A *credal set*, is a set of probability distributions π

$$\mathcal{P}_m = \left\{\pi \text{ defined on } \Omega : \left(\forall a \subseteq \Omega : \pi(a) \geq Bel(a)\right)\right\}.$$

We will consider special probability distributions representing BPA m in specific situations. We will consider so-called *pignistic transform*, *plausibility transform*, and *maximum entropy transform* defined (respectively)

$$\pi_m(x) = \sum_{a \subseteq \Omega : x \in a} \frac{m(a)}{|a|},$$
$$\lambda_m(x) = \frac{Pl_m(x)}{\sum_{y \in \Omega} Pl_m(y)},$$
$$\mu_m = \arg\max_{\pi \in \mathcal{P}_m} \{\mathcal{H}(\pi)\},$$

where $\mathcal{H}(\pi) = -\sum_{x \in \Omega} \pi(x) \log_2 \pi(x)$ is Shannon entropy of a probability distribution π.

3 Requirements on entropy functions

Belief function theory was designed as a generalization of probability theory to surpass the imperfections of the latter. Therefore, most authors introducing the entropy within the theory of belief functions see it as a generalization of Shannon entropy. This is why they require it to equal Shannon entropy for all Bayesian BPAs. This is also why we accept that the belief function's entropy should have the following property.

Probability consistency property. We say that a function H that assigns a real value to each BPA is *consistent with Shannon entropy* if, for all Bayesian BPAs m, the value $H(m)$ is equal to the Shannon entropy of the corresponding probability function, i.e., $H(m) = -\sum_{x \in \Omega} m(\{x\}) \log_2 m(\{x\})$.

The properties studied in this paper are deduced from the requirement that the belief function entropy makes it possible to define mutual information in the same way as in probability theory. Recall that in probabilistic information theory

$$\mathcal{MI}_\pi(X;Y) = \mathcal{H}(\pi(X)) - \mathcal{H}(\pi(X|Y)) = \mathcal{H}(\pi(X)) + \mathcal{H}(\pi(Y)) - \mathcal{H}(\pi(X,Y)).$$

Since mutual information is always non-negative and equals zero if and only if the variables are independent, we get that in the ideal case, the belief function entropy should have the following property:

Strict subadditivity property. We say H is *strictly subadditive*, if

$$H(m) \leq H(m^{\downarrow X}) + H(m^{\downarrow Y}) \tag{1}$$

for all BPA m defined on Ω, with the equality in (1) if and only if $m = m^{\downarrow X} \oplus m^{\downarrow Y}$.

Thus, for the strictly subadditive function H, we could apply Shannon's idea of defining *mutual information* by the formula

$$MI_H(m[X;Y]) = H(m^{\downarrow X}) + H(m^{\downarrow Y}) - H(m). \tag{2}$$

Then MI_H would be symmetric, always non-negative, and equal to 0 if and only if X and Y are independent under BPA m, i.e., if $m = m^{\downarrow X} \oplus m^{\downarrow Y}$. However, we admit that we do not know such a strictly subadditive function. None of the entropies listed in Table 1 is strictly subadditive; none can be used as a basis for introducing information theory in the framework of belief functions. For this, one should find another strictly subadditive function or another way to introduce information-theoretic notions that do not follow the Shannon idea.

One reason for developing information theory for belief functions is to transfer the successful model learning algorithms from probability theory to the theory of belief functions. However, for this purpose, one can heuristically use a function that often manifests mutual information's properties. Therefore, we will study weaker properties in the following section. Namely, it is evident that every strictly subadditive function also has the following two properties.

Additivity property. We say H is *additive*, if $H(m_X \oplus m_Y) = H(m_X) + H(m_Y)$ for any pair of one-dimensional BPAs m_X, m_Y defined on Ω_X, Ω_Y, respectively.

Subadditivity property. H is said to be *subadditive*, if $H(m) \leq H(m^{\downarrow X}) + H(m^{\downarrow Y})$ for all BPA m defined on Ω.

The last property we require is based on the intuition that the entropy function H should measure the informativeness of BPAs. Again, using the analogy with probability theory, the more information there is in a BPA, the lower its entropy. For Bayesian BPAs, this requirement is met by the probability consistency property. For others, we assume that BPA m_1 is not less informative than BPA m_2 (assuming both are defined on the same frame of discernment Ω) if $Bel_{m_2} \leq Bel_{m_1}$, which is equivalent to $Pl_{m_1} \leq Pl_{m_2}$, and also to $\mathcal{P}_{m_1} \subseteq \mathcal{P}_{m_2}$. Note that this situation is very general and covers some other specific cases. In a sense, the simplest case is the following. We say that m_1 is a *simple specification* of m_2 if m_1 is created from m_2 by shifting some of its mass (or all of its mass) from some focal element to its subset. More precisely, there exist subsets $a \subset b \subseteq \Omega$ such that $m_1(a) = m_2(a) + \varepsilon$, and $m_1(b) = m_2(b) - \varepsilon$ (all remaining focal elements of m_1 are the copies of the focal elements of m_2). Since we are moving (a part of) the mass from b to its subset, we see directly from the definition of the belief function that[1] $Bel_{m_1} > Bel_{m_2}$. Thus, in this paper, we use the following notion.

[1]Strict inequality $Bel_{m_1} > Bel_{m_2}$ in this paper means that for all $a \subseteq \Omega$, $Bel_{m_1}(a) \geq Bel_{m_2}(a)$, and for at least one a, $Bel_{m_1}(a)$ is strictly greater than $Bel_{m_2}(a)$.

Simple monotonicity property. We say that a function H which assigns a real value to each BPA is *simple monotonic* if, for any simple specification m_1 of m_2, it holds that $H(m_1) < H(m_2)$.

It should be pointed out that several different types of monotonicity of belief functions entropy were studied in the literature (see e.g., monotonicity with respect to the set inclusion (Ramer, 1987), (Abellan and Moral, 1999), and others (Jiroušek and Shenoy, 2018)). Therefore, for the sake of clarity, we will consistently use only the notion of simple monotonicity. It can be shown that it is equivalent to the implication

$$Bel_{m_1} > Bel_{m_2} \Longrightarrow H(m_1) < H(m_2).$$

4 Survey of entropy functions

Without claiming completeness, we consider 24 definitions of entropy-like functions published in the last four decades (see Table 1). This section is devoted to their evaluation with respect to the requirements described in the previous section. As mentioned, none of them is strictly subadditive, so we will consider the introduced weaker properties: their additivity and subadditivity.

From a theoretical point of view, the Maeda-Ichihashi entropy H_I fits our requirements best. It is the only entropy that satisfies all the required properties except strict subadditivity: probability consistency, simple monotonicity, additivity, and subadditivity. A simple but rather singular counterexample has disproved its strict subadditivity. The main drawback of H_I is its computational complexity. Its computation requires the solution of an optimization problem: the search for the maximum entropy transform of the respective BPA. This precludes the application of this entropy not only to machine learning algorithms but also to the computational experiments we have performed to test the behavior of the entropies considered. Note a difference between this entropy and that of Harmanec-Klir H_H. The latter is not simple monotonic since it can take the same value for a BPA as for its simple specification. Nevertheless, H_H is also of extreme computational complexity, and the computation of the Abellán-Moral entropy H_A is even more complex.

We know of no other entropy that would satisfy all four properties that H_I possesses. The property of probability consistency is not a problem. It is possessed by the vast majority of entropies studied, except H_T, H_D, and H_{GP}. Similarly, most of the studied entropies have the additivity property: H_O, H_D, H_N, H_L, H_R, $H_P, H_B, H_H, H_I, H_J, H_Y, H_\lambda, H_S, H_\pi$. The problems arise with the remaining two properties, simple monotonicity, and subadditivity, which are satisfied by only a few entropies.

Both simple monotonicity and subadditivity properties are simultaneously satisfied only by the Dubois-Prade entropy H_D and the Abellán-Moral entropy H_A

Table 1: Definitions of entropy, chronologically ordered

H_O	Hohle (1982)	$H_O(m) = \sum_{a\subseteq\Omega} m(a) \log\left(\frac{1}{Bel_m(a)}\right)$
H_T	Smets (1983)	$H_T(m) = \sum_{a\subseteq\Omega} \log\left(\frac{1}{Q_m(a)}\right)$
H_D	Dubois, Prade (1987)	$H_D(m) = \sum_{a\subseteq\Omega} m(a) \log(\lvert a\rvert)$
H_N	Nguyen (1987)	$H_N(m) = \sum_{a\subseteq\Omega} m(a) \log\left(\frac{1}{m(a)}\right)$
H_L	Lamata, Moral (1988)	$H_L(m) = H_Y(m) + H_D(m)$
H_R	Klir, Ramer (1990)	$H_R(m) = H_D(m) + \sum_{a\subseteq\Omega} m(a) \log\left(\frac{1}{1-\sum_{b\subseteq\Omega} m(b)\frac{\lvert b\setminus a\rvert}{\lvert b\rvert}}\right)$
H_K	Klir (1991)	$H_K(m) = \sum_{a\subseteq\Omega} Bel_m(a) \log(Pl_m(a))$
H_P	Klir, Parviz (1992)	$H_P(m) = H_D(m) + \sum_{a\subseteq\Omega} m(a) \log\left(\frac{1}{1-\sum_{b\subseteq\Omega} m(b)\frac{\lvert a\cap b\rvert}{\lvert a\rvert}}\right)$
H_B	Pal et al. (1992, 1993)	$H_B(m) = H_D(m) + H_N(m)$
H_I	Maeda, Ichihashi (1993)	$H_I(m) = H_H(m) + H_D(m) = \mathcal{H}(\mu_m) + H_D(m)$
H_H	Harmanec, Klir (1994)	$H_H(m) = \max_{\pi\in\mathcal{P}_m} \mathcal{H}(\pi) = \mathcal{H}(\mu_m)$
H_{GP}	George, Pal (1996)	$H_P(m) = \sum_{a\subseteq\Omega} m(a) \sum_{b\subseteq\Omega} m(b) \left(1 - \frac{\lvert a\cap b\rvert}{\lvert a\cup b\rvert}\right)$
H_M	Maluf (1997)	$H_M(m) = -\sum_{a\subseteq\Omega} Pl_m(a) \log(Bel_m(a))$
H_A	Abellán, Moral (1999)[2]	$H_A(m) = H_I(m) + \min_{\pi\in\mathcal{B}_m} KL(\pi;\mu_m)$
H_J	Jousselme et al. (2006)	$H_J(m) = \mathcal{H}(\pi_m)$
H_Y	Yager (2008)	$H_Y(m) = \sum_{a\subseteq\Omega} m(a) \log\left(\frac{1}{Pl_m(a)}\right)$
H_G	Deng (2016)	$H_G(m) = H_N(m) + \sum_{a\subseteq\Omega} m(a) \log(2^{\lvert a\rvert} - 1)$
H_Z	Zhou et al. (2017)	$H_Z(m) = H_G(m) + \frac{\log(e)}{\lvert\Omega\rvert} \sum_{a\subseteq\Omega} m(a) * (1 - \lvert a\rvert)$
H_λ	Jiroušek, Shenoy (2018)	$H_\lambda(m) = \mathcal{H}(\lambda_m) + H_D(m)$
H_{PD}	Pan, Deng (2018)	$H_{PD}(m) = -\sum_{a\subseteq\Omega} \frac{Bel(a)+Pl(a)}{2} \log\left(\frac{Bel(a)+Pl(a)}{2(2^{\lvert a\rvert}-1)}\right)$
H_S	Jiroušek, Shenoy (2020)	$H_S(Q_m) = \sum_{a\subseteq\Omega} (-1)^{\lvert a\rvert} Q_m(a) \log(Q_m(a))$
H_Q	Qin et al. (2020)	$H_Q(m) = \sum_{a\subseteq\Omega} \frac{\lvert a\rvert}{\lvert\Omega\rvert} m(a) \log(\lvert a\rvert) + H_N(m)$
H_{YD}	Yan, Deng (2020)	$H_{YD}(m) = -\sum_{a\subseteq\Omega} m(a) log \frac{m(a)+Bel(a)}{2(2^{\lvert a\rvert}-1)} e^{\frac{\lvert a\rvert-1}{\lvert\Omega\rvert}}$
H_π	Jiroušek et al. (2022)	$H_\pi(m) = \mathcal{H}(\pi_m) + H_D(m)$

(and the already mentioned Maeda-Ichihashi entropy H_I). However, H_A is dis-

[2] KL denotes the famous Kullback-Leibler divergence of two probability measures defined $KL(\kappa;\pi) = \sum_{x\in\Omega} \kappa(x) \log_2 \frac{\kappa(x)}{\pi(x)}$, and $\mathcal{B}_m$ is a borderline of the convex set $\mathcal{P}_m$.

Table 2: Characteristics of entropy

		probability consistency	simple monotonicity	additivity	subadditivity	computational complexity
H_O	Hohle (1982)	■		■		low
H_T	Smets (1983)					high
H_D	Dubois, Prade (1987)		■	■	■	low
H_N	Nguyen (1987)	■		■		low
H_L	Lamata, Moral (1988)	■		■		low
H_R	Klir, Ramer (1990)	■		■		low
H_K	Klir (1991)	■	■			high
H_P	Klir, Parviz (1992)	■		■		low
H_B	Pal et al. (1992, 1993)	■		■		low
H_I	Maeda, Ichihashi (1993)	■	■	■	■	extreme
H_H	Harmanec, Klir (1994)	■		■	■	extreme
H_{GP}	George, Pal (1996)					low
H_M	Maluf (1997)	■				high
H_A	Abellán, Moral (1999)	■	■		■	extreme
H_J	Jousselme et al. (2006)	■		■		low
H_Y	Yager (2008)	■		■		low
H_G	Deng (2016)	■				low
H_Z	Zhou et al. (2017)	■				low
H_λ	Jiroušek, Shenoy (2018)	■	▒	■	▒	low
H_{PD}	Pan, Deng (2018)	■				high
H_S	Jiroušek, Shenoy (2020)	■		■		high
H_Q	Qin et al. (2020)	■				low
H_{YD}	Yan, Deng (2020)	■				low
H_π	Jiroušek et al. (2022)	■	▒	■	▒	low

qualified because of its computational complexity, and H_D is zero for all Bayesian BPAs, which disqualifies it as a basis for measuring the dependence of variables in the context of belief functions. Note, however, that it is precisely the H_D entropy whose inclusion makes H_I simple monotonic.

Considering that mutual information is usually used in machine learning only to control heuristic approaches, we can use a criterion that does not have all the

[3]Black stripe - proven property, Dark-grey stripe - in random experiments the property manifested in more than 99 % cases, Light-grey stripe - in random experiments the property manifested in more than 98 % cases.

theoretically required properties for this purpose. Since its computation is repeated many times, we need it to be fast and simple. If we allow its heuristic use, we can use a criterion that sufficiently satisfies the required property. To see if this happens, we randomly generated thousands of BPAs and tested how often the four properties under consideration were satisfied. The results are shown in Table 2, which not only summarizes the properties that were theoretically proven for the entropies considered (black stripes), but also shows that some of the properties were manifested sufficiently often (gray stripes). These results suggest that if we were asked to recommend an entropy for defining mutual information used in efficient machine learning algorithms, we would recommend one based on H_π. Although we know that H_π is neither simply monotonic nor subadditive, the computational experiments showed that the situations in which these negative properties can mislead the model learning process are not frequent.

5 Conclusions

Although we gave a kind of recommendation in the last sentences of the previous paragraph, the paper raised more questions than answers to open questions. The idea behind the Maeda-Ichihashi entropy H_I seems to be fruitful. Although the combination of the Dubois-Prade entropy H_D with neither the pignistic nor the plausibility transform yields an entropy with the required properties, the question remains whether some other transform would serve this purpose as well as the maximum entropy transform. Of course, even the choice of properties may not be optimal. Therefore, we regret that we do not have the opportunity to discuss all such questions with Prof. George J. Klir, who had a deep insight into this problem and the intuition to find answers. If he knew all the results published in the last eight years, he might take a different, more promising path.

Acknowledgement

This study was financially supported by the Czech Science Foundation under Grant No. 19-06569S. A part of this paper previously appeared as Jiroušek and Kratochvíl (2021).

References

Glenn Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.

Arthur Ramer. Uniqueness of information measure in the theory of evidence. *Fuzzy Sets and Systems*, 24(2):183–196, 1987.

Joaquin Abellan and Serafin Moral. Completing a total uncertainty measure in

the dempster-shafer theory. *International Journal Of General System*, 28(4-5): 299–314, 1999.

Radim Jiroušek and Prakash P Shenoy. A new definition of entropy of belief functions in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 92:49–65, 2018.

Ulrich Hohle. Entropy with respect to plausibility measures. In *Proc. of 12th IEEE Int. Symp. on Multiple Valued Logic, Paris, 1982*, 1982.

Philippe Smets. Information content of an evidence. *International Journal of Man-Machine Studies*, 19(1):33–43, 1983.

Hung T Nguyen. On entropy of random sets and possibility distributions. *The Analysis of Fuzzy Information*, 1:145–156, 1987.

George J Klir. Generalized information theory. *Fuzzy sets and systems*, 40(1): 127–142, 1991.

Nikhil R Pal, James C Bezdek, and Rohan Hemasinha. Uncertainty measures for evidential reasoning i: A review. *International Journal of Approximate Reasoning*, 7(3-4):165–183, 1992.

Nikhil R Pal, James C Bezdek, and Rohan Hemasinha. Uncertainty measures for evidential reasoning ii: A new measure of total uncertainty. *International Journal of Approximate Reasoning*, 8(1):1–16, 1993.

David A. Maluf. Monotonicity of entropy computations in belief functions. *Intelligent Data Analysis*, 1(3):207–213, 1997.

A-L Jousselme, Chunsheng Liu, Dominic Grenier, and Éloi Bossé. Measuring ambiguity in the evidence theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 36(5):890–903, 2006.

Ronald R Yager. Entropy and specificity in a mathematical theory of evidence. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pages 291–310. Springer, 2008.

Yong Deng. Deng entropy. *Chaos, Solitons & Fractals*, 91:549–553, 2016.

Deyun Zhou, Yongchuan Tang, and Wen Jiang. A modified belief entropy in dempster-shafer framework. *PloS one*, 12(5):e0176832, 2017.

Miao Qin, Yongchuan Tang, and Junhao Wen. An improved total uncertainty measure in the evidence theory and its application in decision making. *Entropy*, 22(4):487, 2020.

Radim Jiroušek, Václav Kratochvíl, and Prakash P Shenoy. Entropy for evaluation of Dempster-Shafer belief function models. *International Journal of Approximate Reasoning*, 151:164–181, 2022.

Didier Dubois and Henri Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy sets and systems*, 24(2):161–182, 1987.

Maria T Lamata and Serafín Moral. Measures of entropy in the theory of evidence. *International Journal of General System*, 14(4):297–305, 1988.

George J Klir and Arthur Ramer. Uncertainty in the Dempster-Shafer theory: a critical re-examination. *International Journal of General System*, 18(2):155–166, 1990.

George J Klir and Behzad Parviz. A note on the measure of discord. In *Uncertainty in Artificial Intelligence*, pages 138–141. Elsevier, 1992.

Yutaka Maeda and Hidetomo Ichihashi. An uncertainty measure with monotonicity under the random set inclusion. *International Journal of General System*, 21(4): 379–392, 1993.

David Harmanec and George J Klir. Measuring total uncertainty in Dempster-Shafer theory: A novel approach. *International Journal of General System*, 22 (4):405–419, 1994.

Thomas George and Nikhil R Pal. Quantification of conflict in dempster-shafer framework: a new approach. *International Journal Of General System*, 24(4): 407–423, 1996.

Lipeng Pan and Yong Deng. A new belief entropy to measure uncertainty of basic probability assignments based on belief function and plausibility function. *Entropy*, 20(11):842, 2018.

Radim Jiroušek and Prakash P Shenoy. On properties of a new decomposable entropy of Dempster-Shafer belief functions. *International Journal of Approximate Reasoning*, 119:260–279, 2020.

Hangyu Yan and Yong Deng. An improved belief entropy in evidence theory. *IEEE Access*, 8:57505–57516, 2020.

Radim Jiroušek and Václav Kratochvíl. Approximations of belief functions using compositional models. In *European Conference on Symbolic and Quantitative Approaches with Uncertainty*, pages 354–366. Springer, 2021.

On Optimizing Worker Assignment in Multi-stage Production Lines

Ryuhei Koura
Aoyama Gakuin University
a5721042@aoyama.jp

Xiaowen Zhao
Aoyama Gakuin University
zhao@ise.aoyama.ac.jp

Shao-Chin Sung*
Aoyama Gakuin University
son@ise.aoyama.ac.jp

Abstract

In this study, we are concerned with a worker assignment problem in multi-stage production lines based on the *restricted-cycle model with multiple periods*, in which workers may have different skill levels, and the operation time of each worker at any stage is stochastic. An outcome of the problem is an assignment of workers to stages, where each worker is assigned to exactly one stage of the production line, and each stage is assigned exactly one worker. Each worker-to-stage assignment is evaluated based on the expected cost of idle and expected cost of delay at all stages, and the goal is to find an assignment that minimizes the total expected cost among all assignments. Here, we propose a fast heuristic algorithm for finding a near-optimal assignment, and numerical experiments evaluate the heuristic. Moreover, we propose a new branch and bound based algorithm for finding an optimal worker assignment with an initial solution obtained by the heuristic, and its effectiveness is evaluated using numerical experiments as well.

1 Introduction

One of the main issues in production management is worker assignment in production line [2], and many different approaches have been proposed to solve the related problems [1, 5]. We are dealing with a worker assignment problem in multi-stage production lines, where workers may have different skill levels. Series production lines are under consideration [3]. The underlying model is called the *restricted-cycle*

*Shao-Chin Sung is currently known as So Akiyama.

model with multiple periods (see [4]), and an analytical study of a restricted case of the model is studied in [7, 9].

2 Problem Formulation

Suppose there are n workers $w_1, w_2, \ldots, w_n$ to be assigned to an n-stage production line that repeatedly generates a single product. In the following, stages in the production line are indexed by i and workers by j.

2.1 Production Line

- Each stage is planned or designed to be operated by one worker and to be completed within a target operation time. Since the production line generates a single product repeatedly, some stages may become idle if the operation time of a stage is shorter than others. Therefore, all stages are assumed to have the same target operation time, denoted by Z. However, by human or machine error, operation times at stages may not equal the target operation time Z, i.e., early and late completion at stages may occur. Of course, earliness and tardiness incur extra costs.

- When an operation is completed early, the corresponding stage becomes idle. A unit-time earliness incurs an earliness cost of C_E.

- On the other hand, a late completion at some stage causes a waiting time at the next stage, and if late completions occur at several consecutive stages, the waiting time for later stages will be much longer.

  Let $i \in \{1, 2, \ldots, n\}$ and $k \leq i$. The completion at the i-th stage is called a *k-consecutive late* if

  - the completions at the $(i-k+1)$-th stage, the $(i-k+2)$-th stage, $\ldots$, the i-th stage are all late, and
  - no late completion occur immediately before the $(i-k+1)$-th stage, i.e., the completion at the $(i-k)$-th stage is not late if $k < i$, or otherwise (if $k = i$), the completions at the first i stages are all late.

  Our model's tardiness cost may differ depending on the number of consecutive late completions. A unit-time of tardiness incurs a cost of $C_T^{(k)}$ when the corresponding completion is a k-consecutive late.

2.2 Workers

There are n workers $w_1, w_2, \ldots, w_n$ to be assigned to the n stages in the production line, and they may have different skill levels. Thus, the following values are associated with each worker w_j $(j = 1, 2, \ldots, n)$:

- $p_j \in [0,1]$: The probability of worker w_j completing the operation early at her or his assigned stage. Equivalently, $1-p_j$ is the probability of worker w_j completing the operation late at her or his assigned stage.
- $T_j \in \mathbb{R}_{\geq 0}$: The expected tardiness of worker w_j at her or his assigned stage.
- $E_j \in \mathbb{R}_{\geq 0}$: The expected earliness of worker w_j at her or his assigned stage.

2.3 Assignments and Their Evaluation

An outcome of our problem is a bijective mapping $\pi : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ of workers to stages in the production line. In this mapping, $\pi(i) = j$ means worker w_j is assigned to the i-th stage. We call such a mapping an *assignment*, and each assignment is evaluated based on the expected earliness cost and expected tardiness cost at all stages.

Let π be an assignment and $i \in \{1,2,\ldots,n\}$. Hence, under the assignment π, worker $w_{\pi(i)}$ is assigned to the i-th stage. Hence,

- the expected earliness cost incurred at the i-th stage is $C_E \cdot E_{\pi(i)}$, and
- the expected tardiness cost incurred at the i-th stage is the sum of the expected tardiness cost incurred by 1-consecutive late completion, 2-consecutive late completion, ..., i-consecutive late completion, i.e.,

$$\sum_{k=1}^{i} \left(\varphi(\pi, i, k) \cdot C_T^{(k)} \cdot T_{\pi(i)} \right)$$

where $\varphi(\pi, i, k)$ is the probability of a k-consecutive late completion occurred at the i-th stage. For every $i, k \in \{1,2,\ldots,n\}$ satisfying $i \geq k$,

$$\varphi(\pi, i, k) = \begin{cases} p_{\pi(i-k)} \cdot \prod_{j=i-k+1}^{i} (1 - p_{\pi(j)}) & \text{if } i > k \\ \prod_{j=1}^{i} (1 - p_{\pi(j)}) & \text{otherwise (i.e., if } i = k). \end{cases}$$

Therefore, the total expected cost under assignment π is

$$\textsc{total}(\pi) = \sum_{i=1}^{n} \left(C_E \cdot E_{\pi(i)} + \sum_{k=1}^{i} \left(\varphi(\pi, i, k) \cdot C_T^{(k)} \cdot T_{\pi(i)} \right) \right).$$

The goal is to find an assignment that minimizes the total expected cost $\textsc{total}(\pi)$ among all assignments π.

3 Previous Works

A straightforward way to find an optimal assignment is by brute force. More precisely, by applying a backtracking-based algorithm in which

- assignments are constructed by assigning workers one by one from the first stage, then the second stage, ..., and finally, the n-th stage;
- whenever a new assignment is entirely constructed, calculate the total expected cost and keep track of the best-constructed assignment.

In the end, return the best-constructed assignment as optimal, i.e., the assignment whose total expected cost is the smallest. The running time of such an algorithm is extremely long.

Yamamoto *et al.* [6] pointed out that the total expected cost of an assignment can be calculated incrementally, stage by stage. By definition, the total expected cost of an assignment is the sum of the expected costs incurred at all stages. For each $i \in \{1, 2, \ldots, n\}$, the expected cost incurred at the i-th stage is

$$\textsc{stage}(\pi, i) = C_E \cdot E_{\pi(i)} + \sum_{k=1}^{i} \left(\varphi(\pi, i, k) \cdot C_T^{(k)} \cdot T_{\pi(i)} \right),$$

and hence,

$$\textsc{total}(\pi) = \sum_{i=1}^{n} \textsc{stage}(\pi, i).$$

Observe that the value of $\textsc{stage}(\pi, i)$ does not depend on $\pi(i+1), \ldots, \pi(n)$, i.e., workers assigned after the i-th stage. This fact allows us to calculate the total expected cost of an assignment incrementally, stage by stage, and based on this fact, Yamamoto *et al.* [6] proposed a backtracking-based algorithm which improves the running time compared with brute force.

The running time for calculating $\textsc{stage}(\pi, i)$ based on the above definition is $\mathcal{O}(n^2)$, where

- the running time for calculating all $\varphi(\pi, i, k)$s, the probabilities of consecutive late completions, is $\mathcal{O}(n^2)$, and
- the running time for the remaining calculation is $\mathcal{O}(n)$.

Recently, Zhao *et al.* [8] pointed out that the values of $\varphi(\pi, i, k)$s, the probabilities of consecutive late completions, can also be calculated incrementally, stage by stage. For every $i, k \in \{2, 3, \ldots, n\}$ satisfying $i \geq k$,

$$\varphi(\pi, i, k) = \varphi(\pi, i-1, k-1) \cdot (1 - p_{\pi(i)}).$$

It follows that, with the values of all $\varphi(\pi, i-1, 1), \varphi(\pi, i-1, 2), \ldots, \varphi(\pi, i-1, i-1)$ provided, the values of all $\varphi(\pi, i, 1), \varphi(\pi, i, 2), \ldots, \varphi(\pi, i, i)$ can be obtained in $\mathcal{O}(n)$ time. Hence, the running time for calculating $\textsc{stage}(\pi, i)$ can also be reduced from $\mathcal{O}(n^2)$ to $\mathcal{O}(n)$, which is a huge reduction from the one by Yamamoto *et al.* [6]. Zhao *et al.* [8] proposed another backtracking-based algorithm with a much shorter running time.

4 Our Proposed Algorithm

In order to achieve the task of reducing the running time of backtracking-based algorithms for finding an optimal assignment, the approaches by Yamamoto *et al.* [6] and Zhao *et al.* [8] are to simplify the calculation for evaluating assignments.

Another way to achieve the task is by pruning branches in backtracking. Our proposed algorithm is a branch and bound-based algorithm for finding an optimal assignment, which is a minor modification from the algorithm proposed in [8].

Suppose, at some point of time in the search process of backtracking, a partial assignment σ is obtained, and the best-constructed assignment is π^*. Let $\ell \in \{1, 2, \ldots, n\}$ be such that only workers assigned to the first k stages are decided by σ. Then, for every (full) assignment π which can be obtained by extending σ, $\textsc{total}(\pi)$ is at least $\sum_{i=1}^{\ell} \textsc{stage}(\sigma, i)$. If $\sum_{i=1}^{\ell} \textsc{stage}(\sigma, i) \geq \textsc{total}(\pi^*)$, it implies $\textsc{total}(\pi) \geq \textsc{total}(\pi^*)$, and equivalently, none of the branches from σ leads to an assignment that updates the best-constructed assignment, and therefore, all branches from σ can be pruned. This argument provides one obvious way of pruning branches.

One can do better. Observe that the sum of all expected earliness costs at all stages is the same for all assignments. That is, for each assignment π, we have

$$\sum_{i=1}^{n} C_E \cdot E_{\pi(i)} = \sum_{j=1}^{n} C_E \cdot E_j.$$

Hence, $\textsc{total}(\pi)$ can be rewritten as follows.

$$\textsc{total}(\pi) = \sum_{j=1}^{n} C_E \cdot E_j + \sum_{i=1}^{n} \sum_{k=1}^{i} \left(\varphi(\pi, i, k) \cdot C_T^{(k)} \cdot T_{\pi(i)} \right)$$

It follows that our goal is essentially to minimize $\sum_{i=1}^{n} \sum_{k=1}^{i} \left(\varphi(\pi, i, k) \cdot C_T^{(k)} \cdot T_{\pi(i)} \right)$ among all assignments. Moreover, for every (full) assignment π which can be obtained by extending the partial assignment σ mentioned above, we have

$$\textsc{total}(\pi) \geq \sum_{j=1}^{n} C_E \cdot E_j + \sum_{i=1}^{\ell} \sum_{k=1}^{\ell} \left(\varphi(\pi, i, k) \cdot C_T^{(k)} \cdot T_{\pi(i)} \right) \geq \sum_{i=1}^{\ell} \textsc{stage}(\sigma, i).$$

One can prune more branches based on this argument except for extreme cases.

5 A Heuristic Algorithm

Generally, when applying a branch and bound-based algorithm, whether one can prune a branch depends on the partial solution under construction and the quality of the best-constructed solution. Hence, finding a solution as good as possible with some fast algorithms, i.e., heuristic, and using such a solution as the first

best-constructed solution may reduce the time of the branch and bound-based algorithm.

We examine local search as a fast heuristic for finding a good assignment, and we find that the qualities of local optimal assignments are surprisingly good. We implement local search based on the following operations.

- SWAP(i, j): Exchange the workers in two specified stages i and j.

- ROTATE(i, j): When $i < j$, extract the worker w at the i-th stage and, for each $i < k \leq j$, move the worker at the k-th stage to the $(k-1)$-th stage, and finally put the worker w at the j-the stage. When $i > j$, extract the worker w at the i-th stage and, for each $j \leq k < i$, move the worker at the k-th stage to the $(k+1)$-th stage, and finally put the worker w at the j-the stage.

This heuristic halts within 1 ms for instances with 15 stages in numerical experiments and produces high-quality solutions. Figure 1 shows that for more than 100,000 instances with stages ranging from 5 to 14, the average rate of the total expected cost of solutions obtained by our heuristic compared with the total expected costs of the corresponding optimal assignments. In the worst case, the rate is 100.538 %.

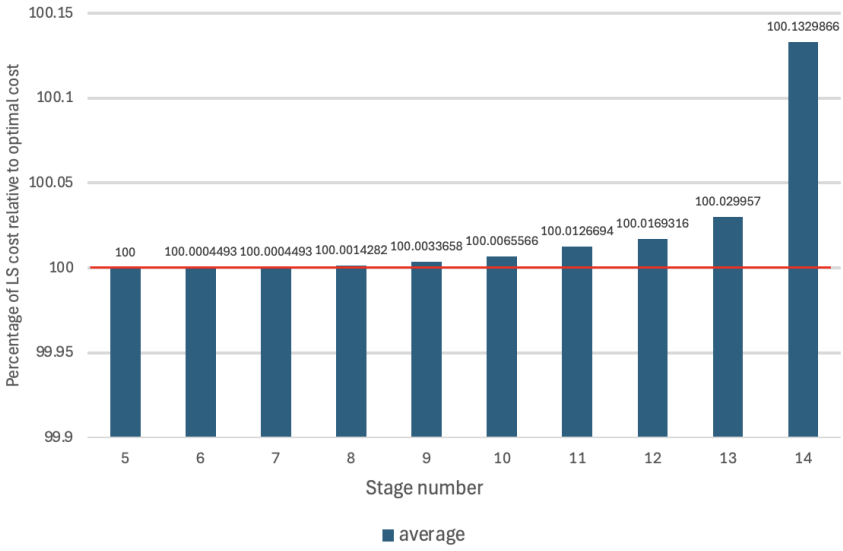


Figure 1: Qualities of heuristic solutions

6 Numerical Experiments

We implement our proposed algorithms with Java, and we conduct numerical experiments to test the effectiveness of our proposed algorithms. We apply the algorithms to more than 100,000 randomly generated instances with stages ranging from 5 to 14. Figure 2 and Table 1 and figures shows the comparison of running times of our proposed algorithms and algorithms proposed by Yamamoto *et al.* [6], Zhao *et al.* [8].

Table 1: The comparison of running times of algorithms

n	B&B + LS (ms)	B&B (ms)	Zhao [8](ms)	Yamamoto [6](ms)
5	0.0151	0.0058	0.0057	0.0072
6	0.0512	0.0027	0.0333	0.0310
7	0.2236	0.1761	0.1925	0.4404
8	1.4145	1.3296	1.6074	4.0879
9	11.6150	11.4700	15.0546	43.5208
10	109.4531	109.2086	148.6765	-
11	1,145.0325	1,147.2280	1,572.9247	-
12	13,311.5551	13,319.8945	18,361.5413	-
13	166,906.6154	167,772.6538	228,823.8077	-
14	1,962,642.3330	1,970,956.0000	2,519,959.6670	-

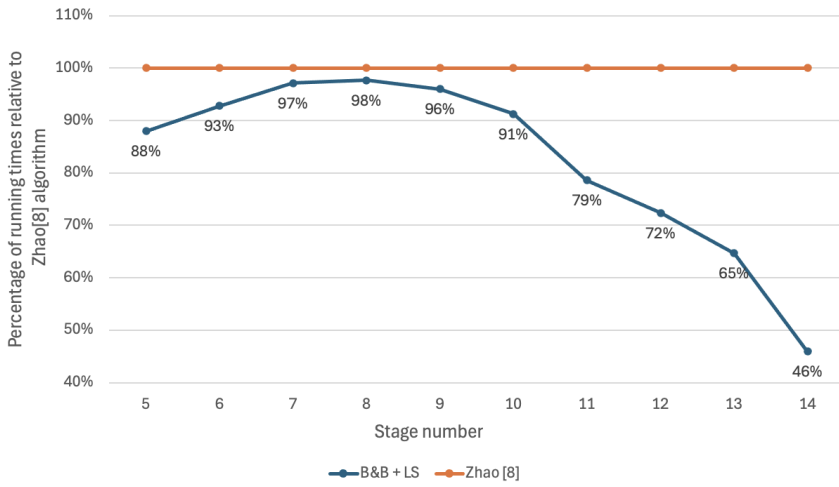


Figure 2: Comparison of running times

References

[1] H. Al-Khazraji. Comparative study of whale optimization algorithm and flower pollination algorithm to solve workers assignment problem. *International Journal of Production Management and Engineering*, 10(1):91–98, 2022.

[2] S. Bouajaja and N. Dridi. A survey on human resource allocation problem and its applications. *Operational Research*, 17:339–369, 2017.

[3] M. Hillier and F. Hillier. Simultaneous optimization of work and buffer space in unpaced production lines with random processing times. *IEE Transaction*, 38:39–51, 2006.

[4] X. Kong, H. Yamamoto, P. Song, J. Sun, and M. Matsui. Special workers' assignment optimization under the limited-cycled model with multiple periods. *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, 10(3), 2016.

[5] G. Ross and R. Soland. A branch and bound algorithm for the generalized assignment problem. *Mathematical Programming*, 8:91–103, 1975.

[6] H. Yamamoto, J. Sun, and M. Matsui. A study on limited-cycle scheduling problem with multiple periods. *Computers & Industrial Engineering*, 59(4):675–681, 2010.

[7] H. Yamamoto, J. Sun, M. Matsui, and X. Kong. A study of optimal arrangement in a reset limited-cycle problem with multiple periods –with fewer special workers (in japanese). *Journal of Japan Industrial Management Association*, 62(5):239–246, 2011.

[8] X. Zhao, R. Ogawa, and S.-C. Sung. On searching optimal worker assignment in multistage production lines. In *Proceedings of Industrial Engineering and Management (Lecture Notes in Mechanical Engineering)*, pages 205–212. Springer, 2024.

[9] X. Zhao, H. Yamamoto, J. Sun, and R. OOOKA. Optimal assignment with three groups in reset limited-cycle multiple periods (in japanese). *Journal of Japan Industrial Management Association*, 71:111–122, 2020.

Improving Stability of the Multi-class Support Vector Machine Maximizing Geometric Margins

Yoshifumi Kusunoki
Department of Interdisciplinary Informatics
Osaka Metropolitan University
yoshifumi.kusunoki@omu.ac.jp

Abstract

In this paper, we study a multi-class support vector machine proposed in our previous study, in which class-pair geometric margins are maximized. It is formulated by a convex approximation of the optimization problem which minimizes the p-th power of a p-norm of reciprocal class-pair geometric margins. However, this problem may not be solved stably because of the p-th power. Hence, we propose a more stable formulation which reduces the power of the objective function.

1 Multi-class Linear Classification and Class-pair Margins

In this paper, we study the multi-class support vector machine proposed in [1]. In this section, we introduce the multi-class classification task and the linear model. After that, we define class-pair geometric margins of the linear model.

The input space $\mathcal{X}$ is a subset of n-dimensional real space $\mathbf{R}^n$, and the set of class labels is defined by $\mathcal{Y} = \{1, 2, \ldots, c\}$. A learning problem is to find a classifier $\mathcal{C} : \mathcal{X} \to \mathcal{Y}$, using m training instances: $(x_1, y_1), \ldots, (x_m, y_m) \in \mathcal{X} \times \mathcal{Y}$. The objective of the learning problem is to correctly classify not only the training instances but also unseen ones. A linear classifier is given in the form: $\mathcal{C}(x) = \mathrm{argmax}_{y \in \mathcal{Y}} \{w_y^\top x + b_y\}$ for every $x \in \mathcal{X}$, where $(w_1, b_1), \ldots, (w_c, b_c) \in \mathbf{R}^{n+1}$ are parameters.

We define class-pair (geometric) margins for the multi-class linear classification model. Let $\mathcal{Y}^{\bar{2}} = \{(y, y') \mid y, y' \in \mathcal{Y}, y < y'\}$ be the set of class label pairs. We assume that its elements are indexed as follows: $\mathcal{Y}^{\bar{2}} = \{e_1, e_2, \ldots, e_{c(c-1)/2}\}$. Let $e_k = (y, y') \in \mathcal{Y}^{\bar{2}}$. We express the training set whose labels are y or y' as $x_1^k, \ldots, x_{m_k}^k$. For each x_i^k, we define a new class label $y_i^k = +1$ if $y_i = y$ and define $y_i^k = -1$ if $y_i = y'$. Let Z_k be the set of training instances $(x_1^k, y_1^k), \ldots, (x_{m_k}^k, y_{m_k}^k)$.

The hyperplane classifying y and y' is defined by $\{x \in \mathbf{R}^n \mid v_k^\top x + d_k = 0\}$, where $v_k = w_y - w_{y'}$ and $d_k = b_y - b_{y'}$. For simplicity, we assume that this hyperplane correctly divides Z_k, namely $y_i^k(v_k^\top x_i^k + d_k) > 0$ for all $(x_i^k, y_i^k) \in Z_k$. Using these definitions, the margin of the hyperplane for class pair e_k is defined as follows:

$$\mu_{Z_k}(v_k, d_k) = \frac{\min_{(x_i^k, y_i^k) \in Z_k} y_i^k(v_k^\top x_i^k + d_k)}{\|v_k\|}. \tag{1}$$

2 Stable Formulation of Margin Optimization

In our previous study [1], we have considered a support vector machine, in which a p-norm of reciprocal class-pair margins is minimized. To solve the optimization problem, we have proposed a convex approximation, which is formulated as follows:

$$\begin{aligned} \underset{(v,d)\in\mathcal{V},s}{\text{minimize}} \quad & \sum_{k\in\mathcal{Y}^{\bar{2}}} \|v_k\|^p / s_k^{p-1} \\ \text{subject to} \quad & y_i^k(v_k^\top x_i^k + d_k) \geq s_k \geq 1,\ (x_i^k, y_i^k) \in Z_k,\ k \in \mathcal{Y}^{\bar{2}}, \end{aligned} \tag{2}$$

where $\mathcal{V}$ is the set of elements (v_k, d_k) defined above. However, this formulation has the drawback that the objective function becomes too small if $\|v_k\|/s_k^{1-1/p}$ are less than one or too large if they are more than one. Hence, to reduce the power of the objective function, we use the following fact: $r^{1/p} = \inf_{s>0}\{sr + (p-1)(1/s)^{1/(p-1)}\}$. That is, we replace the objective function with $s_0 \sum_{k\in\mathcal{Y}^{\bar{2}}} \|v_k\|^p / s_k^{p-1} + (p-1)(1/s_0)^{1/(p-1)}$. Moreover, we replace v_k, d_k, s_k with $v_k/s_0, d_k/s_0, s_k/s_0$. Then, we obtain the following convex optimization problem:

$$\begin{aligned} \underset{(v,d)\in\mathcal{V},s,s_0}{\text{minimize}} \quad & \sum_{k\in\mathcal{Y}^{\bar{2}}} \|v_k\|^p / s_k^{p-1} + (p-1)(1/s_0)^{1/(p-1)} \\ \text{subject to} \quad & y_i^k(v_k^\top x_i^k + d_k) \geq s_k \geq s_0,\ (x_i^k, y_i^k) \in Z_k,\ k \in \mathcal{Y}^{\bar{2}}. \end{aligned} \tag{3}$$

3 Concluding Remarks

In this study, we have proposed a stable formulation to minimize reciprocal class-pair geometric margins. In our presentation, we will show the effect of the proposed formulation by experimental experiments.

References

[1] Y. Kusunoki and K. Tatsumi. Multi-class support vector machine based on minimization of reciprocal-geometric-margin norms. https://doi.org/10.21203/rs.3.rs-3507410/v1, 2023.

MDP-based Approach to Multi-Agent Pathfinding (extended abstract)

Jakub Mestek
Faculty of Mathematics and Physics
Charles University
jakub.mestek@mff.cuni.cz

1 Introduction to Multi-Agent Pathfinding

Multi-agent pathfinding (MAPF) is a problem of finding collision-free paths for agents moving in shared environment. An instance of MAPF [3] is a tuple (G, A, s, t), where $G = (V, E)$ is a graph, A a set of agents, $s\colon A \mapsto V$ and $g\colon A \mapsto V$ are start and goal locations of the agents. MAPF works with discrete time – at each timestep, every agent is located in a vertex of G. Between consecutive timesteps, each agent performs an action – either stays at its current vertex (*wait* action) or moves to a neighboring vertex (*move* action). A solution to classical MAPF is a set of plans $\{\pi_a\}_{a \in A}$, one plan for each agent. A plan for an agent a is a sequence of actions that, once executed, leads the agent from its start location $s(a)$ to its goal location $g(a)$.

Solution $\{\pi_a\}_{a \in A}$ is valid if it does not contain any conflict between any pair of agents. We consider vertex conflicts (two agents occupy same vertex at the same time) and swapping conflicts (two agents swap their positions between consecutive timesteps, i.e. at the same time travel over the same edge in opposing directions) [3].

2 Stochastic Environment

The above-described *classical* MAPF deals with deterministic environments. However, many real-world problems, e.g. various robots-involved scenarios, exhibit stochastic behavior. For example, robot wheels may slip, resulting in not executing the desired action – robot might stay at current location instead of moving, or may continue in a straight direction instead of turning.

Therefore, in our work, we focus on solving MAPF in a *stochastic* environment where a set of possible outcomes $\{(p_i, v_i, l_i)\}_i$ is assigned to each action [1]. An

outcome (p_i, v_i, l_i) means that after executing the action, with probability p_i, agent ends up in vertex v_i (which might be different from v) and executing of the action will take l_i timesteps. Such a model allows to specify a possibility of just getting delayed (i.e., with some probability, traversing the edge might take longer) as well as ending in a different vertex (e.g., continuing straight instead of turning left).

A solution to MAPF in such an environment is a set of policies $\{p_a\}_{a \in A}$. For each agent $a \in A$, a (single-agent) policy $p_a \colon V \times T \mapsto \mathit{Actions}$ prescribes which action the agent should execute in its current situation, given by its location and current time. As in classical MAPF, a collision-free solution is required. Specifically, we look for a solution that is conflict-free for any possible scenario (any possible outcomes of actions). Furthermore, we want to minimize *expected SOC* (sum of expected total travel times of individual agents).

3 Algorithm

To compute the (optimal) policy-based solution of MAPF in the stochastic environment, we proposed an algorithm DeltaPolicyCBS (DPCBS) [1]. The algorithm is a modification of the Conflict-based Search algorithm [2] for classical MAPF. CBS (and so DPCBS) is a two-level algorithm that decomposes the problem of solving MAPF into obtaining optimal single-agent solutions, for each agent individually, (low level) and deconflicting those single-agent solutions (high level).

The low level part of DPCBS is based on modeling the single-agent shortest-path problem in the stochastic environment as Markov Decision Process. If there are no constraints, the problem can be easily modeled as MDP over states corresponding to vertices V of the graph. If there are some constrains, it is necessary to use states $V \times T$ where T is a set of timesteps. However, it is possible to reuse the without-constrains solution for states (v, t) with t greater than the timestep t_{max} of latest constraint, and compute actions for the remaining states ($t \leq t_{max}$) using dynamic programming [1].

4 Improvements – Work in Progress

Experimental evaluation [1] of DPCBS showed some promising results in terms of decreasing real cost of the solution (compared to existing techniques of achieving some robustness of MAPF solution) as well as quite high complexity – the algorithm was not able to solve highly stochastic environments with higher number of agents.

We will present improvements addressing this issue, such as restricting the area (locations) involved in the policy computing, and an online modification of DPCBS that considers only conflicts in near future and handles later conflicts via replanning during execution.

Acknowledgments

Research is partially supported by the Czech Science Foundation, project No 23-05104S, and the Charles University, project GAUK No 36124 and by SVV project number 260 698.

References

[1] J. Mestek. and R. Barták. Multi-agent path finding: Policies instead of plans. In *Proceedings of the 16th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART*, pages 95–104. INSTICC, SciTePress, 2024.

[2] G. Sharon, R. Stern, A. Felner, and N. Sturtevant. Conflict-Based Search For Optimal Multi-Agent Path Finding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1):563–569, Sep. 2012.

[3] R. Stern, N. Sturtevant, A. Felner, S. Koenig, H. Ma, T. Walker, J. Li, D. Atzmon, L. Cohen, T. K. S. Kumar, E. Boyarski, and R. Barták. Multi-Agent Pathfinding: Definitions, Variants, and Benchmarks. *Symposium on Combinatorial Search (SoCS)*, pages 151–158, 2019.

On Pareto Optimal Stable Partitions in Altruistic Hedonic Games

Erika Momo
Infocom Corporation
e.momo@infocom.co.jp

Souta Ikegame
Aoyama Gakuin University
a5720006@aoyama.jp

Xiaowen Zhao
Aoyama Gakuin University
zhao@ise.aoyama.ac.jp

Shao-Chin Sung*
Aoyama Gakuin University
son@ise.aoyama.ac.jp

Abstract

In this study, we are dealing with a model of coalition formation games called altruistic hedonic coalition formation games, an extended model of hedonic games with the altruism of players imposed. We focus on Pareto optimal partitions in altruistic hedonic games and propose two searching algorithms for finding Pareto optimal partitions. By numerical experiments, we confirm the effectiveness of the proposed algorithms.

1 Introduction

Hedonic coalition formation games, or hedonic games, are models of cooperative games such that each player has preferences over all possible coalitions to which she or he can join, and an outcome of the game is a partition of the grand coalition into disjoint coalitions. As an extended model, Kerkmann *et al.* [4] (also see [3]) introduced an extended model called *altruistic hedonic games*, in which players care not only about coalitions they belong to but also the status of all their friends, regardless of whether those friends are in the same coalition to them.

Generally, a partition is *stable* if no player has an incentive to leave her or his coalition and join another one. Deviations from partitions are defined depending on the imposed stability concept. Nash stability, individual stability, core stability, strict core stability, Pareto optimality, and popularity are stability concepts commonly studied in hedonic games (e.g. see [6]).

*Shao-Chin Sung is currently known as So Akiyama.

In the following, after introducing the model of altruistic hedonic games, we will discuss how to find Pareto optimal partitions in altruistic hedonic games. More precisely, we propose two searching algorithms for finding Pareto optimal partitions. By numerical experiments, we confirm the effectiveness of the proposed algorithms.

2 Preliminaries

Let $N = \{1, 2, \ldots, n\}$ be a finite set of players. A coalition $X \subseteq N$ is a non-empty subset of N. For each $i \in N$, the collection of all coalitions in which player i belongs is denoted by $\mathcal{N}_i = \{X \subseteq N \mid i \in X\}$. A partition π of N is a collection of coalitions such that every player in N belongs to at least one coalition in π, i.e., $\bigcup_{X \in \pi} X = N$, and every player in N belongs to not more than one coalition in π, i.e., all coalitions in π are pairwise disjoint. We denote the set of all partitions of N by Π_N. For each partition $\pi \in \Pi_N$ and each player $i \in N$, we denote the coalition in π containing i by $\pi(i)$. In other words, $\pi(i)$ is the only coalition in $\pi \cap \mathcal{N}_i$.

A hedonic coalition formation game, or *hedonic game* is a pair (N, v) of the set N of players and an evaluation profile $v = (v_i)_{i \in N}$, where each $v_i : N \to \mathbb{R}$ is an evaluation function of player $i \in N$, where $v_i(j)$ is player i's evaluation of player j being together in the same coalition. It is assumed that $v_i(i) = 0$ for each $i \in N$. Each player i's utility be a member of a coalition $X \in \mathcal{N}_i$ is $v_i(X) = \sum_{j \in X} v_i(j)$.

An outcome of a hedonic game (N, v) is a partition of N, i.e., a member of Π_N. For each partition $\pi \in \Pi_N$, the utility of each player $i \in N$ under a partition π is $v_i(\pi) = v_i(\pi(i))$. From this definition of utilities, players in hedonic games care not only about the coalitions to which they belong. This model is commonly referred to as *additive hedonic games* (e.g. see [1, 2, 5]).

Recently, Kerkmann *et al.* [4] (also see [3]) introduced an extended model called *altruistic hedonic games*. Players in altruistic hedonic games care not only about the coalitions to which they belong but also the status of all their friends, regardless of whether those friends are in the same coalition. More precisely, for each $i \in N$, we denote the set of player i's friends by $F_i \subseteq N \setminus \{i\}$. For each partition $\pi \in \Pi_N$, the utility $\mu_i(\pi)$ of each player $i \in N$ under partition π is defined as follows.

$$\mu_i(\pi) = v_i(\pi) + \alpha \sum_{j \in F_i} v_j(\pi),$$

where α is a non-negative real number called the *degree of altruism*.

An altruistic hedonic game is a tuple (N, v, F, α) of the set N of players, an evaluation profile $v = (v_i)_{i \in N}$, a friend profile $F = (F_i)_{i \in N}$, and a degree of altruism $\alpha \in \mathbb{R}_{\geq 0}$. An altruistic hedonic game becomes a hedonic game when $\alpha = 0$. Kerkmann *et al.* [4] called the model, *selfish first* if $\alpha \geq 1/M$, *equal treatment* if $\alpha = 1$, and *altruistic treatment* if $\alpha \geq M$, where $M = \sum_{i \in N} \sum_{j \in N} \max\{v_i(j), 0\}$.

Let us point out that, in the model introduced by Kerkmann *et al.*, each F_i is the set of players positively evaluated by player i, and here F_i can be any subset of $N \setminus \{i\}$. In this sense, the model introduced in the paper is slightly extended from the model introduced by Kerkmann *et al.* [4].

3 Pareto Optimality

This section discusses how to construct Pareto optimal partitions in altruistic hedonic games. Aziz *et al.* [1] showed that the problem of constructing a Pareto optimal partition for additive hedonic games is weakly NP-hard.

Let $\pi, \sigma \in \Pi_N$ be two arbitrary partitions. We say that π *Pareto dominate* σ in (N, v, F, α) if $\mu_i(\pi) \geq \mu_i(\sigma)$ for each $i \in N$ and $\mu_i(\pi) > \mu_i(\sigma)$ for some $i \in N$. A partition is called *Pareto optimal* in (N, v, F, α) if it can not be Pareto dominated in (N, v, F, α) by any partition in Π_N.

Proposition 1 *Every altruistic hedonic game has at least one Pareto optimal partition.*

Proof. Let (N, v, F, α) be an arbitrary altruistic hedonic game, and let $\pi_0 \in \Pi_N$ be an arbitrary partition. If π_0 is Pareto optimal, then we are done. Otherwise, a partition π_1 exists in Π_N Pareto dominating π_0. According to the definition of Pareto domination, we have $\sum_{i \in N} \mu_i(\pi_1) > \sum_{i \in N} \mu_i(\pi_0)$. Again, if π_1 is not Pareto optimal, another partition π_2 exists in Π_n Pareto dominating π_1 and satisfying $\sum_{i \in N} \mu_i(\pi_2) > \sum_{i \in N} \mu_i(\pi_1)$.

We can repeat this process whenever the last partition π_k, for some positive integer k, is not Pareto optimal, and all partitions $\pi_0, \pi_1, \pi_2, \ldots, \pi_k$ in such a Pareto domination sequence are different. Moreover, we will find a Pareto optimal partition in a finite step since the number $|\Pi_N|$ of partitions is finite. □

A straightforward way to construct a Pareto optimal partition is to test each partition $\pi \in \Pi_N$, whether π can be Pareto dominated or not. Given a partition $\pi \in \Pi_N$, the running time for calculating the utilities $\mu_i(\pi)$ of all player $i \in N$ is $\mathcal{O}(n^2)$. The running time for performing a Pareto domination test is $O(n)$ if utilities of all players under each test subject. Then, the running time of the straightforward algorithm is $\mathcal{O}(n^2 \cdot |\Pi_N| + n \cdot |\Pi_N|^2 + E_N) = \mathcal{O}(n \cdot |\Pi_N|^2 + E_N)$, where E_N is the running time for enumerating all partitions in Π_N.

In the following, we propose an algorithm for constructing a Pareto optimal partition.

ParetoSearch(N, v, F, α):

- Let $\sigma \in \Pi_N$ be an arbitrary partition.
- Enumerate all partitions in Π_N one by one.
  - Let $\pi \in \Pi_N$ be a newly constructed partition.
  - If π Pareto dominates σ, then set $\sigma := \pi$.
- return σ.

Observe that the running time of this algorithm is $\mathcal{O}(n^2|\Pi_N| + E_N)$.

Theorem 1 *The algorithm* ParetoSearch *always returns a Pareto optimal partition.*

Proof. Let $\sigma_0 \in \Pi_N$ be an arbitrary partition used as the initial partition in the algorithm. Suppose the algorithm enumerates all the partitions in the order $(\pi_1, \pi_2, \ldots, \pi_{|\Pi_N|})$, and the partition σ is updated $m \geq 0$ times by $\pi_{j_1}, \pi_{j_2}, \ldots, \pi_{j_m}$ in the algorithm.

If $m = 0$, the initial partition is Pareto optimal, which the algorithm returns. Suppose $m > 0$, and in the following we show that π_{j_m} is Pareto optimal. To do so, we show that, for each $\ell \in \{1, 2, \ldots, m\}$, π_{j_ℓ} is not Pareto dominated by π_k for any $1 \leq k < j_\ell$.

Since π_{j_1} is the first enumerated partition by which the initial partition σ_0 is Pareto dominated, we have $\mu_i(\pi_{j_1}) \geq \mu_i(\sigma_0)$ for each $i \in N$ and $\mu_i(\pi_{j_1}) > \mu_i(\sigma_0)$ for some $i \in N$. Moreover, if $1 \leq k < j_1$, either there exists $i \in N$ such that $\mu_i(\pi_k) < \mu_i(\sigma_0)$ or $\mu_i(\pi_k) = \mu_i(\sigma_0)$ for every $i \in N$. Hence, if $1 \leq k < j_1$, there exists $i \in N$ such that $\mu_i(\pi_k) < \mu_i(\pi_{j_1})$, and hence, π_{j_1} is not Pareto dominated by π_k.

Now suppose $\ell \in \{1, 2, \ldots, k-1\}$ and π_{j_ℓ} is not Pareto dominated by π_k for any $1 \leq k < \ell$. Recall that $\pi_{j_{\ell+1}}$ is the next enumerated partition by which π_{j_ℓ} is Pareto dominated, and hence, π_{j_ℓ} is not Pareto dominated by π_k for any $1 \leq k < j_{\ell+1}$. By the same arguments above, we can conclude that $\pi_{j_{\ell+1}}$ is not Pareto dominated by π_k for any $1 \leq k < j_{\ell+1}$.

Finally, since π_{j_m} is also not Pareto dominated by π_k for any $j_m < k < |\Pi_N|$, we can conclude that π_{j_m} is Pareto optimal. □

4 Pareto Optimality and Contractual Individual Stability

Now, let us introduce the concept of contractual individual stability in hedonic games. This stability concept helps reduce the running time (not the complexity) for finding Pareto optimal partitions in altruistic hedonic games.

Let $\pi \in \Pi_N$ be an arbitrary partition, $i \in N$, and $X \in \pi \cup \{\emptyset\}$. Observe that $\pi(j) = X$ for each $j \in X$ (if $X \neq \emptyset$). We define the partition $\pi_{(i,X)}$ as follows.

$$\pi_{(i,X)} = \begin{cases} \big(\pi \setminus \{\pi(i), X\}\big) \cup \{\pi(i) \setminus \{i\}, X \cup \{i\}\} & \text{if } X \neq \pi(i), \\ \pi & \text{otherwise.} \end{cases}$$

In other words, $\pi_{(i,X)}$ is the partition, where the only change from *pi* is that player i moves from $\pi(i)$ to X. We say that (i, X) is a contractual individual deviation from π in (N, v) if

- player i has an incentive to deviate from π to $\pi_{(i,X)}$, i.e., $v_i(\pi_{(i,X)}) > v_i(\pi)$,

- player i has permission to leave $\pi(i)$ from all another players in $\pi(i)$, i.e., $v_j(\pi_{(i,X)}) \geq v_j(\pi)$ for each $j \in \pi(i) \setminus \{i\}$, and
- player i has permission to join X from all players in X, i.e., $v_j(\pi_{(i,X)}) \geq v_j(\pi)$ for each $j \in X$.

A partition $\pi \in \Pi_N$ is contractual individually stable in (N, v) if π has no contractual individual deviation.

Proposition 2 *Every hedonic game has at least one contractual individual stable partition.*

Proof. One can prove this claim by a similar argument as Proposition 1. The length of any sequence of contractual individual deviations starting from any partition is finite, and the last partition of a maximal contractual individual deviation sequence is contractual individual stable. □

The relation between Pareto optimality in altruistic hedonic games and contractual individual stability in hedonic games is as follows.

Lemma 1 *Pareto optimality in every altruistic hedonic game implies contractual individual stability in the corresponding hedonic game.*

Proof. Let $\pi \in \Pi_N$ be a partition, and (i, X) be a contractual individual deviation from π in (N, v). Hence, π is not contractual individual stable in (N, v). By definition, we have $v_j(\pi_{(i,X)}) \geq v_j(\pi)$ for every $j \in N$, and $v_i(\pi_{(i,X)}) > v_i(\pi)$. Moreover, from the non-negativity of α, we have $\mu_j(\pi_{(i,X)}) \geq \mu_j(\pi)$ for every $j \in N$, and $\mu_i(\pi_{(i,X)}) > \mu_i(\pi)$ as well. Therefore, π is not Pareto optimal in (N, v, F, α). □

We are ready to introduce a new version of our proposed algorithm for constructing a Pareto optimal partition based on enumerating contractual individual stable partitions. In the following, by Π_N^{CIS}, we denote the set of all contractual individual stable partitions.

ParetoSearch_CIS(N, v, F, α):

- Let $\sigma \in \Pi_N$ be an arbitrary partition.
- Enumerate all partitions in Π_N^{CIS} one by one.
  - Let $\pi \in \Pi_N^{CIS}$ be a newly constructed contractual individual stable partition.
  - If π Pareto dominates σ, then set $\sigma := \pi$.
- return σ.

Observe that the running time of this algorithm is $\mathcal{O}(n^2 \cdot |\Pi_N^{CIS}| + E_N^{CIS})$, where E_N^{CIS} is the running time for enumerating all partitions in Π_N^{CIS}. Generally, the number $|\Pi_N^{CIS}|$ of all contractual individual stable partitions is much smaller than the number $|\Pi_N|$ of all partitions. Therefore, the running time of this algorithm is shorter than the previous version based on enumerating all partitions in Π_N.

Theorem 2 *The algorithm* ParetoSearch_CIS *always returns a Pareto optimal partition.*

Proof. By a similar argument of Theorem 1, the algorithm returns a contractual individual stable partition σ^*, not Pareto dominated by any other contractual individual stable partition.

It follows that σ^* is not Pareto dominated by any partition, which is not contractual individual stable. Let π be a partition that is not contractual individual stable. Then, find the last partition π' of a maximal contractual individual deviation sequence starting from π. Recall that such a partition π^* is a contractual individual stable, and π^* Pareto dominates π.

Since σ^* is not Pareto dominated by π^*, σ^* is also not Pareto dominated by π. Therefore, σ^* is not Pareto dominated by any other partition, and equivalently, σ^* is Pareto optimal. □

5 Numerical Experiments

We implement our proposed algorithms with Java, and we conduct numerical experiments to test the effectiveness of our proposed algorithm. The following table and figure show the comparison of running times of ParetoSearch and ParetoSearch_CIS.

Table 1: Comparison of straightforward algorithm (ST), ParetoSearch (PS), and ParetoSearch_CIS (PSC).

n	instance	ST(ms)	PS(ms)	PSC(ms)
5	10000	0.01	0.01	0.01
6	10000	0.05	0.04	0.03
7	10000	0.19	0.15	0.11
8	10000	1.23	0.78	0.39
9	10000	12.58	3.81	1.98
10	10000	235.74	21.32	8.00
11	3083	8447.74	122.84	45.04
12	166	72245.60	788.05	217.87
13	16	1123291.29	6029.06	1276.00

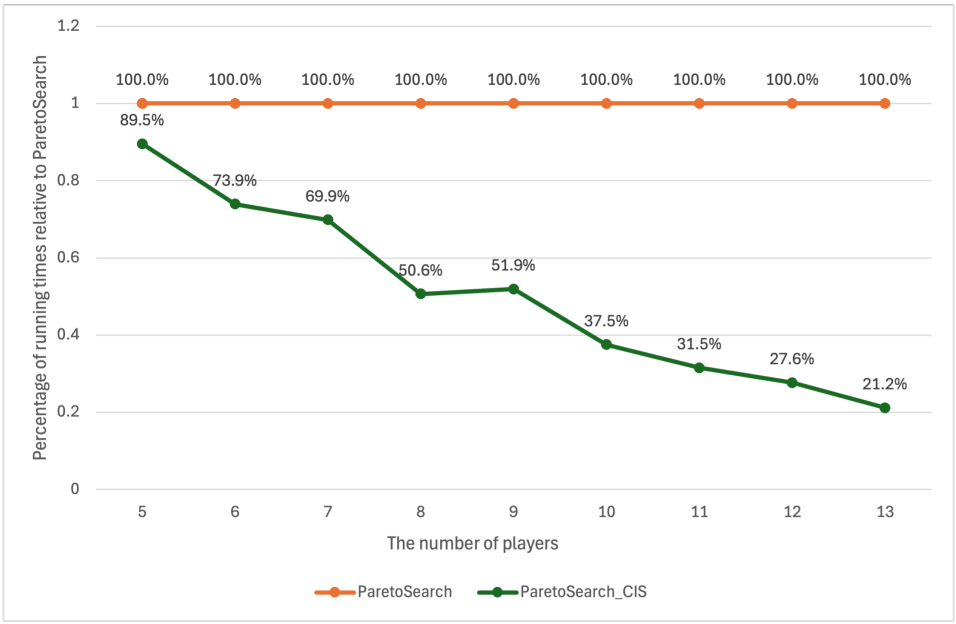


Figure 1: Comparison of ParetoSearch and ParetoSearch_CIS

References

[1] H. Aziz, F. Brandt, and H. G. Seedig. Computing desirable partitions in additively separable hedonic games. *Artificial Intelligence*, 195:316–334, 2013.

[2] A. Bogomolnaia and M. . Jackson. The stability of hedonic coalition structures. *Games and Economic Behavior*, 38(2):201–230, 2001.

[3] A. M. Kerkmann, S. Cramer, and J. Rothe. Altruism in coalition formation games. *Annals of Mathematics and Artificial Intelligence*, 92:601–629, 2024.

[4] A. M. Kerkmann, N.-T. Nguyen, A. Rey, L. Rey, J. Rothe, L. Schend, and A. Wiechers. Altruistic hedonic games. *Journal of Artificial Intelligence Research*, 75:129–169, 2022.

[5] S. C. Sung and D. Dimitrov. On core membership testing for hedonic coalition formation games. *Operations Research Letters*, 35(2), 2007.

[6] S. C. Sung and D. Dimitrov. On myopic stability concepts for hedonic games. *Theory and Decision*, 62, 2007.

Intermediate quantifiers and the problems of non-monotonic logic

Vilém Novák and Petra Murinová
Institute for Research and Applications of Fuzzy Modeling
University of Ostrava, 30. dubna 22 701 03 Ostrava 1, Czech Republic
e-mail: {vilem.novak,petra.murinova}@osu.cz

1 Introduction

In the previous publications, we focused on a formal theory of intermediate quantifiers e.g. "Most", "Several", "Almost all", etc., and non-trivial syllogisms with them (see [6]) that are part of *fuzzy natural logic* (FNL). In this paper we will show that our theory is also capable of solving problems of non-monotonic logic (cf. [2]).

Non-monotonicity is the following feature of commonsense reasoning: we formulate a theory determined by axioms that are "normally true". However, a new formula(s) may occur that we apparently consider to be true as well but which leads to a conclusion contradicting the original theory. We want to update the original theory to contain both the original conclusions as well as the new ones. The latter, however, contradict the original ones. What should we do?

A typical example of nonmonotonic reasoning is the classical *bird-penguin problem*. We start with the commonsense knowledge "All birds can fly". A new information tells us that "Some birds are penguins" and we know that "No penguins can fly". Joining the latter with the former leads in classical logic to a contradiction.

We argue that "All" in commonsense reasoning, in fact, means "Most" or "Almost all". Classical logic suggests no convincing solution how to express it. In this paper we prove that replacing the classical quantifier $\forall$ by a vague intermediate quantifier *Most* the above problem disappears. Hence, if we modify the original default knowledge into (*Most* x)(Bird x, CanFly x) ("*most* birds can fly") then we can prove formally in FNL that the additional information saying that "Penguins are birds which cannot fly" does not lead to a contradiction and so, the resulting theory is consistent. The reason consists in the vagueness of linguistic expressions used in common-sense reasoning and their appropriate mathematical model within FNL. We will show that also non-trivial intermediate syllogisms allow consistent revision by a new information. Note that nonmonotonic reasoning is related to the properties of sub/contrary in the graded square of opposition analyzed in [5].

2 Preliminaries

The theory of intermediate quantifiers is a special formal theory T^{IQ} in fuzzy type theory (FTT), see [7].

An intermediate quantifier of type $\langle 1,1\rangle$ is one of the following formulas:

$$(Q^{\forall}_{Ev}\, x)(B,A) \equiv (\exists z)[(\forall x)((B|z)\, x \Rightarrow Ax) \wedge Ev((\mu B)(B|z))], \tag{1}$$

$$(Q^{\exists}_{Ev}\, x)(B,A) \equiv (\exists z)[(\exists x)((B|z)x \wedge Ax) \wedge Ev((\mu B)(B|z))]. \tag{2}$$

where Ev is an evaluative linguistic expression (see [8] and Figure 1), μ is a measure on fuzzy sets and $B|z$ is an operation of a *cut* of a fuzzy set B using z.

In this paper, we will consider some (arbitrary) model $\mathcal{M}$ of T^{IQ} and a selected universe M. We will use special notation for the interpretation of the formula $Ev(\mu(B)(B|z))$, namely, if $\mathcal{M}$ is a model and $B, Z \subsetneq M$ are are fuzzy sets then

$$F^{R}_{Ev}(B, B|Z) = \mathcal{M}(Ev(\mu(B)(B|z))).$$

where the cut $B|Z$ is a fuzzy subset of B consisting of all singletons ${}^{a}/_{x}$, $x \in M$ that are equal to the corresponding singletons of Z.

A semantic interpretation of intermediate quantifiers (1) and (2) are truth values given by the respective formulas

$$Q^{\forall}_{Ev}(B,A) = \bigvee \left\{ \bigwedge_{u\in N} ((B|Z)(u) \to A(u)) \wedge F^{R}_{Ev}(B,B|Z) \mid Z \in \mathcal{F}(N) \right\}, \tag{3}$$

$$Q^{\exists}_{Ev}(B,A) = \bigvee \left\{ \bigvee_{u\in N} ((B|Z)(u) \wedge A(u)) \wedge F^{R}_{Ev}(B,B|Z) \mid Z \in \mathcal{F}(N) \right\}. \tag{4}$$

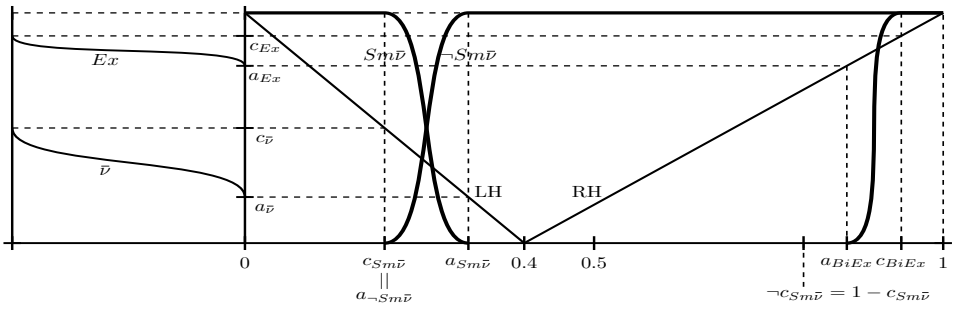


Figure 1: Extensions of evaluative expressions $Sm\,\bar{\boldsymbol{\nu}}$, $\neg\, Sm\,\bar{\boldsymbol{\nu}}$ and $BiEx$ in the context $[0, 0.4]\cup[0.4, 1]$ with marked interpretation of points $\mathbf{a}_{\neg\, Sm\,\bar{\boldsymbol{\nu}}}$, $\mathbf{a}_{BiEx}$ and $\neg\mathbf{c}_{Sm\,\bar{\boldsymbol{\nu}}}$.

In this paper, we will consider the following intermediate quantifiers:

$$\textbf{A: All } B \text{ are } A := (Q^{\forall}_{Bi\boldsymbol{\Delta}}x)(B,A) \equiv (\forall x)(Bx \Rightarrow Ax),$$

$$\textbf{E: No } B \text{ are } A := (Q^{\forall}_{Bi\boldsymbol{\Delta}}x)(B,\neg A) \equiv (\forall x)(Bx \Rightarrow \neg Ax),$$

$$\textbf{P: Almost all } B \text{ are } A := (Q^{\forall}_{Bi\,Ex}x)(B,A)$$

$$\textbf{B: Almost all } B \text{ are not } A := (Q^{\forall}_{Bi\,Ex}x)(B,\neg A)$$

$$\textbf{T: Most } B \text{ are } A := (Q^{\forall}_{Bi\,Ve}x)(B,A)$$

$$\textbf{D: Most } B \text{ are not } A := (Q^{\forall}_{Bi\,Ve}x)(B,\neg A)$$

$$\textbf{K: Many } B \text{ are } A := (Q^{\forall}_{\neg Sm}x)(B,A)$$

$$\textbf{G: Many } B \text{ are not } A := (Q^{\forall}_{\neg Sm}x)(B,\neg A)$$

$$\textbf{I: Some } B \text{ are } A := (Q^{\exists}_{Bi\boldsymbol{\Delta}}x)(B,A) \equiv (\exists x)(Bx \wedge Ax),$$

$$\textbf{O: Some } B \text{ are not } A := (Q^{\exists}_{Bi\boldsymbol{\Delta}}x)(B,\neg A) \equiv (\exists x)(Bx \wedge \neg Ax)$$

where $\boldsymbol{\Delta}$ is the Baaz delta keeping the membership degree 1 and sending all lower membership degrees to 0.

3 Nonmonotonicity and intermediate quantifiers

3.1 Bird-penguin problem

We start with a commonsense knowledge

(i) $(All\ x)(\text{Bird } x, \text{CanFly } x)$ ("all birds can fly").

(ii) $(Some\ x)(\text{Bird } x, \text{Penguin } x)$ ("some birds are penguins").

(iii) $(\forall x)(\text{Penguin } x \Rightarrow \neg\text{CanFly } x)$ ("no penguins can fly").

Axiom (ii) is a new information causing that such a theory is contradictory.

However, we can modify the original default knowledge (i) by

(i') $(Most\ x)(\text{Bird } x, \text{CanFly } x)$ ("*most* birds can fly")

which is a natural modification since we know that most birds fly but not all. Then we can prove that the new theory is consistent.

Theorem 1 *Let us consider an extension T of the theory of intermediate quantifiers T^{IQ} and Bird , Penguin , CanFly $\in Form_{o\alpha}$ be formulas. Let*

$$T \vdash (Most\ x_\alpha)(Bird\, x, CanFly\, x), \tag{5}$$

$$T \vdash (\exists x_\alpha)\boldsymbol{\Delta}(Bird\, x \wedge Penguin\, x), \tag{6}$$

$$T \vdash (\forall x_\alpha)(Penguin\, x \Rightarrow \neg CanFly\, x). \tag{7}$$

Then $T \vdash (\exists x_\alpha)(Bird\, x \wedge \neg CanFly\, x)$ and T is consistent.

Hence, we proved formally in our theory that additional information saying that "Penguins are birds which cannot fly" does lead to a contradiction. The reason consists in the vagueness of expressions of natural language that are used in commonsense reasoning. In our case, we replace the precise quantifier $\forall$ by a vague intermediate quantifier *Most* which allows exceptions and so, the above nonmonotonicity problem disappears.

3.2 Default rules

The reasoning above is related to default logic. The defaults are represented as inference rules rather than object language formulas. Reiter in [9] argues that defaults cannot be reasoned about within the logic. For example, he discusses the following example from [3]: "Normally canaries are yellow" and "Yellow things are never green" and he argues that we cannot conclude "Normally canaries are never green". McDermott and Doyle's suggestion is to have the following sequence of inferences:

- $(\forall x)(\mathrm{CANARY}(x)\&M\,\mathrm{YELLOW}(x) \Rightarrow \neg\mathrm{YELLOW}(x))$,
- $(\forall x)(\mathrm{YELLOW}(x) \Rightarrow \neg\mathrm{GREEN}(x))$,
- we can infer that $(\forall x)(\mathrm{CANARY}(x)\&M\,\mathrm{YELLOW}(x) \Rightarrow \neg\mathrm{GREEN}(x))$

where MA is intended to mean "A is consistent". Reiter in [9] casts doubts, whether the last formula can legitimately be interpreted to mean "Normally canaries are not green".

We argue that in our theory of intermediate quantifiers and syllogisms this is possible, has a good sense and does not lead to a contradiction. Indeed, what does it mean

$$\text{"Normally } Ax \text{ are } Bx\text{"?} \tag{8}$$

This sentence immediately suggests that there can be abnormal situations in which Ax are not Bx. But only exceptionally and otherwise we expect Ax followed by Bx to happen in *most* cases. Hence, we naturally understand sentence (8) either as "Most Ax are Bx", or "Almost all Ax are Bx", but *not* as "All Ax are Bx".

A short analysis reveals that the previous example can be modeled in our theory as follows:

$$\begin{array}{ll} \mathbf{E}: & (\forall x)(\mathrm{Yellow}\,x \Rightarrow \neg\mathrm{Green}\,x), \\ \mathbf{T}: & (\mathrm{Most}\,x)(\mathrm{Canary}\,x, \mathrm{Yellow}\,x), \\ \hline \mathbf{D}: & (\mathrm{Most}\,x)(\mathrm{Canary}\,x, \neg\mathrm{Green}\,x). \end{array}$$

This is a *special case* of the *valid syllogism* **ETD**-I of the first figure. Its validity was proved in [4].

3.3 Sorites/falakros paradoxes

We argue that the classical sorites/falakros paradoxes are also cases of nonmonotonic reasoning. The following is a solution of the paradoxes using intermediate quantifier "Not many".

Theorem 2 *Let $\sigma = o(o\epsilon)$ be a type for natural numbers in fuzzy type theory (cf. [1, Chapter 6]) and T be a consistent extension of the theory of intermediate quantifiers T^{IQ} that contains Peano arithmetics. Let 0, $n, m_0 \in Form_\sigma$ represent natural numbers and $\mathbb{FN} \in Form_{o\sigma}$. If the following formulas are provable*

$$T \vdash \mathbb{FN}\,0, \tag{9}$$

$$T \vdash (Not\ many\ n)((\mathbb{FN}\,n), (\mathbb{FN}(n+1) \mathbin{\mathsf{V}} (n \equiv m_0))), \tag{10}$$

$$T \vdash (\exists n)\neg(\mathbb{FN}\,n). \tag{11}$$

then T is consistent.

It is interesting in the proof of this theorem that we may consider $\mathbb{FN}$ to be a *finite set* of natural numbers, i.e., a set which contains all numbers n saying, that, e.g., a person having n hairs is bald. Moreover, the implication

$$(\mathbb{FN}\,n) \Rightarrow (\mathbb{FN}(n+1) \mathbin{\mathsf{V}} (n \equiv m_0))$$

is true for all $n \in \mathbb{N}$. The number m_0 is a number of, e.g., hairs on one's head for which we know that he/she is *not bald*. This is the number assured by axiom (11).

Interpretation of formula (10) can be rewritten as

$$\bigvee_{Z \subseteq \mathbb{FN}} \Big(\bigwedge_{n \in \mathbb{N}} ((\mathbb{FN}\,|Z)(n) \to \mathbb{FN}(n+1) \vee [n \equiv m_0]) \wedge F^R_{Sm\,\bar{\boldsymbol{\nu}}}(\mathbb{FN}, \mathbb{FN}\,|Z)) \tag{12}$$

where $[n \equiv m_0] \in \{0, 1\}$ is a truth value of the equality $n = m_0$.

We can see from (12) that the quantifier "Not many" contains the value

$$F^R_{Sm\,\bar{\boldsymbol{\nu}}}(\mathbb{FN}, \mathbb{FN}\,|Z) \in [0, 1]$$

which accompanies each step n. If the number of steps is too large, i.e., $|\,\mathbb{FN}\,| \geq m_0$, then $F^R_{Sm\,\bar{\boldsymbol{\nu}}}(\mathbb{FN}, \mathbb{FN}\,|Z) = 0$. This value can be taken as a characterization of how much dubious is the implication $\mathbb{FN}(n) \Rightarrow \mathbb{FN}(n+1)$. Namely, $F^R_{Sm\,\bar{\boldsymbol{\nu}}}(\mathbb{FN}, \mathbb{FN}\,|Z) = 1$ means that the latter implication is not dubious at all and $F^R_{Sm\,\bar{\boldsymbol{\nu}}}(\mathbb{FN}, \mathbb{FN}\,|Z) = 0$ means, that it is completely dubious.

4 Conclusion

In this paper we pointed out some known problems of nonmonotonic logic and demonstrated that they can be overcome in the theory of intermediate quantifiers. Namely, it follows from vagueness of their semantics that existence of exceptions to default rules does not lead to contradiction and we need no special modification of the formal fuzzy logic. We also demonstrated that the classical sorites/falakros paradoxes are also examples of nonmonotonic reasoning.

Acknowledgement

This article has been supported by the project Research of Excellence on Digital Technologies and Wellbeing CZ.02.01.01/00/22-008/0004583 which is co-financed by the European Union.

References

[1] P. Andrews. *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*. Kluwer, Dordrecht, 2002.

[2] A. Bochman. *A Logical Theory of Nonmonotonic Inference and Belief Change*. Springer, Berlin, 2001.

[3] D. J. McDermott, D. Nonmonotonic logic I. *Artif. Intell.*, 25:41–72, 1980.

[4] P. Murinová and V. Novák. A formal theory of generalized intermediate syllogisms. *Fuzzy Sets and Systems*, 186:47–80, 2012.

[5] P. Murinová and V. Novák. Analysis of generalized square of opposition with intermediate quantifiers. *Fuzzy Sets and Systems*, 242:89–113, 2014.

[6] P. Murinová and V. Novák. The theory of intermediate quantifiers in fuzzy natural logic revisited and the model of "Many". *Fuzzy Sets and Systems*, 388:56–89, 2020.

[7] V. Novák. On fuzzy type theory. *Fuzzy Sets and Systems*, 149:235–273, 2005.

[8] V. Novák. A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets and Systems*, 159(22):2939–2969, 2008.

[9] R. Reiter. Nonmonotonic reasoning. *Ann. Rev. Comput. Sci.*, 2:147–186, 1987.

Inverse F-Transform in Light of Kolmogorov's Superposition Theorem

Irina Perfilieva
University of Ostrava
Institute for Research and Applications of Fuzzy Modeling
irina.perfilieva@osu.cz

1 Introduction

The purpose of this contribution is to show the connection and, in a sense, the similarity between the known classical and non-classical results on the representation and approximation of a continuous function.

2 Preliminaries

Theorem 1 (Kolmogorov, Arnold, Kahane, Lorentz, Sprecher) *For any $n \in \mathbb{N}$, $n \geq 2$, there exist real numbers $\lambda_1, \ldots, \lambda_n$ and continuous functions $\phi_k : \mathbb{I} \to \mathbb{R}$, $k = 1, \ldots, 2n+1$, where $\mathbb{I} = [0,1]$, with the property that for every continuous function $f : \mathbb{I}^n \to \mathbb{R}$ there exists a continuous function $g : \mathbb{R} \to \mathbb{R}$ such that for each $(x_1, \ldots, x_n) \in \mathbb{I}^n$,*

$$f(x_1, \ldots, x_n) = \sum_{k=1}^{2n+1} g(\lambda_1 \phi_k(x_1) + \cdots + \lambda_n \phi_k(x_n)). \tag{1}$$

The cited above theorem (see it in e.g., [1, 3, 4]) is an answer to Hilbert's 13th problem, in which he conjectured that there is a continuous function $f : \mathbb{I}^3 \to \mathbb{R}$, where $\mathbb{I} = [0,1]$, which cannot be expressed in terms of composition and addition of continuous functions from $\mathbb{R}^2 \to \mathbb{R}$, that is, as composition and addition of continuous real-valued functions of two variables. It took more than 50 years to prove Hilbert's conjecture false.

2.1 Kolmogorov superposition theorem - Necessary Details

In this section, we present one of the generalizations of Kolmogorov superposition theorem and outline its proof proposed by the Swedish mathematician Torbjörn

Hedberg, which uses the work of George Gunter Lorentz, Jean-Pierre Kahane and David A. Sprecher.

Below, we reproduce some necessary details of the proof for the $n = 2$ variable case taken from [4]. Precisely, the proof will be given for a continuous function $f : \mathbb{I}^2 \to \mathbb{R}$ and the following representation formula:

$$f(x_1, x_2) = \sum_{k=1}^{5} g(\phi_k(x_1) + \lambda\phi_k(x_2)), \tag{2}$$

where only the function g depends on f, in contrast to λ and ϕ_k.

We recall that the set $C(\mathbb{I})$ of all continuous functions from $\mathbb{I}$ into $\mathbb{R}$ is a complete metric space and it is a Banach space with the norm $\|f\| = \sup_{x\in[0,1]} f(x)$. Denote $\Phi_k(x_1, x_2) = \phi_k(x_1) + \lambda\phi_k(x_2)$ and rewrite (2) as follows:

$$f(x_1, x_2) = \sum_{k=1}^{5} g(\Phi_k(x_1, x_2)). \tag{3}$$

Lemma 1 *There exists a real number λ such that for any $x_1, x_2, y_1, y_2 \in \mathbb{Q}$[1],*

$$x_1 + \lambda y_1 = x_2 + \lambda y_2, \Rightarrow x_1 = x_2, y_1 = y_2.$$

The following lemma [4] gives a constructive (albeit preliminary) description of the functional parameters, which, after substituting these parameters into the right-hand side of the equality (2), makes it approximate.

Lemma 2 *We fix λ satisfying Lemma 1 and choose a function $f \in C(\mathbb{I}^2)$ such that $\|f\| = 1$. We define a set $U_f \subseteq [C(\mathbb{I})]^5$ such that $(\phi_1, \ldots, \phi_5) \in U_f$ if and only if there exists $g \in C(\mathbb{R})$, such that*

$$|g(t)| \leq \frac{1}{7} \text{ for } t \in \mathbb{R}, \tag{4}$$

and

$$|f(x, y) - \sum_{k=1}^{5} g(\phi_k(x) + \lambda\phi_k(y))| < \frac{7}{8} \text{ for } (x, y) \in \mathbb{I}^2. \tag{5}$$

Then U_f is an open dense subset of $[C(\mathbb{I}^5)]$.

Below we repeat (following [4]) all necessary details of the proof.

2.1.1 Cover of the set $\mathbb{I}$

Let us start with the describing special functions $(\phi_1, \ldots, \phi_5) \in [C(\mathbb{I})]^5$.

1) First, we will choose a sufficiently large $N \in \mathbb{N}$, which we will refine later.

[1] $\mathbb{Q}$ denotes the set of rational numbers

2) We then create a *cover* of $\mathbb{I}$ using five sets $\mathcal{P}_1, \ldots, \mathcal{P}_5$ of subintervals:

$$\mathbb{I} = \bigcup_{k=1}^{5} \mathcal{P}_k, \tag{6}$$

where $\mathcal{P}_k$ consists of all subintervals of $\mathbb{I}$ that remain after all intervals $[\frac{s}{N}, \frac{s+1}{N}]$ with $0 \leq s < N$, $s \equiv k-1 \,(\mathrm{mod}\, 5)$ are deleted. These remaining N_k intervals, which we will consider closed, will be called *intervals of rank k*.

3) For each $k \in \{1, \ldots, 5\}$, we define $\phi_k : \mathbb{I} \to \mathbb{R}$, as any continuous function satisfying the following requirements:

- ϕ_k is a constant equal to a rational number on each interval of rank k in $\mathcal{P}_k$;
- $\phi_k(x) \neq \phi_k(y)$ for x and y in different intervals of rank k;
- $\phi_k(x) \neq \phi_j(z)$ for x in any interval of rank k and z in any interval of rank j, $k \neq j$.

4) The set of all quintuples $(\phi_1, \ldots, \phi_5)$ of functions satisfying the conditions, given in steps 1 - 3, will be denoted by U_N. It is clear that the union $\bigcup_{n>N} U_n$ is an open dense subset of $[C(\mathbb{I})]^5$.

2.1.2 Cover of the set $\mathbb{I}^2$

Let $\mathbb{I}$ be covered by subintervals according to (6), such that for each $k = 1, \ldots, 5$, $\mathcal{P}_k$ consists of the intervals of rank k. The length of any interval of rank k does not exceed the value of $\frac{4}{N}$. The Cartesian product of two intervals of rank k (one lying in $\{0 \leq x \leq 1\}$ and one lying in $\{0 \leq y \leq 1\}$, will be called a *rectangle of rank k*. Rectangles of rank k will be denoted by $R_{k,1}, R_{k,2}, \ldots$. The (finite) union of all rectangles of all ranks k, $k = 1, \ldots, 5$ forms a *cover* of $\mathbb{I}^2$.

We observe that the squared Euclidean distance between any two points (x, y) and (x', y') in any rectangle of rank k is less than (or equal to) $\frac{32}{N^2}$.

Continuing with the description of functions that fulfill (4) and (5), we define

$$\Phi_k(x, y) = \phi_k(x) + \lambda\phi_k(y),\ k = 1, \ldots, 5, \tag{7}$$

where $\Phi_k : \mathbb{I}^2 \to \mathbb{R}$. By the properties of functions ϕ_k, we have that each function Φ_k is constant on each rectangle $R_{k,j}$, $j = 1, \ldots$, of rank k, and by Lemma 1, the (constant) value of Φ_k on $R_{k,j}$, denoted by $\Phi_{k,j}$, does not equal to the (constant) value $\Phi_{k',j'}$ of $\Phi_{k'}$ on $R_{k',j'}$, if $k \neq k'$ and $j \neq j'$.

Now to prove (4) and (5) we choose $N_f \in \mathbb{N}$ such that

$$|f(x, y) - f(x', y')| < \frac{1}{56} \text{ if } (x - x')^2 + (y - y')^2 \leq \frac{32}{N_f^2}. \tag{8}$$

Let us substitute the value N_f for N at step 1) of the procedure for describing special functions $(\phi_1, \ldots, \phi_5) \in [C(\mathbb{I})]^5$.

Finally, we define $g : \mathbb{R} \to \mathbb{R}$ on the set $\{\Phi_{k,j} \mid k = 1, \ldots, 5; j = 1, \ldots\}$:

$$g(\Phi_{k,j}) = \begin{cases} \frac{1}{7}. & \text{if } f(x,y) > 0 \text{ for all } x, y \in R_{k,j}, \\ -\frac{1}{7}. & \text{if } f(x,y) < 0 \text{ for all } x, y \in R_{k,j}, \end{cases} \tag{9}$$

and extend g to the whole $\mathbb{R}$ in a piecewise-linear fashion so that $|g(t)| \leq \frac{1}{7}$ for all $t \in \mathbb{R}$. Obviously, (4) is satisfied. Moreover, it can be proved (see e.g.,[4]) that (5) is satisfied as well.

It is important to remark that (9) defines function g on a finite set of points, i.e., on the set $\{\Phi_{k,j} \mid k = 1, \ldots, 5; j = 1, \ldots\}$ using only two values: $\frac{1}{7}$ and $-\frac{1}{7}$.

2.1.3 Eliminating the dependence of Φ_k on f

The next step in proving the validity of the representation (2) (in its equivalent form (3)) is to confirm inequalities similar to (4) and (5), but in which the functions $\phi_1, \ldots, \phi_5$ do not depend on f. For this purpose, a technique based on the Baire category theorem [2] is used. The following lemma is proved in [4]:

Lemma 3 *Let λ satisfy Lemma 1. There exist functions $\phi_1, \ldots, \phi_5 \in C(\mathbb{I})$, such that given $f \in C(\mathbb{I}^2)$ there exists $g \in C(\mathbb{R})$, such that*

$$|g(t)| \leq \frac{1}{7}\|f\|, \, t \in \mathbb{R}, \tag{10}$$

and

$$\|f - \sum_{k=1}^{5} g \circ \Phi_k\| < \frac{8}{9}\|f\|, \tag{11}$$

where $\Phi_k(x,y) = \phi_k(x) + \lambda\phi_k(y)$, $k = 1, \ldots, 5$.

To prove Lemma 3, we again assume that $\|f\| = 1$ and (following [4]) choose a sequence of functions $h_1, h_2, \ldots$ from $C(\mathbb{I}^2)$ (from the unit sphere of $C(\mathbb{I}^2)$) such that the set $\{h_j : j \in \mathbb{N}\}$ is dense in the unit sphere of $C(\mathbb{I}^2)$. Therefore, there exists $m \in \mathbb{N}$ such that $\|f - h_m\| \leq \frac{1}{72}$.

By the assertion in section 2.1, each function h_j defines the set $U_{h_j} \subseteq (C(\mathbb{I}))^5$. Since each such U_{h_j} is a dense open subset of the complete metric space $(C(\mathbb{I}))^5$, then by the Baire category theorem their intersection is non-empty. Denote this intersection as V and choose $(\phi_1, \ldots, \phi_5) \in V$. For the chosen function h_m and for $(\phi_1, \ldots, \phi_5) \in V \subseteq U_{h_m}$ there exists a continuous function g, $\|g\| \leq \frac{1}{7}$ such that

$$\|h_m - \sum_{k=1}^{5} g \circ \Phi_k\| < \frac{7}{8}.$$

[2] *The Baire category theorem, see [?].* Let (X, d) be a complete metric space. If $X_1, X_2, \ldots$ is a sequence of open dense subsets of X, then the set $\bigcap_{n=1}^{\infty} X_n$ is also dense in X.

Therefore,

$$\|f - \sum_{k=1}^{5} g \circ \Phi_k\| \leq \|f - h_m\| + \|h_m - \sum_{k=1}^{5} g \circ \Phi_k\| < \frac{1}{72} + \frac{7}{8} = \frac{8}{9}.$$

2.1.4 Completing the proof of Kolmogorov superposition theorem

Having all the necessary technical details, we can formulate the main statement (for the case of $n = 2$ variables), known as Kolmogorov superposition theorem [3]. For the original proofs, we refer to [1, 3, 4].

Theorem 2 (A. N. Kolmogorov) *There exist a real number λ and continuous functions $\phi_k : \mathbb{I} \to \mathbb{R}$, $k = 1, \ldots, 5$, where $\mathbb{I} = [0, 1]$, which have the property that for every continuous function $f : \mathbb{I}^2 \to \mathbb{R}$ there is a continuous function $g : \mathbb{R} \to \mathbb{R}$ such that for all $(x_1, x_2) \in \mathbb{I}^2$,*

$$f(x_1, x_2) = \sum_{k=1}^{5} g(\phi_k(x_1) + \lambda\phi_k(x_2)). \tag{12}$$

3 Inverse F-Transform in the Kolmogorov Superposition Theorem

The purpose of this section is to combine two areas of research: classical functional analysis and a part of fuzzy modeling based on the functional representation of fuzzy sets with $[0, 1]$-valued membership functions. In particular, we aim to show that a constructive proof of Lemma 2 can be given in a functional space with a fuzzy partition, using analytical form known as the inverse fuzzy transform [5].

3.1 Fuzzy Partition of $\mathbb{I}$

The concept of *fuzzy partition* does not have a unique definition in the fuzzy literature. Without going into details, we will focus on the evolution of this concept in connection with the theory of fuzzy transform (see [5, 2, 8]).

A fuzzy partition of the real interval $[a, b]$ with the *Ruspini condition* was introduced in [5] as a collection of bell-shaped fuzzy sets $A_1, \ldots A_n : [a, b] \to [0, 1]$, $n \geq 2$, with continuous membership functions, such that for all $x \in [a, b]$,

$$\sum_{k=1}^{n} A_k(x) = 1.$$

This partition can be characterized as a "partition-of-unity".

A fuzzy partition with the *generalized Ruspini condition* was introduced in [8]. The generalization consists in replacing the "partition-of-unity" by a "fuzzy

r-partition", $r \geq 2$, where for all $x \in [a, b]$, the above given condition changes to

$$\sum_{k=-r+2}^{n+r-1} A_k(x) = r.$$

In [7] a *generalized fuzzy partition* without the Ruspini condition was proposed to obtain a better approximation using an inverse fuzzy transform. Below, in Definition 1, we reproduce this definition and apply it to the interval $\mathbb{I} = [0, 1]$.

Definition 1 *Let $\mathbb{I} = [0, 1]$, $n \geq 2$, and let $x_1, \ldots, x_n$ be nodes such that $0 \leq x_1 < \ldots < x_n \leq 1$. Let $\mathbb{I}$ be covered by the intervals $[x_k - h'_k, x_k + h''_k] \subseteq [0, 1]$, $k = 1, \ldots, n$, such that their left and right margins $h'_k, h''_k \geq 0$ fulfill $h'_k + h''_k > 0$.*

The basic functions[3] *$A_1, \ldots, A_n : \mathbb{I} \to [0, 1]$ constitute a* generalized fuzzy partition *of $[0, 1]$ (with nodes $x_1, \ldots, x_n$ and margins h'_k, h''_k, $k = 1, \ldots, n$) if for every $k = 1, \ldots, n$, the following three conditions are fulfilled:*

1. *(locality) — $A_k(x) > 0$ if $x \in (x_k - h'_k, x_k + h''_k)$, and $A_k(x) = 0$ if $x \in [0, 1] \setminus (x_k - h'_k, x_k + h''_k)$;*
2. *(continuity) — A_k is continuous on $[x_k - h'_k, x_k + h''_k]$;*
3. *(covering) — for $x \in \mathbb{I}$, $\sum_{k=1}^{n} A_k(x) > 0$.*

Below we will use the covering (6) of the interval $\mathbb{I}$ proposed in section 2.1.1 and establish the corresponding fuzzy partition of the interval $\mathbb{I}$. To do this, let us return to section 2.1.1, where (for sufficiently large $N \in \mathbb{N}$) the mentioned covering was described. Recall that it consists of five sets $\mathcal{P}_1, \ldots, \mathcal{P}_5$ such that each set $\mathcal{P}_k$, $k = 1, \ldots, 5$, is a finite set of closed subintervals of rank k. For technical reasons, we are making slight changes to the definition of these sets. In what follows, the set $\mathcal{P}_k$ will consist of all subintervals of $\mathbb{I}$ remaining after removing all open intervals $(\frac{s}{N}, \frac{s+1}{N})$ from $0 \leq s < N$, $s \equiv k - 1 \,(\mathrm{mod}\, 5)$. The remaining closed intervals will be called intervals of rank k. For convenience, we denote intervals of rank k as $\mathbb{I}_{k,1}, \mathbb{I}_{k,2}, \ldots$, assuming that all elements of $\mathbb{I}_{k,i}$ are less than elements of $\mathbb{I}_{k,j}$ if $i < j$. It is worth noting that if $\mathbb{I}_{k,j} = [l_{k,j}, r_{k,j}]$, then it is possible that $l_{k,j} = r_{k,j}$. Lastly, the number of intervals of rank k, denoted N_k, is finite, but different for different k.

Lemma 4 *Suppose $N \in \mathbb{N}$ and $\mathcal{P}_1, \ldots, \mathcal{P}_5$ are the sets of closed subintervals of rank k defined above. For each $k = 1, \ldots, 5$ we take $\mathcal{P}_k$ and define the corresponding set $\mathcal{A}_k$ of fuzzy sets $A_{k,j} : \mathbb{I} \to [0, 1]$, $j = 1, 2, \ldots, N_k$ using the following assignment:*

- *$A_{k,j} = 1$, if $x \in \mathbb{I}_{k,j}$;*
- *if $\mathbb{I}_{k,j} = [l_{k,j}, r_{k,j}]$ and $l_{k,j} \neq 0$, then on the interval $[l_{k,j} - \frac{1}{N}, l_{k,j}]$, $A_{k,j}(x)$ coincides with a linear function that goes through the points $(l_{k,j} - \frac{1}{N}, 0)$ and $(l_{k,j}, 1)$;*

[3]Here, a basic function is a membership function of the corresponding fuzzy set in a fuzzy partition.

- *if* $\mathbb{I}_{k,j} = [l_{k,j}, r_{k,j}]$ *and* $r_{k,j} \neq 1$, *then on the interval* $[r_{k,j}, r_{k,j} + \frac{1}{N}]$, $A_{k,j}(x)$ *coincides with a linear function that goes through the points* $(r_{k,j}, 1)$ *and* $(r_{k,j} + \frac{1}{N}, 0)$;
- $A_{k,j} = 0$, *if* $x \in \mathbb{I} \setminus [\max(l_{k,j} - \frac{1}{N}, 0), \min(r_{k,j} + \frac{1}{N}, 1)]$.

Then

(i) for each $k = 1, \ldots, 5$, *the set* $\mathcal{A}_k = \{A_{k,j} : \mathbb{I} \to [0,1],\ j = 1, 2, \ldots, N_k\}$ *constitute a generalized fuzzy partition of* $[0,1]$ *with the Ruspini condition;*

(ii) the set $\bigcup_{k=1}^{5} \mathcal{A}_k = \{A_{k,j} : \mathbb{I} \to [0,1],\ k = 1, \ldots, 5,\ j = 1, 2, \ldots, N_k\}$ *constitute a generalized "fuzzy* 5*-partition" of* $[0,1]$.

In both cases, the fuzzy partition nodes $x_{k,j}$, $k = 1, \ldots, 5$, $j = 1, 2, \ldots, N_k$ *are calculated as follows:* $x_{k,i} = (r_{k,j} - l_{k,j})/2$; *while margins (length up to* $\frac{1}{N}$*) are described in the assumptions above.*

3.1.1 Inverse F-Transform in the Kolmogorov superposition theorem

Having all the necessary technical details, we can formulate the main statement (for the case of $n = 2$ variables).

Theorem 3 *For each continuous function* $f : \mathbb{I}^2 \to \mathbb{R}$ *and each* $\varepsilon > 0$ *there are five generalized fuzzy partitions* $\mathcal{A}_1, \ldots, \mathcal{A}_5$ *of the interval* $\mathbb{I}$ *and five square real matrices* $V^1, \ldots, V^5$ *where a* $(N_k \times N_k)$*-matrix* $V^k = (v^k_{j,\ell})$, $k = 1, \ldots, 5$, *such that the sum of the inverse F-transforms* $\hat{V}^k$ *of the matrices* V^k, *with respect to the corresponding fuzzy partitions* $\mathcal{A}^2_k$ *approximates* f *with the accuracy of* ε, *i.e. for any* $(x, y) \in \mathbb{I}^2$,

$$|f(x_1, x_2) - \sum_{k=1}^{5} \sum_{j=1}^{N_k} \sum_{\ell=1}^{N_k} v^k_{j,\ell} A_{k,j,\ell}(x, y)| < \varepsilon. \tag{13}$$

Acknowledgement

The work was partially supported by the project AIMet4AI No. CZ.02.1.01/0.0/0.0/17-049/0008414.

References

References

[1] V. I. Arnold, On functions of three variables (Russian). Dokl. Akad. Nauk SSSR 114 (1957), 679 - 681.

[2] M. Holčapek, I. Perfilieva, V. Novák, V. Kreinovich, Necessary and sufficient conditions for generalized uniform fuzzy partitions, Fuzzy Sets Syst. 277 (2015) 97 - 121.

[3] A. N. Kolmogorov, On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition (Russian). Dokl. Akad. Nauk SSSR, 114 (1957) 953 - 956.

[4] S. A. Morris, Hilbert 13: Are There any Genuine Continuous Multivariate Real-Valued Functions? Bulletin (New Series) of the American Mathematical Society, 58/1 (2021) 107 - 118.

[5] I. Perfilieva, Fuzzy transforms: theory and applications. Fuzzy Set Syst, 157(8) (2006) 993 - 1023.

[6] I. Perfilieva, M. Holèapek, V. Kreinovich, A new reconstruction from the F-transform components. Fuzzy Sets Syst. 288 (2016) 3 - 25.

[7] I., Perfilieva, P. Hurtik, F.D. Martino, et al, Image reduction method based on the F-transform. Soft Comput 21 (2017) 1847 - 1861.

[8] L. Stefanini, F-transform with parametric generalized fuzzy partitions, Fuzzy Sets Syst. 180 (2011) 98 - 120.

Ambiguity as a norm: A research report

Klára Šimůnková
Faculty of Management, Prague University of Economics and Business
Klara.simunkova@vse.cz

ABSTRACT

This research delves into the second phase of an experiment initially conducted by Jiroušek and Kratochvíl (2020), which sought to measure ambiguity aversion and explore the underlying psychological mechanisms behind it. While the first phase demonstrated the existence of ambiguity aversion and identified unexpected biases, such as positive attitudes toward ambiguity and ambiguity-seeking behaviors, this second phase adopts a qualitative approach to deepen the understanding of these findings. Data collection involved semi-structured and in-depth interviews with twelve participants from the original experiment. The interviews analyzed using thematic analysis in Atlas.ti identified four key themes: ambiguity as a norm, situational ambiguity tolerance, coping strategies, and the limits of ambiguity tolerance.

RESEARCH INTRODUCTION

This manuscript reports on the second phase of an experiment, following the first phase previously published by Jiroušek and Kratochvíl (2020). The primary goal of their research was to measure the strength of ambiguity aversion, extending beyond Ellsberg's (1961) demonstration of its existence to estimate a personal coefficient of ambiguity aversion. The study sought to move from simply proving that ambiguity aversion exists, as Ellsberg did, to understanding how strongly individuals are affected by it and whether this strength can be quantified.

In this context, the concept of expected value was employed to describe the value computed when uncertainty is formalized in a normative way, such as by probability measures or belief functions. However, Jiroušek and Kratochvíl (2020) also introduced the notion of "personal expected value," which takes into account the intensity of an individual's subjective ambiguity aversion. While the former concept is rooted in mathematical theory, the latter crosses into the psychological domain, reflecting how personal attitudes and biases influence decision-making under uncertainty.

During the experiments, participants displayed a range of behaviors, including unexpected positive attitudes towards ambiguity and instances of ambiguity-seeking behavior—patterns that deviate from traditional theories and align with findings by other researchers. To explore these behaviors, the study hypothesized that participants' preferences would be reflected in their willingness to pay more for preferred lotteries, thus indicating the strength of their ambiguity aversion.

Data were collected from 192 participants across eleven sessions. However, it was observed that not all participants engaged seriously with the task, as some exhibited irrational or inconsistent betting behaviors. Despite these anomalies, all data records were analyzed, with only those whose behavior precluded meaningful computation being excluded. Among the participants, 128 displayed behavior that aligned with the study's rational criteria.

The research findings suggest that while the goal of estimating a personal coefficient of ambiguity aversion from a single experimental session remains challenging and may not fully predict behavior in decision-making under ambiguity, the model effectively captures collective trends in human behavior. Additionally, the study uncovered an unexpectedly high number of participants who exhibited positive attitudes toward ambiguity, alongside observed instability in these attitudes. These findings indicate that while progress has been made in understanding ambiguity aversion, the goal of precisely quantifying it and predicting individual behaviors remains a work in progress.

Jiroušek and Kratochvíl's (2020) findings led to a second research phase with a qualitative design, which is the focus of the present report. The qualitative phase aimed to complement classical mathematical models of decision-making with phenomenological methods to triangulate and enhance the validity of the results. Specifically, this phase sought to: 1) delve deeper into the participants' thinking and strategies, and 2) explore their attitudes towards ambiguity using various research methods, capturing these attitudes in the broader context of everyday life rather than a singular moment (the moment of the game). This approach aimed to extend the focus from a specific decision-making instance to habitual decision-making and behavior, stepping out of the scientific bubble to uncover the underlying "why".

RESEARCH METHODOLOGY

The qualitative phase of the study involved semi-structured interviews, conducted due to COVID-19 restrictions online, with participants from a previous experiment. The sample was saturated with 12 respondents, and each interview lasted approximately 30 minutes. The sample was homogeneous, composed of university students aged 19 to 24, corresponding to Generation Z. The interviews were conducted in two stages. The first phase focused on strategies and decision-making processes during the experiment, aiming to deepen insights gained from the quantitative research. This phase specifically explored whether participants' decision-making differed between experimental settings and their everyday lives. Contrary to concerns about lab vs. field bias, participants consistently reported that they behaved and made decisions as they would in real-life situations. The second phase seamlessly built on the first, drawing inspiration from Budner's (1962) scale of ambiguity intolerance.

Budner (1962) developed a scale (see Figure 1) to measure tolerance-intolerance of ambiguity, aiming to explore how individuals perceive and react to ambiguous situations. He defined "intolerance of ambiguity" as the tendency to see ambiguous situations as threatening, while "tolerance of ambiguity" was seen as the tendency to find such situations desirable.

Ambiguous situations, according to Budner (1962), are those that cannot be easily structured or categorized due to a lack of sufficient cues. He identified three types of ambiguous situations: those that are completely new with no familiar cues, those that are complex with numerous cues to consider, and those that are contradictory, where different cues suggest conflicting interpretations. These situations are characterized by novelty, complexity, or contradiction, making them difficult for individuals to process.

Budner's concept of perceiving ambiguity as a threat is central to his scale. He proposed that individuals respond to ambiguous stimuli on two levels: the phenomenological, which relates to perception and feelings, and the operative, which relates to actions and behaviors. When ambiguity is perceived as a threat, individuals may respond in ways that can be grouped into submission or denial. Submission might manifest as anxiety or discomfort on the phenomenological level or avoidance

behavior on the operative level. Denial, on the other hand, could involve repression or denial of the ambiguity on the phenomenological level, or attempts to alter the situation, either destructively or reconstructively, on the operative level.

If an individual exhibits any of these responses in situations characterized by novelty, complexity, or contradiction, it is reasonable to infer that they are intolerant of ambiguity. To measure this, Budner (1962) developed a 16-item scale, with each item designed to capture a specific mode of response to different types of ambiguous situations. This scale provides a comprehensive tool for assessing how individuals cope with uncertainty and ambiguity across various contexts, offering insights into their tolerance or intolerance of ambiguity.

FIGURE 1: Budner's scale of tolerance-intolerance of ambiguity

TABLE 1
THE SCALE OF TOLERANCE–INTOLERANCE OF AMBIGUITY

	Designed to tap*	
	Type of Response	Type of Situation
Positive Items		
1 An expert who doesn't come up with a definite answer probably doesn't know too much	PD	I
2. There is really no such thing as a problem that can't be solved	PD	I
3 A good job is one where what is to be done and how it is to be done are always clear	OS	C
4 In the long run it is possible to get more done by tackling small, simple problems rather than large and complicated ones	OS	C
5 What we are used to is always preferable to what is unfamiliar	PS	N
6 A person who leads an even, regular life in which few surprises or unexpected happenings arise, really has a lot to be grateful for	PS	N
7 I like parties where I know most of the people more than ones where all or most of the people are complete strangers	PS	N
8 The sooner we all acquire similar values and ideals the better	OD	C
Negative Items		
9 I would like to live in a foreign country for a while.	PS	N
10 People who fit their lives to a schedule probably miss most of the joy of living	PS	C
11 It is more fun to tackle a complicated problem than to solve a simple one	PS	C
12 Often the most interesting and stimulating people are those who don't mind being different and original	PS	C
13 People who insist upon a yes or no answer just don't know how complicated things really are	PD	C
14 Many of our most important decisions are based upon insufficient information	PD	I
15 Teachers or supervisors who hand out vague assignments give a chance for one to show initiative and originality	OS	C
16 A good teacher is one who makes you wonder about your way of looking at things	OS	C

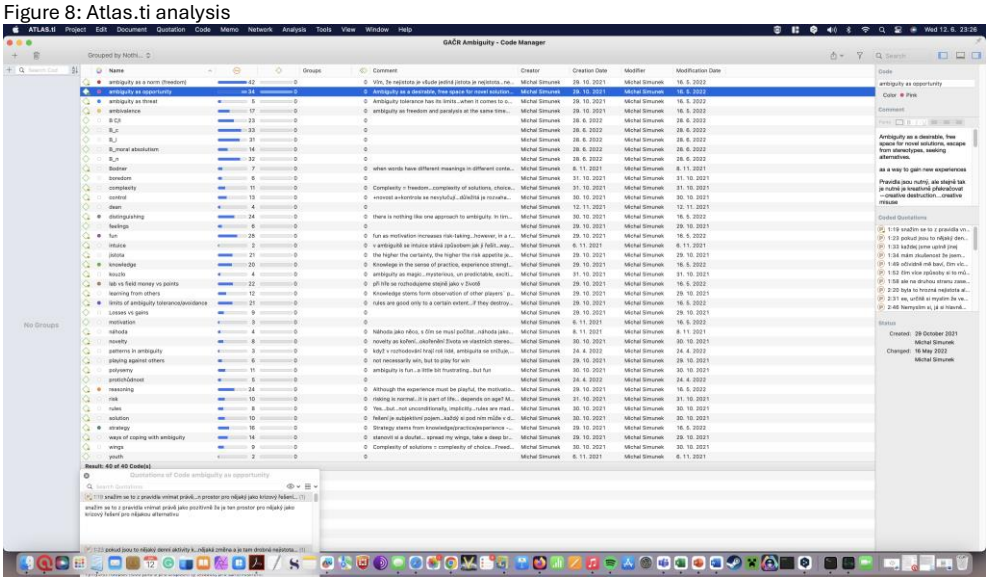
* Codes are as follows
Type of response: PD—phenomenological denial; OS—operative submission; PS—phenomenological submission; OD—operative denial
Type of situation: I—insolubility; C—complexity; N—novelty

Source: Budner (1962)

Unlike Budner (1962), who presented his scale as a questionnaire, our approach was qualitative. Instead of expecting simple "agree" or "disagree" responses, we posed the questionnaire statements during interviews and engaged in discussions about their meaning for each participant. This allowed us to obtain a more nuanced and comprehensive understanding of their attitudes and levels of tolerance or intolerance towards ambiguity. The interviews were conducted via

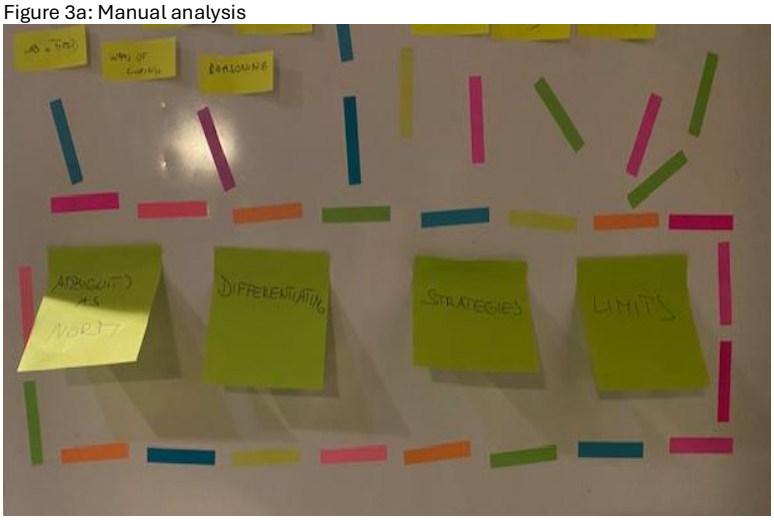
video calls, recorded, transcribed, and analyzed using Atlas.ti as well as manually (see Figures 2 and 3).

Figure 8: Atlas.ti analysis



Source: Author

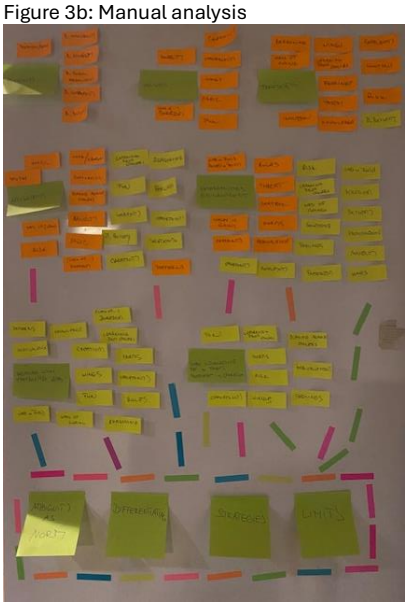
Figure 3a: Manual analysis



Source: Author

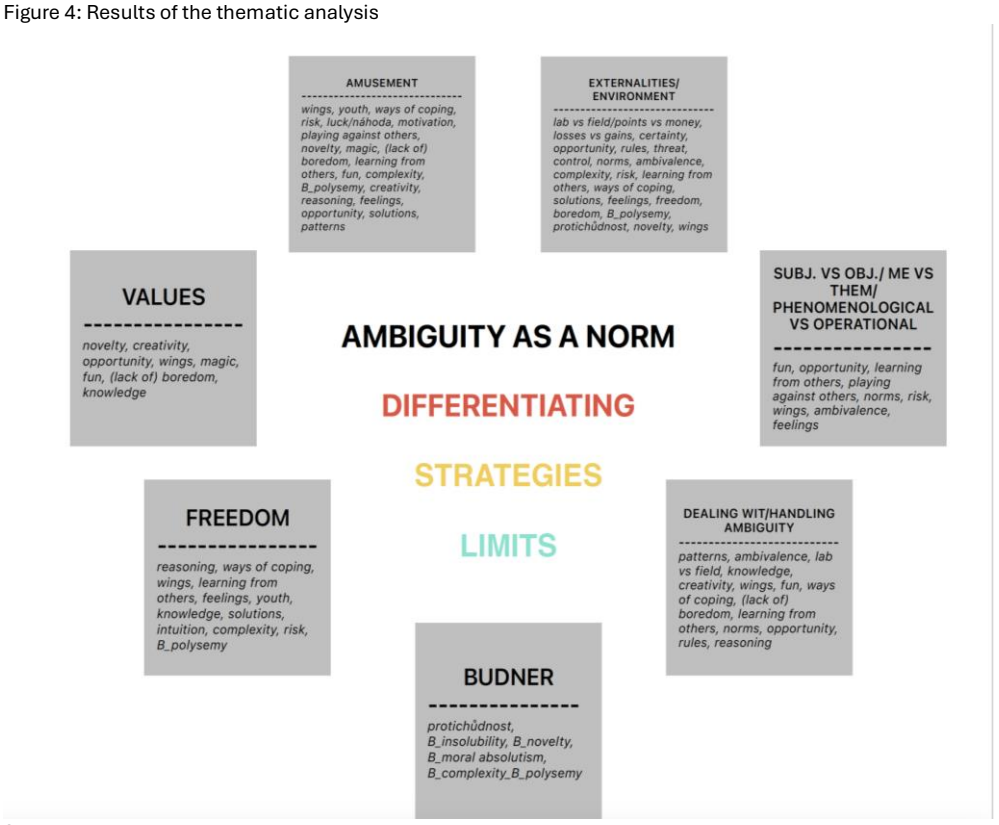
RESEARCH RESULTS

Figure 3b: Manual analysis



Thematic analysis identified four main themes: ambiguity as a norm, differentiating (situational ambiguity tolerance), strategies (ways for coping with ambiguity), and the limits of ambiguity tolerance (see Figure 4). The prevailing theory suggests that humans naturally avoid ambiguity. However, this study challenges the narrow understanding of ambiguity as a unique decision-making moment. It considers what happens when ambiguity becomes the norm and a regular context for decision-making.

Figure 4: Results of the thematic analysis



Source: Author

Reflecting on Budner's (1962) definition of tolerance and intolerance of ambiguity—where "intolerance of ambiguity" is described as the tendency to perceive ambiguous situations as threatening, and "tolerance of ambiguity" as the tendency to find such situations desirable—it can be concluded that our participants demonstrate a high tolerance for ambiguity. They do not see it as a threat but rather as a norm, an opportunity, and even a form of freedom. This pragmatic inclination towards ambiguity is evident in how they approach it as a natural part of life, likening it to "air and wings" that provide freedom and opportunities for new experiences. They differentiate, accommodate, and develop strategies to navigate ambiguity, treating it as a standard condition rather than an exception. However, their acceptance of ambiguity does have its limits, indicating that while they embrace it as a norm, they are also mindful of the boundaries within which they operate.

Ambiguity as a norm

Ambiguity is perceived as a desirable and liberating space where novel solutions can be explored and where individuals can break free from stereotypes, seeking alternatives and new experiences.

DH: *"A black-and-white world would just be too simple compared to a complex one, and there wouldn't be anything new waiting for me there. I wouldn't have anything to explore, and it would just slide into this kind of routine, and that wouldn't be fun for me… Like, trying something new, you know, and if it doesn't work out, it just doesn't work out."*

ZV: *"Well, I take it as a challenge, but yeah, I do feel some pressure or stress when there's a problem instead of being in a space where I'm comfortable and know I can handle it. But it's not like I'd break down and want to go back to something familiar—that's not me. I like pushing myself forward… and I get that a lot at work because we have different projects, and each one comes with its own set of problems to solve. So, I've kind of gotten used to it, and I think you can learn to deal with these kinds of feelings."*

This embrace of ambiguity transforms it into an adventure, offering a creative playground where rules are necessary but are also meant to be creatively bent or even broken, reflecting the concept of "creative destruction" or "creative misuse."

KJ: *"Obviously, I enjoy it—the more complicated, the better ((laughs)), and even better if I can manage the whole process myself ((laughs))."*

MS: *"I mean, yeah, sticking to some kind of procedure, sure, but most revolutionary results in history came from people breaking the rules (..), and that's how they moved forward. So, I guess a certain level of safety and some kind of procedure to get the right outcome, yes, but strictly set rules? Not really."*

Ambiguity is seen as engaging and dynamic, while simplicity is viewed as dull and uninspiring. The complexity and challenge inherent in ambiguous situations are likened to a game—risky and potentially painful, but ultimately rewarding, as they lead to personal growth and strength. The vibrancy and diversity that ambiguity brings are essential for progress, offering a spectrum of possibilities and choices that symbolize freedom.

MH: *"It would be easier to live in a world where there are only contradictory opinions... But, I just had a flash of George Orwell and his novel, so I'm not sure if it would be good to take away these means of expressing our uncertainty or the uncertainty of the world around us. I would choose complexity because conflicting solutions are the same principle as the black-and-white versus colorful world. I don't want to have only 'this way or that way' choices. I'm someone who needs freedom, free decision-making. I want to have a wide range of options to choose from, to find my own path, and to decide on my own terms, in my own way."*

This perspective frames ambiguity and complexity not just as obstacles to be overcome, but as opportunities for enjoyment and discovery. It is through solving problems and unraveling mysteries that individuals experience the true richness of ambiguity, which is understood as a multiplicity of solutions, opinions, and interpretations—a polyvocal symphony that represents freedom.

MD: *"Well, you can't let everything get to you too much, I think that's just part of life. You don't have to completely accept it, but you should recognize that these things exist—we're not always going to know everything. And maybe if we did have answers and solutions for everything, it wouldn't be as fun... We don't need a clear yes or no answer for everything; even some level of not knowing moves us forward. In some way, it's exciting because we try to reduce that*

uncertainty ourselves, to search for answers on our own instead of having everything spelled out for us."

Moreover, complexity increases the likelihood of making a good decision, because when there are only good or bad options (contradictory), the probability decreases, while in complexity, the likelihood of solutions that are not entirely good but also not entirely bad increases. Additionally, what may seem like a bad solution now could turn out to be a quite good one tomorrow. Ambiguity and uncertainty allow for the consideration of unconventional solutions, which might ultimately prove to be the saving grace.

MH: *"In a colorful, complex world, even if something bad happens, it's not 100% bad, so I can still find something in it that will benefit me in the future. But in a black-and-white world, what's bad is just bad in every way. So, another advantage of the colorful world is that what seems bad at first glance doesn't have to be entirely bad or bad in every aspect."*

The presence of threat is acknowledged by our participants, but it is seen as an intrinsic part of the opportunities that ambiguity provides, reinforcing the idea that challenge and risk are integral to the pursuit of growth and innovation. Ambiguity cannot be avoided; it must be dealt with.

JZ: *"I see it as a challenge, something you need to overcome instead of feeling bad about it because anxiety or fear doesn't help in the moment. I just accept that ambiguity exists and always will, to some degree, whether more or less, around me, and I have to accept it as part of my reality. You can't let yourself get thrown off by what you don't know—you have to take it as it is, that you often don't know everything, and still make decisions even when you don't have all the answers."*

MS: *"I think it's largely because of the sheer amount of information thrown at us daily. With fake news and everything, we never really know what's true because we receive millions of pieces of information from different sources every day."*

VF: *"For me, ambiguity feels like an adventure with a bit of threat mixed in ((laughs)). I think you can't fully eliminate that threat, and I'm okay with the fact that it will always be there."*

Differentiating

Ambiguity cannot be avoided; it must be confronted and managed. It adds spice to life, akin to a seasoning that enhances the flavor of experiences. However, this "spice" can become too much, leading to discomfort and uncertainty. This ambivalence highlights the paradox of ambiguity: it offers freedom and potential for new possibilities but can also paralyze decision-making, as individuals struggle between the desire for freedom and the need for security.

While ambiguity can be empowering, allowing for multiple interpretations and outcomes, it can also lead to indecision and anxiety. This ambivalence is particularly evident when individuals face situations where the consequences of their choices carry significant weight. When faced with serious situations, even those who usually embrace ambiguity tend to prefer clear, binary (contradictory) choices over multiple (complex) options.

JK: *"For me, it's kind of more complex, I'd say. When it comes to work processes, I prefer to follow established rules because they're proven to work. But when it's something that's purely personal, like studying, traveling, and so on, I like to dive into new things because, in that case, the responsibility falls solely on me. If something goes wrong, it's basically just my own fault."*

There is no singular approach to dealing with ambiguity. In a world where ambiguity has become the norm, individuals must navigate it daily, adapting their strategies to different contexts. Ambiguity can be a desirable state, full of potential, or a condition to be mitigated, depending on the situation. Navigating ambiguity in form of differentiation is therefore situational. This distinction is evident in various aspects of life: in private life versus work, in online anonymity versus face-to-face authenticity, and in the comfort of home versus the uncertainty of travel.

ZV: *"In an unfamiliar situation with strangers, I probably wouldn't feel completely comfortable. It also depends on the environment—whether it's an online setting or a social one. Social situations are much more stressful for me compared to being thrown into an online environment or a program or work setting. Being in a crowd of people and having to network, like at fairs or events, is something I wouldn't actively seek out, but I think I'd manage it if I had to."*

For some, the workplace represents a zone of responsibility where they avoid risk to protect others, while personal life remains a space for exploration and fun. For others, the reverse is true—personal life is prioritized, and everything else is secondary.

DH: *"The first rule for me is health, so making sure I don't hurt myself or face any health issues is my main priority (laughs). Other than that, I don't really follow any specific rules... survival is the only rule I stick to—nothing else matters beyond that."*

MH: *"I'm definitely the type of person who prefers to think things through three times before acting impulsively. My spontaneity is more present in my personal life, but in my professional life, which I currently view as a short-term business where maximum profit is the goal, I'm quite cautious."*

MD: *"In my professional life, I definitely account for uncertainty more and am able to endure it better. I analyze it more because these are problems that don't affect me personally as much as personal life issues, where I need more certainty and stability."*

Although easier options may be less stimulating, they are often chosen when peace of mind is the priority, illustrating that simplicity, while sometimes boring, is often the preferred choice when the need for calm outweighs the desire for challenge.

JZ: *"Before you mentioned that I should forget whether I enjoy it or not, I wanted to say that it would depend on the situation. If I enjoyed it, I would definitely choose the complex world with more options. But if it were something I found repulsive, I would definitely opt for the simpler world with just option A or B. In that case, I'd prefer the simpler choice to avoid spending time on something I don't find enjoyable. A straightforward choice with no need for deep thinking."*

MD: *"I have a bit of a mix. I can't really agree or disagree entirely. There are things where I like to hold onto certain certainties; I need some level of certainty. On the other hand, I couldn't live in that certainty all the time. I need to explore, discover new information, and have new experiences."*

Strategies

In navigating ambiguity, participants demonstrated a range of strategies that reveal both their adaptability and the inherent limits of their tolerance. Their approach to ambiguity is largely pragmatic, rooted in a recognition that uncertainty is an inevitable part of modern life. Rather than perceiving ambiguous situations as threats, they often view them as opportunities—spaces for freedom, exploration, and creative problem-solving. This attitude reflects a broader cultural shift where ambiguity is not just tolerated but embraced as a norm, especially in a world where complexity and unpredictability are ever-present. The strategies can be summarized as follows: knowledge/information, intuition/chance, and mental peace.

VF: *"Today, it doesn't matter so much what a person immediately knows, but rather how capable they are of finding that information. I believe everything is available; it's just that people are often too lazy to look it up. So, I try to gather as much information as possible about the options to help me make a choice."*

Perhaps somewhat surprisingly, the participants mentioned intuition and chance as ways to handle situations in ambiguity, seeing them not as a form of risk-taking but as a way to engage other cognitive faculties beyond pure logic. This intuitive approach allows them to navigate uncertainty without feeling overwhelmed, leveraging their instincts to make decisions in situations where information is incomplete or contradictory.

MH: *"First impressions and intuition—I'm not sure if everyone links these concepts together, but I do rely on both. Experience shows that in uncertain situations, using first impressions and intuition often leads to decisions that are less likely to be regretted."*

Chance is an inevitable part of life, something that must be accounted for. In fact, chance can be seen as a form of certainty—the only sure thing in an uncertain world. Embracing chance as a factor in decision-making acknowledges the inherent unpredictability of life, and instead of resisting it, individuals can use it as a tool for navigating ambiguity.

DH: *"When it comes to important decisions, I want to gather as much information as possible and hear different perspectives before making a choice. For more routine decisions, I don't need as much information—I just try it and take the risk."*

However, when ambiguity becomes too complex or challenging, and when the stakes are high, participants tend to revert to simpler, more straightforward solutions. This retreat to simplicity often arises from a desire for mental peace rather than certainty. In situations where effort outweighs the perceived benefits of complexity, they prefer clear, decisive actions that allow them to focus on other priorities.

ZV: *"It always depends on how significant the potential loss is. When the risk is small, I'm okay with taking a bit of a gamble, like adding something extra in an offline experiment setting and taking a chance. However, the greater the potential loss, the more cautious I become. I even used to follow a list where, if the expense was over a thousand crowns, I would wait a week before making the purchase. This approach helped me to deal with such decisions rationally and avoid impulsive buys that could later affect me."*

Strategy is rooted in knowledge, practice, and experience, and it plays a crucial role in reducing ambiguity. Effective strategies are not about eliminating ambiguity entirely but about managing it in a way that aligns with personal goals and comfort levels. Coping with ambiguity often involves setting intentions and trusting in the process—spreading one's wings, taking a deep breath, and embracing the uncertainty with a touch of fatalism. Each person must navigate their limits individually, establishing personal boundaries that reflect their unique thresholds for ambiguity. Ultimately, the strategies for coping with ambiguity among participants reflect a delicate balance between embracing uncertainty and recognizing the need for boundaries. They differentiate between situations where ambiguity can be harnessed for creativity and growth, and those where it must be managed to avoid potential pitfalls. Their approach is fluid, adapting to the context and the stakes involved, illustrating a sophisticated understanding of ambiguity as both a challenge and a resource.

Limits

The limits of ambiguity tolerance among our participants are influenced by both external conditions and personal characteristics. While they generally exhibit a high tolerance for ambiguity, several boundaries exist. When ambiguity threatens their sense of ontological security—such as in life-or-death situations or when they are responsible for others—their tolerance diminishes, and they prefer clearer, more defined choices. In

these moments, the desire for stability and certainty outweighs their usual openness to ambiguity.

KJ: *"I'm always aiming towards new things and constantly creating, but I also occasionally need some certainty and anchors."*

Additionally, participants distinguish between contexts where ambiguity is welcome, such as in creative or non-critical situations, and those where it becomes overwhelming or undesirable. They tend to embrace ambiguity as a source of freedom, exploration, and innovation in their personal lives but often seek more clarity and simplicity in professional or high-stakes environments. Their tolerance also fluctuates depending on their level of knowledge and experience; ambiguity becomes more manageable and less threatening when they feel confident and informed. In essence, while our participants are generally comfortable with ambiguity, their tolerance is not limitless. It is shaped by the context in which they encounter ambiguity, the stakes involved, and their own level of preparedness and responsibility.

Discussion

This research challenges the conventional view that humans inherently avoid ambiguity, instead suggesting that ambiguity, when encountered regularly, becomes a normative context for decision-making. By bridging mathematical and psychological perspectives, this study provides a richer understanding of how individuals navigate ambiguity in both experimental and real-life contexts.

Traditional theories and literature on ambiguity primarily focus on internal factors—such as the personal characteristics of decision-makers—and the specific types of situations in which decisions are made, particularly the level of ambiguity inherent in those situations. This approach assumes that ambiguity is a situational variable and that individual responses to it are influenced by personal traits and the immediate context of decision-making.

However, when ambiguity transcends individual situations and becomes a pervasive aspect of the human condition, it transforms from an exception into the norm. In a world increasingly characterized by rapid technological changes, global uncertainties, and hyperproduction of information, ambiguity is no longer an isolated phenomenon but a regular feature of everyday life. This shift necessitates a broader perspective on ambiguity. When ambiguity becomes a general condition of existence, it

is essential to include this broader context in studies of ambiguity tolerance. This means recognizing that ambiguity is not merely an attribute of specific situations but a fundamental aspect of the environment in which individuals not only decide and operate, but into which people are born, in which they learn, develop, and socialize.

We propose considering not only personal and situational aspects but also the phenomenological dimension of ambiguity in terms of the degree of ambiguity in the lived environment and being in the world. This approach requires multidisciplinary collaboration and the sharing of insights from specialized discourse with other scientific perspectives and fields. However, it is essential for understanding the complexity of contemporary life and our "not-human-only" decision-making within it. Indeed, our participants demonstrate a tolerance for ambiguity. The following lines outline potential explanations, which should also be considered as suggestions for further research.

The contemporary world is characterized by rapid change, complexity, and frequent disruptions, which means that ambiguity has become a common and often expected part of life. The availability of vast amounts of information and diverse perspectives through digital media might lead to an increased familiarity with and tolerance for ambiguity. Being accustomed to navigating complex and sometimes conflicting information could help participants manage ambiguity more effectively. Societal shifts towards valuing adaptability and innovation may encourage individuals to embrace ambiguity as a space for creativity and growth rather than a threat. This cultural orientation can make ambiguity seem more manageable and even desirable. Generation Z has grown up in a highly connected, digital environment, which has exposed them to diverse viewpoints and continuous information flow. This constant exposure may contribute to a higher comfort level with ambiguity, as they are used to handling varied and often incomplete information. This generation's familiarity with rapid technological and social changes may foster a more positive attitude towards ambiguity, seeing it as a norm rather than an exception.

It is time to examine the phenomenon of ambiguity and decision-making from different perspectives, as illustrated by the following statement of our participant:

JZ: *"It was actually something we discussed today with a professor who was puzzled by some participants and kept asking where the rationality*

was, noting that they couldn't win the game. I told him, 'Professor, it's not about winning for them; you need to look at it differently...'"

Literature

Budner, S.N.Y. (1962), Intolerance of ambiguity as a personality variable. Journal of Personality, 30, 29-50. https://doi.org/10.1111/j.1467-6494.1962.tb02303.x

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. Quarterly Journal of Economics, 75(2), 643–669.

Jiroušek, R., & Kratochvíl, V. (2020). On subjective expected value under ambiguity. International Journal of Approximate Reasoning, 127, 70-82, https://doi.org/10.1016/j.ijar.2020.09.002

Two Centuries of Linear Programming

Milan Vlach
Theoretical Computer Science and Mathematical Logic
Charles University
vlach@ktiml.mff.cuni.cz

Abstract

Linear programming is concerned with optimizing a linear function over the set of solutions of a system of linear inequalities, or a system of linear equations and inequalities.

The founders of linear programming are often regarded as George B. Dantzig and John Von Neumann. However, the history of linear programming dates back at least as far as to Josephs Fourier's works on linear inequalities in the early 1820's, when Fourier invented an elimination method for solving systems of linear inequalities, formulated the first linear programming problem as a problem of solving a system of linear inequalities, and devised a procedure for solving linear programming problems that can be considered as a prototype of Dantzig's simplex algorithm developed in the second half of 1940's.

This paper is a preliminary (very preliminary and very subjective) brief overview of the development of linear programming and closely related fields from Fourier's initial formulation to the present, including a brief discussion of current open problems.

1 Introduction

Let us start with giving two quotes from the Carathéodory paper *The Beginning of Research in the Calculus of Variations* presented on Aug. 31, 1936, at the meeting of the Mathematical Association of America [1]. This can serve as our motto.

> "I will be glad if I have succeeded in impressing the idea that it is not only pleasant and entertaining to read at times the works of the old mathematical authors, but that this may occasionally be of use for the actual advancement of science."
>
> "It may happen that the work of most celebrated men may be overlooked. If their ideas are too far in advance of their time, and if the general public is not prepared to accept them, these ideas may sleep for centuries on the shelves of our libraries."

Developing of linear programming as a modern formal mathematical discipline started shortly after World War II in the late 1940's and the beginning of 1950's. The very term "linear programming" was used for the first time by G. B. Dantzig on the suggestion of T. J. Koopmans as a name for referring to optimizing a linear function over the set of solutions of a system of linear inequalities, or a system of linear equations and inequalities. In "Reminiscences about the origins of linear programming" [2], Dantzig recalls:

> When I had first analyzed the Air Force planning problem and saw that it could be formulated as a system of linear inequalities, I called my first paper: Programming in a Linear Structure. In the summer of 1948, Koopmans and I visited the RAND Corporation. One day we took a walk near the Santa Monica beach. Koopmans said: "Why not shorten *Programming in a Linear Structure* to *Linear Programming*?". I replied: "That's it! From now on that will be its name."

The problem of minimizing (or maximizing) a linear function of a finite number of real (or rational) variables over the set of solutions of a finite system of linear inequalities may seem trivial from the point of classical mathematics because if a feasible solution exists, then finding an optimal solution only requires examining the objective function at a finite number of points. Nevertheless, we see at present that linear programming is one of the examples of mathematics that has a number of difficult open problems and a wide range of useful applications, and it is now studied and applied by hundreds of thousands of people throughout the world..

Moreover, it is known that the following classes of computational problems are equivalent in the sense that there is an easy way to transform solutions of one to solutions of the other:

- solving linear programming problems,
- solving systems of linear inequalities,
- solving matrix games in mixed strategies,
- finding the best linear approximation in l^∞ space (The discrete Chebyshev approximation problem),
- finding the best linear approximation in l^1 space (The least absolute deviations fit).

The histories of these areas are inextricably linked, even though it may seem that they arose and developed independently.

1.1 Linear programming and linear inequalities

There are many ways to represent instances of the linear programming problems. One of the typical formulation of a linear programming problem is the task to

maximize a linear function

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

of real variables $x_1, x_2, \ldots, x_n$ on the set of nonnegative solutions of the system

$$\begin{aligned} a_{11}x_1 &+ a_{12}x_2 + \cdots + a_{1n}x_n \le b_1 \\ a_{21}x_1 &+ a_{22}x_2 + \cdots + a_{2n}x_n \le b_2 \\ &\cdots \\ a_{m1}x_1 &+ a_{m2}x_2 + \cdots + a_{mn}x_n \le b_m \end{aligned}$$

Briefly:

$$\text{Maximize } cx \text{ subject to } Ax \le b, x \ge 0$$

Inextricably linked to this problem is the task (called the dual problem) of minimizing the linear function

$$y_1b_1 + y_2b_2 + \cdots + y_mb_m$$

of real variables $y_1, y_2, \ldots, y_n$ on the set of nonnegative solutions of the system

$$\begin{aligned} y_1a_{11} + y_2a_{21} + \cdots + y_ma_{m1} &\ge c_1 \\ y_1a_{12} + y_2a_{22} + \cdots + y_ma_{m2} &\ge c_2 \\ &\cdots \\ y_1a_{1n} + y_2a_{2n} + \cdots + y_ma_{mn} &\ge c_n \end{aligned}$$

Briefly:

$$\text{Minimize } yb \text{ subject to } yA \ge c, y \ge 0$$

Another typical formulation is the maximization of a linear function

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

on the set of nonnegative solutions of the system

$$\begin{aligned} a_{11}x_1 &+ a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 &+ a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ &\cdots \\ a_{m1}x_1 &+ a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{aligned}$$

Briefly:

$$\text{Maximize } cx \text{ subject to } Ax = b, x \ge 0$$

Similarly, the dual problem to this problem is the task of minimizing the linear function

$$y_1b_1 + y_2b_2 + \cdots + y_mb_m$$

on the set of solutions of the system

$$\begin{aligned} &y_1a_{11} + y_2a_{21} + \cdots + y_ma_{m1} \geq c_1 \\ &y_1a_{12} + y_2a_{22} + \cdots + y_ma_{m2} \geq c_2 \\ &\cdots \\ &y_1a_{1n} + y_2a_{2n} + \cdots + y_ma_{mn} \geq c_n \end{aligned}$$

Briefly:

$$\text{minimize } yb \text{ subject to } yA \geq c$$

All these linear programming problems are mutually equivalent in the sense that there is an easy way how to transform solutions of one to solutions of the other and vice versa.

Von Neumann shown [3] that the solving the problem

$$\text{maximize } cx \text{ subject to } Ax \leq b$$

can be reformulated into a solution problem of the system linear inequalities

$$Ax \leq b, x \geq 0, yA \geq c, y \geq 0, cx \geq yb.$$

This led directly into the development of duality and existence theorems of linear programming by Gale, Kuhn and Tucker [4], and realization that optimization problems of linear programming and solution problems of systems linear inequalities are equivalent both theoretically and computationally.

1.2 Linear programming and matrix games

A matrix game is given by a real matrix with m rows and n columns

	C_1	C_2	$\cdots$	C_n
R_1	a_{11}	a_{12}	$\cdots$	a_{1n}
R_2	a_{21}	a_{22}	$\cdots$	a_{2n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
R_m	a_{m1}	a_{m2}	$\cdots$	a_{mn}

the rules of the game and the concept of solutions, ...

Entry a_{ij} is the payoff to player R when R uses pure strategy R_i and player C uses pure strategy C_j.

If players use mixed strategies $p = (p_1, p_2, \ldots, p_m)$ and $q = (q_1, q_2, \ldots, q_n)$, then the expected payoff of player R is given by the sum

$$E(p, q) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} p_i q_j$$

and the expected payoff of player C is $-E(p,q)$. Solutions are pairs (p^*, q^*) such that, for all (p,q),

$$E(p,q^*) \leq E(p^*,q^*) \; and \; E(p^*,q^*) \leq E(p^*,q).$$

Player C is interested in finding nonnegative values of $q_1, q_2, \ldots, q_n$ that minimize function

$$0q_1 + 0q_2 + \cdots + 0q_n + 1v$$

subject to

$$q_1 + q_2 + \cdots + q_n = 1$$

and

$$\begin{array}{l} a_{11}q_1 + a_{12}q_2 + \cdots + a_{1n}q_n - v \leq 0 \\ a_{21}q_1 + a_{22}q_2 + \cdots + a_{2n}q_n - v \leq 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}q_1 + a_{m2}q_2 + \cdots + a_{mn}q_n - v \leq 0 \end{array}$$

Player R is interested in finding nonnegative values of $p_1, p_2, \ldots, p_m$ that minimize function

$$0p_1 + 0p_2 + \cdots + 0p_m + 1v$$

subject to

$$p_1 + p_2 + \cdots + p_m = 1$$

and

$$\begin{array}{l} a_{11}p_1 + a_{21}p_2 + \cdots + a_{m1}p_m - v \geq 0 \\ a_{12}p_1 + a_{22}p_2 + \cdots + a_{m2}p_m - v \geq 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}p_1 + a_{2n}p_2 + \cdots + a_{mn}p_m - v \geq 0 \end{array}$$

Note that these two linear programming problems are mutually dual and that it is known nowadays that the theory of linear programming is equivalent to game theory in the sense that:

- The solution (in mixed strategies) of any matrix game can be obtained by solving a suitably chosen linear programming problem and vice versa.
- The basic theorems of the theory of matrix games and linear programming (the existence of solutions in mixed strategies and the duality of linear programming) follow from each other.

1.3 Linear Programming and l^∞ and l^1 linear approximations

An l^∞ linear approximation problem (a discrete Chebyshev approximation problem) is the problem of minimization of the function

$$\max(|f_1|, |f_2|, \ldots, |f_m|)$$

where $f_1, f_2, \ldots, f_m$ are affine functions of n real variables.

This is reducible (in fact equivalent) to solving the linear programming problem:

$$\text{minimize } \ t \text{ subject to } \ -t \leq f_i \leq t \quad (i = 1, \ldots, m).$$

An l^1 linear approximation problem (least-absolute deviations fit) is the problem of minimization of the function

$$\sum_{i=1}^{m} |f_i|,$$

where $f_1, f_2, \ldots, f_m$ are affine functions of n real variables.

This is reducible (in fact equivalent) to solving the linear programming problem:

$$\text{minimize } \ \sum_{i=1}^{m} t_i \text{ subject to } \ -t_i \leq f_i \leq t_i \ (i = 1, \ldots, m).$$

2 From Fourier to Khachian and Karmarkar

The founders of linear programming as a modern mathematical discipline are often regarded as George B. Dantzig, who proposed the simplex algorithm, and John von Neumann, who laid a basis for the duality theory.

As almost always, there are predecessors. Now we know that the history of linear programming dates back at least as far as to Josephs Fourier's works on linear inequalities in the early 1820's, when Fourier invented an elimination method for solving systems of linear inequalities, formulated the first linear programming problem as a problem of optimizing over the solution set of a system of linear inequalities, and devised a procedure for solving linear programming problems that can be considered as a prototype of Dantzig's simplex algorithm.

In this section, we give a brief overview of the development of linear programming and closely related fields from Fourier's initial formulation to the Khachiyan and Karmarkar results on the computational complexity of linear programming.

On this journey, we will orientate ourselves by the following subjectively selected milestones: Fourier (1820's), Farkas (1890's), Minkowski (1890's), Dines (1919), Motzkin (1936), Kantorovich (1939), Dantzig (1947), von Neumann (1948), Khachian (1979), and Karmarkar (1984).

2.1 Before World War II

2.1.1 Fourier

Jean Baptiste Joseph Fourier (1768-1830) was born 21 years before the French Revolution. The events of those unstable and violent times made his life so fascinating that could serve as a script for an exciting movie. There are a number of books

and essays on Fourier life, but we refer to a recent essay by Bernard Maurey [5], which contains enough material and references for interested readers.

Regarding mathematics and physics, Fourier is largely associated with the mathematical analysis of heat diffusion, methods of solving partial differential equations by means of Fourier series, Fourier integrals, and Fourier principle of virtual velocity. But his interests in mathematics included other fields also, and one of his achievements related to linear programming was to invent an elimination method for linear inequalities and the principle underlying Dantzig's simplex algorithm.

Fourier never explained how he came to be interested in this problem. However, it is possible to trace a few possible influences. He was an important member of Bonaparte's Egyptian campaign of 1798-1801, and organized the calculation of the heights of the pyramids at Memphis by the separate measurement of the height of each of their steps. This work might brought him to the problem of minimizing the errors in results deduced from a large number of measurements.

Fourier did not complete his work on inequalities before he died but he did publish a short paper Solution d'une question particuliére du calcule des inégalités (1826) and he inserted two summaries on the subject in l'Histoire de l'Académie Royale des Sciences, one for the year 1823, and the other for 1824. From these publications it is clear that Fourier had a clear geometrical understanding of the solution set to a system of linear inequalities in three variables as a polyhedron in three dimensional space.

> *We consider a set of linear functions in the variables $x, y, z, \ldots$, the numerical values which enter these functions being given. [...] Let us suppose that we assign to $x, y, z, \ldots$ some numerical values and that, after substituting these in each function, we compute the value of the functions, which may be positive or negative.*
>
> *We will regard this result as an error or deviation, and we will take its absolute value as a measure of error. This done we ask which numerical values $X, Y, Z, \ldots$ can be assigned to $x, y, z, \ldots$ in order to minimize the maximum error. [...] It should be stressed that this method can be applied to systems of an arbitrary number of variables. [...]*
>
> *We will use an example in order to describe the method of minimizing the largest absolute error. x and y are the coordinates of some point in the horizontal plane. The vertical coordinate z is used to measure the values of the given linear functions of x and y, and each inequality is represented by a plane in three dimensions whose position is given. This system of planes defines a vase which they delimit or envelop. The shape of this extreme vase is that of a polyhedron which is convex towards the horizontal plane. The lowest point of this vase or polyhedron has coordinates X, Y, Z which are the subject of the enquiry, that is, Z is the least possible value for the largest error and X and Y are the values of x and y which give the absolute minimum. To locate this lowest point of the vase quickly, one erects a vertical line at any point*

of the horizontal plane, for example at the origin, and determines its intersection with the highest plane.

Let m_1 be the point of intersection of this line with the most extreme plane. One descends this plane from m_1 to the point m_2 on an edge of the polyhedron, and following this edge, one descends to m_3, the common summit of three extreme planes. On leaving this summit one continues to descend along a second edge to a new summit m_4, and one continues with the same procedure at each stage choosing that of the two edges which leads to a lower summit. One thus arrives very quickly at the lowest point of the polyhedron. This [geometrical] construction represents the numerical operations which the analytical rule prescribes; it clearly reveals the course of the method, which consists of passing from one extreme function to another and diminishes little by little the value of the largest error.

From this it is clear that (in the present language) Fourier describes the linear programming problem

$$\text{Minimize} \quad 0x + 0y + 1z \quad \text{subject to}$$

$$a_i x + b_i y + c_i - z \leq 0, \quad -a_i x - b_i y - c_i - z \leq 0, \quad i = 1, 2, 3, \ldots,$$

whose solution minimizes the maximum absolute error

$$\max_i |a_i x + b_i x + c_i|.$$

It is worth mentioning that de la Vallée Poussin gives an algebraic analogue to Fourier's method in 1911, see [6] and [7], where it is shown that the de la Vallée Poussin procedure for the solution of linear minimax estimation problems can be adjusted for solving a class of linear programming problems.

In the same summary, Fourier proposed a procedure for solving systems of linear inequalities based on the elimination of variables. It can easily be used to solve linear programming problems as well, but (without suitable modifications) it is not advisable to use it because it could lead to a significant growth of inequalities in each elimination step.

Fourier's elimination procedure was rediscovered, rediscovered, and rediscovered, recall the Carathéodory opening quote. For a long time it was called Motzkin's method of elimination. Now we most often encounter the name Fourier-Motzkin method or Fourier-Dines-Motzkin method.

2.1.2 Farkas

Julius Farkas (1847-1930), a Hungarian theoretical physicist, focused on creating the mathematical basis of the theory of linear inequalities at the end of the 19th century. Farkas' lemma, published at the beginning of the 20th century, did not find real recognition until the 1930s in the works of Minkowski, and its significant

applications had to wait until the development of linear programming theory after World War II.

It is worth mentioning that Farkas' interest in the theory of linear inequalities was originally motivated by Fourier's results on the equilibrium of systems of mass points in analytical mechanics. Originally Farkas viewed linear inequalities as a theoretical foundation for the Fourier inequality form of the principle of virtual power: "A system of points is in equilibrium precisely when the virtual work of forces is not positive".

The development of Farkas' interest from the original motivation to an abstract theory of linear inequalities can be clearly seen in the titles of the following sequence of publications:

- (1897) Die Algebraischen Grundlagen der Anwendungen des Fourier'schen Principes in der Mechanik.
- (1899) Die Algebraische Grundlage der Anwendungen des Mechanischen Princips von Fourier.
- (1901) Theorie der einfachen Ungleichungen.

In the introduction to his 1901 paper Farkas wrote:

> Während der letzten sieben Jahre habe ich mich wiederholt mit Ungleichungen beschäftigt und aus dem Gesichtspunkte der Anwendungen eine vorläufig abgeschlossene Theorie derselben zu Stande gebracht. Meine vereinzelt erschienenen Publicationen sind aber nicht geeignet, ein klares und zusammenhängendes Bild des Ganzen zu liefern, und davon abgesehen, ist ihre Zugänglichkeit auch beschränkt. Darum nehme ich mir die Freiheit, hier eine systematische Darstellung, nebs einigen Ergänzungen, dieser Arbeiten vorzulegen

(*During the last seven years I have worked repeatedly on inequalities and from the point of view of applications I have developed a provisional self-contained theory. However my scattered publications are not suited to give a clear and coherent picture of the whole, and apart from that their accessibility are limited. Therefore I here take the liberty to present a systematic presentation of these works and some additions to them.*)

2.1.3 Minkowski

Hermann Minkowski (1864–1909) is primarily known for his work on number theory and mathematical physics. Minkowski developed his theory of linear inequalities in connection with investigations in pure mathematics. His goal was to obtain results in number theory. Linear inequalities appeared as a natural element in Minkowski's new geometry in the algebraic expression of supporting hyperplanes. Minkowski's geometrical intuition led him to characterize different kind of solutions to systems of linear inequalities and the basic notion in his theory play the extreme solutions.

2.1.4 Dines

The work of Lloyd Lyne Dines (1885–1964) on linear inequalities was a part of this tendency towards generalizations, first as a generalization from systems of linear equations to systems of linear inequalities and second, as a generalization from finitely many unknowns to the continuum. Dines developed his theory not because it was the theoretical foundation for other kinds of problems or because he needed it as a tool for investigations in other branches of mathematics but because he found an interest in linear inequalities for their own sake. He seems to have been motivated by the lack of a theory for such systems, which he compared with the matrix theory of systems of linear equations.

2.1.5 Motzkin

Theodore Motzkin (1908-1970) was born in Germany and studied in Göttingen and Berlin. His thesis [9] *Beitraege zur Theorie der Linearen Ungleichungen*, published in 1936 is important historically for at least two reasons: On the one hand, it summarized the whole development of a theory of systems of linear inequalities up to 1935. On the other hand, it became an important inspiration for the development of the theory of linear programming that took place after the second world war.

The reader interested only in the theory of linear inequalities finds in [8] an excellent detailed description of contributions of Farkas, Minkowski, Dines, Motzkin.

2.1.6 Kantorovich

Leonid Vitalievich Kantorovich (1912-1986), the father of linear programming in the USSR, is well known in the mathematical community not so for his achievements in linear programming but mainly for his work in functional analysis, approximation theory, and operator theory. The classic textbook by Kantorovich and Akilov is still on the desk of many theoretical and applied scientists.

Kantorovich's booklet [11] *Mathematical Methods in the Organization and Planning of Production* which appeared in 1939 is a convincing evidence of the birth of linear programming. In the Editor's Foreword, A. R. Marchenko, writes:

> This work is interesting from a purely mathematical point of view since it presents an original method, going beyond the limits of classical mathematical analysis, for solving extremal problems. On the other hand, this work also provides an application of mathematical methods to questions of organizing production which merits the serious attention of workers in different branches of industry.

Here Kantorovich gave (in current terminology) a dual-simplex-type method for solving two kinds of transportation problems, one of which is equivalent to the general linear programming problem.

During the Second World War Kantorovich completed the first version of his book [12] *The Best Use of Economic Resources*, which led to the Nobel Prize awarded to him and Tjalling C. Koopmans in 1975.

2.2 After World Word II

A similar line of research on inequality constrained optimization problems took place in the USA independently of the work of the Russians during the Second World War.

2.2.1 Von Neumann

We have already mentioned in the Introduction that von Neumann was at the birth of the theory of linear programming duality by formulating the dual problem to the problem of the form

$$\text{maximize } cx \text{ subject to } Ax \leq b, x \geq 0$$

and claiming that the original problem is equivalent to a solution problem of a system of linear inequalities. Harold W. Kuhn reports in [**?**]:

> After G. B. Dantzig visited John von Neumann in Princeton in May, 1948, von Neumann circulated privately a short typewritten note that was published fifteen years later [3]. This note formulated the dual for a linear program and gave a flawed proof of the equality of optimal objective values based on an invalid inhomogeneous form of Farkas' Lemma. (This error is corrected in the published version [3].)

Von Neumann was also among the first who observed that, for any matrix game, there is a dual pair of feasible linear programming problems whose saddle points yield equilibria of the game, which makes it possible to show that solving linear programming problems is equivalent to solving matrix games (in mixed strategies). A detailed description of this connection is carefully described in an essay by Kuhn and Tucker, see [10].

2.2.2 Dantzig

G, B. Dantzig (1914-2005), like Kantorovich, viewed linear programming not just as a qualitative tool in the analysis of economic phenomena, but also as a method that could be used to compute actual answers to concrete real world optimization problems.

One of Dantzig's key contribution to linear programming is the development of the simplex algorithm. This algorithm and its modifications are based on local improvements technique by pivoting between basic feasible solutions. Dantzig's original formulation proposed in 1947 uses a natural pivoting rule. Namely, in each step, the variable with the most negative reduced cost is chosen to enter the basis.

There are several research directions that arise from the behavior of the algorithm and the simplex algorithm and his various modifications remain an important computational tool in linear and integer programming.

The story about the development of the simplex algorithm and comments about its historical significance are nicely captured by Dantzig himself in [14].

2.2.3 Khachiyan

The ellipsoid method was originally developed by Yudin and Nemirovski (1976) and Shor (1977) for the solution of convex function minimization problems. The method can be considered also as modification of the "gravity center method" proposed in 1965 by A. Ju. Levin (Soviet Mat. Dokl. 6, 286-290), or a special representative of methods with space dilatation in the direction of the subgradient.

The method become famous when Khachiyan (1979, 1980) used it to obtain a weakly polynomial-time algorithm for linear programming. E. Lawler remembers in the paper [15] "The Great Mathematical Sputnik of 1979":

> In January 1979, Rainer Burkard brought the reprint of a *Doklady* paper to Oberwolfach. The author, somebody named Khachian, purported to have a polynomial-time algorithm for solving linear programming problems. The next week in Amsterdam, sitting at the elbow of Milan Vlach, I scribbled out a translation. I sent this to a long list of people, and before too long, Peter Gács and László Lovász had supplied Khachian's missing proofs and verified his results.

For further details, we refer to the following papers:

- Khachiyan (1979), A polynomial algorithm in linear programming, Doklady A. Nauk SSSR, 244, 1093-1096.
- Khachiyan (1980), A polynomial algorithms in linear programming, Zhurnal Vychislitelnoi Matematiki i Matematicheskoi Fiziki 20, 51-58.
- Gács and Lovász (1981), Khachiyan's algorithm for linear programming, Mathematical Programming Study 14, 61-68.

2.2.4 Karmarkar

The idea of moving through the interior of the feasible region goes back at least to Frisch in 1955 who proposed using a logarithmic barrier function.

The modern version of these methods is due to Karmarkar, who established the polynomial-time boundedness of the projective method in 1984.

In fact, the affine scaling method of Karmarkar had in fact been considered already by Dikin, a student of Kantorovich, who proposed the basic affine scaling algorithm in the Soviet Mathematics Doklady in 1967, and published a proof of convergence in Upravlyaemye Sistemi in 1974.

3 Open problems

Most of the unsolved problems in the field of linear programming are concened with the computational complexity. For example:

- How fast is the Simplex Algorithm? Is it a (weakly) polynomial-time method?
- Similarly for other deterministic pivoting strategies.
- Is there a strongly polynomial method for LP?
- Polynomial Hirsh conjecture: The diameter of the graph of a d-polytope with m facets is bounded above by a polynomial in d and m.

At the end of the last century, V. I. Arnold, on behalf of the International Mathematical Union, asked a number of mathematicians to describe important yet unsolved mathematical problems for research in the coming century. American mathematician S. Smale proposed 18 problems for the 21st century [16]. The ninth problem concerns linear programming. Namely: Is there a polynomial algorithm (given by a real number machine) that decides whether the system of linear inequalities $Ax \geq b$ has a solution?

References

[1] Carathéodory C. (1937). The Beginning of Research in the Calculus of Variations. *Osiris*, 3, 224-240.

[2] Dantzig G. B. (1982). Reminiscences about the origin of linear programming. *Operations Research Letters*, 1, 43–48.

[3] Von Neumann J. (1963). Discussion of a maximum problem. *John von Neumann, Collected Works*, Vol. VI, Theory of Games, Astrophysics, Hydromechanics and Meteorology, Pergamon, New York, 89–95.

[4] Gale D., Kuhn, H. W. and Tucker, A. W. (1951). Linear Programming and Theory of Games. *Cowles Commission for Research in Economics*, No 13, Part 3, 317–329.

[5] Maurey, B. (2019). Fourier, One Man, Several Lives. *EMS Newsletter*, September, 8–20.

[6] de laVallée Poussin, C. J. (1911). Sur la méthode de l'approximation minimum. *Ann. Soc. Sci. Bruxelles*, 35, 1–16. (English translation by H. E. Salger, National Bureau of Standards, USA).

[7] Farebrother, R. W. (2006). A linear programming procedure based on de la Vallée Poussin's minimax estimation procedure. *Computational Statistics and Data Analysis*, 51, 453–456.

[8] Kjeldsen, T. H. (2002). Different Motivations and Goals in the Historical Development of the Theory of Systems of Linear Inequalities. *Arch. Hist. Exact Sci*, 56, 469–538.

[9] Motzkin, T. (1936). Theorie der linearen Ungleichungen. Jerusalem: Azriel.

[10] Kuhn, H. W. and Tucker, A. W. (1958). John von Neumann's work in the theory of games and mathematical economics. *Bull. Amer. Math. Soc.*, 64, 100–122,

[11] Kantorovich, L. V. (1939). Mathematicheskie Metody Organizatsii i Planirovania Proizvodstva. *Leningrad State University Publishers*. Translated as "Mathematical Methods in the Organization and Planning of Production" (1960) in Management Science 6, 4, pp, 366–422.

[12] Kantorovich, L. V. (1960), Ekonomičeskij rasčot najlučševo ispolzovanija resursov. *Acad. of SC., USSR*. Translated as "The Best Use of Economic Resources." Cambridge, Mass., Harvard U. Press, 1965,

[13] uhn, H. W. (1991). Nonlinear programming: A historical note. *History of Mathematical Programming*, 82–96, CWI Amsterdam. (Edited by J. K. Lentra, A. H. G. Rinnooy Kan and A. Schrijver.

[14] Dantzig, G. B. (1991). Linear Programming. *History of Mathematical Programming*, 19–31, CWI Amsterdam. (Edited by J. K. Lentra, A. H. G. Rinnooy Kan and A. Schrijver.

[15] Lawler, E.L. (1980). The great mathematical Sputnik of 1979. *The Mathematical Intelligencer* 2, 191–198. https://doi.org/10.1007/BF03028602.

[16] Smale, S. (1998). Mathematical problems for the next century. *The Mathematical Intelligencer* 20, 7–15. https://doi.org/10.1007/BF03025291.

Optimizing Worker Allocation for Enhanced Efficiency in Parallel Production Lines

Xiaowen Zhao
Department of Industrial and Systems Engineering
Aoyama Gakuin University
zhao@ise.aoyama.ac.jp

Shao-Chin Sung*
Department of Industrial and Systems Engineering
Aoyama Gakuin University
son@ise.aoyama.ac.jp

Jing Sun
Architecture, Civil Engineering, and Industrial Management Engineering
Nagoya Institute of Technology
sun.jing@nitech.ac.jp

Hisashi Yamamoto
Department of Electrical Engineering and Computer Science
Tokyo Metropolitan University
yamamoto@tmu.ac.jp

Abstract

Potential factors such as worker errors, variations in operating times, parts shortages, and machine failures can impact the scheduled delivery dates of individual production stages. Delays in consecutive stages may subsequently defer overall manufacturing production. This research proposes optimizing worker assignments to meet product delivery deadlines and minimize the total expected production cost. By examining a parallel production line, this research addresses how to optimally allocate workers across various stages to reduce costs when each production line contains a different number of stages.

*Shao-Chin Sung is currently known as So Akiyama

1 Introduction

In the field of production management, addressing issues related to production scheduling and line balancing is crucial for enhancing productivity, reducing production costs, and meeting delivery deadlines. In a serial production line operating under uncertain conditions, the performance and efficiency of a given period are often influenced not only by the activities within that period but also by the risks encountered in preceding periods. Taking a standard serial production line as an example, the ability of a specific stage (or period) to meet its delivery deadline depends on the performance of the previous stage (or period). This scenario is referred to as the restricted-cycle multiple production periods problem, where each cycle's success is contingent on the prior cycle's results.

Despite the assembly line's extensive use over several decades, it continues to offer benefits in reducing operating costs and production time. However, the reliability of machines on the assembly line, despite technological advancements, cannot fully compensate for human error. Mistakes or failures by workers are inevitable and can lead to delays in the entire production stage. The variability in workers' abilities means that the time taken for the same task differs from person to person. Particularly, new recruits, who are less skilled, are more likely to cause delays. Therefore, appropriately managing these new recruits, referred to as untrained workers in this research, is crucial. The first mathematical formalization of assembly line balancing is established by Salveson [3] 60 years ago, and during the assembly line balancing problem develops, Yamamoto *et al.* [5] produce the restricted-cycle model with multiple periods (RCMwMP). This model is set with constraint condition (e.g., the target operating time), which is repeated in every multiple period. If the constraint condition is broken through, an expected cost (e.g., penalty cost) will be increased.

Over the decades, numerous efforts have been made to tackle the scheduling issues in flow shops. The primary goal of such scheduling is often the minimization of make span. Many researchers have devoted their efforts to developing heuristics that provide optimal or near-optimal solutions for large-scale problems. Another significant objective in scheduling is the minimization of total flow time, or the sum of job completion times, which leads to reduce in-process inventory. Heuristics aimed at achieving this objective have been developed by Liu *et al.* [1] and , Rajendran *et al.* [2].

The RCMwMP is categorized into reset and non-reset models based on whether the constraint condition (target operating time) is reset. In the domain of the reset model, Yamamoto *et al.* [4] proposed a recursive formula for calculating the total expected cost and an algorithm for determining optimal assignments. This algorithm employs the branch and bound method to achieve efficient solutions.

Zhao *et al.* [7] introduced an inline series production line and investigated the optimal assignment problem for a configuration where the workforce is divided into three groups consisting of 1, 2, and $n-3$ members respectively. They further proposed a locally optimal assignment strategy specifically for scenarios where an

untrained worker is assigned in the first stage, which operates under a constant operating time.

In a global production network characterized by diverse product varieties, small batch sizes, and workers with varying production capacities, it is crucial to optimize worker assignments for different target operating times. Zhao *et al.* [6] addressed this by proposing a model for the optimal assignment of three groups of workers, specifically in scenarios where the target operating time follows a continuous distribution.

In previous research, we examined an inline series production line and explored how reassigning untrained and well-trained workers could reduce costs. In this research, we shift our focus to a parallel production line, which more closely mirrors real-world production environments and can accommodate more complex scenarios. We propose an optimal assignment problem for three groups of workers aimed at minimizing expected cost, particularly when the number of stages varies across each line.

This research is structured in the prescribed order as follows: Initially, the reset model, a simplified version of RCMwMP, is introduced. Subsequently, the propositions for optimal assignment are demonstrated through detailed derivation. Finally, the optimization rules for assignments concerning scenarios with minority untrained or well-trained workers are discussed through numerical analyses.

2 Model Explanation

In this section, we consider the reset model which is a simple model of the RCMwMP.

2.1 Reset Model of RCMwMP

The model is based on definitions provided by Zhao *et al.* [6], as illustrated in Figure 1. This figure depicts the relationship between expected cost and operating time for a single series production line within a parallel configuration.

- We consider a production line with m parallel lines of n stages.
- The production (we call it job in the following article) is processed in a rotation of stage 1, stage 2, ..., stage n of each line and m productions will be processed by all mn stages.
- Z is the cycle time of all of the stages, which can be also considered as a target operating time. All of the jobs should be accomplished in current stage and moved to next stage by time Z.
- However, because of the various operating abilities of workers, the actual operating time may not always obey the limit of target operating time Z. Idle and delay should be also concerned in this model. For $1 \leq w \leq n$, the operating time of stage w is donated by T_w.

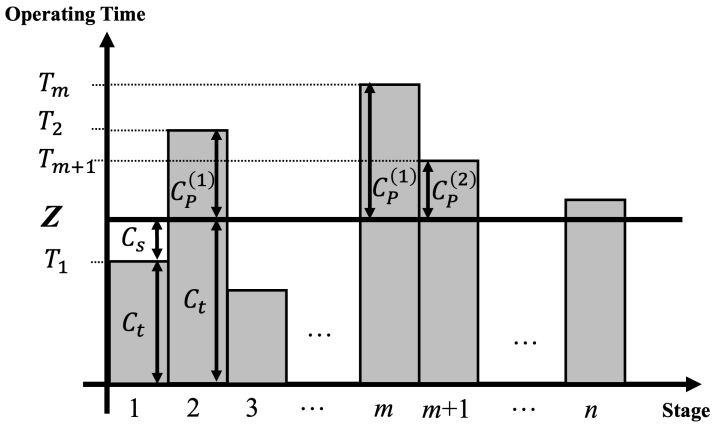


Figure 1: Description of Idle and Delay Cost in Reset Model of RCMwMP

In this model, a regular operating cost C_t (> 0) per unit time will permanently occur during a target operating time Z, regardless whether it is idle or delay. It is for the reason that although the job is accomplished prematurely in current stage, the next stage may be occupied by another job. The job must wait for its start. As a result, an idle cost per unit time, C_S (≥ 0), arises. On the other hand, if the operating time exceeds Z, it is assumed that delays in process time can be compensated by implementing overtime work or deploying spare workers at this stage. Consequently, overtime or additional resources will be required to meet the target operating time Z.

Thus, a delay cost per unit time, $C_P^{(h)}$ (> 0) , arises (that is why we call the model a "Reset model"). As a summary of above, we suppose the following costs:

- The operating cost per unit time, C_t (> 0), for the target operating time limit occurs in each stage.
- When $T_w < Z$, the idle cost per unit time, C_S (≥ 0) occurs.
- When $T_w > Z$, the delay cost per unit time, $C_P^{(h)}$ (≥ 0), occurs in the stage if delay occurs in consecutive h stages before its stage, for $h = 1, 2, 3, ..., n$. If the delay continues for several stages, it can be considered that it will cost more for recover the delay. It is supposed that $C_P^{(h)}$ is increasing in h, which can be expressed as $0 \leq C_P^{(1)} \leq C_P^{(2)} < \cdots \leq C_P^{(n)}$.

2.2 The Assumption of Operating Abilities of Workers

This research aims to search the optimal assignment of workers for different operating abilities, so the reasonable assumption of the property of workers is particularly important. The assumption is the following:

- Only one worker can be assigned to each stage on each line, for a total of mn workers. Each stage must be assigned with one worker.
- The operating time of workers is self-dependent. The operating ability is decided by the property of worker own and is not influenced by the operating status such as idle or delay.
- In this research, the workers are distinguished into three groups of workers by the operating abilities, marking as A, B and C. Worker A (1 worker) represents untrained worker whose operating ability is lower than others, C (1 worker) represents well-trained worker and B ($mn-2$ workers) represents the general worker. In this research, we refer to both untrained and well-trained workers collectively as "special workers".
- ℓ represents the workers in a production each line. For $\ell = 1, 2, \ldots, m$. These workers have different probability of idle or delay. If the operating time of worker is marked as T_ℓ, where $\ell \in \{A, B, C\}$.
  - P_ℓ: The probability of worker ℓ becoming idle,
  - Q_ℓ: The probability of the worker ℓ becoming delayed,
  - TS_ℓ: The expected idle time of the worker ℓ,
  - TL_ℓ: The expected delay time of the worker ℓ.

  Formally, we have

$$\begin{aligned} P_\ell &= \Pr\{T_\ell \le Z\}, \\ Q_\ell &= \Pr\{T_\ell > Z\}, \\ TS_\ell &= \mathrm{E}\big[(Z - T_\ell) \cdot I(T_\ell \le Z)\big], \\ TL_\ell &= \mathrm{E}\big[(T_\ell - Z) \cdot I(T_\ell > Z)\big], \end{aligned}$$

  where $I(O)$ is an index function and given as follows:

$$I(O) = \begin{cases} 1 & \text{if } O \text{ is true,} \\ 0 & \text{otherwise.} \end{cases}$$

2.3 Optimal Assignment Problem under Reset Model

One of the most critical issues in this research is determining how to allocate workers to stages to minimize the total expected cost. We refer to this issue as the optimal assignment problem. To describe the optimal assignment problem for parallel production lines, we define the following notations:

- By (i, j), we denote the j-th stage in the i-th line, where $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$.

- By $\Omega = \{(i,j) \mid i = 1,2,...,m, j = 1,2,...,n\}$, we denote the set of all stages.

Let $\mathbf{A}, \mathbf{C} \subseteq \Omega$ satisfying $\mathbf{A} \cap \mathbf{C} = \emptyset$.

- By $\pi(\mathbf{A}, \mathbf{C})$, we denote an assignment in which workers A are assigned to stages in the set $\mathbf{A}$, workers C are assigned to stages in the set $\mathbf{C}$, and workers B are assigned in stages in the set $\Omega \setminus (\mathbf{A} \cup \mathbf{C})$.

- By $TC^{(p)}(mn; \pi(\mathbf{A}, \mathbf{C}))$, we denote the total expected cost of all mn stages from stage 1 to stage mn of a parallel production line when special workers are assigned the assignment $\pi(\mathbf{A}, \mathbf{C})$.

$$TC^{(p)}(mn; \pi(\mathbf{A}, \mathbf{C})) = mnC_t + f^{(p)}(mn; \pi(\mathbf{A}, \mathbf{C})),$$

  where $f^{(p)}(mn; \pi(\mathbf{A}, \mathbf{C}))$ is the sum of total idle cost and total delays cost in all mn stages from stage 1 to stage mn as a result of special workers being assigned as specified in $\pi(\mathbf{A}, \mathbf{C})$.

By using these notations, the optimal assignment problem with multiple periods becomes the problem of obtaining an assignment in the following equation:

$$TC^{(p)}(mn; \pi(\mathbf{A}^*, \mathbf{C}^*)) = mnC_t + \min_{\mathbf{A},\mathbf{C} \subseteq \Omega, \mathbf{A} \cap \mathbf{C} = \emptyset} f^{(p)}(mn; \pi(\mathbf{A}, \mathbf{C})).$$

In this research, we call $\pi(\mathbf{A}^*, \mathbf{C}^*)$ the optimal assignment.

3 Preliminary Analysis in a Parallel Production Line

In this research, we assume that three groups of workers based on their operating abilities: one untrained worker A, one well-trained worker C and $mn-2$ general workers B. This section presents preliminary analysis results of the optimal assignment for configurations such as $m = 2$, $n_1 = 2$, $n_2 = 3$ (where the parallel production line consists of two series lines, one with 2 stages and the other with 3 stages) and $m = 2$, $n_1 = 2$, $n_2 = 4$ (where one line has 2 stages and the other has 4 stages). We propose the optimal assignment propositions for these scenarios as Proposition 1 and Proposition 2.

3.1 The Case of Two Series Lines, One Line has 2 Stages, the other Line has 3 Stages

Here, the preliminary analysis results regarding the optimal assignment are described for $m = 2$, $n_1 = 2$, $n_2 = 3$.

Proposition 1 *Asumme that* $Q_A > Q_B > Q_C$ *and* $TL_B/Q_B > TL_C/Q_C$.

1. *If*

$$\begin{aligned}
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &< \frac{(Q_A - Q_B) \cdot (TL_B - TL_C)}{(Q_B - Q_C) \cdot TL_B \cdot Q_B} - \frac{1}{Q_B}, \\
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &< \frac{Q_A - Q_B}{Q_B \cdot Q_B}, \\
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &> \frac{Q_B - Q_C}{Q_A \cdot Q_C - Q_B \cdot Q_B},
\end{aligned}$$

Then the optimal assignment is $\pi(\{(1,1)\};\{(1,2)\})$.

2. *If*

$$\begin{aligned}
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &> \frac{(Q_A - Q_B) \cdot (TL_B - TL_C)}{(Q_B - Q_C) \cdot TL_B \cdot Q_B} - \frac{1}{Q_B}, \\
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &< \frac{(Q_B - Q_C) \cdot TL_B}{(TL_B \cdot Q_C - TL_C \cdot Q_B) \cdot Q_B}, \\
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &< \frac{TL_B - TL_C}{TL_B \cdot Q_C},
\end{aligned}$$

Then the optimal assignment is $\pi(\{(1,1)\};\{(2,2)\})$.

3. *If*

$$\begin{aligned}
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &> \frac{Q_A - Q_B}{Q_B \cdot Q_B}, \\
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &> \frac{(Q_B - Q_C) \cdot TL_B}{(TL_B \cdot Q_C - TL_C \cdot Q_B) \cdot Q_B}, \\
\frac{C_P^{(3)} - C_P^{(2)}}{C_P^{(2)} - C_P^{(1)}} &> \frac{TL_B \cdot (Q_A - Q_C) + TL_C \cdot (Q_B - Q_A)}{TL_B \cdot Q_A \cdot Q_C - TL_A \cdot Q_B \cdot Q_B},
\end{aligned}$$

Then the optimal assignment is $\pi(\{(1,1)\};\{(2,3)\})$.

4. If

$$\begin{aligned}
\frac{C_P^{(3)}-C_P^{(2)}}{C_P^{(2)}-C_P^{(1)}} &< \frac{Q_B-Q_C}{(Q_A\cdot Q_C-Q_B\cdot Q_B)},\\
\frac{C_P^{(3)}-C_P^{(2)}}{C_P^{(2)}-C_P^{(1)}} &> \frac{TL_B-TL_C}{TL_B\cdot Q_C},\\
\frac{C_P^{(3)}-C_P^{(2)}}{C_P^{(2)}-C_P^{(1)}} &< \frac{TL_B\cdot(Q_A-Q_C)+TL_C\cdot(Q_B-Q_A)}{TL_B\cdot Q_A\cdot Q_C-TL_C\cdot Q_B\cdot Q_B},
\end{aligned}$$

Then the optimal assignment is $\pi(\{(2,1)\};\{(2,2)\})$.

3.2 The Case of Two Series Lines, One Line has 2 Stages, the other Line has 4 Stages

Here, the preliminary analysis results regarding the optimal assignment are described for $m=2$, $n_1=2$, $n_2=4$.

Proposition 2 *Asumme that* $Q_A>Q_B>Q_C$ *and* $TL_B/Q_B>TL_C/Q_C$.

1. If

$$\begin{aligned}
&\frac{TL_B-TL_C}{(Q_C-Q_B)\cdot TL_B}<\\
&\quad\frac{(C_P^{(2)}-C_P^{(1)})+(C_P^{(3)}-C_P^{(2)})\cdot Q_B+(C_P^{(4)}-C_P^{(3)})\cdot Q_B\cdot Q_B}{(C_P^{(2)}-C_P^{(1)})\cdot(Q_B-Q_A)+(C_P^{(3)}-C_P^{(2)})\cdot Q_B\cdot Q_B},\\
&\frac{Q_A-Q_B}{Q_B\cdot Q_B}>\frac{(C_P^{(3)}-C_P^{(2)})+(C_P^{(4)}-C_P^{(3)})\cdot Q_B}{(C_P^{(2)}-C_P^{(1)})},\\
&\frac{Q_B-Q_C}{Q_A\cdot Q_C-Q_B\cdot Q_B}<\frac{(C_P^{(3)}-C_P^{(2)})+(C_P^{(4)}-C_P^{(3)})\cdot Q_B}{(C_P^{(2)}-C_P^{(1)})+(C_P^{(3)}-C_P^{(2)})\cdot Q_B},
\end{aligned}$$

Then the optimal assignment is $\pi(\{(1,1)\};\{(1,2)\})$.

2. If

$$\begin{aligned}
&\frac{TL_B - TL_C}{(Q_C - Q_B) \cdot TL_B} > \\
&\quad \frac{(C_P^{(2)} - C_P^{(1)}) + (C_P^{(3)} - C_P^{(2)}) \cdot Q_B + (C_P^{(4)} - C_P^{(3)}) \cdot Q_B \cdot Q_B}{(C_P^{(2)} - C_P^{(1)}) \cdot (Q_B - Q_A) + (C_P^{(3)} - C_P^{(2)}) \cdot Q_B \cdot Q_B}, \\
&TL_C \cdot Q_B - TL_B \cdot Q_C > \\
&\quad \frac{(C_P^{(2)} - C_P^{(1)}) \cdot (Q_C - Q_B) \cdot TL_B + (C_P^{(3)} - C_P^{(2)}) \cdot (TL_C - TL_B) \cdot Q_B \cdot Q_B}{(C_P^{(3)} - C_P^{(2)}) \cdot Q_B + (C_P^{(4)} - C_P^{(3)}) \cdot Q_B \cdot Q_B}, \\
&\frac{TL_B \cdot Q_A \cdot Q_C - TL_C \cdot Q_B \cdot Q_B}{Q_B - Q_A} > \\
&\quad \frac{(C_P^{(2)} - C_P^{(1)}) \cdot (TL_C - TL_B) - (C_P^{(4)} - C_P^{(3)}) \cdot TL_B \cdot Q_B \cdot Q_C}{C_P^{(3)} - C_P^{(2)}}
\end{aligned}$$

Then the optimal assignment is $\pi(\{(1,1)\};\{(2,3)\})$.

3. If

$$\begin{aligned}
&\frac{Q_A - Q_B}{Q_B \cdot Q_B} < \frac{(C_P^{(3)} - C_P^{(2)}) + (C_P^{(4)} - C_P^{(3)}) \cdot Q_B}{C_P^{(2)} - C_P^{(1)}}, \\
&TL_C \cdot Q_B - TL_B \cdot Q_C < \\
&\quad \frac{(C_P^{(2)} - C_P^{(1)}) \cdot (Q_C - Q_B) \cdot TL_B + (C_P^{(3)} - C_P^{(2)}) \cdot (TL_C - TL_B) \cdot Q_B \cdot Q_B}{(C_P^{(3)} - C_P^{(2)}) \cdot Q_B + (C_P^{(4)} - C_P^{(3)}) \cdot Q_B \cdot Q_B}, \\
&(TL_C \cdot Q_B \cdot Q_B - TL_B \cdot Q_A \cdot Q_C)(C_P^{(4)} - C_P^{(3)}) < \\
&\quad \left(C_P^{(1)} \cdot Q_A - C_P^{(2)}(Q_A - Q_B)\right) \cdot (TL_C - TL_B) + C_P^{(2)} \cdot (Q_B - Q_C) \cdot TL_B + \\
&\quad \left((C_P^{(3)} - C_P^{(2)}) \cdot Q_B - C_P^{(1)}\right) \cdot (TL_C \cdot Q_B - TL_B \cdot Q_C) + \\
&\quad (C_P^{(3)} - C_P^{(2)}) \cdot (Q_B \cdot Q_B - Q_A \cdot Q_C) \cdot TL_B
\end{aligned}$$

Then the optimal assignment is $\pi(\{(1,1)\};\{(2,4)\})$.

4. If

$$\frac{Q_B - Q_C}{Q_A \cdot Q_C - Q_B \cdot Q_B} > \frac{(C_P^{(3)} - C_P^{(2)}) + (C_P^{(4)} - C_P^{(3)}) \cdot Q_B}{(C_P^{(2)} - C_P^{(1)}) + (C_P^{(3)} - C_P^{(2)}) \cdot Q_B}$$

$$\frac{TL_B \cdot Q_A \cdot Q_C - TL_C \cdot Q_B \cdot Q_B}{Q_B - Q_A} <$$

$$\frac{(C_P^{(2)} - C_P^{(1)}) \cdot (TL_C - TL_B) + (C_P^{(4)} - C_P^{(3)}) \cdot TL_B \cdot Q_B \cdot Q_C}{C_P^{(3)} - C_P^{(2)}}$$

$$TL_C \cdot Q_B - TL_B \cdot Q_C >$$

$$\frac{(C_P^{(2)} - C_P^{(1)}) \cdot (Q_C - Q_B) \cdot TL_B + (C_P^{(3)} - C_P^{(2)}) \cdot (TL_C - TL_B) \cdot Q_B \cdot Q_B}{(C_P^{(3)} - C_P^{(2)}) \cdot Q_B + (C_P^{(4)} - C_P^{(3)}) \cdot Q_B \cdot Q_B},$$

$$(TL_C \cdot Q_B \cdot Q_B - TL_B \cdot Q_A \cdot Q_C)(C_P^{(4)} - C_P^{(3)}) <$$

$$\big(C_P^{(1)} \cdot Q_A - C_P^{(2)}(Q_A - Q_B)\big) \cdot (TL_C - TL_B) + C_P^{(2)} \cdot (Q_B - Q_C) \cdot TL_B +$$

$$\big((C_P^{(3)} - C_P^{(2)}) \cdot Q_B - C_P^{(1)}\big) \cdot (TL_C \cdot Q_B - TL_B \cdot Q_C) +$$

$$(C_P^{(3)} - C_P^{(2)}) \cdot (Q_B \cdot Q_B - Q_A \cdot Q_C) \cdot TL_B$$

Then the optimal assignment is $\pi(\{(2,1)\};\{(2,2)\})$.

4 Numerical Experiments

Note that the propositions mentioned previously hold for any operating time distribution. However, for numerical experiments, it is necessary to assume a specific operating time distribution. In this case, we assume that the operating time of each worker follows an exponential distribution. Throughout this section, we will refer to $f_\ell(t) = \mu e^{-\mu_\ell t}$ as the probability density function of worker ℓ's operating time for $\ell \in \{A, B, C\}$. The following parameters are used in the numerical experiments:

- The Number of Stages: $n_1 = 2$, $n_2 = 3, 4, 5$.
- Delay Cost: $C_P^{(1)} = 40$, $C_P^{(2)} = 50$, $C_P^{(3)} = 60$, $C_P^{(4)} = 70$, $C_P^{(5)} = 80$.
- Target Operating Time: $Z = 2$.
- Idle Cost: $C_S = 20$.

In this section, we describe the outcomes of numerical experiments conducted for Propositions 1 and 2. These experiments involve varying the parameters μ_A and μ_B, which represent the operating times of different worker groups. The results demonstrate that the experimental findings align with the theoretical propositions previously established, confirming their applicability and accuracy in modeling real-world scenarios.

Table 1: Optimal assignment for varying the operating ability μ_A of worker A ($\mu_A = 0.1 \sim 0.8$)

μ_A	μ_B	μ_C	$n_1 = 2, n_2 = 3$	$n_1 = 2, n_2 = 4$	$n_1 = 2, n_2 = 5$
0.1	0.9	1.0	AC, BBB	$AC, BBBB$	$AC, BBBBB$
0.2	0.9	1.0	AC, BBB	$AC, BBBB$	$AC, BBBBB$
0.3	0.9	1.0	AC, BBB	$AC, BBBB$	$AC, BBBBB$
0.4	0.9	1.0	AC, BBB	$AC, BBBB$	$AC, BBBBB$
0.5	0.9	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$
0.6	0.9	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$
0.7	0.9	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$
0.8	0.9	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$

Table 2: Optimal assignment for varying the operating ability μ_B of worker A ($\mu_B = 0.2 \sim 0.9$)

μ_A	μ_B	μ_C	$n_1 = 2, n_2 = 3$	$n_1 = 2, n_2 = 4$	$n_1 = 2, n_2 = 5$
0.1	0.2	1.0	BB, ACB	$BB, ACBB$	$AB, BBCBB$
0.1	0.3	1.0	BB, ACB	$BB, ACBB$	$AB, BBCBB$
0.1	0.4	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$
0.1	0.5	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$
0.1	0.6	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$
0.1	0.7	1.0	BB, ACB	$BB, ACBB$	$BB, ACBBB$
0.1	0.8	1.0	AC, BBB	$AC, BBBB$	$AC, BBBBB$
0.1	0.9	1.0	AC, BBB	$AC, BBBB$	$AC, BBBBB$

5 Conclusion

In this research, we proposed an optimal assignment strategy aimed at minimizing the total expected cost by utilizing a restricted-cycle model with multiple periods, tailored to scenarios where the number of stages in each line varies. Additionally, through numerical experimentation, we evaluated the performance of the optimal assignment rule under conditions that deviate from those initially proposed, providing a broader perspective on the rule's applicability and effectiveness.

As a future research, we propose an optimal assignment rule of parallel production line with m lines and n stages.

References

[1] J. Liu and C. Reeves. Constructive and composite heuristic solutions to the $P||\sum C_i$ scheduling problem. *European Journal of Operational Research*, 132(2):439–452, 2001.

[2] C. Rajendran and H. Ziegler. Ant-colony algorithms for permutation flowshop scheduling to minimize makespan/total flowtime of jobs. *European Journal of Operational Research*, 155(2):426–438, 2004.

[3] M. E. Salveson. The assembly-line balancing problem. *Journal of Industrial Engineering*, 6(3):18–25, 1955.

[4] H. Yamamoto, M. Matsui, and X. Bai. A branch and bound method for the optimal assignment during a limit-cycle problem with multiple periods (in japanese). *Journal of Japan Industrial Management Association*, 58(1):37–43, 2007.

[5] H. Yamamoto, M. Matsui, and J. Liu. A basic study on a limited-cycle problem with multi periods and the optimal assignment problem (in japanese). *Journal of Japan Industrial Management Association*, 57(1):23–31, 2006.

[6] X. Zhao, H. Yamamoto, and J. Sun. Optimal assignment of three special workers in reset limited-cycle multiple periods with variable target processing time. In *The 19th Asia Pacific Industrial Engineering and Management System Conference*, Hong Kong, China, 2018.

[7] X. Zhao, H. Yamamoto, and J. Sun. Optimal assignment with three special workers in a reset limited-cycle problem with multiple periods. In *The Fourteenth International Conference on Industrial Management*, Hangzhou, China, 2018.

List of Authors

24th Czech-Japan Seminar on Data Analysis and Decision Making
September 9 - 12, 2024
Telč, Czech Republic

Editors:
Václav Kratochvíl, Masahiro Inuiguchi, Ondřej Čepek, Yoshifumi Kusunoki

Organized by:
ÚTIA - Institute of Information Theory and Automation, Czech Academy of Sciences

Sponsors:
Institute of Information Theory and Automation, Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University
Institute for Research and Applications of Fuzzy Modeling, University of Ostrava
Faculty of Management, Prague University of Economics and Business
Centre Telč, Institute of Theoretical and Applied Mechanics, Czech Academy of Sciences

http://cjs.utia.cas.cz

ÚTIA

ISBN: 978-80-905688-0-8