UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

Guilherme Dean Pelegrina

Multicriteria decision making: Dealing with criteria interactions by means of latent variable analysis

Tomada de decisão multicritério: Lidando com interações entre critérios por meio da análise de variáveis latentes

Campinas
2020

Guilherme Dean Pelegrina

Multicriteria decision making: Dealing with criteria interactions by means of latent variable analysis

Tomada de decisão multicritério: Lidando com interações entre critérios por meio da análise de variáveis latentes

Thesis presented to the Faculty of Electrical and Computer Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering, in the area of Telecommunications and Telematics, under the Cotutelle Agreement between Unicamp and Université Paris I Panthéon-Sorbonne.

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica, na Área de Telecomunicações e Telemática, no âmbito do Acordo de Cotutela firmado entre a Unicamp e a Université Paris I Panthéon-Sorbonne.

Supervisors/*Orientadores*: Prof. Dr. João Marcos Travassos Romano (Unicamp) e Prof. Dr. Michel Grabisch (Université Paris I Panthéon-Sorbonne)
Co-supervisor/*Co-orientador*: Prof. Dr. Leonardo Tomazeli Duarte

Este exemplar corresponde à versão final da tese defendida pelo aluno Guilherme Dean Pelegrina, e orientada pelo Prof. Dr. João Marcos Travassos Romano e pelo Prof. Dr. Michel Grabisch

---

Campinas
2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

P362m Pelegrina, Guilherme Dean, 1989-
Multicriteria decision making : dealing with criteria interactions by means of latent variable analysis / Guilherme Dean Pelegrina. – Campinas, SP : [s.n.], 2020.

Orientadores: João Marcos Travassos Romano e Michel Grabisch.
Coorientador: Leonardo Tomazeli Duarte.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.
Em cotutela com: Université Paris I Panthéon-Sorbonne.

1. Processo decisório. 2. Processamento de sinais. 3. Estimativa de parâmetro. 4. Variáveis latentes. I. Romano, João Marcos Travassos, 1960-. II. Grabisch, Michel. III. Duarte, Leonardo Tomazeli, 1982-. IV. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. VI. Título.

Informações para Biblioteca Digital

Título em outro idioma: Tomada de decisão multicritério : lidando com interações entre critérios por meio da análise de variáveis latentes
Palavras-chave em inglês:
Decision making
Signal processing
Parameter estimation
Latent variables
Área de concentração: Telecomunicações e Telemática
Titulação: Doutor em Engenharia Elétrica
Banca examinadora:
João Marcos Travassos Romano [Orientador]
Denis Bouyssou
Michel Grabisch
Christian Jutten
Aline de Oliveira Neves Panazio
Data de defesa: 27-11-2020
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-7301-6167
- Currículo Lattes do autor: http://lattes.cnpq.br/9984556081239908

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidato: Guilherme Dean Pelegrina RA: 093821
Data de defesa: 27 de novembro de 2020

Título da tese: "*Multicriteria decision making: Dealing with criteria interactions by means of latent variable analysis*"

Prof. Dr. João Marcos Travassos Romano (Presidente, FEEC/UNICAMP)
Prof. Dr. Michel Grabisch (Université Paris I Panthéon-Sorbonne)
Prof. Dr. Christian Jutten (Université Grenoble Alpes)
Prof. Dr. Denis Bouyssou (Université Paris Dauphine)
Prof. Dr. Aline de Oliveira Neves Panazio (Universidade Federal do ABC)

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

Université Paris 1 Panthéon - Sorbonne

PROCES VERBAL DE SOUTENANCE DU 27/11/2020 A 15h00

ANNEE UNIVERSITAIRE 2019/2020

Etudiant : M. GUILHERME DEAN PELEGRINA né le : 02/10/1989

Diplôme : Doctorat Mathématiques

Titre des travaux : Multicriteria Decision Making: Dealing with Criteria Interactions by Means of Latent Variable Analysis

Ecole doctorale : Économie Panthéon-Sorbonne

Directeur : M. MICHEL GRABISCH

Codirecteur : M. JOAO MARCOS TRAVASSOS ROMANO

Lieu de soutenance : UNIVERSIDADE ESTADUAL DE CAMPINAS (BRESIL)

La soutenance est publique.

Résultat : Admis

Mention :

Avis de reproduction : En l'état

Membres du Jury

Nom	Qualité	Etablissement	Rôle	Signature
M. DENIS BOUYSSOU	DIRECTEUR DE RECHERCHE CNRS	UNIVERSITE PARIS 9 - DAUPHINE	Rapporteur	
Mme ALINE NEVES	PROFESSEUR DES UNIVERSITES	Federal University of São Carlos	Rapporteur	
M. MICHEL GRABISCH	PROFESSEUR DES UNIVERSITES	Université Paris 1 Panthéon-Sorbonne	Directeur	
M. CHRISTIAN JUTTEN	PROFESSEUR DES UNIVERSITES	UNIVERSITE GRENOBLE 1 JOSEPH FOURIER	Président	
M. CHRISTOPHE LABREUCHE	INGENIEUR DE RECHERCHE	TRT	Examinateur	
M. LEONARDO TOMAZELI DUARTE	MAITRE DE CONFERENCES	UNIVERSIDADE ESTADUAL DE CAMPINAS	Co-directeur	Leonardo Tomazeli Duarte
M. JOAO MARCOS TRAVASSOS ROMANO	PROFESSEUR DES UNIVERSITES	UNIVERSIDADE ESTADUAL DE CAMPINAS	Co-directeur	

To my wife, Thais,
to my parents, Rosemary and José Carlos,
and to my grandparents, Catarina and Artemio.

Acknowledgments

Firstly, I would like to thank my wife, Thais, my parents, Rosemary and José Carlos, and my grandparents, Catarina and Artemio, for the constant support and motivation during my Ph.D.

I also would like to thank Prof. João Romano for having accepted and guided me on this new journey. I am also grateful to all my colleagues and professors at DSPCom laboratory for their friendship and joyful moments.

Special thanks to Prof. Leonardo Duarte for his friendship and exceptional guidance since 2011. His advices contributed significantly to my academic background. I also thank all my colleagues and professors at FCA for their friendship and valuable discussions.

I owe great thanks to Prof. Michel Grabisch, for receiving me into his research group and accepted to be my supervisor in the cotutelle program. The research internship at CES contributed enormously to the development of my thesis. I also thank all colleagues at CES for their kind reception and friendship.

Thanks to the committee members, professors Christian Jutten, Denis Bouyssou, Christophe Labreuche and Aline Neves for their availability and valuable suggestions that contributed to the improvement of this manuscript.

I also would like to thank all the employees of FEEC and CES for their support.

This study was financed in part by the grants #2016/21571-4 and #2017/23879-9, São Paulo Research Foundation (FAPESP).

Resumo

Um problema típico em tomada de decisão multicritério consiste em ordenar um conjunto de alternativas de acordo com suas avaliações em um conjunto de critérios. A fim de lidar com tal problema, diversos métodos foram desenvolvidos na literatura. No entanto, grande parte deles não foram construídos de tal forma a considerar informações estruturais contidas nos dados de decisão coletados. Por exemplo, redundâncias entre critérios são frequentemente observadas em situações práticas. Consequentemente, a presença de relações entre critérios pode influenciar o ordenamento obtido.

Nesta tese de doutorado, propomos novos métodos que podem ser usados para lidar com critérios redundantes. Um aspecto interessante é que a redundância pode ser explicada por fatores latentes que estão associados a dois ou mais critérios simultaneamente. Em outras palavras, as avaliações coletadas consistem em uma mistura de dados latentes. Portanto, podemos formular o problema abordado como um problema de separação cega de fontes e extrair as informações relevantes para recuperar esses dados. Essas informações são usadas para aprimorar alguns métodos existentes a fim de mitigar resultados enviesados e, assim, alcançar uma classificação mais justa das alternativas. Além disso, também usamos essas informações para ajustar os parâmetros da média aritmética ponderada. Nossos experimentos atestam que a abordagem proposta penaliza critérios redundantes, diminuindo seus respectivos pesos. Portanto, o impacto da redundância na classificação obtida é amenizado.

Também revisitamos duas funções de agregação que modelam relações intercritérios: a integral de Choquet e o modelo multilinear. No entanto, notamos que poucos trabalhos foram conduzidos no contexto desta última função. Sendo assim, neste estudo, exploramos tanto resultados teóricos, formulando o modelo multilinear 2-aditivo, quanto resultados experimentais, aplicando abordagens supervisionadas e não supervisionadas para identificação de capacidade. Nas abordagens supervisionadas, consideramos termos de regularização no modelo de otimização, o que pode levar a uma capacidade próxima da aditiva ou da 2-aditiva. Com relação às abordagens não supervisionadas, associamos alguns parâmetros à medidas de similaridade entre critérios e a uma medida que indica o impacto de um conjunto de critérios na saída do modelo, chamada índice de Sobol. Os parâmetros encontrados através das abordagens propostas levaram à avaliações globais imparciais.

Além do problema multicritério, também podemos nos deparar com situações envolvendo vários tomadores de decisão. Nesse caso, podem ocorrer interações entre critérios e entre decisores. Neste estudo, investigamos métodos que são capazes de ajustar os pesos usados na média aritmética ponderada a fim de penalizar tanto os critérios correlacionados quanto os decisores “correlacionados”, ou seja, indivíduos que podem não estar agindo independentemente dos demais. Outra análise que realizamos é a aplicação da integral de

Choquet para modelar interações entre critérios e entre decisores. Nesse caso, exploramos uma representação alternativa para a integral de Choquet multinível e investigamos se há comutatividade no procedimento de agregação em duas etapas.

Com base em nossos experimentos e nos resultados interessantes alcançados, pretendemos contribuir para a discussão sobre o uso de técnicas de análise de variáveis latentes em problemas de tomada de decisão multicritério. Além disso, pretendemos motivar o desenvolvimento de trabalhos futuros sobre este assunto.

Palavras-chaves: Processo decisório; Processamento de sinais; Estimativa de parâmetro; Variáveis latentes.

Abstract

A typical problem in multicriteria decision making consists in ranking a set of alternatives according to their evaluations by a set of criteria. Aiming at dealing with this problem, several methods were developed in the literature. However, most of them were not conceived in order to consider structural information contained within the collected decision data. For instance, redundancies among criteria is frequently observed in practical situations, which may bias the achieved ranking.

In this Ph.D. thesis, we propose novel methods that can be used to deal with redundant criteria. An interesting aspect is that the redundancy may be explained by latent factors that are associated with two or more criteria simultaneously. In other words, the collected evaluations consist in a mixture of latent data. Therefore, we may formulate the addressed problem as a blind source separation one and extract the relevant information used to recover these data. We use this information to improve existing methods in order to mitigate biased results and achieve a fairer ranking of alternatives. Moreover, we also use this information to adjust the weighted arithmetic mean parameters. Our experiments attest that the proposed approach penalizes redundant criteria by decreasing their associated weights. Therefore, the impact of such a redundancy in the obtained ranking is softened.

We also revisit two aggregation functions that model intercriteria relations, the Choquet integral and the multilinear model. However, we could note that few works have been conducted in the context of the latter one. Therefore, in this study, we provide both theoretical results, by formulating the 2-additive multilinear model, and experimental results, by applying supervised and unsupervised approaches for capacity identification. In supervised approaches, we consider regularization terms in the optimization model, which may lead to a capacity close to the additive or 2-additive ones. With respect to the unsupervised approaches, we associate some parameters to similarity measures among criteria and to a measure of how a set of criteria impacts on the output function, called Sobol' index. The parameters obtained by our proposals led to unbiased overall evaluations.

Other than the multicriteria problem, we also may be concerned with situations involving multiple decision makers. In this case, one may have interactions among criteria and among individuals. In this study, we investigate methods that are able to adjust the weights used in the weighted arithmetic mean in order to penalize both correlated criteria and "correlated" decision makers, i.e., individuals that may not be acting independently from the other ones. Another analysis that we conduct is the application of the Choquet integral to model interactions among criteria and among decision makers. In this case, we exploit an alternative representation for the multilevel Choquet integral and investigate if the 2-step aggregation procedure commutes.

Based on our experiments and the interesting achieved results, we aim at contributing to the discussion on the use of latent variable analysis techniques in multicriteria decision making problems and motivating the development of future works on this subject.

Keywords: Decision making; Signal processing; Parameter estimation; Latent variables.

Résumé

Un problème typique de la prise de décision multicritère consiste à classer un ensemble d'alternatives en fonction de leurs évaluations par un ensemble de critères. Afin de traiter ce problème, plusieurs méthodes ont été développées dans la littérature. Cependant, la plupart d'entre elles n'ont pas été conçues pour prendre en compte les informations structurelles contenues dans les données de décision collectées. Par exemple, des redondances entre critères sont fréquemment observées dans des situations pratiques. En conséquence, la présence de ces relations inter-critères peut biaiser le classement obtenu.

Dans cette thèse, nous proposons de nouvelles méthodes qui peuvent être utilisées pour traiter les critères redondants. Un aspect intéressant dans cette situation est que la redondance peut être expliquée par des facteurs latents qui sont associés à deux ou plusieurs critères simultanément. En d'autres termes, les évaluations collectées consistent en un mélange de données latentes. Par conséquent, nous formulons le problème de décision comme un problème de séparation aveugle des sources et extrayons les informations pertinentes qui seront utilisées pour récupérer les données latentes. Nous utilisons ces informations pour améliorer quelques méthodes existantes afin d'atténuer les résultats biaisés et d'obtenir un classement plus juste des alternatives. En outre, nous utilisons également ces informations pour ajuster les paramètres de la moyenne arithmétique pondérée. Les expériences réalisées attestent que l'approche proposée pénalise les critères redondants en diminuant leurs poids associés. En conséquence, l'impact d'une telle redondance dans le classement obtenu est atténué.

Nous revisitons également quelques méthodes existantes qui sont capables de modéliser les relations inter-critères, comme l'intégrale de Choquet et le modèle multilinéaire. Cependant, on peut noter que peu de travaux ont été dévéloppés dans le cadre du modèle multilinéaire. Par conséquent, dans cette étude, nous fournissons des résultats théoriques, en formulant le modèle multilinéaire 2-additif, et des résultats expérimentaux, en appliquant des approches supervisées et non supervisées pour l'identification des capacités. Dans les approches supervisées (basées sur les données de décision et les évaluations globales comme données d'apprentissage), nous avons considéré des termes de régularisation dans le modèle d'optimisation, ce qui peut conduire à des capacités proches de celles additives ou 2-additives. En ce qui concerne les approches non supervisées, où la capacité est estimée uniquement sur des informations extraites de la base de données et d'une hypothèse sur le problème de décision que nous voulons traiter (redondances, par exemple), nous avons associé certains paramètres du modèle à des mesures de similarité entre les critères et à une mesure de l'impact d'un ensemble de critères sur la sortie du modèle, appelée indice de Sobol. En utilisant cettes approches, nous pourrions définir un ensemble de paramètres pouvant conduire à des évaluations globales non biaisées.

Outre le problème de la prise de décision multicritère, nous pouvons également être concernés par des problèmes de décision avec plusieurs décideurs. Dans cette situation, on peut avoir des interactions entre les critères et aussi entre les décideurs. Dans cette étude, nous étudions des méthodes capables d'ajuster les poids utilisés dans la moyenne arithmétique pondérée afin de pénaliser les critères corrélés et les décideurs "corrélés", c'est-à-dire les individus qui n'agissent pas indépendamment des autres. Une autre analyse que nous menons est l'application de l'intégrale de Choquet pour modéliser les interactions entre les critères et entre les décideurs. Dans le but d'exploiter cette intégrale de Choquet à plusieurs niveaux, nous effectuons deux investigations. Dans le première, nous revisitons un résultat fourni dans la littérature qui comprend une représentation alternative de l'intégrale de Choquet à plusieurs niveaux. Dans une seconde analyse, nous vérifions si la procédure d'agrégation en deux étapes commute.

Sur la base de nos expériences et des résultats intéressants obtenus, nous souhaitons contribuer à la discussion sur l'utilisation des techniques d'analyse des variables latentes dans les problèmes de prise de décision multicritère. En outre, nous avons l'intention de motiver le développement de futurs travaux sur ce sujet.

Mots-clé: Prise de décision; Traitement du signal; Estimation des paramètres; Variables latentes.

List of Figures

List of Tables

List of Acronyms and Abbreviations

ANOVA *ANalysis Of VAriance*

AWGN *Additive White Gaussian Noise*

BSS *Blind Source Separation*

CRITIC *CRiteria Importance Through Intercriteria Correlation*

DM *Decision Maker*

ELECTRE *ELimination Et Choix Traduisant la REalité*

FITradeoff *Flexible and Interactive Tradeoff*

GDP *Gross Domestic Product*

GNI *Gross National Income*

HDMR *High-Dimensional Model Representation*

ICA *Independent Component Analysis*

JADE *Joint Approximate Diagonalization of Eigenmatrices*

LVA *Latent Variable Analysis*

MAUT *Multi-Attribute Utility Theory*

MAVT *Multi-Attribute Value Theory*

MCDA *MultiCriteria Decision Aiding*

MCDM *MultiCriteria Decision Making*

NIA *Negative Ideal Alternative*

OWA *Ordered Weighted Averaging*

PCA *Principal Component Analysis*

PIA *Positive Ideal Alternative*

SNR *Signal-to-Noise Ratio*

TOPSIS *Technique for Order Preference by Similarity to Ideal Solution*

WAM *Weighted Arithmetic Mean*

Contents

1 Introduction

Several practical situations that one may face in either personal life or industrial environments can be modeled as Multicriteria Decision Making (MCDM) problems (Figueira et al., 2016). A typical one consists in defining a ranking of a set of alternatives according to the consequences of choosing each one of them. These consequences are measured in terms of decision criteria, which, among other properties, are expected to be independent. However, most of the methods used to deal with MCDM problems do not take into account the data structural information and, therefore, redundancies among criteria are frequently neglected in the decision analysis. For example, by using the weighted arithmetic mean, if one does not take into account intercriteria relations, the ranking may be biased towards alternatives with good performances only in the dependent criteria. Therefore, one should be aware of the data structural information in order to adopt decisional parameters that can avoid unfair results. For instance, if we consider the application of the weighted arithmetic mean, one should penalize redundant criteria.

In the literature, one may find some methods that were conceived to deal with dependent criteria in MCDM problems. An example is the TOPSIS-M method (Vega et al., 2014), which is an extended version of the classical TOPSIS (Hwang and Yoon, 1981). TOPSIS-M extracts second-order statistics from the decision data and use this information to mitigate biased effects provided by redundant criteria. Although it has been applied in several situations, the use of second-order statistics to deal with dependencies in the decision data is questionable. Since the considered measure leads to a set of uncorrelated data, one may not ensure that we deal with statistical dependence. Recall that the concept of statistical dependence is stronger than correlation (Papoulis and Pillai, 2002). Therefore, an approach that is based on higher-order statistics could be used to extract relevant information that leads to independent data.

Moreover, other well-known methods that can be cited are the ones based on capacity coefficients, such as the Choquet integral and the multilinear model (Grabisch, 2016). They can deal with correlated data by modeling interactions among criteria. In the case in which the criteria are positively correlated, one may model a redundant effect between them. On the other hand, if the criteria are negatively correlated, one may consider a complementary effect. However, in order to use these aggregation functions, one needs to define a set of parameters whose number increases exponentially with the number of criteria. Therefore, one frequently adopts an automatic approach to define (or estimate) these parameters. In the literature, several works exploit the capacity identification problem with respect to the Choquet integral (Grabisch et al., 2008). However, few researches have been done in the context of the multilinear model.

Besides the redundancy among criteria, we also may be concerned with group MCDM problems in which some decision makers have correlated opinions. In such scenarios, for example, they may combine their evaluations in order to favor specific alternatives. Since we consider that this should be avoided, it would be interesting to model interactions between decision makers or penalize redundant opinions.

Based on the aforementioned motivations, the main goal of this Ph.D. consists to deal with redundant criteria in MCDM problems. For this purpose, we further investigate the application of capacity-based aggregation functions in such situations. Since the literature on the use of multilinear model in MCDM problems is less abundant in comparison with the Choquet integral, we focus on the former aggregation function in some of our contributions. Moreover, we also consider the application of latent variable analysis (LVA) methods in order to extract relevant information from the decision data. This information will be used to define the decisional parameters that could lead to fairer rankings of alternatives.

We would like to highlight that we tackled decision problems in which the set of criteria as well as how the information will be aggregated have already been defined. However, we proposed adjustments on the aggregation procedure (by improving existing methods or defining the set of parameters), based on either the preference provided by the decision maker or the relevant information extracted from the decision data, in order to overcome biased results. Therefore, we may say that our proposals cannot be seen as decision aiding methods, since we do not consider that the decision maker helped to build the whole model rationality and construction. For such a reason, some proposed approaches which are based on sophisticated statistical analysis may be difficult to be interpreted by the decision makers.

1.1 Objective and thesis organization

In the sequel, we present the thesis organization and our expected objectives:

- In Chapter 2, we present the theoretical foundation of multicriteria decision making. As a central point in our analysis, we discuss the case of redundant criteria. Moreover, a focus will be devoted to the MCDM methods, especially the ones that will be considered in our contributions.

- Chapter 3 describes the latent variable analysis techniques that will be used to compose our proposals. We address three techniques, namely principal component analysis (PCA) (Jolliffe, 2002), independent component analysis (ICA) (Hyvärinen et al., 2001) and sensitivity analysis (Saltelli et al., 2008).

- Our contributions begin in Chapter 4. In that chapter, we present a theoretical discussion on TOPSIS-M method. Moreover, we propose an ICA-based approach to deal with dependencies in the decision data. For this purpose, we combine an ICA technique to existing TOPSIS methods.

- As a second contribution, presented in Chapter 5, we also exploit ICA techniques to extract the information that will be used to mitigate biased effects provided by correlated criteria. However, in this chapter, we propose an approach that allows us to adjust the weights used in the weighted arithmetic mean by taking into account intercriteria relations. Moreover, other than ICA, we also verify the application of principal component analysis.

- In Chapter 6, we conduct both theoretical and experimental analysis on the use of the multilinear model in MCDM problems. We formulate the 2-additive multilinear model expression and address the problem of capacity identification in a supervised fashion with regularization.

- Similarly as in Chapter 6, Chapter 7 also tackles the problem of capacity identification. However, we consider the unsupervised case, in which we only have the decision matrix as learning data. Therefore, by assuming a characteristic on the decision data that we aim at dealing with, such as correlations among criteria, we propose approaches which can be used to estimate capacity coefficients, leading to a fairer ranking of alternatives.

- In our last contribution, presented in Chapter 8, we conduct preliminary experiments on the application of the Choquet integral in group MCDM problems. We revisit an interesting result in the literature and provide a discussion on the commutativity property in a 2-step Choquet integral.

- Finally, in Chapter 9 we conclude this Ph.D. thesis and present our future perspectives.

1.2 Publications

In this section, we describe the works originated from this Ph.D. thesis that were published in scientific journals and conferences.

Articles in Journals:

1. Pelegrina, G. D., Duarte, L. T., and Romano, J. M. T. Application of independent component analysis and TOPSIS to deal with dependent criteria in multicriteria decision problems. *Expert Systems with Applications*, 122:262–280, 2019.

2. Pelegrina, G. D., Duarte, L. T., Grabisch, M., and Romano, J. M. T. The multilinear model in multicriteria decision making: The case of 2-additive capacities and contributions to parameter identification. *European Journal of Operational Research*, 282:945–956, 2020.

Conference Papers:

1. Pelegrina, G. D., Duarte, L. T., and Romano, J. M. T. Muticriteria decision making based on independent component analysis: A preliminary investigation considering the TOPSIS approach. In Deville Y., Gannot S., Mason R., Plumbley M., and Ward D., editors, *Latent Variable Analysis and Signal Separation (LVA/ICA 2018). Lecture Notes in Computer Science*, volume 10891, pages 568–577. Springer, Cham, 2018.

2. Pelegrina, G. D., Duarte, L. T., and Romano, J. M. T. Análise de componentes independentes para ajuste não supervisionado dos pesos em decisão multicritério. In *LI Simpósio Brasileiro de Pesquisa Operacional (SBPO 2019)*. Galoá, Limeira, Brazil, 2019. Available at: https://proceedings.science/sbpo-2019/papers/analise-de-componentes-independentes-para-ajuste-nao-supervisionado-dos-pesos-em-decisao-multicriterio.

3. Pelegrina, G. D., Duarte, L. T., Grabisch, M., and Romano, J. M. T. Multilinear model: New issues in capacity identification. In *From Multiple Criteria Decision Aid to Preference Learning (DA2PL'2018) Workshop*. Poznan, Poland, 2018. Available at: http://da2pl.cs.put.poznan.pl/programme/detailed-programme/da2pl2018-abstract-10.pdf.

4. Pelegrina, G. D., Duarte, L. T., Grabisch, M., and Romano, J. M. T. An unsupervised capacity identification approach based on Sobol' indices. In Torra V., Narukawa Y., Nin J., Agell N., editors, *Modeling Decisions for Artificial Intelligence (MDAI 2020). Lecture Notes in Computer Science*, volume 12256, pages 66–77. Springer, Cham, 2020.

Part I

Foundations

2 Multicriteria decision making

Suppose we are planning to visit Paris during our next vacation and we need to book a hotel. There are several options in Paris, so it will be interesting to take into account some characteristics associated with the hotels in order to choose a proper one according to our preference. For example, we may look for a cheap hotel, with a comfortable room, in a safe neighborhood and with a good location (next to the downtown, for example). This is a typical multicriteria decision making (MCDM) problem, in which we need to make a decision (which hotel to choose?) based on a set of consequences associated with our choice. Such kind of situations are also faced in either private or public organizations. For instance, aiming at improving the public transportation between a suburb and an employment zone[1], policy makers may evaluate the construction of a highway linking these two areas. In this context, several issues are involved, such as the number and location of the stations, the investment and operating costs, station's accessibility, well-being of the users and environmental impacts.

A characteristic frequently observed in multicriteria decision making problems is that the consequences are conflicting. Moreover, one may have situations in which the consequences are not known with certainty. Therefore, in order to help the decision maker (DM) to handle these complex situations, it is convenient to adopt an appropriate method that can, rationally, lead to a satisfactory recommendation.

In this chapter, we present the theoretical foundation of multicriteria decision making, especially the concepts and techniques that will be exploited in our experiments. In Section 2.1, we start with the basic definitions about decision aiding and the different problems that one may face. Section 2.2 discusses what is a multicriteria decision making problem. We also define the ingredients involved in such problems and what one should expect when defining the set of criteria and the associated evaluations. For instance, an important aspect is the redundancy among them, which will be addressed in most of our contributions. Thereafter, in Section 2.3, we present the basic notions on preference modeling and how the decision maker opinions can be translated into a value function. In Section 2.4, we describe some MCDM methods. A special attention will be devoted to the aggregation functions, mainly the ones based on capacities. Section 2.5 addresses the problem of determining the parameters used in the MCDM methods. By this task, known as elicitation process, we present some different approaches that can be applied depending on the information provided by the decision maker. Finally, in Section 2.6, we outline our final remarks about decision aiding and the MCDM problems that will be tackled in the contributions of this work.

[1] This example was borrowed from (Roy, 1996).

2.1 Decision aiding: Concepts and definitions

In this section, we discuss some basic concepts involved in decision making. We start with the definition of decision and what is the goal of decision aiding. We also describe the decision aiding approaches and the steps that are followed until achieving a recommendation. Finally, we present the different classes of problems involving multiple criteria and highlight which one will be addressed in this study.

2.1.1 Decision, decision aiding and decision making

A first question that arises when thinking about decision making is what do we understand by decision? According to Roy and Bouyssou (1993), "*the decision is often presented as the fact of an isolated individual (the decision maker) freely exercising a choice between several possibilities of actions at a given moment in time*". In our daily life, we make several decisions, whether consciously or unconsciously, which may be simple (e.g., choosing a pizza among the options in the menu) or complex (e.g., buying a new car). Complex decisions are frequently faced by organizations and the adopted ones should take into account the impacts on people lives. For example, in order to decide the location of a new hospital or where to build a nuclear power plant will affect positively or negatively the local people. Moreover, a university that must choose a set of students to participate in an exchange program may impact the future of the selected candidates.

Therefore, the consequences of the chosen option must be considered when making a decision in a complex situation. In some cases, one does not know with certitude what are the consequences, but we assume probability distributions to be known on the set of consequences. Moreover, the consequences are often conflicting, and one may not have an option that is the best one for all of them. Besides, one also may have other concerns when the decision problem involves several stakeholders, who interact between them during the decision process. In such situations, even if a particular one (the decision maker) is considered in the evolution of the decision process, the stakeholders may have different interests and conflicting objectives[2]. Aiming at helping the DM (or the set of stakeholders) to answer his/her concerns and providing a recommendation to the addressed decision problem, an activity, called decision aiding, is conducted by an analyst. A definition stated by Roy (1996) is the following:

Definition 2.1. ***(Decision aiding):*** *is the activity of the person who, through the use of explicit but not necessarily completely formalized models, helps obtain elements of responses to the questions posed by a stakeholder of a decision process. These elements work towards clarifying the decision and usually towards recommending, or simply favoring, a*

[2] Most of this work considers a single decision maker. However, we will also tackle group decision problems and, therefore, more than one (with different opinions) will be involved in the decision process.

behavior that will increase the consistency between the evolution of the process and this stakeholder's objectives and value system.

Other than the decision maker, one may note in Definition 2.1 the presence of another actor, the analyst. As further discussed by Roy (1996) and Bouyssou et al. (2006), although the DM is the actor that has the knowledge of the situation and will make the final decision, he/she may not have the background to conduct the decision aiding process. Therefore, the presence of the analyst (an expert on operations research, for example), which has the methodological knowledge, is fundamental to translate the concerns and knowledge of the DM into a rational meaningful model of recommendation.

It is important to highlight that, in the literature, one may find distinctions between decision aiding and decision making. According to Bouyssou et al. (2006), decision aiding implies the presence of at least the aforementioned two actors (decision maker and analyst). On the other hand, decision making is referred to a situation in which the DM adopts a decision theory tool in order to support his/her decision. Therefore, there is no distinction between analyst and decision maker, and the latter one is assumed to have the appropriate knowledge to deal with the addressed problem. In the contributions of this study, we propose some methods that can be used by the DM to deal with inconveniences in the decision data. Therefore, since we do not emphasize the presence of these two actors, for convenience, we will often adopt the term decision making.

2.1.2 Decision aiding approaches

This section presents an overview of the different decision aiding approaches, namely normative, descriptive, prescriptive and constructive. Our description was based on the works of Bouyssou et al. (2006) and Tsoukiàs (2007, 2008), who provide further discussion on these terminologies.

Normative approach

The normative approach assumes a rationality stated by norms. Hypotheses and axioms, which are independent of the DM and the addressed problem (they intended to be universal) are then postulated as necessary for rational behavior. Therefore, it is an approach that brings the rationality from "outside" the situation. A typical example is the utility theory used in economics. Classical references are the works of von Neumann and Morgenstern (1944), Savage (1954) and Fishburn (1970).

Descriptive approach

Different from the normative approach, the descriptive one derives the rationality based on empirical behavior, i.e., by observing the behavior of DMs facing decision problems. Therefore, once learned by observation, the model can be applied on similar situations. Interested readers may refer to Kahneman and Tversky (1979) and Poulton (1994) for more details.

Prescriptive approach

In both normative and descriptive approaches, the model of rationality is imposed to the decision maker. Contrastedly, in the prescriptive approach, the DM provides information related to his/her preference and the analyst attempts to fit the model the best way possible at the moment. Therefore, the model is not general, but adjusted according to the DM preference. One may say that the rationality is discovered "within" the decision problem. A further discussion on this approach is presented by Tversky (1977) and Keeney (1992).

Constructive approach

The constructive approach is similar to the prescriptive one in the sense that the DM is solicited to derive the rationality of the model. However, it does not only answer questions about his/her preference, but is also helped by the analyst to build the rationality of the model. It is assumed that the construction of the model cannot be performed only by the analyst (the decision maker should also be involved). Moreover, it is expected to reach a consensus in the final model, which satisfies the perceptions of the decision maker and the knowledge of the analyst in terms of achieving a rational meaningful model. For more discussions about this approach, see the works of Roy (1996) and Dias and Tsoukiàs (2003).

According to Bouyssou et al. (2006), analysts do not follow, in practice, any aforementioned approach as a guideline. In fact, they usually consider interactions between the approaches. Furthermore, the authors also indicate that there are several multiple criteria decision support methods that follow a prescriptive approach and multiple criteria decision analysis that explicitly refers to a constructive approach.

2.1.3 The decision aiding process

We provide, in this section, a general description of the decision aiding process based on a constructive approach. We attach to the steps presented by Bouyssou et al. (2006) and Tsoukiàs (2008), who consider the presence of at least two actors (i.e., the decision maker and the analyst), the concerns of the DM, the methodological knowledge of the analyst and a converging objective. The steps are described in the sequel.

1. **Problem representation**: In this step, decision maker and analyst interact in order to clarify the situation and provide a representation of the problem. Therefore, the understanding of what is the problem, who has this problem and who takes the consequences of the decision is crucial in this step. With this description in mind,

both actors have a better understanding of their position and role within the decision process.

2. **Problem formulation**: Based on the problem representation, the analyst will translate the concerns of the DM into formal problems. Therefore, in this step, the points of view from which the potential actions (alternatives) can be analyzed are taken into account to associate the addressed decision problem with one of the classes that will be discussed in Section 2.1.4. Once the problem is formulated, it can be tackled by using an existing (or adapted) decision aiding method. Another aspect in this step is that some conclusions about the decision process can be anticipated. Therefore, the DM may validate or revise the formulation according to his/her concerns.

3. **Evaluation model**: In this step, given the problem formulation, the analyst provides a construction of an evaluation model that will be used to obtain a formal answer for the decision problem. In order to do so, the alternatives must be described in terms of a set of dimensions (attributes, with associated scales) and measured according to a set of criteria, the latter one reflecting the DM preference[3]. The aim is, therefore, to synthesize the evaluations into a global relation on the set of alternatives, which is used to provide the final recommendation. Several algorithms, methods or aggregation operators can be used to derive this global relation. As it will be noted by the reader, most of this work will focus on this step.

4. **Final recommendation**: Before recommending the solution of the decision problem, it is important to perform a sensibility analysis of the decision model. In that respect, we verify the robustness of the evaluation model when different scenarios and/or different parameters values are considered. Moreover, since the output of the evaluation model may be incomprehensible by the decision maker, the analyst must also translate the conclusions of the decision process into a format that can be interpreted and be convincing by the DM and also for other possible actors.

2.1.4 Classes of problems in decision aiding

In the problem formulation step described in the previous section, we mentioned that the decision situation is associated with a class of problem in decision aiding[4]. The classification depends on the answer that the DM wants to obtain. Three typical classes are considered (Bana e Costa, 1996; Roy, 1996), namely *choice*, *sorting* and *rank-*

[3] A formal discussion about what we call alternatives, attributes and criteria will be conducted in Section 2.2.3.

[4] The original name, in French, is *problématique*. However, in order to avoid English grammar mistakes, we will simply refer to this term as "class of problem".

ing. The difference between these approaches, given a set of potential alternatives, is presented as follows:

- **Choice problem**: the aim in the choice problem is to choose one action (the "best" one) among all possible alternatives[5]. Optimization methods are generally used to tackle this situation. An example is the problem of selecting one neighborhood (among all possible alternatives in the city) to build a new hospital.
- **Sorting problem**: the sorting problem consists in allocating the alternatives into different predefined categories. Therefore, each alternative will be associated with exactly one category, whose definition must not be in conflict with other categories. As an example we may consider the problem of sorting a set of research projects according to its impact degree on the environment (e.g., low, medium or high impact).
- **Ranking problem**: in the ranking problem, the aim is to rank the set of possible actions from the best to the worst one. Therefore, one may think in a simple competition between the alternatives in which they will be ordered according to transitive binary relations that indicate the preference of the DM. However, since one often deal with conflicting criteria, determining the ranking becomes a complex task. Moreover, depending on the preference structure provided by the DM, one may achieve a ranking either complete or partial (we will clarify this point in Section 2.3.1). As an example we may consider the problem of ranking a set of students according to their academic performances.

The focus of this work is to deal with ranking problems in scenarios with a set of consequences describing the alternatives. In the next section, we provide detailed discussions on the fundamental aspects related to such problems.

2.2 General aspects of multicriteria decision making

This section presents basic concepts in multicriteria decision making. Firstly, we define what are the addressed multicriteria decision problems and discuss two different terminologies, which are devoted to the American and the French schools on this subject. Then we provide some formal definitions to the basic ingredients involved in multicriteria decision problems, such as the actors, alternatives and criteria. With respect to the latter, we also present some desirable properties.

[5] In this class of problem, instead of choosing the best one, some situations require to select a subset of the possible actions. For example, the human resources department needs to hire a subset (greater than one) of the candidates that have applied for a job.

2.2.1 What is multicriteria decision making problems?

Let us recall the hotel booking problem illustrated in the beginning of Section 2. We may note that some of the intended desires may not be simultaneously achieved. Certainly, the cheapest hotel in Paris will not be the one with the better location, for example. In this case, if we choose the cheapest one, we probably will be far from the downtown. Otherwise, if we opt for the better one in terms of location, we probably will pay more for this benefit. Therefore, a common characteristic in multicriteria decision making problems is that the objectives are conflicting, and dealing with these trade-offs is a central point. We provide a general definition of such problems as follows (adapted from (Pomerol and Barba-Romero, 2000)):

Definition 2.2. ***(Multicriteria decision making problem):*** *is the situation in which the decision maker needs to make a choice among several alternatives (or ranking them), the set of these alternatives forming what is called the choice set. To make his choice from this set, the decision maker adopts several points of view, often contradictory, which we call criteria. These criteria are at least partially contradictory in that, if the decision maker adopts one of the points of view, he will not choose the same alternative as he would from the standpoint of another criterion.*

The points of view mentioned in Definition 2.2 (also called criteria, which will be formalized in Section 2.2.3) are associated with the consequences of choosing each alternative. How much we know about the consequences of each alternative defines if the addressed decision making problem is under certainty, risk or uncertainty. The knowledge about what kind of decision problem we are dealing with is important to define which method will be used (different problems are tackled by different methods). In a certainty scenario, we assume that the DM have full information on the alternatives, i.e., the consequences are known precisely. Conversely, decision under risk and uncertainty deals with situations in which the consequences of his/her possible actions are uncertain. In the case of decision under risk, we assume that probabilistic information on the consequences is available. However, in decision under uncertainty, we suppose that the DM has its own personal probability measure on the consequences. In this study, we address decision problems whose mathematical models assume a scenario under certainty. For more details and discussion about the other scenarios, the interested reader may see (Keeney and Raiffa, 1976; Gilboa, 2010; Wakker, 2010; Takemura, 2014).

It is important to mention that, in this study, we focus on discrete decision making, i.e., the case in which the number of alternatives is finite. Therefore, the continuous case (with infinite alternatives), will not be addressed here.

2.2.2 MCDM and MCDA

In the literature, we are frequently faced with two different definitions: multicriteria decision making (MCDM) and multicriteria decision aiding (MCDA). Their differences are discussed in several classical works (Roy, 1990a; Lootsma, 1990; Munda, 1993; Roy and Vanderpooten, 1996), so we try to highlight some of them in this section.

MCDM is emblematic of the American school on decision making and follows a prescriptive approach (see Section 2.1.2). It is based on the assumption that a utility function, which is used to obtain the optimal action, already exists, and that the role of the analyst is to discover this entity and prescribe the solution to the decision maker. In that respect, it is assumed that the information provided by the DM (preference, for example) is complete and sufficient for the analyst to verify the conditions (axioms, theorems or normative hypotheses) that guarantee the existence of such an entity. Once it is defined, the DM should agree with the obtained solution. In this framework, one may cite the famous textbook of Keeney and Raiffa (1976), in which the authors discuss the classical concepts of Multi-Attribute Utility Theory (MAUT) and the additive utility/value functions.

On the other hand, MCDA comes from the European (mainly French) school on decision aiding and is based on a constructive approach. It assumes that an ideal entity does not completely pre-exist and that the role of the analyst is to construct it together with the decision maker during the decision aiding process. The analyst guides the understanding about what is important in the provided information and help the DM to analyze and, maybe, reshape his preference in order to achieve a solution in accordance with their goals. Therefore, the interaction between these two actors in the decision aiding process is of extreme importance. As a representative example of a method developed by the French school, one may cite the series of the ÉLECTRE (*ELimination Et Choix Traduisant la REalité*) methods developed by Roy (1968, 1990b), who is the father of the French school.

In this work, we do not deal with situations in which there are interactions between analysts and decision makers. The information used in the considered methods is assumed to be given by the DM in a specific moment and that they do not evolve along with the decision process. Therefore, given that our contributions can be associated with a prescriptive approach, we often prefer to use MCDM instead of MCDA.

2.2.3 Basic elements in MCDM

Although we have already introduced some ingredients that are involved in MCDM problems, we provide, in this section, a formalization of them. We also indicate how these elements will be represented along this work. They are described as follows:

Actors: decision maker and analyst

As mentioned in Section 2.1.1, one may identify two actors involved in the decision problem (at least in the prescriptive and constructive approaches): the decision maker and the analyst. The decision maker is the one in charge of making a decision in a given situation. Therefore, it is the person who has the knowledge of the situation but may not have the background to lead the decision process and solve the problem. He/she is the one that has more concerns about the consequences of choosing an action, since the responsibility of the decision is on him/her. For this reason, he/she may need the help of a second actor to solve the problem. This actor, called analyst, is an expert who has the technical knowledge on decision theory and will conduct the decision process in order to solve the problem according to the DM goals. An important remark is that the analyst must be neutral in terms of the preference provided by the DM. His/her role is to adjust the model used to take the preference into account, ensure that the associated axioms are satisfied and provide a recommendation. In situations with more than one decision maker, the analyst may also work as a facilitator in the group decision, providing an aggregation of the individual preferences. In most of this work we consider a single DM, represented by d. However, in some situations, we will consider that more than one decision maker is involved in the addressed decision problem (see Section 2.1.4). In this context, we will refer to the decision maker k as d_k and the set of t DMs as $\mathcal{D} = \{d_1, d_2, \ldots, d_t\}$.

Actions, objects of interest or alternatives

Actions, objects of interest or alternatives are possible solutions for the decision problem. In order to define them, one should ensure that all possible actions are considered and that the consequences by choosing each one are different. Depending on the situation, the aim of the DM is to choose the best one, to rank them or to sort the alternatives into predefined categories (see Section 2.1.4 for more details about the different classes of decision making problems). Locations, candidates and research projects are examples of alternatives. The set of n possible alternatives will be represented by $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$.

Objectives and criteria

Consider the problem of deciding the location where to build a new hospital. The purpose of constructing this new hospital is generally associated with a main objective, in this case, to provide a better medical care service to the people. This objective may be subdivided into lower-level objectives. For example, in order to provide a better medical care service, we should maximize the number of patients that can be accommodated in a day, increase the speed of medical cares in emergency and maximize the accessibility of the population. If it is the case, this lower-level objectives can also be subdivided into other lower-level objectives and this procedure may be repeated several times until we achieve what we call lowest-level objectives, which will be measured in terms of attributes (Keeney and Raiffa, 1976) or criteria[6]. For instance, objectives like maximizing the speed of medical cares in emergency and minimizing the distance to get in the hospital may be measured in terms of the average waiting time in emergency and the distance from the new hospital location and a central point in the urban zone, respectively.

Therefore, in the multicriteria decision problem, the consequences of the alternatives are associated with the (generally conflicting) criteria describing them. In this work, the set of m criteria ($m \geq 2$) will be represented by $\mathcal{C} = \{c_1, c_2, \ldots, c_m\}$. For the sake of convenience, we may also refer to the index set of criteria $C = \{1, 2, \ldots, m\}$.

Decision matrix

As mentioned in the previous paragraph, each alternative a_j is described according to a set of m criteria. Let us assume that the vector that describes the alternative a_j is represented by $\mathbf{x}_j = (x_{j,1}, \ldots, x_{j,m})$, where $x_{j,i} = x_i\,(a_j)$ is the consequence of alternative a_j with respect to criterion c_i (hereafter, we will also use $\mathbf{x}_j$ to refer to the alternative a_j). For all $j = 1, \ldots, n$, $\mathbf{x}_j \in X \subseteq X_1 \times \ldots \times X_m$, where X_i contains all possible consequences that can be achieved by any alternative with respect to the criterion c_i. It is important to mention that, although it is easy to think that all X_i contain numerical elements (average waiting time or number of patients that can be accommodated, for example), they may also be qualitative, e.g., degrees of satisfaction of local people (partly satisfied, satisfied, more than satisfied, etc.). Therefore, in order to translate $x_{j,i}$ into a real number, we consider a (marginal) value function[7] $u_i(x_{j,i})$ ($x_{j,i} \in X_i$), which represents the evaluation of the alternative $\mathbf{x}_j$ with respect to the criterion c_i. $u_i(x_{j,i})$ may also be interpreted as the DM satisfaction on the alternative a_j regarding the associated consequence $x_{j,i}$. From now on, for simplicity of notation, we will often refer to $u_i(x_{j,i})$ as $u_{j,i}$. In MCDM problems, we generally represent the decision data in a matrix $\mathbf{M} = (u_{j,i})_{n \times m}$, defined by

$$\mathbf{M} = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_m \end{array} \\ \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,m} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n,1} & u_{n,2} & \cdots & u_{n,m} \end{bmatrix} \end{array}. \tag{2.1}$$

It is worth mentioning that most MCDM problems consider that the decision matrix comprises a set of evaluations collected from a dataset (grades of a set of students, for example). However, situations in group decision may consider that each $\mathbf{M}_k$ is provided by the decision maker d_k. This case is further discussed in Chapter 8.

[6] It is worth mentioning that, in this study, we use "attribute" and "criteria" as synonymous. However, in the literature one may find other works that make a distinction between them. For further details, interested readers may refer to (Roy, 1996).

[7] Further discussions about the value functions are presented in Sections 2.3.2 and 2.4.1. Moreover, since we here tackle decision problems with certainty, we prefer to use value function instead of utility function, the latter being commonly used in situations under uncertainty and risk.

2.2.4 Desirable properties on the set of criteria and the structural dependence

In order to define the set of criteria, some aspects should be taken into account. For instance, we expect that the following properties are ensured (Keeney and Raiffa, 1976):

- **Comprehensive**: since the DM will take the decision based on the consequences of each alternative, it is important to him/her to have a clear understanding about what each criterion represents, its implications on the selected alternative (or alternatives) and how much the objective is achieved by considering the associated values.
- **Measurable**: in order to associate a value for each criterion and be able to assess the DM's preference, they must be measurable.
- **Completeness**: the set of criteria is said to be complete if they are sufficient to represent all possible concerns with respect to the main objective.
- **Minimum size**: dealing with a large number of criteria may be difficult for the DM's understanding. Therefore, it is important to consider a set of criteria as small as possible but, surely, guaranteeing the completeness of this set.
- **Non-redundancy**: when defining the set of criteria, one should avoid double-counting of consequences. If this property is not ensured, a redundancy may be introduced in the decision data.

The last property (non-redundancy) is a central concern that we deal with in our contributions and will be further discussed in this study. In (Keeney and Raiffa, 1976), the authors provide an example that illustrates such a situation. If, for instance, we are analyzing some portfolios of investments in companies A and B, we may consider the "income from company A" and the "income from both companies" as two criteria. However, it is clear that the income from both companies consists in the sum of the income from company A and the income from company B. Therefore, we are double-counting the income from company A, which introduces redundancy in the decision data.

Besides the situation in which we are double-counting consequences (in the sense that we add twice the same variable), we may also consider that redundancies come from the existence of (latent) factors that are associated with two or more criteria. Under the name of structural or statistical dependence, some works in the literature have already provided a discussion on this subject (Roy, 1983, 1996, 2009; Roy and Bouyssou, 1993). An example presented by Roy (1983) illustrates this scenario. Consider that we are evaluating possible locations for the construction of a highway linking two areas. Moreover, assume that, among the set of criteria, we have

- $x_1(a)$: difference between the cost of location a and a reference location a_0;
- $x_2(a)$: average time saved for the user that travels through the location a instead of a reference location a_0.

One may note that both criteria are associated with the length of the highway. Generally, the longer the highway, the higher will be the cost and the slower will be the travel. Therefore, one may assume that the criteria are correlated. Moreover, note that, in this case, the relation between the length of the highway and both $x_1(a)$ and $x_2(a)$ is not a sum of variables. In fact, the length can be seen as a latent factor associated with both criteria.

As argued by Roy (1983) (see, also, (Roy and Bouyssou, 1993; Roy, 1996) for further details), the statistical dependence in such a scenario can be analyzed in two different ways:

- If the set of criteria have already been defined, the decision maker may simply remove one of the two criteria or substitute both of them by a third one, e.g., the difference between the length of location a and a reference location a_0 ($x_3(a)$).
- If one considers a constructive approach in which the actors involved in the decision problem have worked together to define the set of criteria, one may accept the statistical dependence without performing any changes on the set of criteria. This can be justified, for instance, if the criteria represent different points of view that deserve to be taken into account. Moreover, they may also represent different interests, since $x_1(a)$ and $x_2(a)$ are more related to investors and users concerns, respectively.

Remark that, since criteria $x_1(a)$ and $x_2(a)$ are not exactly the same, by considering the first approach one possibly loses some information. On the other hand, the second approach preserves the information in the decision data. However, we neglect the statistical dependence, which can be acceptable (or not) in some situations (Roy, 2009).

In this thesis, we also consider that latent factors may be associated with more than one criterion and, as a consequence, one may have redundancies in the decision data. However, we attempt to deal with such a redundancy by means of something between the aforementioned approaches. In other words, instead of neglecting the statistical dependence or simply removing a criterion (or substituting by another one), our aim is to adjust the mechanism used to aggregate the collected information in such a way that the bias introduced by dependent criteria is softened. We consider, therefore, situations in which neither the loss of information nor the acceptation that the statistical dependence does not deserve to be taken into account are convenient for the addressed decision problem.

A typical situation frequently mentioned in the literature (Grabisch, 1996; Marichal, 2000) in which redundant criteria may bias the derived ranking is the problem of ranking students according to their grades in a set of disciplines (see Example 2.3 presented in Section 2.4.1.2.1). If we consider, for example, grades in calculus, physics and literature, we possibly have a redundancy between the two first, since they are associated with a performance on scientific skills. It is not interesting to remove the grades of either calculus or physics, since we lose a performance measure. However, accepting the redundancy may lead to a ranking biased towards the students that have good grades only in both calculus and physics. Therefore, we consider that a mechanism that aggregates the collected information and softens the inconvenience introduced by dependent criteria is important in order to provide fairer ranking of alternatives.

2.3 Preference modeling

In the ranking problem described in Section 2.1.4, we mentioned that it can be solved by ordering the alternatives based on the preference provided by the decision maker. In this section, we formalize the concept of preference structures and discuss how the DMs may express their opinion (in terms of preference) about the alternatives. We start by describing the notion of binary relation and some basic properties, which are used to define preference structures and construct orders. The knowledge about the preference structure that we handle is an important aspect, since it may admit a numerical representation. We also address this topic in this section. It is important to mention that most of the definitions presented here were based on the works of Roubens and Vincke (1985) and Moretti et al. (2016).

2.3.1 Binary relations and preference structures

Since the decision maker provides a comparison between pairs of alternatives, it is natural to bring the concept of binary relations into the decision problem. A binary relation is defined as follows:

Definition 2.3. ***(Binary relation):*** *Let W be a finite set of elements $a, b, \ldots$. A binary relation R on the set W is a subset of the Cartesian product $W \times W$, that is, a set of ordered pairs (a, b) such that a and b are in W : $R \subseteq W \times W$.*

We may denote by aRb the fact that $(a, b) \in R$. Moreover, consider the complement $\overline{R}$, defined by $a\overline{R}b$ iff (if and only if) $not(aRb)$. The binary relation R is called

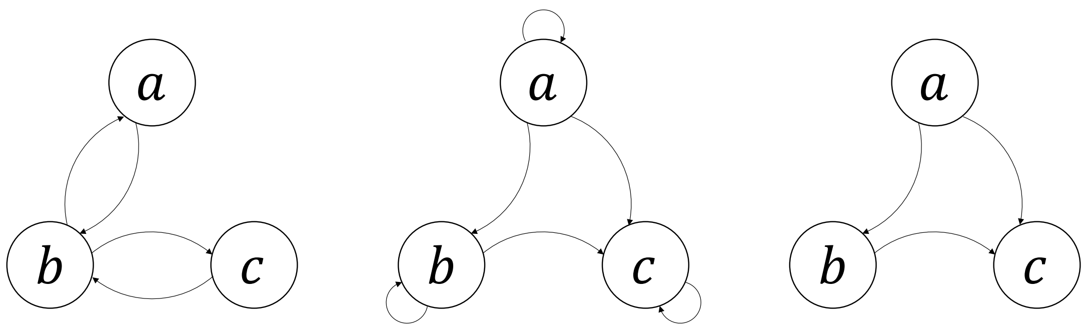
- reflexive if aRa, $\forall a \in W$;
- irreflexive if $a\overline{R}a$, $\forall a \in W$;

- symmetric if $aRb \Rightarrow bRa$, $\forall a, b \in W$;
- antisymmetric if (aRb and bRa) $\Rightarrow a = b$, $\forall a, b \in W$;
- asymmetric if $aRb \Rightarrow b\overline{R}a$, $\forall a, b \in W$;
- transitive if (aRb and bRc) $\Rightarrow aRc$, $\forall a, b, c \in W$;
- complete if aRb or bRa, $\forall a \neq b \in W$;
- strongly complete if aRb or bRa, $\forall a, b \in W$ (note that this implies reflexivity).

Moreover, some operations on two binary relations R and R' on the same set W are the following:

- Inclusion: $R \subseteq R'$ if $aRb \Rightarrow aR'b$;
- Union: $a(R \cup R')b$ iff aRb or (inclusive) $aR'b$;
- Intersection: $a(R \cap R')b$ iff aRb and $aR'b$;
- Relative product: $a(R.R')b$ iff $\exists\, c \in W \;:\; aRc$ and $cR'b$.

Binary relations are commonly represented (and easily interpreted) through graphs, where W and R are the set of nodes and arcs, respectively. In that respect, an arc from a to b exists iff aRb. Figure 1 illustrates three binary relations (R_1, R_2, R_3) on the set of elements (a, b, c). One may note that R_1 is irreflexive and symmetric, R_2 is asymmetric, transitive and strongly complete (therefore, reflexive) and R_3 is asymmetric, transitive and complete.



(a) Binary relation R_1. (b) Binary relation R_2. (c) Binary relation R_3.

Figure 1 – Examples of binary relations.

In decision problems, binary relations serve as the basis for the definition of preference relations. As pointed out by Moretti et al. (2016), the most traditional preference relations provided by the decision maker over a set of alternatives consist in the following:

- Strict preference (P): the decision maker may strictly prefer a to b. We represent this binary relation by aPb or $a \succ b$;
- Indifference (I): the decision maker may be indifferent between a and b. In this case, it has no preference between this pair of alternatives. We represent this binary relation by aIb or $a \sim b$.

Therefore, in the case where the DM provides all his/her preference relations between all pairs of alternatives, one achieves what is called *preference structure*. Formally, it is defined as follows:

Definition 2.4. ***(Preference structure):*** *A preference structure is a collection of binary relations defined on a set W and such that, for each pair of alternatives $a, b \in W$, one and only one relation is satisfied.*

Based on both P and I relations, one may define the well-known $\langle P, I\rangle$ structure:

Definition 2.5. ***($\langle P, I\rangle$ structure):*** *A $\langle P, I\rangle$ structure on the set W is a pair $\langle P, I\rangle$ of binary relations on W such that:*

- *P is asymmetric;*
- *I is reflexive and symmetric.*

The $\langle P, I\rangle$ structure is important in ranking problems since it can be associated with the construction of different *orders*. Depending on the order that one has by taking the preference structure provided by the DM, one knows if it is possible to rank all the alternatives. Moreover, one also knows if an *ex æquo*[8] is allowed. Among the most important orders in decision making, the weak order (or complete preorder) and total order are directly associated with the $\langle P, I\rangle$ structure. They are defined as follows (adapted from (Roubens and Vincke, 1985) and (Moretti et al., 2016)):

Definition 2.6. ***(Total order):*** *$\langle P, I\rangle$ is a total order structure iff*

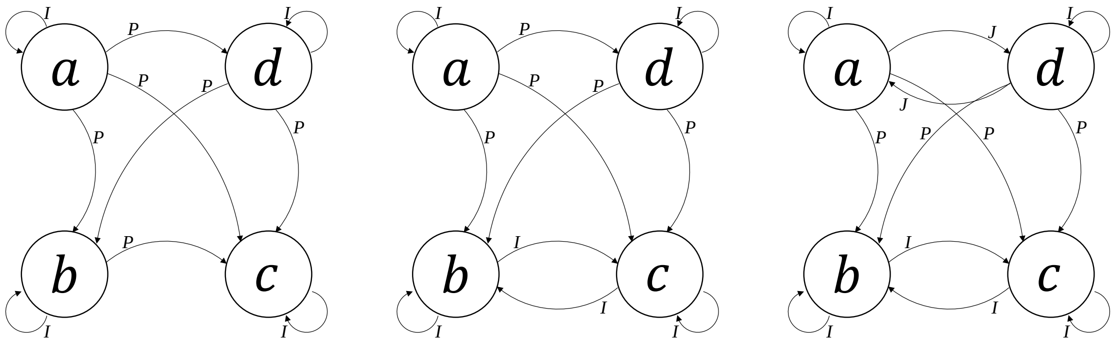
$$\begin{cases} I = \{(a,a), \forall a \in W\} \\ P \text{ is transitive} \\ P \cup I \text{ is reflexive and complete} \end{cases}.$$

[8] In preference structures, the existence of an *ex æquo* means that there are different alternatives that cannot be distinguished in terms of preference. In other words, among these alternatives, one cannot attest that one alternative is preferable over another one.

Definition 2.7. ***(Weak order):*** *$\langle P, I\rangle$ is a weak order structure iff*

$$\begin{cases} I \text{ is transitive} \\ P \text{ is transitive} \\ P \cup I \text{ is reflexive and complete} \end{cases}.$$

One may note that the main difference between the total order and the weak order structures is that, by assuming the former, the indifference is not allowed for different alternatives. Therefore, the total order structure leads to a ranking of alternatives without any *ex æquo*. Figures 2a and 2b illustrate a total order and a weak order structures, respectively, based on a set of $n = 4$ alternatives. One may note that, by taking the total order structure, the ranking (from the best to the worst) is given by $a \succ d \succ b \succ c$. However, by taking the weak order structure, the obtained ranking is $a \succ d \succ b \sim c$, which indicates that an *ex æquo* is allowed in this case.



(a) Total order. (b) Weak order. (c) Partial preorder.

Figure 2 – Particular cases of preference structures.

Besides P and I, there is another binary relation that may be considered by the decision maker, called incomparability (J). Therefore, other than expressing preference and indifference over pair of elements, in the $\langle P, I, J\rangle$ structure, the DM may also consider that two alternatives are incomparable. We represent this binary relation by aJb or $a?b$.

At first sight, incomparability may not be an ideal relation provided by the DM, since we normally expect to have comparisons between all pairs of alternatives in order to obtain the ranking. However, although some methods neglect this relation, other ones can handle incomparability when providing a recommendation (see, e.g., Section 2.4.2). With respect to the construction of orders, by including incomparability on the $\langle P, I\rangle$ leads to a partial preorder (or quasi-order), defined as follows (also adapted from (Roubens and Vincke, 1985) and (Moretti et al., 2016)):

Definition 2.8. ***(Partial preorder):*** *$\langle P, I, J \rangle$ is a partial preorder structure iff*

$$\begin{cases} I \text{ is reflexive, symmetric and transitive} \\ P \text{ is asymmetric and transitive} \\ J \text{ is irreflexive and symmetric} \\ (P.I \cup I.P) \subset P \end{cases}.$$

Figure 2c illustrates a partial preorder structure. In this case, one may note that $a \succ b \sim c$ and $d \succ b \sim c$. For more details on both $\langle P, I \rangle$ and $\langle P, I, J \rangle$ structures, see (Roubens and Vincke, 1985; Roy, 1996; Moretti et al., 2016).

2.3.2 Modeling preference through a value function

The contributions of this study assume that the preference structure provided by the decision maker is a weak order. Therefore, we only consider the $\langle P, I \rangle$ structure, which excludes incomparability. Moreover, since in the literature of multicriteria decision making it is common to use $\succsim$ to indicate a $\langle P, I \rangle$ structure, from now on, we will adopt the former notation. For instance, $\succ$ and $\sim$ denote the asymmetric (strict preference) and symmetric (indifference) parts of $\succsim$, respectively. $a \succsim b$ indicates that a is at least as good as b, and is also called an outranking relation (Roy, 1996).

An interesting and relevant aspect concerning preference structures is under what conditions we can use a value function $U(\cdot)$ to represent them. By representing each element through a numerical value, the ranking is easily obtained by ordering the alternatives from the highest to the lowest value. For instance, if we consider the binary relation $\succsim$, we have the following result[9] (adapted from (Moretti et al., 2016)):

Theorem 2.1. ***(Value function for $\langle P, I \rangle$ structure and weak order):*** *Let $\succsim$ be a binary relation on a finite set W. The following definitions are equivalent:*

(i) $\succsim$ is a weak order;

(ii) $\exists U : \succsim \mapsto \mathbb{R}^+$ *satisfying* $\forall a, b \in W : \begin{cases} a \succ b \text{ iff } U(a) > U(b) \\ a \sim b \text{ iff } U(a) = U(b) \end{cases}$;

(iii) $\exists U : \succsim \mapsto \mathbb{R}^+$ *satisfying* $\forall a, b \in W : a \succsim b \text{ iff } U(a) \geq U(b)$.

Note that the function $U(\cdot)$ is not unique. For instance, all strictly increasing transformation of $U(\cdot)$ still give a representation function. In this study, one of the main interests lies on the task of building an overall value function that fits the preference relations provided by the decision maker in a scenario with multiple criteria. We address this topic with more details in Section 2.5.1.2.

[9] For more details and a complete proof, see (Krantz et al., 1971).

2.4 Multicriteria decision making methods

Section 2.3 discussed the most typical forms of how a DM can express his preference and how it can be translated into a numerical representation. However, we did not distinguish the task of ranking alternatives based on a preference structure when several points of view must be taken into account. In fact, comparing alternatives and ordering them is easier when they are evaluated in terms of a single criterion. When we consider a set of criteria, which are generally conflicting, we need to formalize how we aggregate the preference provided by the decision maker for each point of view (or numerical representations of them) in order to achieve the ranking.

In this section, we present some of the most common MCDM methods (Roy and Bouyssou, 1993; Triantaphyllou, 2000; Greco et al., 2016), which can be divided in two families, each one associated with a different school on decision making (see Section 2.2.2). The first family is related to the American school and comprises the methods based on multi-attribute value theory (MAVT) (Keeney and Raiffa, 1976). By these methods, we make explicit use of the numerical representation of preference in order to aggregate the criteria evaluations (known with certainty) and obtain an overall evaluation for each alternative. Therefore, the ranking is achieved based on the comparison of these unique values representing each alternative. This family also includes the multi-attribute utility theory (MAUT) methods, which differ from MAVT since they consider uncertainty when measuring the consequences of the alternatives. With respect to the second family, associated with the European school on decision making (Roy and Vanderpooten, 1996), it encompasses methods based on outranking relations (Roy, 1990b). Differently from MAVT, these methods are based on a relational preference system, in which the preference between alternatives is defined under the light of each criterion individually. Thereafter, the "global" preference about pairs of alternatives are achieved based on concepts of concordance and discordance, which comprise arguments that attest or refuse the preference of an alternative over another one.

The focus of this study is on the first family. Therefore, we address the methods based on a unique synthesizing criterion with more details in the next section. For instance, we present some aggregation functions, such as the additive ones and the functions based on capacities, and another well-known method in MCDM, called TOPSIS (Technique for Order Preference by Similarity to Ideal Solution). Moreover, in Section 2.4.2, we also provide an overview of outranking methods, which include the well-known method called ELECTRE (ELimination Et Choix Traduisant la REalité).

2.4.1 Methods based on a unique synthesizing criterion

The most traditional methods in MCDM are the ones based on a synthesizing criterion (Roy, 2016). There are several reasons for their popularity. For instance, since people are used to compare objects through scores associated with each of them (students evaluated in terms of grades in a set of disciplines or the performance of soccer players measured according to the number of scored goals or assists), these methods are easy to be accepted, interpreted and manipulated by DMs. Moreover, being supported by axioms that take into account a rational behavior of DMs, the application of these approaches is justified. Therefore, given the relevance of such methods, some of them will be detailed in this section.

As stated by Theorem (2.1), in a weak order structure (restricted to $\langle \succ, \sim \rangle$, since $? = \emptyset$),

$$\mathbf{x} \succsim \mathbf{x}' \Leftrightarrow U(\mathbf{x}) \geq U(\mathbf{x}'),\ \forall \mathbf{x}, \mathbf{x}' \in X, \tag{2.2}$$

where $U(\cdot)$ is the (overall) value function. Since $U(\mathbf{x})$ depends on the set of evaluations with respect to each criterion, a first task is to build the marginal value functions $u_i(x_i)$, for all $i = 1, \ldots, m$. Based on these values (which we assume to be provided by the DM), in this thesis, we consider the monotone decomposable model of Krantz (Krantz et al., 1971; Blackorby et al., 1978) in which $U(\cdot)$ is defined by

$$U(\mathbf{x}) = F\left(u(\mathbf{x})\right),\ \forall \mathbf{x} \in X. \tag{2.3}$$

where $u(\mathbf{x}) = (u_1(x_1), \ldots, u_m(x_m))$ and $F(\cdot)$ is an aggregation function, which is nondecreasing in its arguments. Therefore, a higher satisfaction with respect to the criterion value x_i indicates a higher marginal value u_i, which leads to a higher positive impact on $U(\mathbf{x})$. For more details about this model, see the works of Greco et al. (2004) and Bouyssou and Pirlot (2004).

Although Theorem (2.1) guarantees the existence of a numerical representation of the preference, it does not indicate how to define $U(\cdot)$, which, in this case, is dependent on the form of $F(\cdot)$. Generally, the aggregation function is defined based on the hypotheses about the decision problem that we are dealing with and some characteristics that we would like to model (e.g., interactions among criteria). We address this subject in the sequel.

2.4.1.1 Additive aggregation functions and preferential dependence

One of the most traditional and largely used additive aggregation function is the weighted arithmetic mean (WAM), defined by

$$F_{WAM}(u(\mathbf{x})) = \sum_{i=1}^{m} w_i u_i, \tag{2.4}$$

where the weights w_i ($w_i \geq 0$, for all $i = 1, \ldots, m$, and $\sum_{i=1}^{m} w_i = 1$) represent the relative importance of each criterion in the decision problem. In situations in which the data are not on the same scale (and are not normalized), the sum presented in Equation (2.4) makes no sense. Therefore, the weights may also be used to compensate the different scales on which the criteria are represented.

Several reasons justify the popularity of WAM in multicriteria decision making. For instance, it is easy to be explained and used by the decision maker. He/she may clearly see the compromise between each criterion evaluation and its impact on the overall value. Moreover, since childhood, we are accustomed to use such function. For example, the final grade in a discipline is generally calculated based on a weighted sum of the performances that one achieves in a set of tests. However, although it is a simple function, the WAM has some limitations. These limitations are associated with some aspects that one assumes about the addressed decision problem:

- The criteria evaluations are linearly aggregated: since the WAM is an additive function that linearly aggregates the criteria evaluation, this characteristic must fit the DM opinions. Let us consider the following example:

Example 2.1. ***(Ranking hotels):*** *Suppose that we would like to book a hotel and we evaluate the set of options according to two criteria: price and location. The values representing these criteria are presented in Table 1. It is worth mentioning that, for both values, the higher the better (i.e., a higher value in price indicates a cheaper hotel).*

Table 1 – Criteria evaluations for Example 2.1.

Hotels	**Evaluations**	
	Price (u_1)	Location (u_2)
Hotel 1 (a_1)	8.0	4.0
Hotel 2 (a_2)	5.8	5.8
Hotel 3 (a_3)	4.0	8.0
Hotel 4 (a_4)	9.5	1.0
Hotel 5 (a_5)	2.0	2.5

If we are not willing to give up satisfaction in both price and location, maybe the best hotel should be the one with "more balanced" values in both criteria. Therefore, in the considered example, we should prefer Hotel 2 instead of the other ones. However, it is impossible to find a set of weights $\mathbf{w} = (w_1, w_2)$*, with* $w_1 + w_2 = 1$*, that leads to an overall value for Hotel 2 greater than the ones for both Hotels 1 and 3. Mathematically speaking,* $\nexists\, (w_1, w_2), w_1 + w_2 = 1$*, such that* $5.8w_1 + 5.8w_2 \geq 8.0w_1 +$

$4.0w_2$ *and* $5.8w_1 + 5.8w_2 \geq 4.0w_1 + 8.0w_2$. *In fact,* $5.8w_1 + 5.8w_2 \geq 8.0w_1 + 4.0w_2 \rightarrow 2.2w_1 \leq 1.8w_2 \rightarrow w_1 \leq 0.45$ *and* $5.8w_1 + 5.8w_2 \geq 4.0w_1 + 8.0w_2 \rightarrow 1.8w_1 \geq 2.2w_2 \rightarrow w_1 \geq 0.55$*, which is impossible.*

The aforementioned example shows a typical scenario in multiobjective optimization, illustrated in Figure 3. The dashed lines linking a_1 to a_3 and a_4 (the Pareto optimal solutions) indicate the convex hull. Since Hotel 2 lies in the interior of this convex region, the maximization of WAM can never attain this point. Therefore, Hotel 2 could not be the first one in the ranking by using the WAM.

- The aggregation is totally compensatory: the compensatory effect inherent to the WAM means that a bad performance on a single criterion may be compensated by a good one on another criterion. This effect may be measured in terms of the substitution rates $v_{i,i'} = w_{i'}/w_i$, which indicates the increase (resp., decrease) in the evaluation $u_i(\mathbf{x})$ that compensates a decrease (resp., an increase) of one unity in $u_{i'}(\mathbf{x})$. Figure 3 also illustrates this effect in Example 2.1, for two different weighting vectors. For instance, if $w_1 = w_2 = 1/2$ and, therefore, $v_{1,2} = 1$, a_1 and a_3 achieve the same overall value, since a reduction of 4 in the location value of Hotel 1 is compensated by increasing $4v_{1,2} = 4$ in price. By considering $w_1 = 2/3$ and $w_2 = 1/3$, $F_{WAM}(8, 4) = F_{WAM}(9.5, 1)$, i.e., the increase of $3v_{1,2} = 1.5$ in the price value of Hotel 4 compensates the decrease of 3 in location.

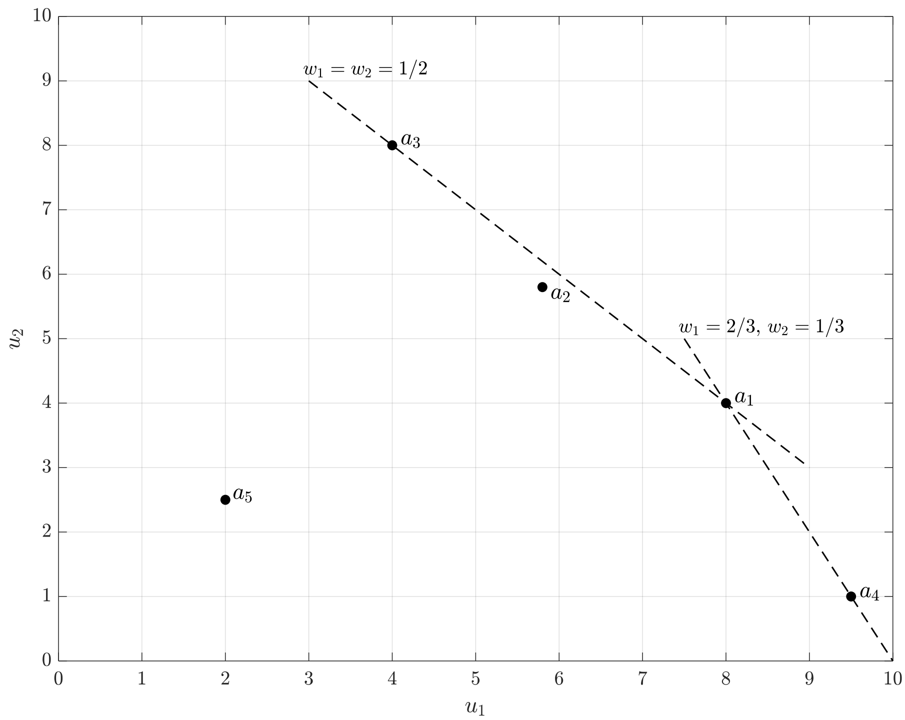


Figure 3 – Compensatory effect of Example 2.1.

- One assumes that the preference between any pair of alternatives with respect to a subset $A \subseteq C$ does not depend on the evaluations on its complement $\bar{A}$: this aspect, called mutual preferential independence and which will be further explained in this section, is illustrated in the following example:

  Example 2.2. ***(Where to build a new hospital?):*** *Let us recall the problem mentioned in Section 2.2.3, which is related to choosing the location where a new hospital will be built. Table 2 presents the possible locations and the criteria evaluations (all to be maximized), namely average waiting time in emergency, number of patients that can be accommodated in a day and the distance from the new hospital location and a central point in the urban zone.*

Table 2 – Criteria evaluations for Example 2.2.

Locations	**Evaluations**		
	Waiting time (u_1)	Patients that can be accommodated (u_2)	Distance (u_3)
Location 1 (a_1)	8.0	7.0	4.0
Location 2 (a_2)	8.0	4.0	7.0
Location 3 (a_3)	4.0	7.0	4.0
Location 4 (a_4)	4.0	4.0	7.0

Suppose that, for the person who is in charge of taking the decision, if an alternative has a bad evaluation in terms of the waiting time, he/she will prefer a better location instead of a higher number of patients that can be accommodated. The justification is that, since the patients will wait a long time in emergency, it is convenient that they could arrive faster in the hospital. Therefore, in this case, $a_4 \succ a_3$. However, if an alternative has a very satisfactory waiting time, i.e., the patients will not take a long time in emergency, the DM will prefer a higher number of patients that can be accommodated instead of a better location. Even if the patients take a longer time to arrive in the hospital, they will be accommodated quickly. In this case, $a_1 \succ a_2$.

One may note that the preferences between the number of attended patients and the distance from the new hospital location and a central point in the urban zone depends on the value achieved on the waiting time. In such situations, the WAM is not able to model the preferences provided by the decision maker. Indeed, $4w_1 + 4w_2 + 7w_3 > 4w_1 + 7w_2 + 4w_3 \rightarrow w_3 > w_2$ and $8w_1 + 7w_2 + 4w_3 > 8w_1 + 4w_2 + 7w_3 \rightarrow w_2 > w_3$, which is impossible.

The last aspect is associated with an important result in the literature. Before presenting it, let us define some concepts about preferential independence:

Definition 2.9. ***(Preferential independence):*** *Let C be the set of all criteria. $A \subseteq C$ is preferentially independent of its complement $\overline{A}$ if, for every $\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}' \in X$,*

$$(\mathbf{x}_A, \mathbf{y}_{\overline{A}}) \succsim (\mathbf{x}'_A, \mathbf{y}_{\overline{A}}) \Leftrightarrow (\mathbf{x}_A, \mathbf{y}'_{\overline{A}}) \succsim (\mathbf{x}'_A, \mathbf{y}'_{\overline{A}}), \tag{2.5}$$

where $(\mathbf{x}_A, \mathbf{y}_{\overline{A}})$ is the compound alternative taking value x_i if $i \in A$ and value y_i otherwise (and similarly for the other alternatives).

Definition 2.10. ***(Mutual preferential independence):*** *Let C be the set of all criteria. The attributes $X_1, \ldots, X_m$ are mutually preferentially independent if every $A \subseteq C$ is preferentially independent of its complement.*

In the sequel, we present the condition for the existence of an additive value function when there are at least three criteria[10] (adapted from (Bouyssou and Pirlot, 2016)):

Theorem 2.2. ***(Additive value function when $m \geq 3$):*** *Let $\succsim$ be a binary relation on a set $X_1 \times X_2 \times \ldots \times X_m$ with $m \geq 3$. If restricted solvability holds on all attributes and at least three attributes are essential then $\succsim$ has a representation*

$$\mathbf{x} \succsim \mathbf{x}' \Leftrightarrow \sum_{i=1}^{m} u_i(x_i) \geq \sum_{i=1}^{m} u_i(x'_i), \tag{2.6}$$

for all $\mathbf{x} = (x_1, x_2, \ldots, x_m)$, $\mathbf{x}' = (x'_1, x'_2, \ldots, x'_m) \in X$, if and only if $\succsim$ is a weak order satisfying the Archimedean condition and the attributes are mutually preferentially independent.

One may note that preferential independence does not hold in Example 2.2. As a consequence, the WAM will not be able to model all preference relations provided by the decision maker. In such a situation, as well as in the scenario described in Example 2.1, we say that there exist interaction among criteria. Therefore, in these cases, we need to adopt another aggregation function in order to be able to consider the DM opinion over the alternatives. We address this subject in the next section.

2.4.1.2 Choquet integral and multilinear model

This section addresses the construction of two non-additive aggregation functions, namely Choquet integral (Choquet, 1954) and multilinear model (Owen, 1972). Since these functions can model interactions among criteria, they will be used in some contributions of this work that deal with redundancies in the decision problem. In order to represent these interactions, instead of weights $w_1, \ldots, w_m$ on individual criteria (the case of WAM), both Choquet integral and multilinear model are based on a

[10] For more details about Theorem theo:exaddvf such as the definition of restricted solvability, essentiality and Archimedean condition, the interested readers may refer to (Krantz et al., 1971; Bouyssou and Pirlot, 2016).

set of parameters (regarded as generalizations of the weighting vectors) associated with all possible coalitions of criteria, called capacity (Choquet, 1954). A capacity[11] $\mu = [\mu(\emptyset), \mu(\{1\}), \ldots, \mu(\{m\}), \mu(\{1,2\}), \ldots, \mu(\{m-1,m\}), \ldots, \mu(C)]$ (also known as fuzzy measure (Sugeno, 1974)) defined on a set $C = \{1, \ldots, m\}$ of m criteria is a set function $\mu : 2^C \to \mathbb{R}$, satisfying the following axioms:

- $\mu(\emptyset) = 0$ and $\mu(C) = 1$ (normalization),
- if $A \subseteq B \subseteq C$, $\mu(A) \leq \mu(B) \leq \mu(C)$ (monotonicity).

Before defining the Choquet integral and the multilinear model, we start by constructing these functions through an interpolation problem (Grabisch, 2016). For instance, consider that $\mathbb{O}$ and $\mathbb{I}$ represent attribute values giving lower and upper level of satisfaction, respectively. Moreover, suppose that, for all $i = 1, \ldots, m$, $u_i(\mathbb{O}_i) = 0$ and $u_i(\mathbb{I}_i) = 1$. Aiming at determining value functions using the MACBETH method[12] (Bana e Costa and Vansnick, 1997; Bana e Costa et al., 2012), in the context of the weighted arithmetic mean, we have that

$$U(\mathbb{I}_i, \mathbb{O}_{-i}) = F(u_i(\mathbb{I}_i), u_{-i}(\mathbb{O}_{-i})) = F(1_i, \mathbf{0}_{-i}) = w_i, \ \ \forall i \in C, \tag{2.7}$$

where $u_{-i}(\mathbb{O}_{-i})$ indicates that $u_{i'}(\mathbb{O}_{i'}) = 0$ for all $i' \neq i$.

Since our interest is to build aggregation functions whose parameters are associated with all possible coalitions of criteria, we may extend (2.7) to any $A \subseteq C$. Let us fix $U(\mathbb{O}_C) = 0$ and $U(\mathbb{I}_C) = 1$. For any coalition of criteria, we have the following:

$$U(\mathbb{I}_A, \mathbb{O}_{-A}) = F(u_A(\mathbb{I}_A), u_{-A}(\mathbb{O}_{-A})) = F(\mathbf{1}_A, \mathbf{0}_{-A}), \ \ \forall A \subseteq C, \tag{2.8}$$

where $-A$ represents the complement set of A.

Given that $F(\cdot)$ is nondecreasing in its arguments, for any $A \subseteq B$, we have that $U(\mathbb{I}_A, \mathbb{O}_{-A}) \leq U(\mathbb{I}_B, \mathbb{O}_{-B})$. Therefore, by performing the change of notation $U(\mathbb{I}_A, \mathbb{O}_{-A}) = \mu(A)$, one may note that Equation (2.8) defines a capacity μ. Moreover, based on Equation (2.8), it is also possible to determine the aggregation function $F(\cdot)$ on all vertices of the hypercube $[0,1]^m$. For a value inside the convex closure defined by all these vertices, $F(\cdot)$ can be easily obtained through a linear interpolation.

Figure 4 illustrates the case with $m = 2$ criteria (remark that $u_2(x_2) > u_1(x_1)$). A possible interpolation is to use the three vertices (marked in red) in Figure 4a (recalling that $F(0,0) = 0$), which leads to

$$F(u(\mathbf{x})) = 0 + (\alpha_1 u_1(x_1) + \alpha_2 u_2(x_2))F(0,1) + (\beta_1 u_1(x_1) + \beta_2 u_2(x_2))F(1,1) \tag{2.9}$$

[11] In this study, we refer to the vector μ in a cardinal-lexicographic representation, i.e., a vector whose elements are sorted according to their cardinality and, for each cardinality, based on the lexicographic order.

[12] This thesis does not tackle the problem of constructing value functions. For further details on this subject by using the MACBETH method, see (Grabisch, 2016)

If we take $(u_1(x_1), u_2(x_2)) = (0, 1)$, $F(0,1) = \alpha_2 F(0,1) + \beta_2 F(1,1)$, which implies $\alpha_2 = 1$ and $\beta_2 = 0$. Moreover, for $(u_1(x_1), u_2(x_2)) = (1, 1)$, $F(1,1) = (\alpha_1 + \alpha_2)F(0,1) + (\beta_1 + \beta_2)F(1,1)$, which implies $\alpha_1 + \alpha_2 = 0$ and $\beta_1 + \beta_2 = 1$. Therefore, given the unique solution for these four constraints ($\alpha_1 = -1$, $\alpha_2 = 1$, $\beta_1 = 1$ and $\beta_2 = 0$), Equation (2.9) leads to

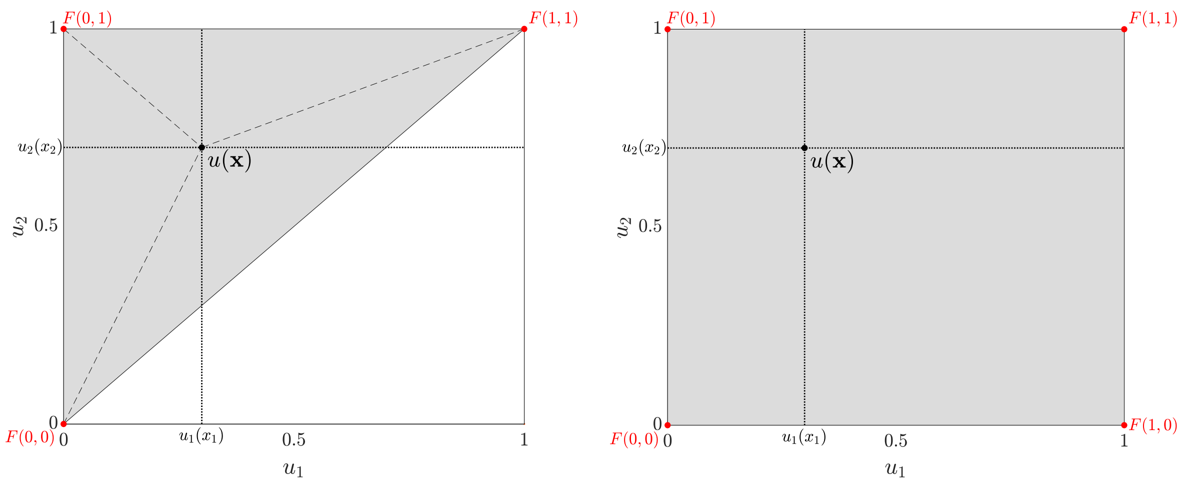
$$F(u(\mathbf{x})) = (u_2(x_2) - u_1(x_1))F(0,1) + u_1(x_1)F(1,1), \quad (0 \leq u_1(x_1) \leq u_2(x_2) \leq 1). \tag{2.10}$$

Since $F(0,1) = \mu(\{2\})$ and $F(1,1) = \mu(\{1,2\})$, Equation (2.10) leads to an aggregation function called (discrete) Choquet integral. The definition of the Choquet integral for m criteria is the following:

$$F_{CI}(u(\mathbf{x})) = \sum_{i=1}^{m} (u_{(i)}(x_{(i)}) - u_{(i-1)}(x_{(i-1)}))\mu(\{(i), \ldots, (m)\}), \tag{2.11}$$

where $u_{(i)}(x_{(i)})$ indicates a permutation of the indices i so that $0 \leq u_{(1)}(x_{(1)}) \leq \ldots \leq u_{(m)}(x_{(m)}) \leq 1$ (with $u_{(0)}(x_{(0)}) = 0$). By using the Choquet integral, it is interesting that (see (Grabisch, 2016) for further details)

- it is obtained through the parsimonious interpolation of $F(\cdot)$ over $(m+1)$ vertices of the hypercube $[0,1]^m$,
- it is equivalent to the Lovász extension of a capacity (Lovász, 1983; Marichal, 2002),
- it comprises a piecewise linear aggregation function, and
- (2.11) only requires the use of a subset of all capacity coefficients, which depends on the order of the evaluations $u_1(x_1), \ldots, u_m(x_m)$.



(a) Parsimonious interpolation. (b) Multilinear interpolation.

Figure 4 – Example of interpolation for $m = 2$ criteria (adapted from (Grabisch, 2016)).

Besides the parsimonious interpolation, another possibility is to use all vertices (marked in red in Figure 4b). In this case, we have the following:

$$F(u_1(x_1),0) = \frac{u_1(x_1)-0}{1-0}F(1,0) + \frac{1-u_1(x_1)}{1-0}F(0,0), \tag{2.12}$$

$$F(u_1(x_1),1) = \frac{u_1(x_1)-0}{1-0}F(1,1) + \frac{1-u_1(x_1)}{1-0}F(0,1), \tag{2.13}$$

$$F(u(\mathbf{x})) = \frac{u_2(x_2)-0}{1-0}F(u_1(x_1),1) + \frac{1-u_2(x_2)}{1-0}F(u_1(x_1),0). \tag{2.14}$$

By using (2.12) and (2.13) in (2.14) and recalling that $F(0,0)=0$, we achieve

$$F(u(\mathbf{x})) = u_1(x_1)(1-u_2(x_2))F(1,0) + u_2(x_2)(1-u_1(x_1))F(0,1) + u_1(x_1)u_2(x_2)F(1,1). \tag{2.15}$$

Since $F(1,0)=\mu(\{1\})$, $F(0,1)=\mu(\{2\})$ and $F(1,1)=\mu(\{1,2\})$, Equation (2.15) leads to an aggregation function called multilinear model. Its definition for any number of criteria is given by

$$F_{ML}(u(\mathbf{x})) = \sum_{A\subseteq C} \mu(A) \prod_{i\in A} u_i(x_i) \prod_{i\in \overline{A}} \left(1-u_i(x_i)\right), \tag{2.16}$$

where $u_i(x_i) \in [0,1]$ and $\overline{A}$ is the complement set of A. Associated with the multilinear model, one has the following properties (see (Grabisch, 2016) for further details):

- it is obtained through the multilinear interpolation of $F(\cdot)$ over all vertices of the hypercube $[0,1]^m$,
- it is equivalent to the Owen extension of a capacity (Owen, 1972),
- it comprises a polynomial aggregation function and
- (2.16) requires the use of all capacity coefficients $\mu(A)$.

Another interesting aspect about the Choquet integral and the multilinear model (see (Grabisch et al., 2003) for further details) is that the former assumes commensurateness between the attributes. This can be easily verified since one needs to compare the evaluations and, therefore, they should lie on the same scale. In this case, it is required to define reference levels. However, commensurateness is not needed in the multilinear model and, therefore, there is no need of reference levels. In turn, the preference relation must satisfy weak difference independence (a condition similar to independence but on pairs of alternatives).

Aiming at verifying the application of both functions to overcome the limitations inherent to the WAM, consider Example 2.1. Assuming $\mu(\{1\})=\mu(\{2\})=0.2$, we have the overall values[13] presented in Table 3. Therefore, either the Choquet integral or

[13] In order to normalize the data and be able to apply the multilinear model, we divided all evaluations by 10.

the multilinear model can be used to represent the preference of Hotel 2 over the other ones.

Table 3 – Application of the Choquet integral and the multilinear model in Example 2.1.

Hotels	$F_{CI}(\cdot)$	$F_{ML}(\cdot)$
Hotel 1	0.4800	0.4320
Hotel 2	0.5800	0.4338
Hotel 3	0.4800	0.4320
Hotel 4	0.2700	0.2670
Hotel 5	0.2100	0.1200

2.4.1.2.1 Interaction indices

A relevant aspect in MCDM methods is how the parameters used to construct the considered aggregation functions can be interpreted. For instance, each weight w_i in the WAM represents the relative importance of criterion i in the decision problem. However, when considering the Choquet integral or the multilinear model, the capacity coefficients $\mu(A)$ have no clear interpretation. In order to overcome this inconvenience and to highlight the interactions that we have modeled, one can consider an alternative representation of a capacity, called interaction index. In the sequel, we present the interactions indices associated with the Choquet integral and the multilinear model. Both have their origins in cooperative game theory (Grabisch and Roubens, 1999).

Consider a singleton i and a subset $B \in C \backslash \{i\}$. The effect that i has in the coalition with B is given by $\mu(B \cup \{i\}) - \mu(B)$. In order to measure the marginal contribution of i alone in all possible coalitions of criteria in C, one may calculate the average of $\mu(B \cup \{i\}) - \mu(B)$ over all $B \in C \backslash \{i\}$. In the context of the Choquet integral, this marginal contribution is called Shapley (power) index (Shapley, 1953), a concept largely used in sharing worth in cooperative game theory. It is defined by

$$\phi_i^{\mathcal{S}} = \sum_{B \subseteq C \backslash i} \frac{(m - |B| - 1)! \, |B|!}{m!} \left[\mu(B \cup \{i\}) - \mu(B) \right], \tag{2.17}$$

where $|B|$ represents the cardinality of the subset B and $\phi_i^{\mathcal{S}} \in [0, 1]$. Therefore, one may note in Equation (2.17) that one averages $\mu(B \cup \{i\}) - \mu(B)$ by considering the cardinality of B. For instance, in a scenario with 3 criteria, the marginal contribution of criterion 1 is given by

$$\phi_1^{\mathcal{S}} = \frac{2}{6} \left[\mu(\{1\}) - \mu(\emptyset) \right] + \frac{1}{6} \left[\mu(\{1,2\}) - \mu(\{2\}) \right] + \frac{1}{6} \left[\mu(\{1,3\}) - \mu(\{3\}) \right] + \frac{2}{6} \left[\mu(C) - \mu(\{2,3\}) \right].$$

The same idea used to calculate the marginal contribution of criterion i can be extended to a pair of criteria i, i'. In this case, the effect that i, i' have in the coalition with B is measured by $\mu(B \cup \{i, i'\}) - \mu(B \cup \{i\}) - \mu(B \cup \{i'\}) + \mu(B)$. If one takes the average over all $B \in C \backslash \{i, i'\}$, one achieves

$$I^{\mathcal{S}}_{i,i'} = \sum_{B \subseteq C \backslash \{i,i'\}} \frac{(m - |B| - 2)!\,|B|!}{(m-1)!} \left[\mu(B \cup \{i, i'\}) - \mu(B \cup \{i\}) - \mu(B \cup \{i'\}) + \mu(B)\right], \tag{2.18}$$

where $I^{\mathcal{S}}_{i,i'} \in [-1, 1]$ and can be interpreted as the interaction degree of coalition of criteria i, i' by taking into account all possible coalitions of criteria in C. Furthermore, the sign of $I^{\mathcal{S}}_{i,i'}$ indicates the type of interaction that we have modeled and what this implicates. For instance, we have the following situations:

- If $I^{\mathcal{S}}_{i,i'} < 0$, we model a negative interaction (also called negative synergy, substitutive or redundant effect) between criteria i, i', which means that, in order to obtain a good overall value, it is sufficient to satisfy either criterion i or criterion i'. This is generally used when the contribution of the coalition of criteria i, i' in the overall evaluation is lower than the sum of their individual contributions. For instance, as will be discussed in some parts of this work, a negative interaction may be used to penalize criteria that are associated with the same latent factor (redundant criteria).

- If $I^{\mathcal{S}}_{i,i'} > 0$, we model a positive interaction (also called positive synergy or complementary effect) between criteria i, i', which means that, in order to obtain a good overall evaluation, the evaluations on both criteria i, i' must be satisfactory. This is generally used when the contribution of the coalition of criteria i, i' in the overall evaluation is greater than the sum of their individual contributions.

- If $I^{\mathcal{S}}_{i,i'} = 0$, no interaction is modeled between criteria i, i'. Therefore, we assume that they act independently and the contribution of this coalition in the overall evaluation is equal to the sum of their individual ones.

In order to illustrate the marginal contribution of a singleton and the interaction degree between a pair of criteria, consider the Example 2.2. Based on the capacity $\mu = [0, 0.2, 0.2, 1]$, one obtains $\phi^{\mathcal{S}}_1 = \phi^{\mathcal{S}}_2 = 0.5$ and $I^{\mathcal{S}}_{1,2} = 0.6$. Although both criteria have the same marginal contribution, they have a positive interaction, which means that higher overall values will be obtained by hotels with good evaluations on both price and location. As a consequence, Hotel 2 achieved the first position in the ranking.

Besides $\phi^{\mathcal{S}}_i$ and $I^{\mathcal{S}}_{i,i'}$, one may also define the interaction index for any $A \subseteq C$. In this case, the (generalized) interaction index[14] is defined by (Grabisch, 1997a)

$$I^{\mathcal{S}}(A) = \sum_{B \subseteq C \backslash A} \frac{(m - |B| - |A|)!\,|B|!}{(m - |A| + 1)!} \left(\sum_{B' \subseteq A} (-1)^{|A| - |B'|} \mu(B \cup B') \right). \tag{2.19}$$

[14] Remark that $I^{\mathcal{S}}(\{i\}) = \phi^{\mathcal{S}}_i$ and $I^{\mathcal{S}}(\{i, i'\}) = I^{\mathcal{S}}_{i,i'}$.

However, one does not have a clear interpretation as for in $\phi_i^{\mathcal{S}}$ and $I_{i,i'}^{\mathcal{S}}$.

It is important to remark that, given the interaction indices $I^{\mathcal{S}}(A)$, one may recover each capacity coefficient $\mu(A)$ through the linear transformation

$$\mu(A) = \sum_{B \subseteq C} \beta_{|A \cap B|}^{|B|} I^{\mathcal{S}}(B), \tag{2.20}$$

where $\beta_{|A \cap B|}^{|B|}$ is defined by

$$\beta_p^{p'} = \sum_{q=0}^{p} \binom{p}{q} \alpha_{p'-q}, \tag{2.21}$$

with

$$\alpha_p = - \sum_{p'=0}^{p-1} \frac{\alpha_{p'}}{p - p' + 1} \binom{p}{p'} \tag{2.22}$$

being the Bernoulli numbers and $\alpha_0 = 1$.

The aforementioned interpretations (marginal contribution of criterion i and interaction degree of coalition of criteria i, i') also hold for the parameters that compose the multilinear model. However, associated with this aggregation function, we have the Banzhaf interaction index (Roubens, 1996), which is based on the Banzhaf value (Banzhaf, 1965). By taking a singleton, one obtains the Banzhaf power index, defined as follows:

$$\phi_i^{\mathcal{B}} = \frac{1}{2^{|C|-1}} \sum_{B \subseteq C \backslash \{i\}} \left[\mu(B \cup \{i\}) - \mu(B) \right]. \tag{2.23}$$

Differently from Equation (2.17), one may note that Banzhaf power index takes the arithmetic mean over all $\mu(B \cup \{i\}) - \mu(B)$. If we consider a scenario with 3 criteria, the marginal contribution of criterion 1 by using the Banzhaf power index is given by

$$\phi_1^{\mathcal{B}} = \frac{1}{4} \left[\mu(\{1\}) - \mu(\emptyset) \right] + \frac{1}{4} \left[\mu(\{1,2\}) - \mu(\{2\}) \right] + \frac{1}{4} \left[\mu(\{1,3\}) - \mu(\{3\}) \right] + \frac{1}{4} \left[\mu(C) - \mu(\{2,3\}) \right].$$

In the case of a pair of criteria i, i', one obtains

$$I_{i,i'}^{\mathcal{B}} = \frac{1}{2^{|C|-2}} \sum_{B \subseteq C \backslash \{i,i'\}} \left[\mu(B \cup \{i,i'\}) - \mu(B \cup \{i\}) - \mu(B \cup \{i'\}) + \mu(B) \right]. \tag{2.24}$$

If one considers the capacity used in Example 2.2, one also achieves $\phi_1^{\mathcal{B}} = \phi_2^{\mathcal{B}} = 0.5$ and $I_{1,2}^{\mathcal{B}} = 0.6$. In fact, in the case of two criteria, Equations (2.18) and (2.24) are identical. By extending to any $A \subseteq C$, we have the Banzhaf interaction index

$$I^{\mathcal{B}}(A) = \frac{1}{2^{|C|-|A|}} \sum_{B \subseteq C \backslash A} \sum_{B' \subseteq A} (-1)^{|A|-|B'|} \mu(B' \cup B), \forall A \subseteq C. \tag{2.25}$$

Moreover, given the set of the Banzhaf interaction indices, the capacity $\mu(A)$ may be retrieved by the expression

$$\mu(A) = \sum_{B \subseteq C} \left(\frac{1}{2} \right)^{|B|} (-1)^{|B \backslash A|} I^{\mathcal{B}}(B), \forall A \subseteq C. \tag{2.26}$$

We could note that, by modeling the interactions among criteria, we can overcome some limitations inherent to the WAM. Indeed, Marichal (2000) discusses three situations in which modeling interactions can be useful. They are described as follows:

- **Preferential dependence**: as discussed in Section 2.4.1.1, this situation violates the condition described in Theorem 2.2 and, therefore, an additive function cannot be used to model the preference information. By modelling interactions and using either the Choquet integral or the multilinear model, one may consider a preference provided by the decision maker which is dependent on the value of a given criterion (or a set of criteria). For instance, if we tackle the problem described in Example 2.2 (dividing the data by 10 for normalization) with a capacity $\mu = [0, 0.40, 0.25, 0.50, 0.85, 0.55, 0.75, 1]$, we obtain the overall evaluations presented in Table 4.

Table 4 – Application of the Choquet integral and the multilinear model in Example 2.2.

Locations	$F_{CI}(\cdot)$	$F_{ML}(\cdot)$
Location 1	0.6950	0.6950
Location 2	0.6050	0.6380
Location 3	0.4750	0.5350
Location 4	0.5500	0.5440

  We note that both functions could model the preferential dependence, i.e., $a_1 \succ a_2$ and $a_4 \succ a_3$. It is also interesting to remark that, in order to model these preferences, we considered a positive interaction between criteria 1 and 2 (both waiting time and the attended patients must have good evaluations to achieve a higher overall value) and a negative interaction between criteria 1 and 3 (even with a bad waiting time, a good location is enough to achieve a good overall value). In a cardinal-lexicographic representation, $I^{\mathcal{S}} = [0.525, 0.325, 0.350, 0.325, 0.200, -0.350, 0, 0]$ and $I^{\mathcal{B}} = [0.538, 0.325, 0.350, 0.325, 0.200, -0.350, 0, 0]$.

- **Substitutiveness and complementarity**: in such situations, the decision maker, based on his/her satisfaction on the relative importance of criteria, may impose to the aggregation procedure a substitutive and/or complementary effect among criteria. The aforementioned example could be also classified into this situation, since it comprises a substitutive effect between waiting time and location and a complementary effect between waiting time and attended patients.

- **Correlation**: another common situation comprises the presence of correlation among criteria. As commented in Section 2.2.4, this characteristic is undesirable, since it introduces redundancies in the set of criteria. However, this inconvenience may also be mitigated by considering an aggregation function that models interactions among criteria. Differently from the previous situation, in which the positive/negative interactions depend on the DM opinions, in this case, they can be modeled based on the information (correlation coefficient, for example) extracted from the decision data. In order to illustrate this case, consider the following example (adapted from (Grabisch, 1996)):

Example 2.3. ***(Ranking students):*** *Suppose that the person in charge with selecting candidates for an exchange program wants to rank a set of students based on their grades in calculus, physics and literature. Table 5 presents the performance of each student.*

Table 5 – Criteria evaluations for Example 2.3.

Students	**Grades**			**WAM**	
	Calculus (u_1)	Physics (u_2)	Literature (u_3)	$F_{WAM}(\cdot)$	Pos.
Student 1 (a_1)	18	16	10	15.6	1
Student 2 (a_2)	10	12	18	12.4	3
Student 3 (a_3)	15	15	16	15.2	2

Consider that, for the person in charge with this task, scientific disciplines have more importance in the aggregation than literature. For instance, when applying the WAM, the decision maker defined $\mathbf{w} = [0.4, 0.4, 0.2]$*. The achieved overall values and ranking positions are also presented in Table 5. In this case, Student 1 achieved the first position in the ranking. However, both calculus and physics may be associated with the same latent factor, e.g., scientific skills. Therefore, this correlation between criteria should be considered when aggregating the evaluations in order to avoid biased results towards students with very good grades in these two correlated disciplines but a weak performance in literature.*

In situations as described in Example 2.3, biased rankings may be avoided by penalizing redundant criteria and/or modeling a complementary effect between criteria. For example, by assuming $\mu = [0, 0.4, 0.4, 0.2, 0.6, 0.7, 0.7, 1]$, which leads to $I^{\mathcal{S}} = I^{\mathcal{B}} = [0.50, 0.35, 0.35, 0.30, -0.20, 0.10, 0.10, 0]$, we obtain the overall values presented in Table 6. One may note that Student 3, which is the one with good performances in all disciplines, achieves the first position. It is worth mentioning that, in order to perform the normalization, we divided the data by 20.

Table 6 – Application of the Choquet integral and the multilinear model in Example 2.3.

Students	$F_{CI}(\cdot)$	$F_{ML}(\cdot)$
Student 1	0.7200	0.7210
Student 2	0.6300	0.6590
Student 3	0.7600	0.7675

2.4.1.2.2 The 2-additive case

Another important concept with respect to the Choquet integral and the multilinear model is the case of a 2-additive capacity[15], defined as follows (Grabisch, 1997b):

Definition 2.11. ***(2-additive capacity):*** *A capacity μ is 2-additive if $I^{\mathcal{S}}(A) = 0$ (or $I^{\mathcal{B}}(A) = 0$) for all $A \subseteq C$, with $|A| > 2$.*

This kind of capacity brings relevant benefits to deal with the MCDM problem (Angilella et al., 2015). One of them is that it considerably reduces the computational complexity of defining the unknown parameters used in both Choquet integral and the multilinear model. Indeed, without any assumptions about the capacity, we need to define $2^m - 2$ coefficients (since two of them are found given the axioms of a capacity). However, if the capacity is 2-additive, this number is reduced to $m(m+1)/2 - 1$, which facilitates their estimation (or identification).

By assuming a 2-additive capacity, the Choquet integral takes a particular form, defined as follows (Grabisch, 1997b):

$$F_{CI}(u(\mathbf{x})) = \sum_i u_i \left(\phi_i^{\mathcal{S}} - \frac{1}{2} \sum_{i'} \left| I_{i,i'}^{\mathcal{S}} \right| \right) + \sum_{I_{i,i'}^{\mathcal{S}} < 0} (u_i \vee u_{i'}) \left| I_{i,i'}^{\mathcal{S}} \right| + \sum_{I_{i,i'}^{\mathcal{S}} > 0} (u_i \wedge u_{i'}) I_{i,i'}^{\mathcal{S}}, \quad (2.27)$$

where $\vee$ and $\wedge$ represent the maximum and the minimum operators, respectively. Moreover, the axioms of a capacity may also be rewritten in terms of the interaction indices:

- $I^{\mathcal{S}}(\emptyset) - \frac{1}{2} \sum_i \phi_i^{\mathcal{S}} + \frac{1}{6} \sum_{i,i'} I_{i,i'}^{\mathcal{S}} = 0$,
- $\sum_i \phi_i^{\mathcal{S}} = 1$,
- $\phi_i^{\mathcal{S}} - \frac{1}{2} \sum_{i' \neq i} \left| I_{i,i'}^{\mathcal{S}} \right| \geq 0, \ \forall i \in C$.

Although the 2-additive Choquet integral has been well studied in the literature, such a type of capacity has not been used in the context of the multilinear model.

[15] It is important to mention that one may define any k-order additive capacity (Grabisch, 1997b). In particular, by assuming a 1-additive capacity (or, simply, an additive capacity), which is the situation in which the criteria act independently, both Choquet integral and multilinear model turn to the weighted arithmetic mean. In other words, they generalize the WAM.

One of the contributions of this work (see Chapter 6) comprises a study of the 2-additive multilinear model.

2.4.1.3 TOPSIS

Another MCDM method that is based on a unique synthesizing criterion is the TOPSIS method, developed by Hwang and Yoon (1981). Although TOPSIS has some drawbacks[16] (Martel and Roy, 2006), it has been applied in several practical problems in the literature (Behzadian et al., 2012; Zavadskas et al., 2016; Yoon and Kim, 2017). For instance, Kaya and Kahraman (2011), Cables et al. (2012) and Kumar et al. (2013) tackled energy planning, factory maintenance and lean performance problems, respectively, by using fuzzy TOPSIS methods. Moreover, Bai et al. (2014) addressed pure-play e-commerce companies evaluation and Hu et al. (2016) deal with the identification of spreading ability of nodes based on modified versions of TOPSIS.

The main idea of TOPSIS is to assign higher (resp., lower) overall evaluations to alternatives closer (resp., farther) to the positive ideal alternative and farther (resp., closer) to the negative ideal one. Both ideal alternatives are synthetic ones, which are defined based on the collected criteria evaluations. The positive ideal alternative is the one composed by the highest values that one may achieve given all criteria evaluations. On the other hand, the negative ideal alternative is the one composed by the lowest values that one may find for each criterion in the decision data. Once these elements are defined, the overall value of each alternative (also called closeness measure) is calculated based on the Euclidean distances between the alternative and the positive and negative ideal alternatives, respectively.

Algorithm 2.1 describes the different steps of TOPSIS. We assume in this study that all criteria are to be maximized. In the case in which lower evaluations are better, the steps of Algorithm 2.1 must be adapted to take into account this characteristic (see, e.g., (Hwang and Yoon, 1981)).

As we can note in Step 5 of Algorithm 2.1, the closeness measure o_j lies in the range $[0, 1]$. Moreover, $o_j \approx 1$ (resp., $o_j \approx 0$) indicates that the alternative a_j is close to the positive (resp., negative) ideal alternative and far from the negative (resp., positive) one. Therefore, the ranking is obtained by ordering the alternatives according to their closeness measure in a descending order.

An interesting aspect that one may visualize in the literature is that some works address the problem of redundancies among criteria by using an extended version

[16] Since the inclusion (or exclusion) of an alternative interferes on the normalization step and, as a consequence, in the distance calculations and the overall values, the meaningfulness of TOPSIS is questionable. For more details and a possible modification in order to avoid this drawback, known as rank reversal, see (García-Cascales and Lamata, 2012).

Algorithm 2.1 TOPSIS

Input: Decision data $\mathbf{M}$ and weighting vector $\mathbf{w}$.

Output: Closeness measure $\mathbf{o} = [o_1, o_2, \ldots, o_n]$.

Step 1: Normalization. For each evaluation $u_{j,i}$, perform the following normalization:

$$\tilde{u}_{j,i} = \frac{u_{j,i}}{\sqrt{\sum_{j=1}^{n} u_{j,i}^2}}, \;\; i = 1, \ldots, m, \; j = 1, \ldots, n.$$

Step 2: Weighted normalization. For each normalized evaluation $\tilde{u}_{j,i}$ and based on the weighting vector $\mathbf{w}$, perform the following weighted normalization:

$$p_{j,i} = w_i \tilde{u}_{j,i}, \;\; i = 1, \ldots, m, \; j = 1, \ldots, n.$$

Step 3: Ideal alternatives determination. Calculate both positive and negative ideal alternatives (PIA and NIA, respectively) through the equation

$$PIA = \mathbf{p}^+ = \left\{p_1^+, p_2^+, \ldots, p_m^+\right\} \text{ and } NIA = \mathbf{p}^- = \left\{p_1^-, p_2^-, \ldots, p_m^-\right\},$$

where $p_i^+ = \max\{p_{j,i} | 1 \leq j \leq n\}$ and $p_i^- = \min\{p_{j,i} | 1 \leq j \leq n\}$, $i = 1, \ldots, m$.

Step 4: Euclidean distances. Calculate the Euclidean distances from each alternative a_j and both PIA and NIA:

$$D_j^+ = \sqrt{\left(\mathbf{p}_j - \mathbf{p}^+\right)^T \left(\mathbf{p}_j - \mathbf{p}^+\right)}, \;\; j = 1, \ldots, n,$$

and

$$D_j^- = \sqrt{\left(\mathbf{p}_j - \mathbf{p}^-\right)^T \left(\mathbf{p}_j - \mathbf{p}^-\right)}, \;\; j = 1, \ldots, n,$$

where $\mathbf{p}_j = [p_{j,1}, p_{j,2}, \ldots, p_{j,m}]$.

Step 5: Closeness measure. For each alternative a_j, calculate its closeness measure:

$$o_j = \frac{D_j^-}{D_j^+ + D_j^-}, \;\; j = 1, \ldots, n.$$

of TOPSIS, called TOPSIS-M (Vega et al., 2014). Differently from the original version of TOPSIS, TOPSIS-M calculates the distances between each alternative and the positive and negative ideal alternatives based on the Mahalanobis distance (Mahalanobis, 1936; De Maesschalck et al., 2000) (instead of the Euclidean one). Therefore, the main idea of TOPSIS-M is to use information about the covariance matrix of the data, extracted through the Mahalanobis distance, to deal with criteria that are dependent. Algorithm 2.2 describes the steps of this approach.

Algorithm 2.2 TOPSIS-M

Input: Decision data $\mathbf{M}$ and weighting vector $\mathbf{w}$.
Output: Closeness measure $\mathbf{o} = [o_1, o_2, \ldots, o_n]$.

Step 1: Normalization. For each evaluation $u_{j,i}$, perform the following normalization:

$$\tilde{u}_{j,i} = \frac{u_{j,i}}{\sqrt{\sum_{j=1}^{n} u_{j,i}^2}}, \;\; i = 1, \ldots, m, \; j = 1, \ldots, n.$$

Step 2: Ideal alternatives determination. Calculate both positive and negative ideal alternatives (PIA and NIA, respectively) through the equation

$$PIA = \tilde{\mathbf{u}}^+ = \left\{\tilde{u}_1^+, \tilde{u}_2^+, \ldots, \tilde{u}_m^+\right\} \text{and } NIA = \tilde{\mathbf{u}}^- = \left\{\tilde{u}_1^-, \tilde{u}_2^-, \ldots, \tilde{u}_m^-\right\},$$

where $\tilde{u}_i^+ = \max\{\tilde{u}_{j,i} | 1 \leq j \leq n\}$ and $\tilde{u}_i^- = \min\{\tilde{u}_{j,i} | 1 \leq j \leq n\}$, $i = 1, \ldots, m$.

Step 3: Covariance matrix. Calculate the covariance matrix $\mathbf{\Sigma}_{\tilde{\mathbf{M}}} \in \mathbb{R}^{m \times m}$ of $\tilde{\mathbf{M}} = (\tilde{u}_{j,i})$, given by

$$\mathbf{\Sigma}_{\tilde{\mathbf{M}}} = \mathbb{E}\left[\left(\tilde{\mathbf{M}} - \mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)\left(\tilde{\mathbf{M}} - \mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)^T\right],$$

where $\mathbb{E}\left[\tilde{\mathbf{U}}\right]$ denotes the expectation of $\tilde{\mathbf{U}}$.

Step 4: Mahalanobis distances. Calculate the Mahalanobis distances from each alternative a_j and both PIA and NIA:

$$DM_j^+ = \sqrt{\left(\tilde{\mathbf{u}}_j - \tilde{\mathbf{u}}^+\right)^T \Delta^T \mathbf{\Sigma}_{\tilde{\mathbf{M}}}^{-1} \Delta \left(\tilde{\mathbf{u}}_j - \tilde{\mathbf{u}}^+\right)}, \;\; j = 1, \ldots, n$$

and

$$DM_j^- = \sqrt{\left(\tilde{\mathbf{u}}_j - \tilde{\mathbf{u}}^-\right)^T \Delta^T \mathbf{\Sigma}_{\tilde{\mathbf{M}}}^{-1} \Delta \left(\tilde{\mathbf{u}}_j - \tilde{\mathbf{u}}^-\right)}, \;\; j = 1, \ldots, n,$$

where $\tilde{\mathbf{u}}_j = [\tilde{u}_{j,1}, \tilde{u}_{j,2}, \ldots, \tilde{u}_{j,m}]$ and $\Delta = diag\left(w_1, w_2, \ldots, w_m\right)$ is the diagonal matrix whose elements are the weights $\mathbf{w}$.

Step 5: Closeness measure. For each alternative a_j, calculate the closeness measure:

$$o_j = \frac{DM_j^-}{DM_j^+ + DM_j^-}, \;\; j = 1, \ldots, n,$$

Several works in the literature have applied the TOPSIS-M to deal with decision problems with dependent criteria. Examples include ranking problems on mutual funds (Chang et al., 2010), areas for building redevelopment (Antuchevičiene et al., 2010), Chinese high-tech industry (Wang and Wang, 2014), insurance companies (Chen and Lu, 2015) investment policies (Chen, 2017) and China energy regulation (Wang et al., 2018). Therefore, given the relevance of both TOPSIS and TOPSIS-M in dealing with decision problems in the literature, a further investigation of such approaches was conducted in this study. See, for instance, the contributions described in Chapter 4.

2.4.2 Methods based on binary relations

The methods based on binary relations (or outranking methods (Bouyssou, 2001)) are frequently associated with the name of Bernard Roy and his works on the development of the family of ELECTRE methods (Roy, 1990b). These methods were designed to overcome some difficulties that were found when dealing with real-world problems by using compensatory approaches (see, e.g., (Roy and Vanderpooten, 1996)). For instance, it may be undesirable for the decision maker that a loss on an individual criterion may be compensated by a gain on another one.

Differently from the methods based on a unique synthesizing criterion, the outranking methods derive the ranking of alternatives according to a system of relational preferences provided by the decision maker. Instead of aggregating the criteria evaluations and, then, establishing the preference relations by comparing the obtained overall evaluations, in outranking methods one first defines the preference about pairs of alternatives looking at each criterion individually and, then, derive the global preference (or outranking relations). Therefore, one may decompose the outranking methods into two steps: (i) construction of the outranking relations and (ii) exploitation of such relations in order to provide a recommendation.

In the first step, the outranking relation, denoted by $\succsim$, is determined upon the concordance-discordance principle, defined as follows:

- Concordance principle: we say that an alternative a outranks another alternative b, represented by $a \succsim b$, if the majority of the criteria is in favor of this statement.
- Non-discordance principle: when the concordance principle hold, we say that an alternative a outranks another alternative b if the criteria that are not in favor of this statement (the minority) are not "strong enough".

Therefore, in summary, given the set of preferences of alternative a over alternative b under the light of each criterion individually, we conclude that a outranks b if there is enough arguments that attest this statement and no reason to refute it.

The second step in binary relations methods consists in exploiting the outranking relations in order to provide a recommendation. Some aspects are important to be highlighted here (Figueira et al., 2016). For instance, such approaches allow indifference and incomparability between pairs of criteria, which lead to an order which is not necessarily complete. Moreover, the outranking relation may lead to situations in which the order is not transitive. Therefore, since providing a recommendation may be a difficult task, specific techniques were developed along the years (Vanderpooten, 1990; Roy and Bouyssou, 1993).

As mentioned in the beginning of this section, the family of ELECTRE methods are examples of outranking approaches. Each one was developed to deal with a specific class of problems in MCDM. ELECTRE I and ELECTRE IS deals with the choice problem. The sorting problem is tackled by ELECTRE TRI. Finally, ELECTRE II, ELECTRE III and ELECTRE IV address ranking problems. Since the contributions of this thesis lie on the unique synthesizing criterion methods, we do not go into detail of the ELECTRE methods. However, for more details about each one and their applications, see (Figueira et al., 2016; Govindan and Jepsen, 2016).

2.5 On the determination of weights and capacity values

The methods presented in Section 2.4.1, which are the focus of this study, require the definition of a set of parameters[17] (e.g., weights in the WAM and TOPSIS, and capacity coefficients in the Choquet integral and multilinear model). For instance, the easiest decision about the weights would be assigning the same values for all w_i, i.e., $w_i = 1/m$. This reasoning is founded on the principle of maximum entropy (MacKay, 2007), which states that, in the absence of information about the criteria weights, the best approach would be assigning the same value for all of them. However, this may be inconsistent with the opinions of the decision maker or may not take into account information that can be extracted from the decision data. Therefore, one generally adopts an approach to determine the decisional parameters. This approach, referred as a preference elicitation process, can be defined as follows (adapted from (Mousseau, 2003)):

Definition 2.12. ***(Preference elicitation process):*** *A preference elicitation process comprises the interaction between the decision maker and the analyst (or a computer software) in order to help the former to express his/her preferential information according to the adopted MCDM method. This information leads to a set of possible values for the preferential parameters used in the MCDM method. At the end of this process, these values should lead, based on the application of the MCDM method, to a result compatible with the decision maker's point of view.*

[17] We recall that we assume that the value functions $u_i(x_i)$ have already been determined.

According to Definition 2.12, therefore, the DM should provide the preferential information used to determine the parameters of the considered MCDM method. However, in some situations, the DM does not have this information or may not be willing to express it. Aiming at dealing with this task, several approaches were developed in the literature. For instance, some sophisticated and well-known ones are the Analytic Hierarchy Process (Saaty, 1987) and the FITradeoff (de Almeida et al., 2016).

Other methods, among which some of them are related with the contributions of this study, will be presented in the next sections. They are divided into two categories, namely subjective and objective approaches (for further details see (Pöyhönen and Hämäläinen, 2001; Wang et al., 2009; Zardari et al., 2015)).

2.5.1 Subjective approaches

Subjective approaches comprise methods that use the information provided by the decision maker in order to determine the decisional parameters. In this case, the DM can directly express the parameters values that he/she wants to consider (the case of a direct elicitation) or the preference between alternatives. In the latter scenario, one attempts to define a set of parameters that, given the application of the considered MCDM method, satisfies the preference provided by the decision maker. We call this approach an indirect elicitation. Both types of elicitation are presented in the next sections. It is worth mentioning that some works also refer to them as aggregation/disaggregation approaches (Mousseau, 2003).

2.5.1.1 Direct elicitation: Defining the set of weights

In direct elicitation, the analyst conducts the decision maker to define the parameters values used in the considered MCDM problem. An important aspect in such situations is that the analyst must guarantee that the DM is aware of the adopted model and the influence that each parameter has in the final aggregation. This knowledge is fundamental in order to avoid discrepancies (with the preference of the DM) in the obtained result. In this case, different weights in the WAM may lead to a ranking with which the DM may not agree.

Another concern when directly eliciting the parameters is that the decision maker may not be aware of possible intercriteria relations. For instance, there are situations (as the one presented in Example 2.3) in which pairs of criteria are correlated, which may introduce a bias in the achieved ranking of alternatives. Therefore, defining the parameters without this characteristic in mind may favour a subset of alternatives whose only the evaluations on the correlated criteria are satisfactory. An approach proposed in this study (see Chapter 5) deals with this scenario.

There are several direct elicitation methods that can be used to help the DM in the parameters values definition. Some of them, specially used in the weights determination, are presented in the sequel. Aiming at illustrating them, we use Example 2.3. In this case, criteria c_1, c_2 and c_3 represent the grades in Calculus, Physics and Literature, respectively.

Direct rating

One of the simplest methods that can be used by the decision maker in parameters elicitation is the direct one. In this method, the DM assigns a number for every criterion (or coalitions of criteria), which indicates its importance in the decision problem under consideration. For example, he/she may indicate numbers from 1 to 10 for each weight and, thereafter, normalize them (to ensure that the sum is equal to one) in order to use the WAM. For instance, in Example 2.3, the decision maker may assign the numbers 8, 8 and 6 for c_1, c_2 and c_3, respectively, which leads to $\mathbf{w} = [0.36, 0.36, 0.28]$.

This procedure may also be used to define the capacity coefficients. However, in this case, the indicated values must be in the interval $[0, 1]$ and satisfy the axioms of a capacity.

Point allocation

In the point allocation method, the DM is asked to divide a certain number of points (e.g., 100) among the criteria. The more points allocated to a criterion, the more important it is. After having assigned all points, one normalizes the weights to ensure that the sum is equal to one. In Example 2.3, one may allocate 40, 35 and 25 points for Calculus, Physics and Literature, respectively, which leads to $\mathbf{w} = [0.40, 0.35, 0.25]$.

A remark in this method is that giving more points to a criterion must be compensated by withdrawing the same amount to one or several criteria. Therefore, the decision maker must be aware of this trade-off during the elicitation process.

SWING

In the SWING method (von Winterfeldt and Edwards, 1986), one asks to the decision maker what is the criterion (the most important for him/her) that he/she would prefer to change from the worst level to the best one. For this criterion, one assigns a value of 100. Thereafter, the DM selects the second preferred criterion and assigns a value less or equal to 100. This is repeated for all criteria and, at the end of the process, one normalizes the weights. For instance, in Example 2.3, the decision maker may assign c_2 with value 100, c_1 with 90 and c_3 with 60. Therefore, in this case, the weights will be $\mathbf{w} = [0.36, 0.40, 0.24]$.

SMART and SMARTER

Differently from SWING, the SMART method developed by Edwards (1977) starts by ranking the criteria from the worst to the best in terms of the importance according to the DM opinion. For the least important criterion, one assigns a value of 10. Then, for the second criterion on the ranking, the decision maker assigns a value greater or equal to 10. This procedure is repeated for all criteria and we obtain the set of weights by normalizing the assigned values. For instance, in Example 2.3, assuming 10 for c_3, 15 for c_1 and 18 for c_2, one achieves $\mathbf{w} = [0.35, 0.42, 0.23]$.

Although both SWING and SMART are different in terms of the order in which the information about each criterion is provided (from the most to the least important in SWING and from the least to the most important in SMART), they require that the DM assigns values for each one. In order to avoid this, Edwards and Barron (1994) proposed an extended version of SMART, called SMARTER. By using this method, the DM only needs to rank the criteria from the best to the worst (in terms of preference) and define the weight $w_{(i)}$ for the i-th criterion on the ranking through the expression

$$w_{(i)} = \frac{1}{m} \sum_{q=i}^{m} \frac{1}{q}. \tag{2.28}$$

Therefore, assuming that the DM has a small preference of c_2 over c_1 and prefers c_1 to c_3, one achieves

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1/2 + 1/3 \\ 1 + 1/2 + 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0.28 \\ 0.61 \\ 0.11 \end{bmatrix}. \tag{2.29}$$

One may note that, even with a small preference between c_2 and c_1, the application of SMARTER led to very different weights values for these two criteria. Therefore, although SMARTER automatically assigns the weights values given the preference provided by the decision maker, the obtained result may be unsatisfactory for a small set of criteria.

2.5.1.2 Indirect elicitation: Supervised approaches for capacity identification

In the previous section, we presented some direct elicitation methods used to define the set of weights according to the DM preference over the set of criteria. In this section, we discuss the indirect elicitation methods used to identify the capacity coefficients.

Differently from the direct elicitation, in the indirect one the analyst conducts the DM to express his/her preference over a subset $\mathcal{T}$ of all possible objects of interest X. $\mathcal{T}$ may be a (generally small) subset of the set of alternatives $\mathcal{A}$ or a set of synthetic alternatives. In the latter case, one must ensure that the synthetic alternatives are enough for the DM to provide his/her preference relations and that we may use this information to define the decisional parameters. Since in this work we consider a numerical representation of the preference relations, the information provided by the DM contains overall scores of the set of alternatives $\mathcal{T}$. Furthermore, in this study, we often refer to the indirect elicitation as a supervised approach and the set of criteria evaluations (the decision data) and their associated overall values as learning data.

In the literature, several methods were developed to address the problem of capacity identification in the context of the Choquet integral (Miranda and Grabisch, 1999; Marichal and Roubens, 2000). For instance, Grabisch et al. (2008) presented an interesting review of such methods. Generally, they comprise an optimization problem whose constraints represent the information provided by the decision maker (preference over alternatives and/or interaction indices) and ensure that the axioms of a capacity are satisfied. However, the main difference between them lies in the adopted cost function. In this study, we explore the least-squares-based approaches[18] (see (Grabisch et al., 2008) for more details), whose goal is to define a capacity that minimizes the difference between the achieved overall evaluations and the ones provided by the DM. Mathematically, this cost function (also called representation error) is given by

$$E = \sum_{j=1}^{n} \left(F_{CI}(u(\mathbf{x}_j)) - y(u(\mathbf{x}_j))\right)^2, \tag{2.30}$$

where $y(u(\mathbf{x}_j))$ is the desired overall evaluation for alternative $\mathbf{x}_j$ (according to the decision maker opinion). With respect to the constraints, the DM may express his preference about the criteria through the interaction indices. For instance, he/she may consider that

- the marginal contribution of criterion i should be greater than (or equal to) the one of criterion k, i.e., $\phi_i^{\mathcal{S}} - \phi_k^{\mathcal{S}} \geq \delta_{i,k}^{\phi}$, and/or
- the interaction degree of criteria i, k should be greater than (or equal to) the one of criteria d, l, i.e., $I_{i,k}^{\mathcal{S}} - I_{d,l}^{\mathcal{S}} \geq \delta_{ik,dl}^{I}$ and $I_{d,l}^{\mathcal{S}} \geq 0$,

where $\delta_{i,k}^{\phi}$ and $\delta_{ik,dl}^{I}$ are nonnegative indifference thresholds. Therefore, the optimization model, which also includes the constraints associated with the axioms of a capacity, can

[18] Besides the least-squares-based approaches, there are capacity identification models that are based on linear programming. For instance, the approach proposed by Marichal and Roubens (2000) consists in a linear programming whose aim is to maximize the minimal difference between the overall evaluations of pairs of alternatives in $\mathcal{T}$ based on a ranking provided by the decision through a partial preorder.

be defined by

$$
\begin{array}{ll}
\min\limits_{\mu} & \sum_{j=1}^{n} \left(F_{CI}(u(\mathbf{x}_j)) - y(u(\mathbf{x}_j))\right)^2 \\
\text{s.t.} & \phi_i^{\mathcal{S}} - \phi_k^{\mathcal{S}} \geq \delta_{i,k}^{\phi}, \\
& \vdots \\
& I_{i,k}^{\mathcal{S}} - I_{d,l}^{\mathcal{S}} \geq \delta_{ik,dl}^{I}, \\
& I_{d,l}^{\mathcal{S}} \geq 0, \\
& \vdots \\
& \mu(\{A \cup i\}) - \mu(A) \geq 0, \ \ \forall i \in C, \ \ \forall A \subseteq C \backslash i, \\
& \mu(\emptyset) = 0, \\
& \mu(C) = 1.
\end{array}
\tag{2.31}
$$

It is important to highlight that (2.31) may have no feasible solution, since the opinions provided by the decision maker may not be consistent with the axioms of a capacity. Moreover, the number of capacity coefficients to be determined, given by $2^m - 2$ (excluding $\mu(\emptyset)$ and $\mu(C)$), increases exponentially with the number of criteria. In that respect, aiming at reducing the computational effort in decision problems with a large number of criteria, one commonly adopts a 2-additive capacity. In this case, the optimization problem is given by

$$
\begin{array}{ll}
\min\limits_{\mathbf{I}^{\mathcal{S}}} & \sum_{j=1}^{n} \left(F_{CI}(u(\mathbf{x}_j)) - y(u(\mathbf{x}_j))\right)^2 \\
\text{s.t.} & \phi_i^{\mathcal{S}} - \phi_k^{\mathcal{S}} \geq \delta_{i,k}^{\phi}, \\
& \vdots \\
& I_{i,k}^{\mathcal{S}} - I_{d,l}^{\mathcal{S}} \geq \delta_{ik,dl}^{I}, \\
& I_{d,l}^{\mathcal{S}} \geq 0, \\
& \vdots \\
& \phi_i^{\mathcal{S}} - \frac{1}{2} \sum_{i \neq k} \left| I_{i,k}^{\mathcal{S}} \right| \geq 0, \ \ \forall i \in C \\
& I^{\mathcal{S}}(\emptyset) - \frac{1}{2} \sum_i \phi_i^{\mathcal{S}} + \frac{1}{6} \sum_{i,k} I_{i,k}^{\mathcal{S}} = 0, \\
& \sum_i \phi_i^{\mathcal{S}} = 1,
\end{array}
\tag{2.32}
$$

where the number of variables decreases to $m(m+1)/2 - 1$.

A remarkable aspect in capacity identification is that most works in the literature deal with this task in the context of the Choquet integral. Therefore, as will be discussed in Chapter 6, a contribution of this thesis also addresses the problem of capacity identification in the multlinear model.

2.5.2 Objective approaches

In the last sections, we described some subjective approaches used to determine the parameters of the MCDM methods, which require some information provided by the decision maker. Conversely, in this section, we present some objective approaches (Wang et al., 2009; Zardari et al., 2015)). In such methods, the parameters are obtained through mathematical models that extract information through the decision data, only, without the intervention of the decision maker. This characteristic may also be a critic for such methods, since the experience of the DM is not considered in the elicitation process (although some methods may also include this information). However, they can be seen as automatic approaches that avoid subjectivity expressed by the decision maker when defining the decisional parameters.

We describe some objective approaches in the sequel. For instance, we explore some methods used in the weights determination and mathematical models for capacity identification in an unsupervised scenario.

2.5.2.1 Estimating weights from the decision data

This section presents some objective approaches used to determine the weighting vector. Each one was developed according to a specific goal about what one can achieve with the estimated parameters. Aiming at illustrating the application of them, we considered a synthetic dataset composed of 1000 alternatives with 3 decision criteria. Scatter plots of pairs of criteria are illustrated in Figure 5. The data were generated according to a uniform distribution on the interval $[0, 1]$ and with a positive correlation between criteria 2 and 3.

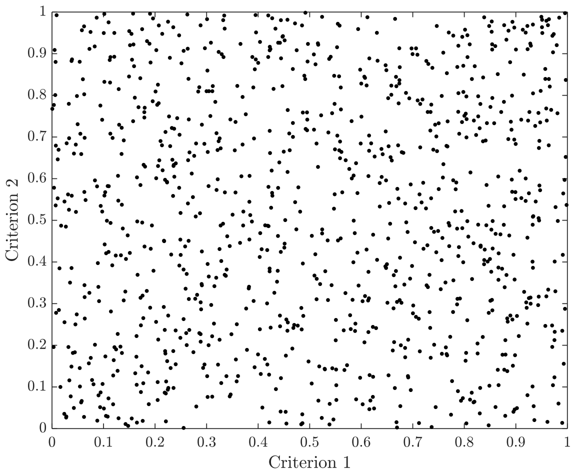
Entropy method

The first method presented in this section is based on a concept in information theory called entropy. Roughly speaking, the Shannon entropy (Shannon, 1948) represents a measure of the uncertainty in a discrete probability distribution. If one compares, for instance, a broad and a sharply peaked distributions, one has more uncertainty in the former than in the latter. Therefore, in the context of MCDM, the reasoning is the following: since the decision matrix carries an amount of information, criteria with more uncertainty in their values (broad distributions) should be assigned with relative weights greater than the ones that does not carry so much information. In that respect, after the normalization of the decision data, the weights can be defined through the following expression (for further details, see (Hwang and Yoon, 1981)):

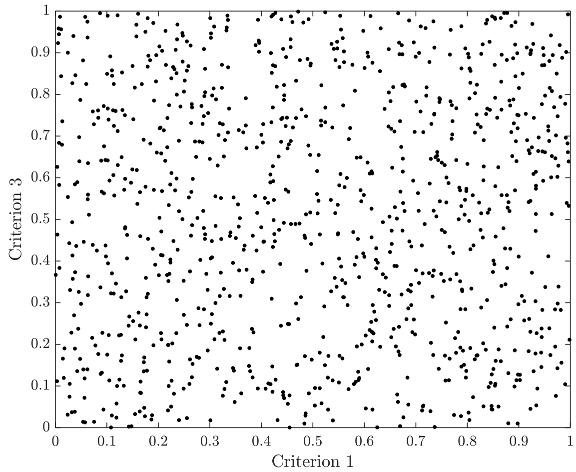
$$u'_{j,i} = \frac{u_{j,i}}{\sum_{j=1}^{n} u_{j,i}}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{2.33}$$

$$H_i = -\frac{\sum_{j=1}^{n} u'_{j,i} \ln(u'_{j,i})}{\ln(n)}, \quad i = 1, \ldots, m, \tag{2.34}$$

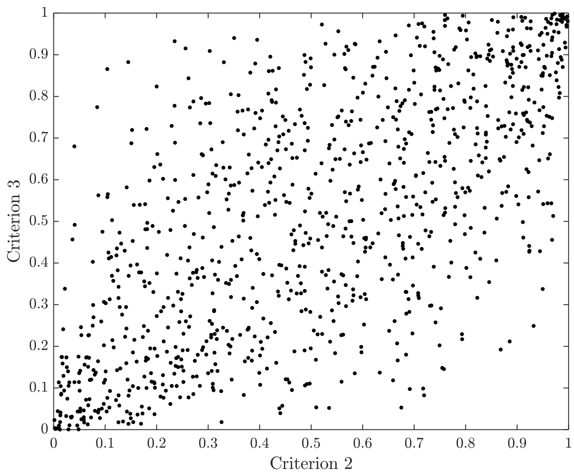
$$w_i = \frac{1 - H_i}{\sum_{i=1}^{m} (1 - H_i)}, \quad i = 1, \ldots, m. \tag{2.35}$$



(a) Criteria 1 and 2.



(b) Criteria 1 and 3.



(c) Criteria 2 and 3.

Figure 5 – Scatter plot of pairs of criteria.

If we apply the entropy method in the dataset illustrated in Figure 5, we achieve $\mathbf{w} = [0.34, 0.33, 0.33]$, i.e., practically the same weight for all criteria. This is because all criteria are generated according to a uniform distribution on the same interval. Moreover, we may remark that the correlation between criteria 2 and 3 does not intervene in the entropy method.

It is important to highlight that, if the decision maker has predefined weights w_i', $i = 1, \ldots, m$, that he/she wants to include in this elicitation process, the weights can be modified by

$$\hat{w}_i = \frac{w_i' w_i}{\sum_{i=1}^{m} w_i' w_i}, \quad i = 1, \ldots, m. \tag{2.36}$$

Standard deviation method

The idea behind the standard deviation method is similar to the one of the entropy method. Weights assigned to the criteria whose evaluations are "more spread" between the lower and upper bounds should be greater than the parameters associated with criteria that do not bring a considerable information in the decision problem. In this context, the weights can be defined by

$$w_i = \frac{\sigma_i}{\sum_{i=1}^{m} \sigma_i}, \quad i = 1, \ldots, m, \tag{2.37}$$

where σ_i, given by

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{n} \left(u_{j,i} - \frac{\sum_{j=1}^{n} u_{j,i}}{n} \right)^2}{n}}, \quad i = 1, \ldots, m, \tag{2.38}$$

is the standard deviation of the evaluations associated with criterion i. By applying this method in the example presented in Figure 5, we also obtain $\mathbf{w} = [0.34, 0.33, 0.33]$. Similarly as in the entropy method, the standard deviation does not exploit interactions between criteria and, therefore, even with a correlation between criteria 2 and 3, we obtained the same weights for all of them.

CRITIC

In both entropy and standard deviation methods, although the intracriteria information was taken into account in the weights determination, one could note that they do not consider the intercriteria relations. Aiming at exploiting both characteristics, Diakoulaki et al. (1995) developed a new method, called CRITIC (CRiteria Importance Through Intercriteria Correlation). This method quantifies the information carried by each criterion and the correlation in the decision data through the following equation:

$$\gamma_i = \sigma_i \sum_{i'=1}^{m} \left(1 - \rho_{i,i'}\right), \quad i = 1, \ldots, m, \tag{2.39}$$

where σ_i is the standard deviation of the (normalized) evaluations associated with criterion i (see Equation (2.38)) and $\rho_{i,i'}$, defined by

$$\rho_{i,i'} = \frac{n \left(\sum_{j=1}^{n} u_{j,i} u_{j,i'} \right) - \left(\sum_{j=1}^{n} u_{j,i} \right) \left(\sum_{j=1}^{n} u_{j,i'} \right)}{\sqrt{\left(n \sum_{j=1}^{n} u_{j,i}^2 - \left(\sum_{j=1}^{n} u_{j,i} \right)^2 \right) \left(n \sum_{j=1}^{n} u_{j,i'}^2 - \left(\sum_{j=1}^{n} u_{j,i'} \right)^2 \right)}}, \quad \forall i, i' \in C, \tag{2.40}$$

is the linear correlation coefficient between the (normalized) evaluations of criteria i, i'. Given all γ_i, $i = 1, \ldots, m$, the weights are calculated as follows:

$$w_i = \frac{\gamma_i}{\sum_{i=1}^{m} \gamma_i}, \quad i = 1, \ldots, m. \tag{2.41}$$

In the dataset illustrated in Figure 5, the application of CRITIC leads to the weighting vector $\mathbf{w} = [0.42, 0.29, 0.29]$. Therefore, the weights associated with the correlated criteria (w_2 and w_3) were lower than the one assigned to the independent criterion (w_1), which attests that the application of CRITIC allows us to consider intercriteria relations.

2.5.2.2 Unsupervised approaches for capacity identification

Similarly as discussed in the previous section, there are some methods in the literature whose goal is to estimate the capacity coefficients in the Choquet integral through the information extracted from the decision data (Kojadinovic, 2008; Rowley et al., 2015; Duarte, 2018). We also refer to this methods as unsupervised approaches for capacity identification, since they do not require the knowledge of the overall evaluations (or the preference relations provided by the DM).

Among these methods, the approach proposed by Duarte (2018) is of interest for this study (see Chapter 7), since it deals with the identification problem by associating the interaction indices $I^{\mathcal{S}}_{i,i'}$ of pairs of criteria to similarity measures (e.g., the Pearson's correlation coefficient (Chen and Popovich, 2002)) between them. Aiming at reducing the number of parameters to be estimated, the author considered a 2-additive Choquet integral. Therefore, other than the interaction indices $I^{\mathcal{S}}_{i,i'}$, one also needs to estimate[19] the power indices $\phi^{\mathcal{S}}_i$. However, in the absence of information about each criterion individually, it was assigned the same marginal contribution for all power indices (principle of maximum entropy (MacKay, 2007)), i.e., $\phi^{\mathcal{S}}_i = 1/m \ \forall i = 1, \ldots, m$.

With respect to the interaction indices, since these values are in the interval $[-1, 1]$, the main idea is to estimate them by means of a similarity measure $\psi(u_i(\mathbf{x}_i), u_{i'}(\mathbf{x}_{i'}))$ (or, simply, $\psi_{i,i'}$) (which is also a value in the interval $[-1, 1]$) between each pair of decision criteria, such as the Pearson's or the Spearman's correlation coefficients (Chen and Popovich, 2002). If a similarity measure $\psi_{i,i'}$ is close to 1 (resp., close to -1), which means that both criteria are positively (resp., negatively) correlated, one penalizes (resp., rewards) this coalition with an interaction index close to -1 (resp., close to 1), which introduces a redundant (resp., complementary) effect between them. In other words, for each pair of criteria i and i', the associated interaction index $I^{\mathcal{S}}_{i,i'} \approx -\psi_{i,i'}$.

Recall that, in order to apply this approach, one needs to satisfy the axioms of a capacity, presented in Section 2.4.1.2.2. For the first two conditions (normalization axioms), assuming that $\phi^{\mathcal{S}}_i = 1/m$, $i = 1, \ldots, m$, both are automatically satisfied. However, the monotonicity condition, given by

$$\phi^{\mathcal{S}}_i - \frac{1}{2}\sum_{i \neq i'} \left| I^{\mathcal{S}}_{i,i'} \right| \geq 0, \ \ \forall i \in C, \tag{2.42}$$

may not always be true for $I^{\mathcal{S}}_{i,i'} = -\psi_{i,i'}$, $\forall i, i' \in C$. Therefore, one searches for $\hat{\psi}_{i,i'}$ that are as close as possible to $\psi_{i,i'}$, for all $i, i' \in C$. This is achieved by solving the following

[19] Recall that $I^{\mathcal{S}}(A) = 0 \ \forall A$ such that $|A| \geq 3$ and $I(\emptyset)^{\mathcal{S}}$ can be retrieved by the axioms of a capacity (see Section 2.4.1.2.2).

optimization problem:

$$\begin{array}{ll}\min\limits_{\hat{\Psi}} & \left\|\hat{\Psi}-\Psi\right\|_{\mathcal{F}} \\ \text{s.t.} & \frac{1}{m}-\frac{1}{2}\sum_{i\neq i'}\left|\hat{\psi}_{i,i'}\right|\geq 0,\ \ \forall i\in C \\ & \hat{\Psi}\geq 0 \\ & \hat{\psi}_{i,i'}-\hat{\psi}_{i',i}=0,\ \ \forall i,i'\in C,\end{array} \tag{2.43}$$

where the matrix Ψ, given by

$$\Psi=\begin{bmatrix}\psi_{1,1} & \psi_{1,2} & \dots & \psi_{1,m} \\ \psi_{2,1} & \psi_{2,2} & \dots & \psi_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{m,1} & \psi_{m,2} & \dots & \psi_{m,m}\end{bmatrix}, \tag{2.44}$$

is obtained from the decision data (e.g., the covariance matrix of $\mathbf{M}$),

$$\hat{\Psi}=\begin{bmatrix}\hat{\psi}_{1,1} & \hat{\psi}_{1,2} & \dots & \hat{\psi}_{1,m} \\ \hat{\psi}_{2,1} & \hat{\psi}_{2,2} & \dots & \hat{\psi}_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\psi}_{m,1} & \hat{\psi}_{m,2} & \dots & \hat{\psi}_{m,m}\end{bmatrix} \tag{2.45}$$

and $\left\|\hat{\Psi}-\Psi\right\|_{\mathcal{F}}$ is the Frobenius norm (Golub and Van Loan, 2013) between $\hat{\Psi}$ and Ψ, i.e.,

$$\left\|\hat{\Psi}-\Psi\right\|_{\mathcal{F}}=\sqrt{\sum_{i=1}^{m}\sum_{i'=1}^{m}\left(\hat{\psi}_{i,i'}-\psi_{i,i'}\right)^2}. \tag{2.46}$$

Algorithm (2.3) describes the steps of the addressed approach in determining the weights according to a decision data stored in matrix $\mathbf{M}$.

Algorithm 2.3 Unsupervised approach proposed by Duarte (2018).

Input: Decision data $\mathbf{M}$.
Output: Interaction indices $I^{\mathcal{S}}(A)$.

Step 1: Estimate Ψ from $\mathbf{M}$.
Step 2: Set $\phi_i^{\mathcal{S}}=1/m$, $\forall i=1,\dots,m$.
Step 3: **if** $I_{i,i'}^{\mathcal{S}}=-\psi_{i,i'}$, $\forall i,i'\in C$, satisfies (2.42),
  Set $I_{i,i'}^{\mathcal{S}}=-\psi_{i,i'}$, $\forall i,i'\in C$,
 else
  Obtain $\hat{\Psi}$ by solving (2.43) and set $I_{i,i'}^{\mathcal{S}}=-\hat{\psi}_{i,i'}$, $\forall i,i'\in C$.
 endif
Step 4: Obtain $I^{\mathcal{S}}(\emptyset)$ from the axioms of a capacity, set $I^{\mathcal{S}}(A)=0$ $\forall A$ such that $|A|\geq 3$ and calculate the capacity μ from (2.20).

2.6 Conclusions

This chapter has presented the theoretical aspects of multicriteria decision making. We discussed the different approaches and classes of problems that decision makers may face. We also described the basic elements involved in multicriteria decision making and which aspects will be considered in our analysis. A special attention was devoted to the case in which two or more criteria are redundant, i.e., when there are correlations among the decision data. We consider that characteristic may introduce biases in the overall evaluations, which should be avoided in order to achieve a fairer ranking of alternatives. For this purpose, our aim consists in adopting aggregation procedures that take into account the redundancies and softened the associated biases.

As mentioned in this chapter, some methods in the literature were conceived, among other goals, to deal with correlation between criteria. An example is the TOPSIS-M, which exploits correlations by means of the Mahalanobis distance. Another method that can be used is the Choquet integral. In this case, the considered capacity can overcome inconveniences provided by redundant criteria by modeling interactions among them. Moreover, the weights used in the weighted arithmetic mean may also be adjusted to penalize positively correlated criteria. However, a concern in such methods is how to define the parameters in order to take into account the observed correlations.

Some contributions of this study deal with the parameter identification problem by using statistical signal processing techniques. The main idea is to tackle this task through statistical measures extracted from the decision data. In the next chapter, we present the latent variable analysis techniques considered in this thesis.

3 Latent variable analysis

The last chapter presented the theoretical foundation of multicriteria decision making. We raised some important aspects associated with MCDM problems, such as the biased results that one may face given the presence of correlations among criteria. In this study, we propose to deal with such situations by means of Latent Variable Analysis (LVA). For instance, we verify the application of three techniques that exploit the information contained within the observed decision data.

The first two techniques presented in this chapter are the Principal Component Analysis (PCA) and the Independent Component Analysis (ICA). Both are detailed in Sections 3.1 and 3.2, respectively. PCA is a technique commonly used in feature extraction and dimension reduction problems. In summary, the goal of PCA is to project the data into a set of uncorrelated variables. Similarly as PCA, the ICA is also used in feature extraction problems. A typical application of this technique is in signal separation. However, a difference between PCA and ICA is that the aim of the latter is to find a transformation process that leads to a set of independent variables.

The other technique discussed in this chapter, called sensitivity analysis, is presented in Section 3.3. The sensitivity analysis is useful to verify the impact that a variable (or a set of them) has on the output model.

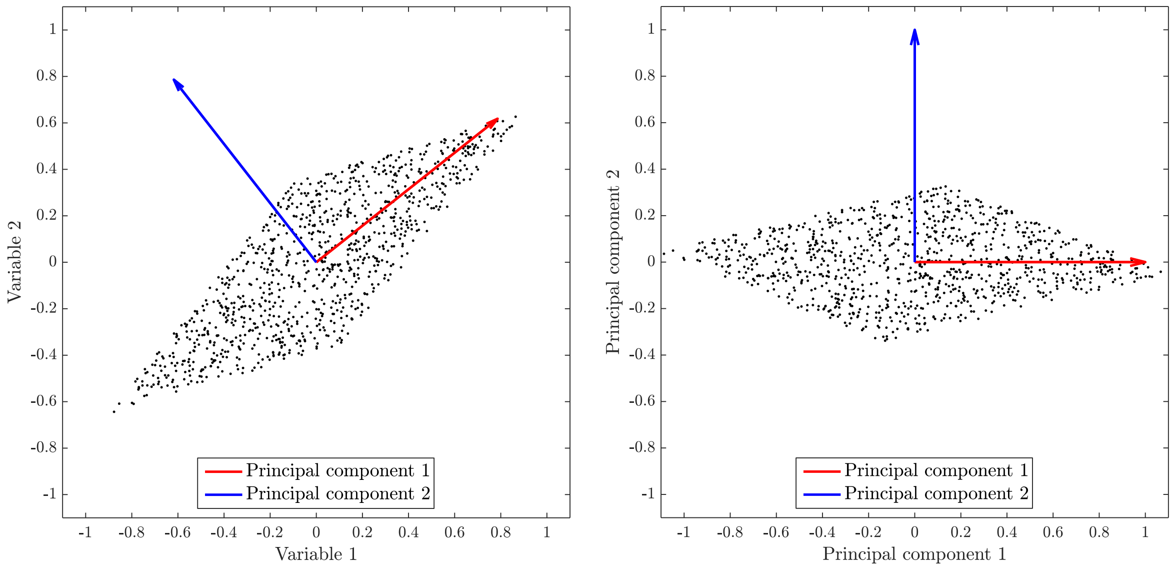
Finally, in Section 3.4, we describe our final remarks about the considered LVA methods. Moreover, we briefly discuss how they will be applied in MCDM problems.

3.1 Principal component analysis (PCA)

The principal component analysis (Jolliffe, 2002), also known as Hotelling transform (Hotelling, 1933), is a well-known technique and largely used in signal processing and machine learning problems. In multicriteria decision making, we also may find some works that combine PCA and others MCDM methods to deal with redundancies in the decision data. One may cite, for example, the works of Wang (2015) and Zhu et al. (2016), which combine TOPSIS with PCA to address correlation between criteria.

The main idea of PCA is to find a linear transformation of the considered dataset into a set of uncorrelated variables (normally, with a reduced dimension) such that each variable (the principal component scores) retains as much information as possible from the original data. In high-dimensional problems, for example, PCA is a useful tool to perform dimension reduction (e.g., data compression), since a few number of principal components may carry enough information about the original data. Moreover, PCA is also applied in feature extraction, since relevant information can be retained while redundancies are eliminated (recall that principal components are uncorrelated).

A geometric interpretation of PCA is presented in Figure 6. One may clearly see that the original data (Figure 6a) was projected (or rotated) into a novel representation (Figure 6b) whose components are uncorrelated. Although it is a simple example, one may see that the projection into the first component is the one that retains more information from the original data.



(a) Original data. (b) Uncorrelated projected data.

Figure 6 – Example of PCA.

Suppose that $\hat{\mathbf{M}} \in \mathbb{R}_{m\times n}$ and $\mathbf{z}_1 = [z_{1,1}, \ldots, z_{1,n}]$ represent the original data[1] and the first principal component scores, respectively. Moreover, consider that $\mathbf{e}_1 = [e_{1,1}, \ldots, e_{m,1}]^T$ is the vector that projects $\hat{\mathbf{M}}$ on the first principal component. Therefore, in order to determine $\mathbf{z}_1$, one needs to find the principal component coefficients $e_{1,1}, \ldots, e_{m,1}$ that maximize the information contained in the original data. This information is measured in terms of the variance of $\mathbf{z}_1 = \mathbf{e}_1^T\hat{\mathbf{M}}$, given by[2]

$$\mathrm{Var}\left[\mathbf{z}_1\right] = \mathrm{Var}\left[\mathbf{e}_1^T\hat{\mathbf{M}}\right] = \mathbf{e}_1^T\hat{\mathbf{M}}\hat{\mathbf{M}}^T\mathbf{e}_1 = \mathbf{e}_1^T\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_1, \tag{3.1}$$

where $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$ represents the covariance matrix of $\hat{\mathbf{M}}$. Based on this measure, the first principal component coefficients can be obtained by solving the following optimization problem:

$$\begin{aligned} \max_{\mathbf{e}_1} \quad & \mathbf{e}_1^T\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_1 \\ \text{s.t.} \quad & \mathbf{e}_1^T\mathbf{e}_1 = 1, \end{aligned} \tag{3.2}$$

[1] Aiming at simplifying the vector/matrix notation, we considered that the rows and the columns in $\hat{\mathbf{M}}$ represent variables and samples, respectively. In other words, $\hat{\mathbf{M}}$ can be seen as the transpose of $\mathbf{M}$ (defined in Chapter 2, Section 2.2.3), i.e., $\hat{\mathbf{M}} = \mathbf{M}^T$.

[2] For simplicity of notation, we assume that $\hat{\mathbf{M}}$ has zero mean. Otherwise, we need to centralize it by subtracting its mean, i.e., $\hat{\mathbf{M}} \leftarrow \hat{\mathbf{M}} - \mathbb{E}\left[\hat{\mathbf{M}}\right]$.

where the constraint $\mathbf{e}_1^T\mathbf{e}_1 = 1$ guarantees that $\mathbf{e}_1$ is a unitary vector. As will be further discussed in this section, this condition is important to obtain an orthogonal matrix which will perform the PCA.

Although the optimization problem (3.2) has a constraint, it can be easily solved by using the Lagrange multipliers (Vanderbei, 2014). In this case, $\mathbf{e}_1$ can be obtained through the solution of

$$\max_{\mathbf{e}_1,\lambda_1} \quad \mathbf{e}_1^T\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_1 + \lambda_1\left(1 - \mathbf{e}_1^T\mathbf{e}_1\right). \tag{3.3}$$

By taking the gradient of this cost function, one obtains that

$$2\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_1 - 2\lambda_1\mathbf{e}_1 = \mathbf{0} \rightarrow \mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_1 = \lambda_1\mathbf{e}_1, \tag{3.4}$$

where one recognizes that $\mathbf{e}_1$ is an eigenvector of $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$ and λ_1 is the associated eigenvalue. Therefore, the first principal component coefficients are the eigenvector associated with the highest eigenvalue of the covariance matrix $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$. Moreover, since $\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_1 = \lambda_1\mathbf{e}_1$,

$$\mathrm{Var}\left[\mathbf{z}_1\right] = \mathbf{e}_1^T\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_1 = \mathbf{e}_1^T\lambda_1\mathbf{e}_1 = \lambda_1\mathbf{e}_1^T\mathbf{e}_1 = \lambda_1, \tag{3.5}$$

i.e., the variance in the projected data $\mathbf{z}_1$ is given by the highest eigenvalue of $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$.

Given the first principal component, one may move to the second one. Other than the constraint that ensures a unitary vector, one also needs to guarantee that the second principal component is orthogonal to the first one. Therefore, a constraint associated with this property must be included into the optimization problem. In this case, the optimization problem to be solved is the following:

$$\begin{aligned} \max_{\mathbf{e}_2} \quad & \mathbf{e}_2^T\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{e}_2 \\ \text{s.t.} \quad & \mathbf{e}_2^T\mathbf{e}_2 = 1, \\ & \mathbf{e}_2^T\mathbf{e}_1 = 0, \end{aligned} \tag{3.6}$$

where the constraint $\mathbf{e}_2^T\mathbf{e}_1 = 0$ guarantees that $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthogonal. One also may tackle (3.6) through the use of Lagrange multipliers and, after some mathematical manipulations, one achieves that $\mathbf{e}_2$ is the eigenvector of $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$ associated with the second highest eigenvalue (see (Jolliffe, 2002) for further details). Moreover, the variance in the projected data $\mathbf{z}_2$ is also equivalent to the second highest eigenvalue of $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$.

The aforementioned technique can be iteratively used to find all principal components. However, in the literature (Shlens, 2014a), one generally sees an approach that finds the principal components through the diagonalization of the covariance matrix $\mathbf{\Sigma}_{\mathbf{Z}}$, where

$$\mathbf{Z} = \mathbf{P}\hat{\mathbf{M}} \tag{3.7}$$

is the uncorrelated data whose rows are composed by $\mathbf{z}_1, \ldots, \mathbf{z}_m$. The idea is quite simple and the trick is to set $\mathbf{P}$ as the transpose of matrix $\mathbf{E}$, whose columns correspond to the eigenvectors of $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$. Moreover, since $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$ is a symmetric matrix, it can be decomposed into its eigenvectors and eigenvalues, i.e., $\mathbf{\Sigma}_{\hat{\mathbf{M}}} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$, where $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$. The diagonalization of $\mathbf{\Sigma}_{\mathbf{Z}}$ can be verified as follows:

$$\mathbf{\Sigma}_{\mathbf{Z}} = \mathbf{P}\hat{\mathbf{M}}\hat{\mathbf{M}}^T\mathbf{P}^T = \mathbf{P}\mathbf{\Sigma}_{\hat{\mathbf{M}}}\mathbf{P}^T = \mathbf{P}\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T\mathbf{P}^T \tag{3.8}$$

and, by setting $\mathbf{P} = \mathbf{E}^T$ and remembering that $\mathbf{E}$ is an orthogonal matrix, i.e., $\mathbf{E}^T = \mathbf{E}^{-1}$,

$$\mathbf{\Sigma}_{\mathbf{Z}} = \mathbf{E}^T\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T\mathbf{E} = \mathbf{I}\mathbf{\Lambda}\mathbf{I} = \mathbf{\Lambda}, \tag{3.9}$$

where $\mathbf{I}$ is the identity matrix. Therefore, $\mathbf{\Sigma}_{\mathbf{Z}}$ was diagonalized and the solution of PCA, which is the same as obtained in the previous approach, is the set of eigenvectors of $\mathbf{\Sigma}_{\hat{\mathbf{M}}}$.

It is important to remark that $\mathbf{\Sigma}_{\mathbf{Z}}$ being diagonal means that all $\mathbf{z}_i$ are uncorrelated (all off-diagonal elements of $\mathbf{\Sigma}_{\mathbf{Z}}$ are zero) and that the first principal components can be selected according to its diagonal elements ordered from the highest to the lowest values. For instance, if one would like to reduce the dimension of the dataset, one may select a subset of the principal components that can represent the original data in a satisfactory degree. For example, suppose a problem with 6 variables and that, after applying PCA, one achieves the eigenvalues $\lambda = [0.95, 0.80, 0.75, 0.60, 0.30, 0.25, 0.10]$ (ordered from the highest to lowest values). If one decide that a transformed data that represents 80% of the original one is enough for the addressed problem, it is sufficient to consider the first four principal components, since

$$\frac{0.95 + 0.80 + 0.75 + 0.60}{0.95 + 0.80 + 0.75 + 0.60 + 0.30 + 0.25 + 0.10} \approx 83\%. \tag{3.10}$$

However, in some situations, one also may want to use all principal components in order to transform the original data into uncorrelated variables. This case is considered in a contribution of this study (see Chapter 5). In such MCDM problems, PCA can be used to obtain relevant information about the redundancies in the observed decision matrix. In the next section, we present another method used in feature extraction, which is also related to PCA, called independent component analysis.

3.2 Independent component analysis (ICA)

In the foregoing section, we presented an approach used to decorrelate a dataset. Based on second-order statistics (covariance, in this case), PCA projects the dataset on a set of uncorrelated variables. However, in some situations, one may be interested in a projection in which the variables are as independent as possible. This can be achieved with the application of a well-known technique in blind source separation (BSS) (Comon and Jutten, 2010), called independent component analysis (Comon, 1994). In this section, we present the BSS problem and how to deal with it through independent component analysis. Moreover, we provide an overview of the ICA techniques considered in this thesis.

3.2.1 The blind source separation problem

The origins of BSS can be associated with the work of Hérault, Jutten and Ans (Hérault et al., 1985) in the context of a neural modeling problem of motion coding[3]. In summary, the blind source separation problem consists in estimating a set of source signals through an observed mixture of these signals. The problem is called blind since both mixing parameters and source signals are unknown. Therefore, hypotheses about the latter must be assumed in order to estimate such signals.

Several situations in signal processing are modeled as BSS problems (Comon and Jutten, 2010). A typical one is the *cocktail party problem* (Cherry, 1953; Haykin and Chen, 2005), illustrated in Figure 7. In this situation, we have a group of people that are speaking simultaneously in a room. In the same room, there are a set of microphones that are recording all voices (and other possible noises). Since the signals recorded by these microphones comprise a mixture of people voices (and noises), the goal in this problem is to separate the sources without the knowledge of the original signals and how they were propagated until be recorded.

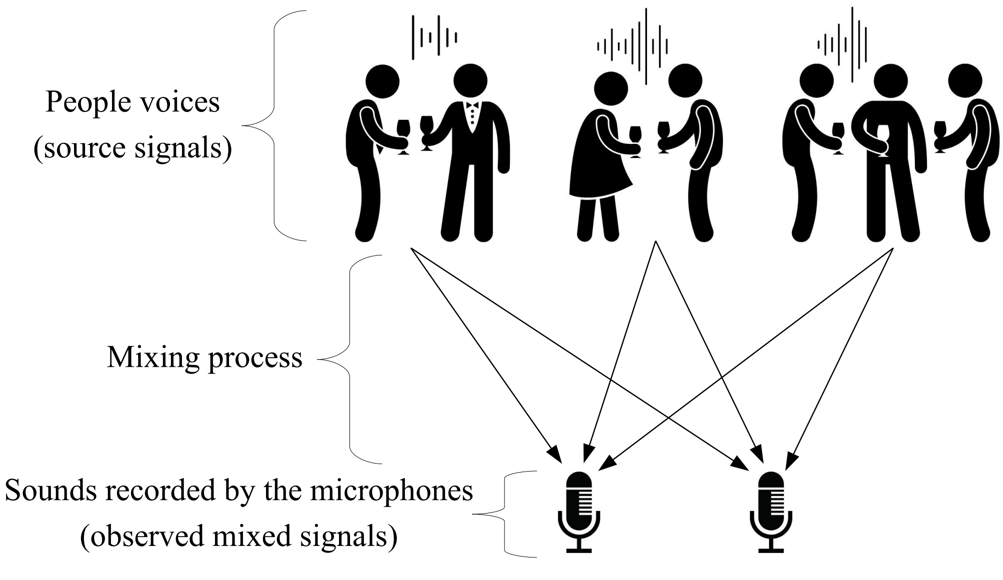


Figure 7 – The cocktail party problem.

Intuitively, people are able to hear several different sounds from several different sources and “separate” them. In fact, we are able to focus on a specific sound (a dog barking, for example) and, through our nervous and auditory systems, “extract” it from the others. A similar mechanism is desirable in several audio signal processing problems (Comon and Jutten, 2010), whose goal is to develop an automatic method to

3 For a complete history of blind source separation, see (Jutten and Taleb, 2000).

separate audio signals. By extending this idea to other applications in signal processing, several works in the literature highlight the interest in the development of automatic systems that are able to deal with blind source separation problems. For instance, one may cite applications in biomedical signals (Rieta et al., 2004), chemical analysis (Duarte et al., 2014) and seismic data (Batany et al., 2016). In the next section, we provide a mathematical formulation of the BSS problem.

3.2.1.1 Mathematical formulation

As mentioned in the previous section, the goal in BSS is to retrieve a set of source signals based on the mixtures and statistics characteristics of the sources (e.g., independence). In this section, we provide a mathematical formulation for such a problem. For instance, consider that $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_h(t)]$ represents the set of h sources, each one with n observations ($t = 1, \ldots, n$). Moreover, suppose that the set of m observed signals $\mathbf{q}(t) = [q_1(t), q_2(t), \ldots, q_m(t)]$ are obtained through the unknown instantaneous linear mixing process[4] defined by

$$\mathbf{q}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{g}(t), \tag{3.11}$$

where $\mathbf{A} \in \mathbb{R}^{m \times h}$ and $\mathbf{g}(t) = [g_1(t), g_2(t), \ldots, g_m(t)]$ represent the mixing matrix and the additive white Gaussian noise (AWGN), respectively. The goal, therefore, is to adjust a separation matrix $\mathbf{B} \in \mathbb{R}^{h \times m}$ such that the estimated signals $\hat{\mathbf{s}}(t) = [\hat{s}_1(t), \hat{s}_2(t), \ldots, \hat{s}_h(t)]$, given by

$$\hat{\mathbf{s}}(t) = \mathbf{B}\mathbf{q}(t), \tag{3.12}$$

are as close as possible to $\mathbf{s}(t)$. One may note in Equations (3.11) and (3.12) that, ideally, one expects that the adjusted separation matrix $\mathbf{B}$ be the inverse of the mixing matrix $\mathbf{A}$, i.e., $\mathbf{B} = \mathbf{A}^{-1}$. It is worth mentioning that, in this study, we consider the determined case, i.e., the situation in which the number of mixtures is equal to the number of source signals. Therefore, both number of sources and number of mixtures will be represented by m.

However, although one may achieve a good separation of the considered signals, there are possible scaling and/or permutation ambiguities involved in this process (Comon and Jutten, 2010). In order to illustrate a BSS problem and these ambiguities, consider the scenario presented in Figure 8. Two source signals are mixed by a mixing process $\mathbf{A}$. After finding $\mathbf{B}$, we achieve the estimates $\hat{\mathbf{s}}(t) = [\hat{s}_1(t), \hat{s}_2(t)]$. Although the separation process were well estimated, due to the permutation and scaling ambiguities, the second retrieved signal is an estimation of the unknown first source with a gain of scale η, i.e.,

[4] In the contributions of this thesis (see Chapters 4 and 5), we consider an instantaneous linear mixing process. However, it is worth mentioning that the classical cocktail party problem presented in the previous section is generally modeled as a convolutive mixture. Therefore, instead of simple scalar weights, the mixing process is represented by filters in order to model the propagation delay.

$\hat{s}_2(t) \approx \eta s_1(t)$. Furthermore, the first retrieved signal is an estimation for the inverse of the unknown second source, i.e., $\hat{s}_1(t) \approx -s_2(t)$.

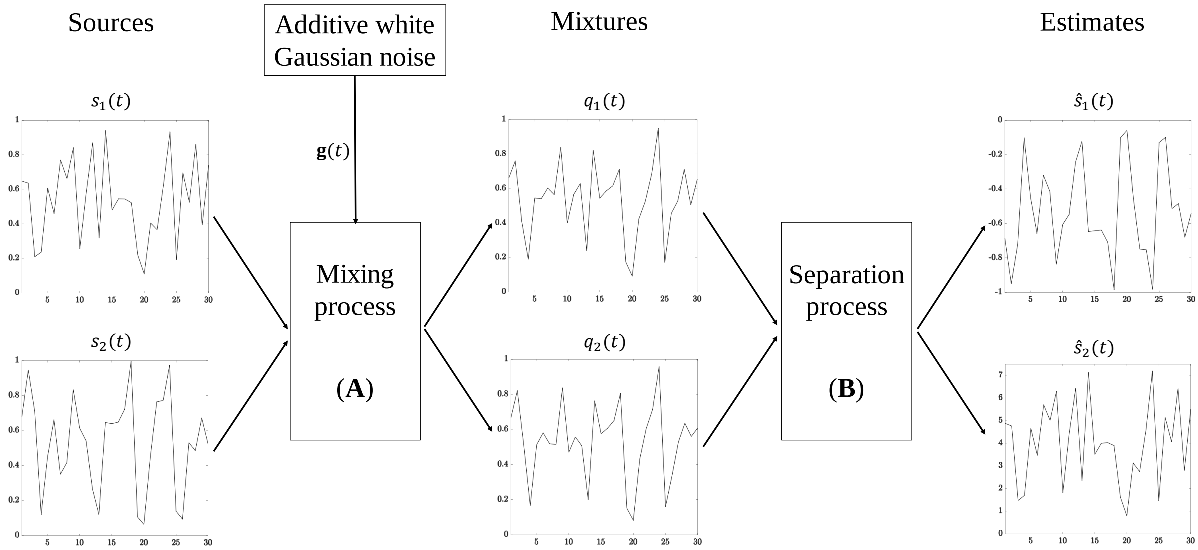


Figure 8 – Illustrative example of a BSS problem.

So far, we presented what is a BSS problem. In the literature, one may find several approaches used to deal with it. In this study, we considered an approach whose aim is to recover independent variables. We discuss this method in the next section.

3.2.2 Dealing with BSS problems through ICA

One of the first studies whose aim was to deal with BSS problems based on statistical independence among the source signals was conducted by Jutten (1987) in the end of the 80s. A few years latter, this approach was formalized by the works of Comon (1992, 1994).

In summary, the independent component analysis methods assume that the source signals $\mathbf{s}(t)$ are statistically independent and identically distributed (iid) and non-Gaussian. Given the mixing process described in (3.11), the observed signals $\mathbf{q}(t)$ are, consequently, dependent and closer to a Gaussian distribution. The variables in $\mathbf{q}(t)$ are dependent since each $q_i(t)$ comprises a linear combination of the source signals. Moreover, according to the central limit theorem (Papoulis and Pillai, 2002), this linear combination also leads to a distribution that is closer to a Gaussian one. Therefore, a strategy used to estimate the signal sources is to adjust a separation matrix $\mathbf{B}$ such that the retrieved signals $\hat{\mathbf{s}}(t)$ are as independent (or less Gaussian) as possible. For this purpose, one formulates an optimization problem in which the cost function $\mathcal{J}(\cdot)$ (usually called contrast function) is associated with a measure of statistical independence of $\hat{\mathbf{s}}(t) = \mathbf{Bq}(t)$. Mathematically, our aim is to deal with the following problem:

$$\max_{\mathbf{B}} \quad \mathcal{J}\left(\mathbf{Bq}(t)\right). \tag{3.13}$$

A natural candidate for $\mathcal{J}(\cdot)$ is the negative of the mutual information of $\hat{\mathbf{s}}$, which is defined as the Kullback-Leibler divergence between the joint and the product of marginal distributions of $\hat{\mathbf{s}}$ (Cover and Thomas, 1991). It is given by

$$I(\hat{\mathbf{s}}) = \int p(\hat{\mathbf{s}}) \log \left(\frac{p(\hat{\mathbf{s}})}{\prod_{i=1}^{m} p(\hat{s}_i)} \right) d\hat{\mathbf{s}}, \tag{3.14}$$

where $p(\hat{\mathbf{s}})$ and $p(\hat{s}_i)$ represent the joint distribution of $\hat{\mathbf{s}}$ and the marginal distribution of $p(\hat{s}_i)$, respectively, and $I(\hat{\mathbf{s}}) \geq 0$. Remark that, if $\hat{\mathbf{s}}$ has mutually independent components (i.e., $p(\hat{\mathbf{s}}) = \prod_{i=1}^{m} p(\hat{s}_i)$), $I(\hat{\mathbf{s}}) = 0$. Therefore, one may deal with BSS problem through ICA by solving the optimization problem (3.13) with the contrast function $\mathcal{J}(\cdot) = -I(\cdot)$ (Comon, 1994).

Note that $I(\cdot)$ requires the estimation of probability density functions, which is a difficult task in practical situations. In order to overcome this inconvenience, some ICA techniques consider optimization models based on higher-order statistics. Section 3.2.2.2.1 describes an algorithm that exploits such an idea. Moreover, in Section 3.2.2.2.2, we also present an algorithm which is based on the maximization of a non-Gaussianity measure of $\hat{s}_i$.

An important aspect in ICA techniques that we have already mentioned in the previous section is that one may face scaling and/or permutation ambiguities. By assuming that the signal sources are independent, it is easy to verify that, for any permutation matrix $\boldsymbol{\Theta}$ and diagonal matrix $\mathbf{D}$ such that $\mathbf{BA} = \boldsymbol{\Theta}\mathbf{D}$, the estimated signals are also independent. In fact, permutation and/or scaling ambiguities do not interfere in the statistical independence property and, therefore, such inconveniences are inherent to the ICA techniques. In our experiments (see Chapter 4), we attempt to mitigate these ambiguities by assuming some hypotheses about the decision problem.

3.2.2.1 PCA and ICA

There are several algorithms in the literature that can be used in ICA. Some of them, which will be considered in this study, are presented in the next section. A common point among them is that a preprocessing step, called *whitening* or *sphering*, is generally used to prepare the data to the application of an ICA algorithm (Hyvärinen et al., 2001). The aim of this step is to find a whitening matrix $\bar{\mathbf{P}}$ such that the transformed variables

$$\bar{\mathbf{z}}(t) = \bar{\mathbf{P}}\mathbf{q}(t) \tag{3.15}$$

are "white", i.e., the vectors $\bar{z}_i(t)$, $i = 1, \ldots, m$, are uncorrelated and with unit variances.

One may note that the transformation (3.15) is similar to PCA (see Equation (3.7)). For instance, in both cases, the obtained variables are uncorrelated. The difference is that, in the whitening step, we must also ensure that they have unit variances. Since in PCA one only rotates the dataset, one does not guarantee this property.

However, what is interesting here is that one only needs to perform a simple modification in the projection matrix used in PCA to achieve white data. For instance, instead of multiplying the observed signals $\mathbf{q}(t)$ in the left by $\mathbf{P} = \mathbf{E}^T$, where $\mathbf{E}$ is the matrix whose columns are the eigenvectors of $\mathbf{\Sigma}_{\mathbf{q}(t)}$, we simply multiply by $\bar{\mathbf{P}} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T$, where $\mathbf{\Lambda}$ is the diagonal matrix whose elements are the eigenvalues of $\mathbf{\Sigma}_{\mathbf{q}(t)}$. In this case, it is easy to verify that the vectors in $\bar{\mathbf{z}}(t) = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{q}(t)$ are uncorrelated and with unit variances (recall that $\mathbf{\Sigma}_{\mathbf{q}(t)} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$):

$$\mathbf{\Sigma}_{\bar{\mathbf{z}}(t)} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{\Sigma}_{\mathbf{q}(t)}\mathbf{E}\mathbf{\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T\mathbf{E}\mathbf{\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{I}\mathbf{\Lambda}\mathbf{I}\mathbf{\Lambda}^{-1/2} = \mathbf{\Lambda}^0 = \mathbf{I}. \tag{3.16}$$

Given the white data $\bar{\mathbf{z}}(t)$, the ICA algorithm will search for a rotation matrix $\mathbf{V}$ such that the variables in $\hat{\mathbf{s}}(t) = \mathbf{B}\mathbf{q}(t) = \mathbf{V}\bar{\mathbf{P}}\mathbf{q}(t) = \mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{q}(t)$ are as independent as possible. It is important to highlight that neither PCA nor whitening are enough to perform source separation when the sources are assumed to be independent. Since the concept of statistical dependence is stronger than correlation, one may have dependencies between variables even if they are uncorrelated. For instance, in the Figure 6b presented in Section 3.1, one may clearly see that, for a value of the first projection close to one, the associated value for the second projection tends to be close to zero. Therefore, there is a dependence between these two variables.

An interesting interpretation of PCA and ICA (see (Shlens, 2014b) for further details) is presented in Figure 9. Since mixing matrix $\mathbf{A}$ can rotate and stretch the source signals $\mathbf{s}(t)$, it can be decomposed as $\mathbf{A} = \bar{\mathbf{E}}\bar{\mathbf{\Lambda}}\bar{\mathbf{V}}^T$, where $\bar{\mathbf{E}}$ is a rotation matrix, $\bar{\mathbf{\Lambda}}$ is a diagonal matrix (which stretches the data) and $\bar{\mathbf{V}}^T$ is another rotation matrix[5]. Therefore, PCA can be used to obtain the first rotation matrix $\mathbf{E}^T$, which leads to uncorrelated variables. In whitening step, $\bar{\mathbf{P}} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T$ will also stretch the data, which leads to uncorrelated variables with unit variances. Finally, the ICA algorithm will adjust the matrix $\mathbf{V}$ that leads to the estimates $\hat{\mathbf{s}}(t)$.

An important remark in whitening step is that $\bar{\mathbf{P}} = \mathbf{\Lambda}^{-1/2}\mathbf{E}^T$ is not the unique matrix that can be used. Another whitening matrix frequently adopted in the literature is the one called the inverse root of $\mathbf{\Sigma}_{\mathbf{q}(t)}$ (or, simply, $\mathbf{\Sigma}_{\mathbf{q}(t)}^{-1/2}$), defined by

$$\bar{\mathbf{P}} = \mathbf{E}\mathbf{\Lambda}^{-1/2}\mathbf{E}^T. \tag{3.17}$$

The idea in such a matrix is that it returns the rotation provided by $\mathbf{E}^T$ in PCA. Moreover, one may also verify that the transformed data

$$\bar{\mathbf{z}}(t) = \mathbf{E}\mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{q}(t) \tag{3.18}$$

is white, since

$$\mathbf{\Sigma}_{\bar{\mathbf{z}}(t)} = \mathbf{E}\mathbf{\Lambda}^{-1/2}\mathbf{E}^T\mathbf{\Sigma}_{\mathbf{q}(t)}\mathbf{E}\mathbf{\Lambda}^{-1/2}\mathbf{E}^T = \mathbf{E}\mathbf{I}\mathbf{E}^T = \mathbf{I}. \tag{3.19}$$

[5] This comprises a singular value decomposition of matrix $\mathbf{A}$, which is possible by assuming that this matrix is invertible (Shlens, 2014b).

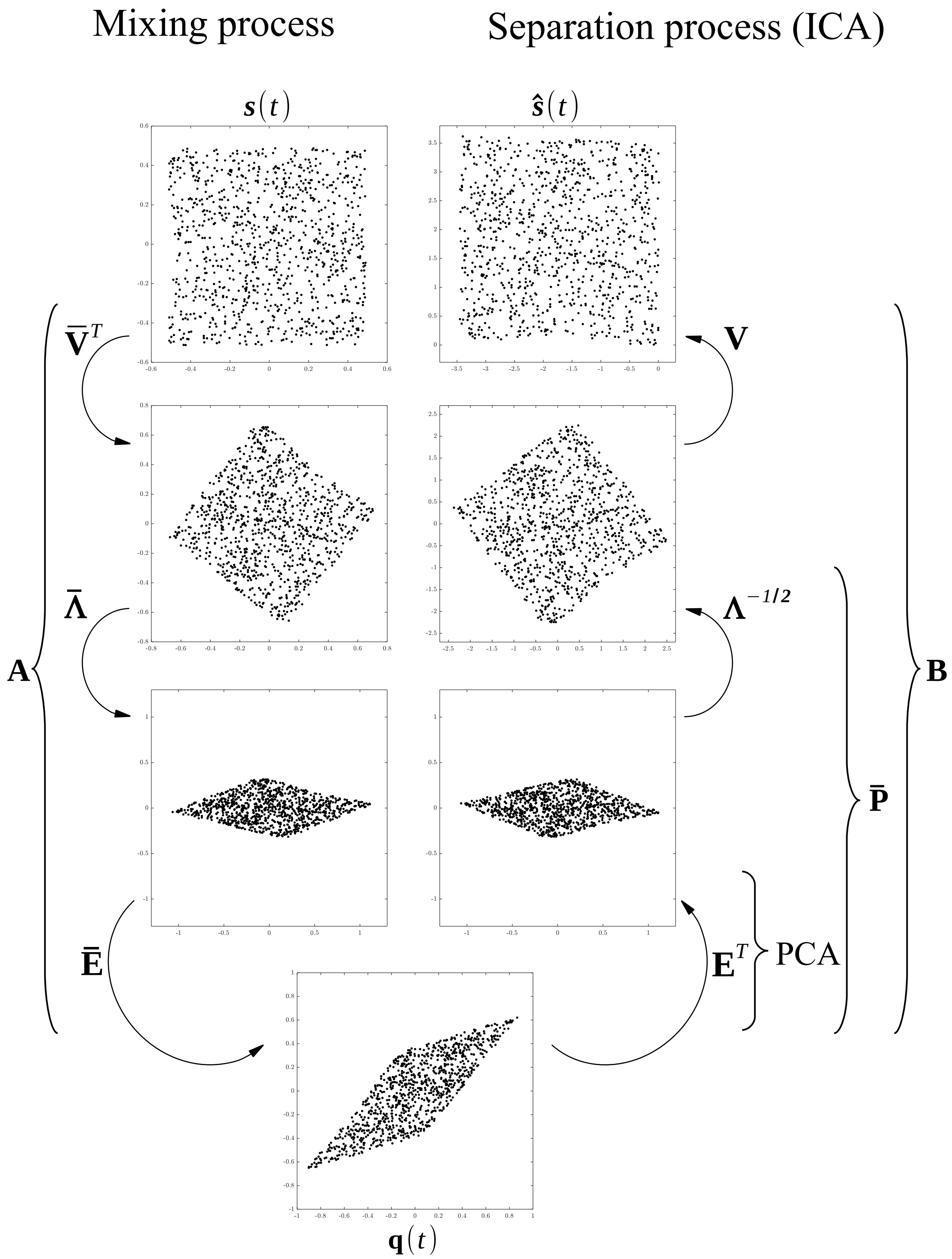


Figure 9 – An interpretation of PCA and ICA (adapted from (Shlens, 2014b)).

In our experiments, we also adopted the $\bar{\mathbf{P}} = \mathbf{E}\mathbf{\Lambda}^{-1/2}\mathbf{E}^T$ as the whitening matrix.

3.2.2.2 ICA techniques

The last section described the whitening step, which is generally used by several ICA algorithms. In this section, we present an overview of two algorithms used to find the matrix $\mathbf{V}$ and, consequently, the separation matrix $\mathbf{B}$. The main idea of such algorithms are described in the sequel. For simplicity of notation, we will refer to $\mathbf{z}(t)$ as $\mathbf{z}$ (similarly for the source signals and estimates).

3.2.2.2.1 JADE

The first algorithm presented in this study, known as JADE (Joint Approximate Diagonalization of Eigenmatrices), was proposed by Cardoso and Souloumiac (1993). Similarly as in PCA, the idea of the algorithm JADE also consists in the diagonalization of a set of matrices. However, instead of considering second-order statistics (covariance, in PCA), this technique takes into account higher-order statistics in the separation matrix estimation. Precisely, JADE consists in adjusting a rotation matrix $\hat{\mathbf{V}}$ that diagonalizes a set of cumulant matrices $Q^{\mathbf{z}}\left(\mathbf{N}^i\right)$, where $\mathbf{N}^i$ is a $m \times m$ matrix (which will be defined later). Each element $\alpha\beta$ in $Q^{\mathbf{z}}\left(\mathbf{N}^i\right)$ is given by

$$\left[Q^{\mathbf{z}}\left(\mathbf{N}^i\right)\right]_{\alpha\beta} = \sum_{\gamma,\delta=1}^{m} \operatorname{cum}\left(z_\alpha, z_\beta, z_\gamma, z_\delta\right) N^i_{\gamma\delta}, \tag{3.20}$$

where

$$\operatorname{cum}\left(z_\alpha, z_\beta, z_\gamma, z_\delta\right) = \mathbb{E}\left[z_\alpha z_\beta z_\gamma z_\delta\right] - \mathbb{E}\left[z_\alpha z_\beta\right]\mathbb{E}\left[z_\gamma z_\delta\right] - \mathbb{E}\left[z_\alpha z_\gamma\right]\mathbb{E}\left[z_\beta z_\delta\right] - \mathbb{E}\left[z_\alpha z_\delta\right]\mathbb{E}\left[z_\beta z_\gamma\right] \tag{3.21}$$

is the fourth-order joint cumulant of $z_\alpha, z_\beta, z_\gamma, z_\delta$, $\forall \alpha, \beta, \gamma, \delta \in \{1, \ldots, m\}$, and $N^i_{\gamma\delta}$ is the element of the γ-th row and δ-th column of $\mathbf{N}^i$. Based on these matrices, the optimization problem to be solved is the following:

$$\min_{\hat{\mathbf{V}}} \quad \sum_i \operatorname{off}\left(\hat{\mathbf{V}}^T Q^{\mathbf{z}}\left(\mathbf{N}^i\right)\hat{\mathbf{V}}\right), \tag{3.22}$$

where the operator off$(\cdot)$ calculates the quadratic sum of the elements that are not in the diagonal. Therefore, since the diagonalization expressed in (3.22) takes into account higher-order statistics, the adjusted $\hat{\mathbf{V}}$ should lead to independent variables $\hat{\mathbf{s}} = \hat{\mathbf{V}}\mathbf{z}$.

$Q^{\mathbf{z}}\left(\mathbf{N}^i\right)$ can be interpreted as the result of the application of the cumulant tensor of $\mathbf{z}$ on $\mathbf{N}^i$. Therefore, natural candidates for $\mathbf{N}^i$ that are considered in the literature are the eigenmatrices of $Q^{\mathbf{z}}\left(\mathbf{N}^i\right)$, i.e., the matrices $\mathbf{N}^i$, $i = 1, \ldots, m$, such that $Q^{\mathbf{z}}\left(\mathbf{N}^i\right) = \lambda_i \mathbf{N}^i$. For further details, see (Cardoso and Souloumiac, 1993).

3.2.2.2.2 FastICA

The previous section presented the JADE algorithm, whose aim is to find a set of estimated signals as independent as possible. Given the independence measure considered in the optimization problem (3.22), its resolution leads to all retrieved signals $\hat{\mathbf{s}}$ simultaneously. In this section, we address an approach that takes into account the maximization of a non-Gaussianity measure of each retrieved signal $\hat{s}_i$. Since this measure is associated with a single signal, i.e., it does not depend on the relation between two or more variables, the optimization model can be iteratively conducted to extract signals that are as non-Gaussian as possible. For instance, we describe two different optimization models used in FastICA algorithm, developed by Hyvärinen (1997, 1999). Each one is associated with different measures of non-Gaussianity.

The first measure presented in this section is the fourth-order cumulant of a variable $\hat{s}_i$, known as kurtosis. In Equation (3.21), we expressed this cumulant for a set of variables, however, for a single variable $\hat{s}_i$ (with zero mean), kurtosis can be defined by

$$\mathrm{kurt}\left(\hat{s}_i\right) = \mathbb{E}\left[\hat{s}_i^4\right] - 3\left(\mathbb{E}\left[\hat{s}_i^2\right]\right)^2. \tag{3.23}$$

Moreover, if one assumes that $\hat{s}_i = \hat{\mathbf{v}}\bar{\mathbf{z}}$, where $\hat{\mathbf{v}}$ is a unit vector (a row of the rotation matrix $\hat{\mathbf{V}}$), and $\bar{\mathbf{z}}$ is white, $\mathbb{E}\left[\hat{s}_i^2\right] = \mathbb{E}\left[\hat{\mathbf{v}}\bar{\mathbf{z}}\bar{\mathbf{z}}^T\hat{\mathbf{v}}^T\right] = 1$. Therefore, Equation (3.23) reduces to $\mathrm{kurt}\left(\hat{s}_i\right) = \mathbb{E}\left[\hat{s}_i^4\right] - 3$, i.e., a normalized version of the fourth moment of $\hat{s}_i$.

The relevant aspect of kurtosis in ICA is that it is equal to zero for Gaussian random variables and non-zero for most of the other distributions (Hyvärinen et al., 2001). For instance, kurtosis is negative for sub-Gaussian variables (e.g., the uniform distribution) and positive for super-Gaussian variables (e.g., the Laplacian distribution). For such a reason, this measure can be used in ICA in the sense that, if we search for a signal $\hat{s}_i$ far from being Gaussian, we should adjust a vector $\hat{\mathbf{v}}$ that maximizes the kurtosis (in absolute value) of $\hat{s}_i = \hat{\mathbf{v}}\bar{\mathbf{z}}$. The optimization model is given by

$$\begin{aligned} \max_{\hat{\mathbf{v}}} \quad & \left|\mathrm{kurt}\left(\hat{\mathbf{v}}\bar{\mathbf{z}}\right)\right| \\ \text{s.t.} \quad & \hat{\mathbf{v}}\hat{\mathbf{v}}^T = 1. \end{aligned} \tag{3.24}$$

Other than the kurtosis, one may also adopt another measure of non-Gaussianity, called negentropy. Negentropy is associated with the entropy, a well-known concept in information theory (Shannon, 1948; MacKay, 2007). Although we have already mentioned this concept in Section 2.5.2.1, a formal mathematical expression of the differential entropy of a random vector $\hat{s}_i$ is provided as follows (Hyvärinen et al., 2001):

$$H\left(\hat{s}_i\right) = -\int p_{\hat{s}_i}\left(\xi\right)\log p_{\hat{s}_i}\left(\xi\right)\mathrm{d}\xi, \tag{3.25}$$

where $p_{\hat{s}_i}\left(\xi\right)$ is the probability density of $\hat{s}_i$. Based on $H\left(\hat{s}_i\right)$, the negentropy is defined by

$$J\left(\hat{s}_i\right) = H\left(\hat{s}_i^{gauss}\right) - H\left(\hat{s}_i\right), \tag{3.26}$$

where $\hat{s}_i^{gauss}$ is a Gaussian random variable whose variance is the same as in $\hat{s}_i$.

It is known that the Gaussian distribution has the largest entropy among all other distributions with the same variance. As a consequence, Equation (3.26) is either equal to zero, if the $\hat{s}_i$ follows a Gaussian distribution, or greater than zero, otherwise.

Therefore, in order to extract a signal far from being Gaussian, one should maximize the negentropy. By restricting $\hat{\mathbf{v}}$ to a unit vector, the optimization problem is given by

$$\begin{aligned} \max_{\hat{\mathbf{v}}} \quad & H\left(\hat{s}_i^{gauss}\right) - H\left(\hat{\mathbf{v}}\bar{\mathbf{z}}\right) \\ \text{s.t.} \quad & \hat{\mathbf{v}}\hat{\mathbf{v}}^T = 1. \end{aligned} \tag{3.27}$$

Both aforementioned measures can be used in the FastICA algorithm. It is important to remark that they lead to a single estimate, which is as non-Gaussian as possible. Therefore, in order to retrieve all source signals, one may adopt either an iterative approach (known as deflationary orthogonalization) or a technique that estimates them simultaneously (known as symmetric orthogonalization). For a further description of these orthogonalization methods as well as the optimization issues associated with FastICA, see the book of Hyvärinen et al. (2001), which was also used to write this section.

3.3 Sensitivity analysis

The third latent variable analysis method considered in this study is the sensitivity analysis (Saltelli et al., 2004, 2008). In short, this method consists of the analysis of the uncertainty on a model output according to the uncertainties associated with the model inputs. Therefore, with such an approach, one may measure the impact that (or the importance of) each input (or a set of inputs) has on the output model. For this purpose, it is usual to carry out a decomposition of the model into terms associated with different input variables Z_i, $i = 1, \ldots, m$. For instance, consider the high-dimensional model representation (HDMR) (Li et al., 2001) of $Y = f(Z_1, Z_2, \ldots, Z_m)$, given by

$$Y = f_\emptyset + \sum_{i=1}^{m} f_i(Z_i) + \sum_{i<i'}^{m} f_{i,i'}(Z_i, Z_{i'}) + \ldots + f_C(Z_C), \tag{3.28}$$

where $f_\emptyset, f_i, f_{i,i'}, \ldots, f_C, \forall i, i' \subseteq C$, represent terms of increasing dimensions. A frequently used candidate for function f is the following[6]:

$$f_\emptyset = \mathbb{E}\left[Y\right] \tag{3.29}$$

$$f_i(Z_i) = \mathbb{E}\left[Y|Z_i\right] - f_\emptyset \tag{3.30}$$

$$f_{i,i'}(Z_i, Z_{i'}) = \mathbb{E}\left[Y|Z_i, Z_{i'}\right] - f_i(Z_i) - f_i(Z_i) - f_\emptyset \tag{3.31}$$

$$\vdots$$

$$f_A(Z_A) = \sum_{B\subseteq A} (-1)^{|A\setminus B|}\mathbb{E}\left[Y|Z_B\right] \tag{3.32}$$

$$\vdots$$

$$f_C(Z_C) = \sum_{B\subseteq C} (-1)^{|C\setminus B|}\mathbb{E}\left[Y|Z_B\right], \tag{3.33}$$

[6] Conditions about the uniqueness of this decomposition can be seen in (Saltelli et al., 2008).

where $\mathbb{E}[Y|Z_B]$ denotes the conditional expectation of Y given the variables Z_i, $\forall i \in B$. Given this decomposition, we may estimate the importance of inputs on the output model. In the next section, we describe the approach considered in our study.

3.3.1 Variance-based approach and the Sobol' indices

An approach commonly used in sensitivity analysis is the variance-based one (Saltelli et al., 2004, 2008). When all variables are free to assume any value, the uncertainty in the output model can be simply calculated from the unconditioned variance $\mathrm{Var}[Y]$. However, if one would estimate the uncertainty of Y associated with a specific variable Z_i, one may fix a value z_i^* of Z_i (i.e., $Z_i = z_i^*$) and calculate the conditional variance $\mathrm{Var}[Y|Z_i = z_i^*]$. For instance, if Z_i has a large impact on the output model, by fixing $Z_i = z_i^*$, one achieves a small value of $\mathrm{Var}[Y|Z_i = z_i^*]$. Therefore, smaller values of $\mathrm{Var}[Y|Z_i = z_i^*]$ indicates larger impacts of Z_i on Y.

However, by fixing $Z_i = z_i^*$, the interpretation of the importance of Z_i will be dependent on the value z_i^*, which is questionable. Therefore, aiming at avoiding incorrect interpretations, one may consider an average value over all possible $z_i \in Z_i$, i.e., $\mathbb{E}[\mathrm{Var}[Y|Z_i]]$. By using this expectation, we have the following properties (Saltelli et al., 2008):

- $\mathbb{E}[\mathrm{Var}[Y|Z_i]] \leq \mathrm{Var}[Y]$ and
- $\mathbb{E}[\mathrm{Var}[Y|Z_i]] + \mathrm{Var}[\mathbb{E}[Y|Z_i]] = \mathrm{Var}[Y]$.

From the above properties, one may also note that $\mathrm{Var}[\mathbb{E}[Y|Z_i]] \leq \mathrm{Var}[Y]$. Moreover, by using the same interpretation as before, larger values of $\mathrm{Var}[\mathbb{E}[Y|Z_i]]$ indicate larger impacts of Z_i on Y. Therefore, $\mathrm{Var}[\mathbb{E}[Y|Z_i]]$ can be used as a measure of the importance of Z_i. By dividing this measure by $\mathrm{Var}[Y]$, one achieves the so-called first-order Sobol' sensitivity index (Sobol, 1990), given by

$$S_i = \frac{\mathrm{Var}[\mathbb{E}[Y|Z_i]]}{\mathrm{Var}[Y]}. \tag{3.34}$$

Other than the first-order Sobol' sensitivity index, one can extend the aforementioned conclusions for higher-order terms. Therefore, for any coalition $A \subseteq C$,

$$S_A = \frac{\mathrm{Var}[f_A(Z_A)]}{\mathrm{Var}[Y]} \tag{3.35}$$

represents a measure of the importance (or the "impact") that the coalition of variables Z_A have in the model output Y.

In the Sobol' indices calculation, we assume that the input variables are independent. In this case, the conditional variances can be decomposed in the ANOVA-HDMR decomposition[7] (Li et al., 2010), given by

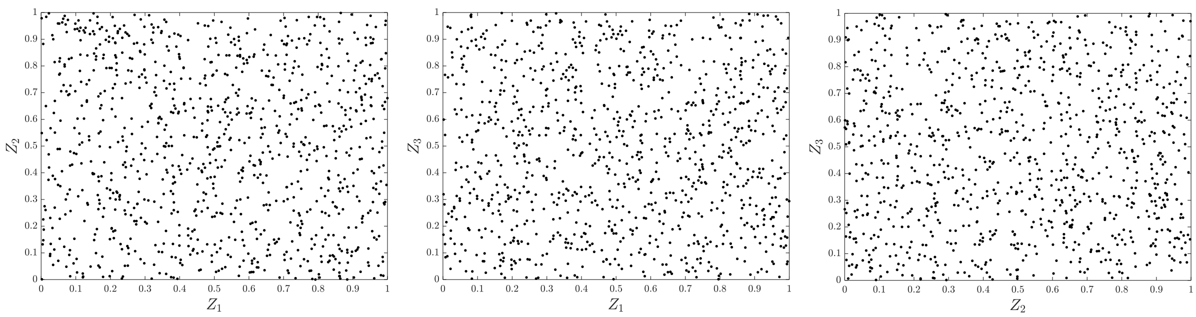
$$\mathrm{Var}\,[Y] = \sum_i \mathrm{Var}_i + \sum_i \sum_{i'>i} \mathrm{Var}_{ii'} + \ldots + \mathrm{Var}_{12\ldots m}, \tag{3.36}$$

where $\mathrm{Var}_A = \mathrm{Var}\,[f_A(Z_A)]$. Moreover, by dividing both sides by $\mathrm{Var}\,[Y]$, one achieves

$$1 = \sum_i S_i + \sum_i \sum_{i'>i} S_{ii'} + \ldots + S_{12\ldots m}. \tag{3.37}$$

3.3.2 An illustrative example

In this section, we illustrate the Sobol's indices calculation through a simple example composed by $m = 3$ variables and $n = 1000$ samples (generated according to a uniform distribution on $[0, 1]$). Figure 10 presents the scatter plot among pairs of variables.



(a) Variables Z_1 and Z_2. (b) Variables Z_1 and Z_3. (c) Variables Z_2 and Z_3.

Figure 10 – Scatter plot of pairs of variables.

Suppose that the output model (generally unknown in practice) is given by

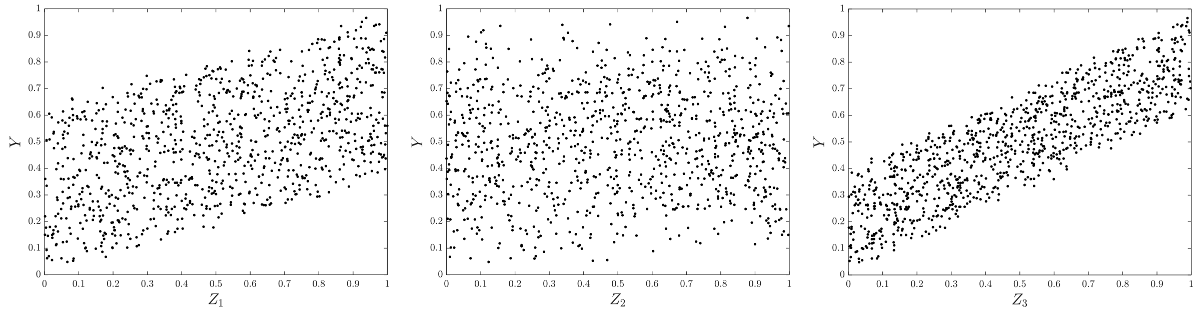
$$Y = 0.35Z_1 + 0.05Z_2 + 0.60Z_3. \tag{3.38}$$

Moreover, suppose that we want to calculate the importance of each variable individually. Figure 11 presents the output conditioned by each variable. One may see that there is a clear pattern in Figure 11c, since Z_3 is the variable that is multiplied by the largest weight in Equation (3.38). On the other hand, no pattern is found in Figure 11b, since Z_2 is multiplied by the smallest weight.

In order to calculate the Sobol' indices, firstly, we need to estimate the conditional expectations. For this purpose, as indicated in (Saltelli et al., 2008), one may estimate $\mathbb{E}\,[Y|Z_i]$ by cutting the Z_i domain into slices and calculating the average value for each one. By considering slices of size 0.01, we obtain the conditional expectations illustrated in Figure 12.

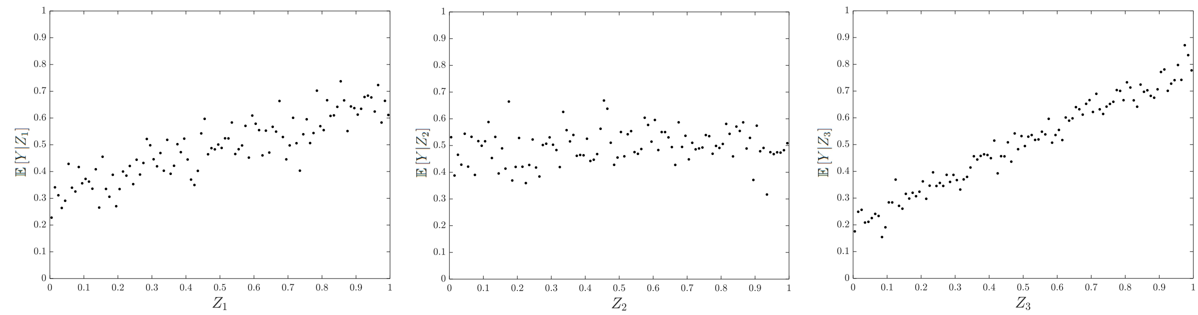
The first order Sobol' indices are obtained, therefore, by taking the variances of the conditional expectations and normalizing them by $\mathrm{Var}\,[Y]$. These values are given

[7] ANOVA stands for analysis of variance.



(a) Y given Z_1. (b) Y given Z_2. (c) Y given Z_3.

Figure 11 – Model output given each input variable.



(a) Conditioned on Z_1. (b) Conditioned on Z_2. (c) Conditioned on Z_3.

Figure 12 – Expected values.

by[8] $S_1 = 0.3553$, $S_2 = 0.1096$ and $S_3 = 0.8056$. As expected, Z_3 is the variable with the highest impact on the output model, followed by Z_1 and, then, Z_2. Note that this result is in accordance with the weights associated with each variable in Equation (3.38).

3.4 Conclusions

This chapter presented the latent variable analysis techniques that will be considered in some contributions of this thesis. In summary, they will be used to extract relevant information from the decision data in order to deal with MCDM situations. Moreover, most of these information will contribute to adjust the decisional parameters in scenarios with dependencies among criteria.

Principal component analysis was presented in Section 3.1. With this technique, we are able to transform the considered dataset into a set of uncorrelated variables. For this purpose, we adjust a transformation matrix that will perform the projection. In a contribution of this work, discussed in Chapter 5, the information contained in this

[8] It is important to remark that the sum of the Sobol's indices calculated in this example is greater than 1. This is due to the estimation of the conditional expectations. Since it is dependent on the data distribution as well as on the size of the considered slices, these elements may not be perfectly estimated. As a consequence, one may achieve results that are not in accordance with Equation (3.37).

matrix will be used to adjust the WAM parameters.

Besides the use of PCA in weights determination, we also consider the information acquired through ICA. This technique, presented in Section 3.2, aims at separating the observed data in order to achieve a set of vectors as independent as possible. Similarly as in PCA, ICA (in a linear mixture) is conducted through the estimation of a transformation matrix. In the contribution described in Chapter 5, we verify the application of ICA and the higher-order information extracted from the adjusted separation matrix in the WAM parameters determination. Moreover, we combine ICA and TOPSIS approaches to deal with dependencies among criteria. We present this contribution in Chapter 4.

The last LVA technique addressed in this chapter (see Section 3.3) was the sensitivity analysis. Such a technique can be used to verify the importance of each input variable (or a set of them) in the model output, without the knowledge of the exact form of this model. Therefore, it is also a method that extract information from the considered dataset. For instance, in a contribution conducted in Chapter 7, sensitivity analysis will be adopted to extract relevant information used in an unsupervised based approach for capacity identification.

Part II

Contributions

4 Application of ICA and TOPSIS to deal with dependent criteria

As pointed out in Section 2.4.1.3, although the meaningfulness of TOPSIS is questionable, this technique has been applied in several practical situations. Moreover, other than the classical TOPSIS, an improved version, called TOPSIS-M, has also been used to deal with decision problems whose criteria are statistically dependent. The main idea of such an approach is to tackle dependencies among criteria by calculating the closeness measure (or overall evaluations) by means of the Mahalanobis distance. However, an interesting remark is that, since the Mahalanobis distance takes into account the covariance matrix extracted from the decision data, i.e., a second-order statistic, this procedure may not be sufficient to capture the complexity behind statistical dependencies.

In view of the practical problems that have been dealt with by means of TOPSIS-based methods and motivated by the aforementioned observation, the aim of this chapter is to further investigate the application of TOPSIS-M to deal with dependent criteria. As will be discussed in Section 4.2, the distances calculated in TOPSIS-M can be seen as Euclidean ones in a transformed space in which the decision data are not correlated anymore. In other words, this approach aims at finding an alternative representation of the data in which the distances calculation should be conducted. However, the decorrelation procedure of the TOPSIS-M approach does not necessarily imply a representation in which the decision data are independent. Therefore, an approach that takes into account the independence among the decision data can better deal with dependent criteria in MCDM problems.

An interesting aspect in the above-mentioned scenario is that the procedure to find an alternative representation of the decision data can be seen as a task of recovering a set of independent latent (hidden) data from the observed criteria. As already discussed in Section 3.2, this can be achieved by means of a signal processing technique called independent component analysis. Since the latent criteria can be seen as the alternative representation of the decision data in which the criteria are independent, one can use the estimates provided by ICA as inputs of TOPSIS methods. In this context, the contribution of this chapter lies in two different approaches to deal with dependent criteria in MCDM problems. The first one, called ICA-TOPSIS and described in Section 4.4.1, comprises the application of an ICA technique to estimate the latent criteria and, thus, perform the TOPSIS on this retrieved data. The other one, called ICA-TOPSIS-M and presented in Section 4.4.2, also applies an ICA technique to estimate the latent criteria, but uses this retrieved data only to determine the positive and negative ideal alternatives, which will be used as inputs in a modified version of TOPSIS-M.

It is worth mentioning that the first conference paper pointed out in Chapter 1 presents preliminary results achieved on this subject. Moreover, the full analysis was published in the journal article 1.

4.1 Problem statement

Let us recall that, for simplicity of notation, we defined $\hat{\mathbf{M}} = \mathbf{M}^T$, where $\mathbf{M} = (u_{j,i})_{n\times m}$, as the decision matrix whose columns represent the alternatives and rows contain the criteria evaluations. Generally, a MCDM method is applied directly on this observed decision data. However, since we are interested in MCDM problems in which $\hat{\mathbf{M}}$ can be seen as a set of dependent criteria, we consider that an alternative representation of the decision data, in which the criteria are independent, should be used as input of TOPSIS method. The procedure of extracting this new representation from $\hat{\mathbf{M}}$ can be seen as a problem of retrieving independent latent factors from a set of mixtures of these factors.

Mathematically, consider that the independent latent data is defined by[1]

$$\mathbf{L} = \begin{bmatrix} l_{1,1} & l_{1,2} & \dots & l_{1,n} \\ l_{2,1} & l_{2,2} & \dots & l_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{m,1} & l_{m,2} & \dots & l_{m,n} \end{bmatrix}, \tag{4.1}$$

where $l_{i,j}$ represents the evaluation of alternative a_j with respect to the independent latent criterion $\mathcal{L}_i$. We define the set of independent latent criteria as $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_m\}$. Therefore, by considering a signal processing formulation, data $\hat{\mathbf{M}}$ can be modeled as

$$\hat{\mathbf{M}} = \mathbf{AL} + \mathbf{G}, \tag{4.2}$$

where $\mathbf{A} \in \mathbb{R}^{m\times m}$ is the mixing matrix (which provides the linear mixture of the independent data) and $\mathbf{G} = (g_{i,j}) \in \mathbb{R}^{m\times n}$ is the additive white Gaussian noise matrix (which includes randomness in the mixing process). In this scenario, each evaluation $u_{j,i}$ comprises a linear combination of the evaluations of alternative a_j with respect to the set of independent latent criteria $\mathcal{L}$.

By assuming that $\mathbf{L}$ is the alternative representation of the decision data that should be used to derive the ranking of alternatives, the application of TOPSIS directly on the dependent criteria $\hat{\mathbf{M}}$ may lead to biased results. Therefore, our main concern is how to adjust a transformation matrix $\mathbf{B}$ that provides the estimates

$$\hat{\mathbf{L}} = \mathbf{B}\hat{\mathbf{M}} \tag{4.3}$$

[1] Recall that our analysis assumes that the number of latent factors (or sources) is equal to the number of observed criteria (or mixtures).

for the independent data $\mathbf{L}$, i.e., the alternative representation of the decision criteria in which the data are independent. Figure 13 illustrates this scenario, in which the procedure used to adjust $\mathbf{B}$ can be seen as a preprocessing step. It is important to mention that, in this thesis, we deal with a linear dependence among criteria. However, our approach can be adapted in order to consider nonlinear relations.

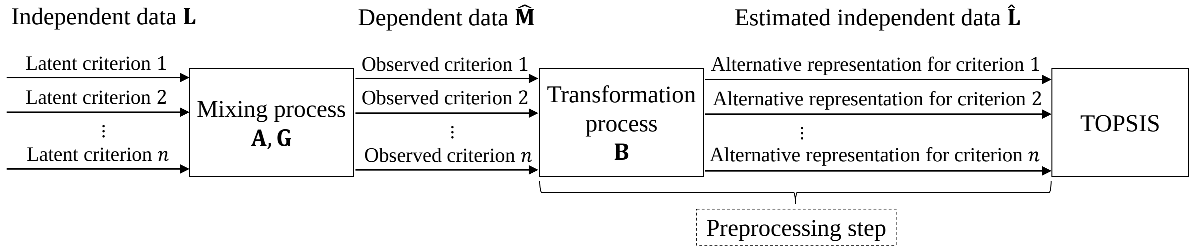


Figure 13 – Application of TOPSIS in an alternative representation of the decision data.

Since we search for an alternative representation of the observed decision criteria $\hat{\mathbf{M}}$, it is not straightforward to assign a meaning for this data. However, the rationale behind such a model is that the observed criteria can be seen as linear combinations of independent (hidden) criteria $\mathbf{L}$. Although the interpretation of these data can be difficult in some situations, there are decision problems in which one may associate each observed criteria with a specific latent factor. For example, if one considers the problem of ranking students (see Example 2.3 mentioned in Section 2.4.1.2.1), one may consider that the grades (dependent decision data $\hat{\mathbf{M}}$) may be represented as a set of estimated latent competences (independent data $\mathbf{L}$) in areas such as mathematics, natural sciences and linguistics. Thus, there may be a correlation between the grades in calculus and physics because both depend on scientific skills. Figure 14 illustrates this scenario.

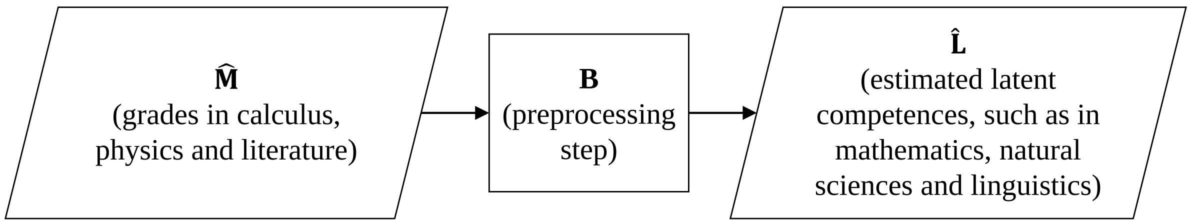


Figure 14 – Example of dependent decision criteria and estimated latent data.

4.2 Theoretical analysis about TOPSIS-M

The classical version of TOPSIS is typically applied in MCDM problems in which the decision criteria are independent. However, if there are correlations among the decision data, an improved version, called TOPSIS-M, could be used. The TOPSIS-M method, described in Algorithm 2.2, addresses the dependence among criteria by exploiting second-order statistics (covariance matrix) extracted from the decision data. An interesting aspect of TOPSIS-M method is that, since the Mahalonobis distance considers information about the covariance matrix of $\hat{\mathbf{M}}$, this method addresses the issue of dependent criteria by performing a decorrelation procedure. This result can be understood under the light of the following theorem:

Theorem 4.1. *Assuming the same importance for all observed criteria (i.e., $w_i = w_{i'}$ for all i and i') and that $\mathbf{\Sigma}_{\tilde{\mathbf{M}}}$ is a positive definite matrix, the Mahalanobis distance calculated in Algorithm 2.2 is equivalent to the Euclidean one in a transformed space in which the data are uncorrelated.*

Proof. Consider the Cholesky factorization (Golub and Van Loan, 2013) of $\mathbf{\Sigma}_{\tilde{\mathbf{M}}}$ given by

$$\mathbf{\Sigma}_{\tilde{\mathbf{M}}} = \mathbf{F}\mathbf{F}^T, \tag{4.4}$$

where $\mathbf{F}$ is an unique lower triangular matrix. In that respect, DM_j^+ may be written as

$$\begin{aligned} DM_j^+ &= \sqrt{\left(\tilde{\mathbf{u}}_j - \tilde{\mathbf{u}}^+\right)^T \Delta^T \left(\mathbf{F}\mathbf{F}^T\right)^{-1} \Delta \left(\tilde{\mathbf{u}}_j - \tilde{\mathbf{u}}^+\right)} \\ &= \sqrt{\left(\Delta\tilde{\mathbf{u}}_j - \Delta\tilde{\mathbf{u}}^+\right)^T \left(\mathbf{F}^T\right)^{-1} \mathbf{F}^{-1} \left(\Delta\tilde{\mathbf{u}}_j - \Delta\tilde{\mathbf{u}}^+\right)} \\ &= \sqrt{\left(\mathbf{F}^{-1}\Delta\tilde{\mathbf{u}}_j - \mathbf{F}^{-1}\Delta\tilde{\mathbf{u}}^+\right)^T \left(\mathbf{F}^{-1}\Delta\tilde{\mathbf{u}}_j - \mathbf{F}^{-1}\Delta\tilde{\mathbf{u}}^+\right)} \\ &= \sqrt{\left(\ddot{\mathbf{u}}_j - \ddot{\mathbf{u}}^+\right)^T \left(\ddot{\mathbf{u}}_j - \ddot{\mathbf{u}}^+\right)}, \quad j = 1, \ldots, n, \end{aligned}$$

where $\ddot{\mathbf{u}}_j = \mathbf{F}^{-1}\Delta\tilde{\mathbf{u}}_j$ and $\ddot{\mathbf{u}}^+ = \mathbf{F}^{-1}\Delta\tilde{\mathbf{u}}^+$. Therefore, DM_j^+ represents the Euclidean distance between the transformed data $\ddot{\mathbf{M}} = \left(\ddot{u}_{i,j}\right)_{m\times n}$ and the transformed positive ideal alternative $\ddot{\mathbf{u}}^+$. The same conclusion may be achieved considering DM_j^-.

In order to show that the transformed data $\ddot{\mathbf{M}}$ are uncorrelated, consider the covariance matrix

$$\begin{aligned} \mathbf{\Sigma}_{\ddot{\mathbf{M}}} &= \mathbb{E}\left[\left(\ddot{\mathbf{M}} - \mathbb{E}\left[\ddot{\mathbf{M}}\right]\right)\left(\ddot{\mathbf{M}} - \mathbb{E}\left[\ddot{\mathbf{M}}\right]\right)^T\right] \\ &= \mathbb{E}\left[\left(\mathbf{F}^{-1}\Delta\tilde{\mathbf{M}} - \mathbf{F}^{-1}\Delta\mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)\left(\mathbf{F}^{-1}\Delta\tilde{\mathbf{M}} - \mathbf{F}^{-1}\Delta\mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)^T\right] \\ &= \mathbb{E}\left[\mathbf{F}^{-1}\Delta\left(\tilde{\mathbf{M}} - \mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)\left(\tilde{\mathbf{M}} - \mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)^T \Delta^T(\mathbf{F}^{-1})^T\right] \\ &= \mathbf{F}^{-1}\Delta\mathbb{E}\left[\left(\tilde{\mathbf{M}} - \mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)\left(\tilde{\mathbf{M}} - \mathbb{E}\left[\tilde{\mathbf{M}}\right]\right)^T\right]\Delta^T(\mathbf{F}^{-1})^T \\ &= \mathbf{F}^{-1}\Delta\Sigma_{\tilde{\mathbf{M}}}\Delta^T(\mathbf{F}^{-1})^T \\ &= \mathbf{F}^{-1}w_i\mathbf{I}_m\Sigma_{\tilde{\mathbf{M}}}w_i^T\mathbf{I}_m^T(\mathbf{F}^{-1})^T \\ &= w_i^2\mathbf{F}^{-1}\Sigma_{\tilde{\mathbf{M}}}(\mathbf{F}^{-1})^T, \end{aligned}$$

where $\mathbf{I}_m$ is the $m \times m$ identity matrix. Since $\mathbf{\Sigma}_{\tilde{\mathbf{M}}} = \mathbf{F}\mathbf{F}^T$, then $\mathbf{F}^{-1}\Sigma_{\tilde{\mathbf{M}}}(\mathbf{F}^{-1})^T = \mathbf{I}_m$. Therefore, $\mathbf{\Sigma}_{\tilde{\tilde{\mathbf{M}}}} = w_i^2\mathbf{I}_m$, i.e., the transformed data $\mathbf{\Sigma}_{\ddot{\mathbf{M}}}$ are uncorrelated.

□

4.3 Graphical interpretation of TOPSIS approaches

In order to illustrate the problems that may arise by applying the TOPSIS methods in situations with correlated criteria, we provide in this section a graphical interpretation of both TOPSIS and TOPSIS-M. For instance, let us consider a MCDM problem with 200 alternatives and 2 observed dependent criteria, whose evaluations are generated according to Equation (4.2). We randomly generate the latent criteria based on a uniform distribution in the range $[0, 1]$ and apply a mixing matrix given by

$$\mathbf{A} = \begin{bmatrix} 1.00 & 0.70 \\ -0.25 & 1.00 \end{bmatrix}.$$

Moreover, we consider a weight vector given by $\mathbf{w} = [0.5, 0.5]$.

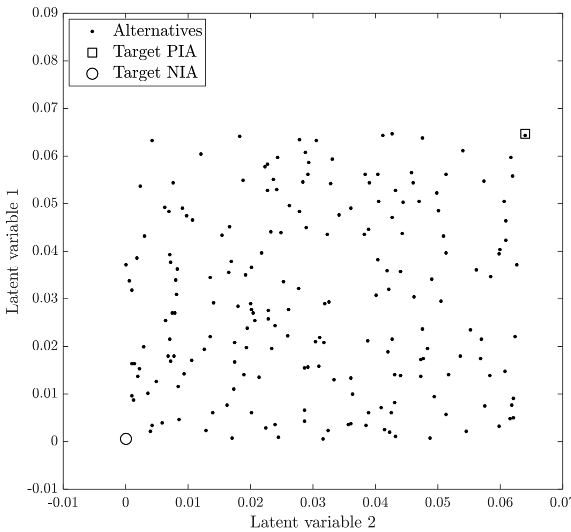
By applying the TOPSIS method on the latent criteria, one finds the target[2] PIA and NIA illustrated in Figure 15a. However, if we apply this approach in the observed decision data (mixed data according to Equation (4.2)), one finds both ideal alternatives illustrated in Figure 15b. One may note that TOPSIS applied on the mixed data does not lead to the desirable PIA and NIA. As a consequence, distance measures will not be correct, which will lead to a ranking of alternatives different from the target one. In other words, the mixing process introduces some bias in the calculation of the PIA and NIA.

If we consider the TOPSIS-M, we achieve both PIA and NIA illustrated in Figure 15c. Although the data are not correlated anymore, both ideal alternatives are also far from the target ones. This is due to fact that in TOPSIS-M both PIA and NIA ($\ddot{\mathbf{u}}^{+}$ and $\ddot{\mathbf{u}}^{-}$, respectively) are derived from the transformation of the ideal solutions ($\tilde{\mathbf{u}}^{+}$ and $\tilde{\mathbf{u}}^{-}$, respectively) obtained in the mixed data. Therefore, if PIA and NIA in mixed data are not the target ones, the application of TOPSIS-M will not lead to the target ideal solutions.

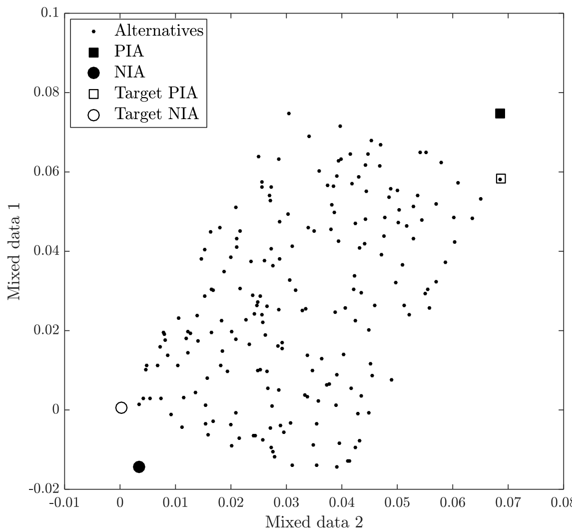
4.4 Proposed approaches

The aforementioned results point out that the covariance information among criteria and the transformation procedure conducted by TOPSIS-M may not be enough to mitigate the biased effect provided by dependent criteria. Since decorrelation does not imply independence (Papoulis and Pillai, 2002), a method that aims at retrieving a set of independent data may lead to better results. In the sequel, we present our proposed approaches.

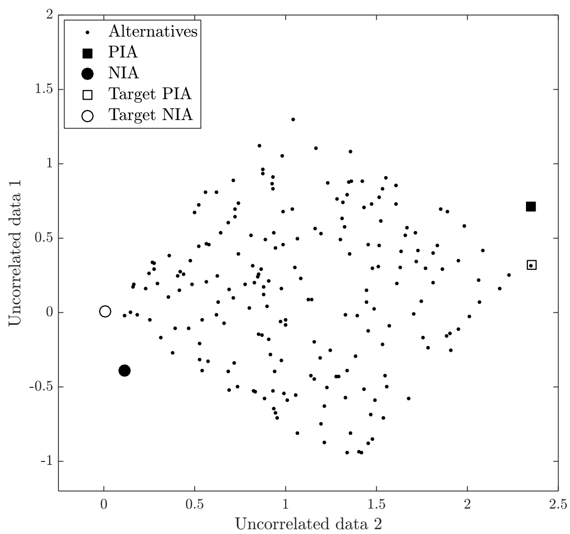
[2] We here refer to “target” PIA and NIA as the ideal solutions that are obtained by considering the independent latent criteria. Conversely, we refer to “target” ranking as the one provided by TOPSIS applied on this independent data.



(a) TOPSIS in latent criteria.



(b) TOPSIS in mixed data.



(c) TOPSIS-M in uncorrelated data.

Figure 15 – Graphical interpretation of TOPSIS and TOPSIS-M approaches.

4.4.1 ICA-TOPSIS

The first approach considered in this chapter, called ICA-TOPSIS, comprises three steps, as described in Algorithm 4.1. In the first one, we assume that the decision data $\hat{\mathbf{M}}$ is composed by a mixture of latent criteria $\mathbf{L}$ according to the mixing process described in Equation (4.2). Therefore, we formulate a BSS problem and apply an ICA technique to adjust the separating matrix $\mathbf{B}$, which gives us the estimates $\hat{\mathbf{L}}$ (alternative representation of $\hat{\mathbf{M}}$) according to Equation (4.3)[3].

One may note that the second step of Algorithm 4.1 aims at performing ad-

[3] If we consider the separating process described in Equation (3.12), $\hat{\mathbf{s}}(t)$ and $\mathbf{q}(t)$ represent, respectively, $\hat{\mathbf{L}}$ and $\hat{\mathbf{M}}$.

Algorithm 4.1 ICA-TOPSIS

Input: Decision data $\hat{\mathbf{M}}$ and set of weights $\mathbf{w}$.
Output: Closeness measure $\mathbf{o} = [o_1, o_2, \ldots, o_n]$.

Step 1: Latent criteria estimation. Based on $\hat{\mathbf{M}}$, apply an ICA technique and estimate the latent criteria $\mathbf{L}$.

Step 2: Ambiguities mitigation. Perform adjustments in both estimated mixing matrix $\hat{\mathbf{A}}$ and estimated latent criteria $\hat{\mathbf{L}}$, derived from the ICA technique, in order to avoid permutation and/or scaling ambiguities.

Step 3: TOPSIS and ranking determination. Apply the TOPSIS approach (Algorithm 2.1) by using the adjusted estimated latent criteria $\hat{\mathbf{L}}^{Adj_c}$ and the set of weights $\mathbf{w}$ as inputs, and determine the ranking of alternatives.

justments in the estimated latent criteria $\hat{\mathbf{L}}$ in order to deal with the ambiguities inherent to ICA (see Section 3.2.1.1 for further details). For this purpose, we assume that (i) the latent criteria $\mathbf{L}$ are in similar range and (ii) the diagonal elements in the mixing matrix $\mathbf{A}$ are positive and greater (in absolute value) than the off-diagonal elements in the same row. Although assumption (ii) may be strong in BSS, we consider that it is feasible in MCDM problems since we expect that each latent data has a positive majority influence in each observed criterion. If we consider the example illustrated in Figure 14, competences in mathematics, natural sciences and linguistics (independent latent criteria) should have a positive majority influence in subjects such as calculus, physics and literature, respectively. Therefore, based on the mixing matrix $\hat{\mathbf{A}} = \mathbf{B}^{-1}$ derived by an ICA technique, the adjustments performed in $\hat{\mathbf{L}}$ are the following ones:

1. For the first row in $\hat{\mathbf{A}}$, we find the column d that contains the greater absolute value. This value represents the influence of the estimated latent criterion $\hat{l}_d$ in the mixed process that results in the set of evaluations $\hat{u}_{1,j}$ $(j = 1, 2, \ldots, n)$ with respect to the first criterion (first row of $\hat{\mathbf{M}}$). Therefore, in order to place $\hat{l}_d$ as the first estimated latent criterion (without invaliding the relation $\hat{\mathbf{M}} = \hat{\mathbf{A}}\hat{\mathbf{L}}$), we permute the first and the d-th columns of $\hat{\mathbf{A}}$ and also the first and the d-th estimates (rows of $\hat{\mathbf{L}}$). This procedure is repeated for all rows in $\hat{\mathbf{A}}$, which leads to both estimated mixing matrix ($\hat{\mathbf{A}}^{Adj_p}$) and estimated latent criteria ($\hat{\mathbf{L}}^{Adj_p}$) partially adjusted, and mitigates the permutation ambiguity provided by the BSS method.

2. For each column in $\hat{\mathbf{A}}^{Adj_p}$, we verify the signal of the diagonal element. If the diagonal element d' is negative (negative contribution of the associated estimated latent criterion $\hat{l}_{d'}$), we multiply all the elements in the same column of d' by -1. As a consequence, we also need to invert the signal of $\hat{l}_{d'}$, since it is necessary to ensure that $\hat{\mathbf{M}} = \hat{\mathbf{A}}\hat{\mathbf{L}}$. After verifying all diagonal elements of $\hat{\mathbf{A}}^{Adj_p}$ and performing the sig-

nal changes, we obtain both estimated mixing matrix ($\hat{\mathbf{A}}^{Adj_c}$) and estimated latent criteria ($\hat{\mathbf{L}}^{Adj_c}$) completely adjusted, which mitigates the signal inversion ambiguity provided by the BSS method. With respect to the scaling ambiguity provided by a positive factor or a negative factor different from -1, it is automatically dealt with in the normalization step of TOPSIS.

Finally, after the application of ICA and the elimination of both permutation and/or scaling ambiguities, the third step of the proposed approach comprises the application of TOPSIS in $\hat{\mathbf{L}}^{Adj_c}$ in order to derive the ranking of alternatives. Therefore, Steps 1 and 2 in Algorithm 4.1 can be seen as a preprocessing step.

In order to illustrate the application of the ICA-TOPSIS approach, consider the example described in Section 4.3. The application of the ICA algorithm lead to the estimated mixing matrix

$$\hat{\mathbf{A}} = \begin{bmatrix} -0.2158 & 0.2864 \\ -0.2888 & -0.0651 \end{bmatrix},$$

which is associated with the estimated latent criteria $\hat{\mathbf{L}} = [\hat{l}_1, \hat{l}_2]^T$. In order to mitigate the permutation ambiguity, the first adjustment leads to both estimated mixing matrix

$$\hat{\mathbf{A}}^{Adj_p} = \begin{bmatrix} 0.2864 & -0.2158 \\ -0.0651 & -0.2888 \end{bmatrix}$$

and estimated latent criteria $\hat{\mathbf{L}}^{Adj_p} = [\hat{l}_2, \hat{l}_1]$ partially adjusted, which comprises the permutation of both columns of $\hat{\mathbf{A}}^{Adj_p}$ and estimates $\hat{l}_1$ and $\hat{l}_2$. With respect to the scaling ambiguity, the second adjustment leads to both estimated mixing matrix[4]

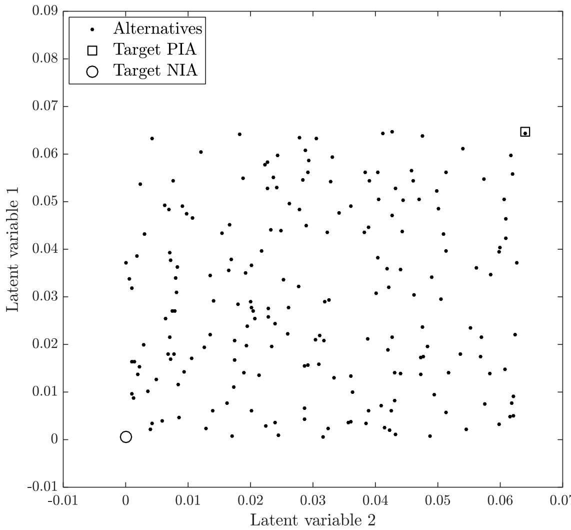
$$\hat{\mathbf{A}}^{Adj_c} = \begin{bmatrix} 0.2864 & 0.2158 \\ -0.0651 & 0.2888 \end{bmatrix}$$

and estimated latent criteria $\hat{\mathbf{L}}^{Adj_c} = [\hat{l}_2, -\hat{l}_1]$ completely adjusted, which comprises the signal inversion of both second column of $\hat{\mathbf{A}}^{Adj_c}$ and second estimates of $\hat{\mathbf{L}}^{Adj_c}$. Therefore, based on $\hat{\mathbf{L}}^{Adj_c}$, we apply the TOPSIS approach, which leads to the PIA and NIA illustrated in Figure 16b. One may note that the ICA algorithm performs a good estimation of the latent criteria, which also provides both PIA and NIA close to the target ones.

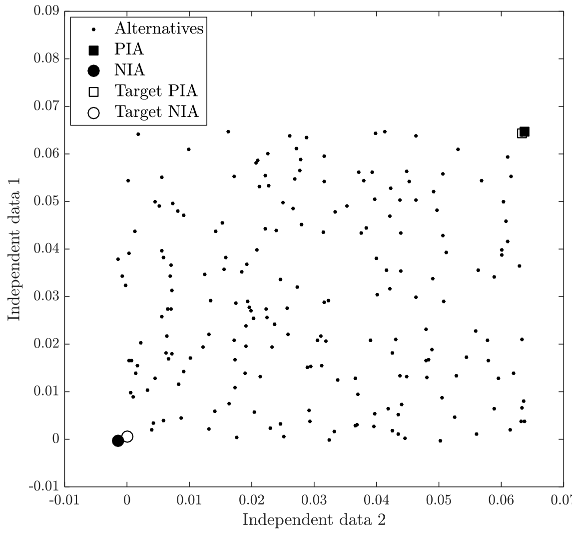
4.4.2 ICA-TOPSIS-M

The second proposed approach, called ICA-TOPSIS-M, combines an ICA technique and a modified version of the TOPSIS-M. It comprises five steps, as described in Algorithm 4.2, in which the first and the second ones are identical to the ICA-TOPSIS

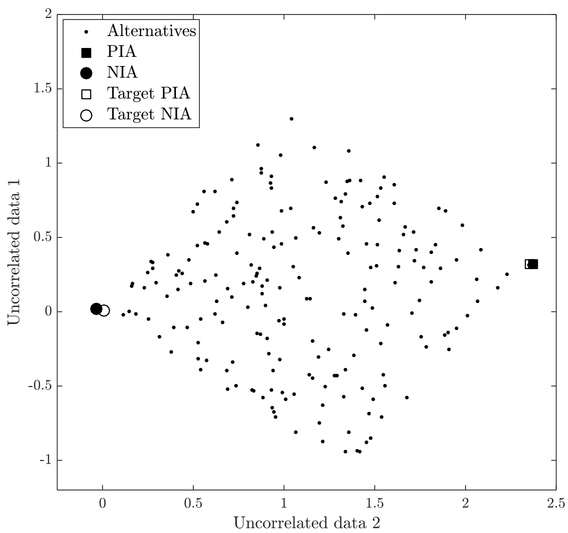
[4] One may note that $\hat{\mathbf{A}}^{Adj_c}$ differs from the correct one (defined on Section 4.3) by a factor of 3.5, approximately, which introduces a scaling ambiguity provided by a positive factor.



(a) TOPSIS in latent criteria.



(b) ICA-TOPSIS in estimated latent criteria.



(c) ICA-TOPSIS-M in uncorrelated data.

Figure 16 – Graphical interpretation of TOPSIS, ICA-TOPSIS and ICA-TOPSIS-M approaches.

approach. However, the third step consists in determining both PIA and NIA directly from the estimated latent criteria completely adjusted $\hat{\mathbf{L}}^{Adj_c}$. In the fourth step, we apply the estimated mixing matrix completely adjusted $\hat{\mathbf{A}}^{Adj_c}$ on both PIA and NIA, which leads to the transformation of these ideal alternatives into the mixed space (the same as the mixed decision data $\hat{\mathbf{M}}$). Since the transformed PIA and NIA were obtained based on the estimated latent criteria, we expect that both ideal alternatives are placed close to the target ones in the mixed space.

Finally, the last step comprises the application of a modified version of TOPSIS-M approach to determine the ranking of alternatives. Instead of deriving PIA and NIA in Step 2 of Algorithm 2.2, we use as inputs the transformed PIA and NIA calculated in Step

Algorithm 4.2 ICA-TOPSIS-M

Input: Decision data $\hat{\mathbf{M}}$ and set of weights $\mathbf{w}$.
Output: Closeness measure $\mathbf{o} = [o_1, r_2, \ldots, o_n]$.

Step 1: Latent criteria estimation. Based on $\hat{\mathbf{M}}$, apply an ICA technique and estimate the latent criteria $\mathbf{L}$.

Step 2: Ambiguities mitigation. Perform adjustments in both estimated mixing matrix $\hat{\mathbf{A}}$ and estimated latent criteria $\hat{\mathbf{L}}$, derived from the ICA technique, in order to avoid permutation and/or scaling ambiguities.

Step 3: Ideal alternatives determination. Determine both positive and negative ideal alternatives (PIA and NIA, respectively) directly from $\hat{\mathbf{L}}^{Adj_c}$:

$$PIA = \hat{\mathbf{l}}^{+} = \left\{\hat{l}_1^{+}, \hat{l}_2^{+}, \ldots, \hat{l}_m^{+}\right\} \text{ and } NIA = \hat{\mathbf{l}}^{-} = \left\{\hat{l}_1^{-}, \hat{l}_2^{-}, \ldots, \hat{l}_m^{-}\right\},$$

where $\hat{l}_i^{+} = \max\{\hat{l}_{i,j} | 1 \leq j \leq n\}$ and $\hat{l}_i^{-} = \min\{\hat{l}_{i,j} | 1 \leq j \leq n\}$, $i = 1, \ldots, m$.

Step 4: Ideal alternatives transformation. Apply $\hat{\mathbf{A}}^{Adj_c}$ on both PIA and NIA and transform these ideal alternatives into the mixed data space.

Step 5: Modified TOPSIS-M and ranking determination. Based on the transformed PIA and NIA calculated in Step 4, apply the modified TOPSIS-M approach (Algorithm 2.2 without the Step 2) and determine the ranking of alternatives.

4 of the ICA-TOPSIS-M approach. Therefore, we expect that these ideal alternatives be placed close to the target ones, which leads to an improvement in the distances calculated in Step 4 of Algorithm 2.2 and, therefore, in the ranking determination.

By considering the example described in Section 4.3 and by applying the ICA-TOPSIS-M approach with the transformed PIA and NIA derived from $\hat{\mathbf{L}}^{Adj_c}$, we obtain the ideal alternatives illustrated in Figure 16c. Since these ideal alternatives are derived directly from the estimated latent criteria instead of the observed mixed data, they are closer to the target PIA and NIA compared with the ones obtained by the TOPSIS-M approach, illustrated in Figure 15c.

4.5 Experiments and results

Aiming at attesting the proposed approaches in MCDM problems with dependent criteria, we perform experiments in both synthetic and real data. They are described in the sequel.

4.5.1 Experiments on synthetic data

In order to verify the robustness of the proposed approaches compared to the traditional TOPSIS and TOPSIS-M methods, we performed numerical experiments on synthetic data. The latent criteria were randomly generated according to a uniform distribution in the range $[0, 1]$ and we considered that all criteria have the same importance on the decision data, i.e., $w_i = w_{i'}$ for all i and i'. Note that this scenario may represent the example described in Figure 14. In this case, the latent criteria and the mixed decision data correspond, respectively, to the competences of students in different subjects (to be estimated) and their grades.

4.5.1.1 TOPSIS-M performance for different mixing matrices

As a first experiment, we verify the performance of TOPSIS-M approach in the decision problem defined on Section 4.1. We considered 100 alternatives, 2 latent criteria and different degrees of mixture. For this purpose, in the mixing process described in Equation (4.2) (without the additive noise), we vary the off-diagonal elements α and β of mixing matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha \\ \beta & 1 \end{bmatrix}$$

in the range $[-0.75, 0.75]$. In order to verify the robustness of TOPSIS-M approach to deal with different mixture intensities, we considered as a performance index the Kendall tau coefficient (Kendall, 1938), defined by

$$\tau = \frac{J_{agre} - J_{disa}}{n(n-1)/2}, \tag{4.5}$$

where J_{agre} and J_{disa} are the number of pairwise agreements and pairwise disagreements between two rankings, respectively, n is the number of alternatives and $n(n-1)/2$ is the total number of pairwise combinations. $\tau = 1$ and $\tau = -1$ indicate that the two rankings are the same and the reverse of the other, respectively. In the case of $\tau \approx 0$, the two rankings are independent. Therefore, in our experiments, the larger τ the better, since it indicates that the retrieved ranking is close to the target one provided by the application of TOPSIS on the latent criteria.

The application of TOPSIS-M method on the mixed decision data leads to the Kendall tau coefficients (averaged over 500 simulations) shown in Figure 17. The results indicate that TOPSIS-M method performs better when both α and β are positive or both α and β are negative and equal, i.e., where there is no negative contribution of only one latent criterion in the mixing process. However, if α and/or β are negative (and different), τ decreases, indicating that there are a great number of disagreements between the ranking obtained by the TOPSIS-M approach applied on the mixed decision data and the target one provided by TOPSIS applied on the latent criteria. This is a consequence of the fact that both PIA and NIA are far from the target values.

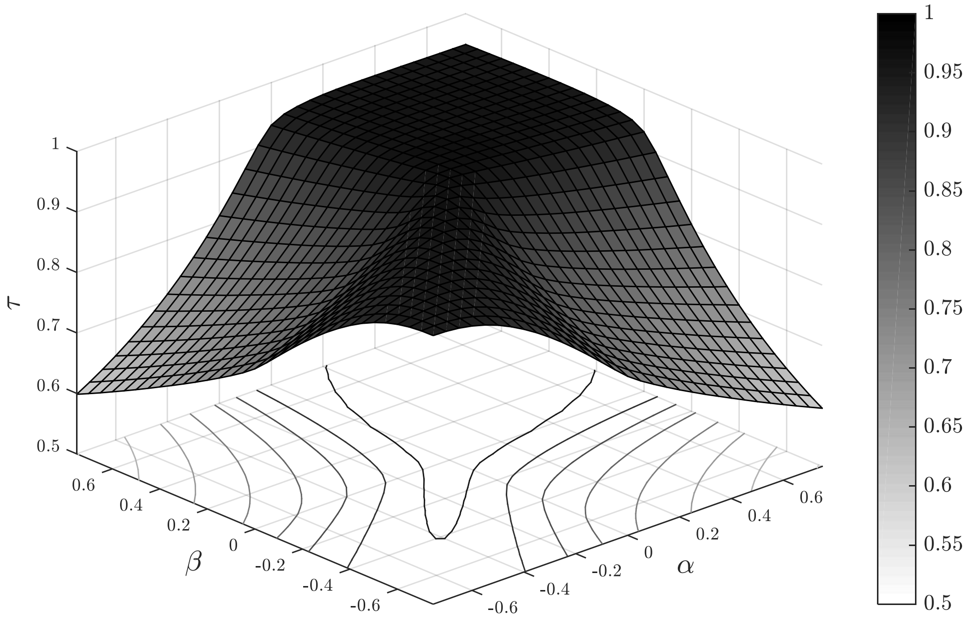


Figure 17 – Robustness of TOPSIS-M for different degrees of mixture.

4.5.1.2 Comparison for different degrees of noise

In this second experiment, we compare the performance of TOPSIS, TOPSIS-M and the proposed ICA-TOPSIS and ICA-TOPSIS-M approaches. They are applied to deal with a MCDM problem with $n = 100$ alternatives, $m = 2$ criteria and for different degrees of noise $\mathbf{G}$ in the mixing process (4.2). With respect to the mixing matrix[5], we fix the diagonal elements equal to 1 and all off-diagonal elements are randomly generated in the interval $[-0.5, -0.2] \cup [0.2, 0.5]$. This analysis is important since it points out how robust the approaches are in situations in which the adopted observed data generative model is not perfect (which is often the case in real situations). In our experiment, we provide results for different Signal-to-Noise Ratio (SNR, in dB), given by

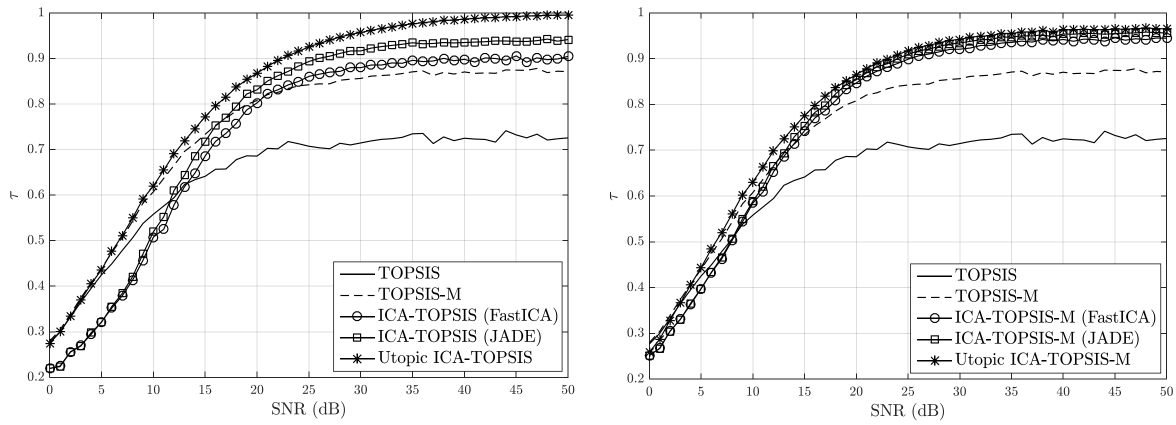
$$SNR = 10 \log_{10} \frac{P^2_{signal}}{P^2_{noise}}, \tag{4.6}$$

where P^2_{signal} and P^2_{noise} are, respectively, the signal ($\mathbf{AL}$ in the mixing process (4.2)) power and the noise ($\mathbf{G}$) power, in the range $(0, 50]$ dB. Therefore, lower values of SNR indicate higher degrees of noise in the mixing process.

Figure 18 presents a comparison of the obtained Kendall tau coefficients (averaged over 500 simulations for each SNR value) for the considered approaches. With

[5] We use the same mixing matrices generation for the experiments conducted in Sections 4.5.1.2, 4.5.1.3 and 4.5.1.4.

respect to ICA-TOPSIS and ICA-TOPSIS-M, we considered both FastICA and JADE algorithms. Figure 18a also illustrates τ for the "Utopic ICA-TOPSIS", which comprises the application of ICA-TOPSIS with the ideal separating matrix $\mathbf{B} = \mathbf{A}^{-1}$ and is used as benchmark for the achieved results. Similarly, in Figure 18b, we use as a benchmark the "Utopic ICA-TOPSIS-M", which comprises the application of ICA-TOPSIS-M with the transformation of the target PIA and NIA derived from the latent criteria.



(a) ICA-TOPSIS approach

(b) ICA-TOPSIS-M approach

Figure 18 – Comparison of Kendall tau coefficients for different SNR values.

Although the TOPSIS-M approach applied to the mixed decision data performed better compared to the TOPSIS, an improvement may be achieved by using either ICA-TOPSIS or ICA-TOPSIS-M. One may clearly see these findings in Figure 18, especially for SNR greater than 15 dB (with a much better performance of ICA-TOPSIS-M in comparison with ICA-TOPSIS). For SNR lower than 15 dB, which leads to a strong noise interference, the ICA techniques could not perform a good separation and, therefore, the considered approaches could not retrieve a proper ranking of alternatives.

With the purpose of further exploiting the performance of the considered approaches, we calculate the Pearson's correlation coefficient between the closeness measure of the estimates ($o_j^{\hat{\mathbf{L}}^{Adjc}}$) and the independent latent criteria ($o_j^{\mathbf{L}}$), given by

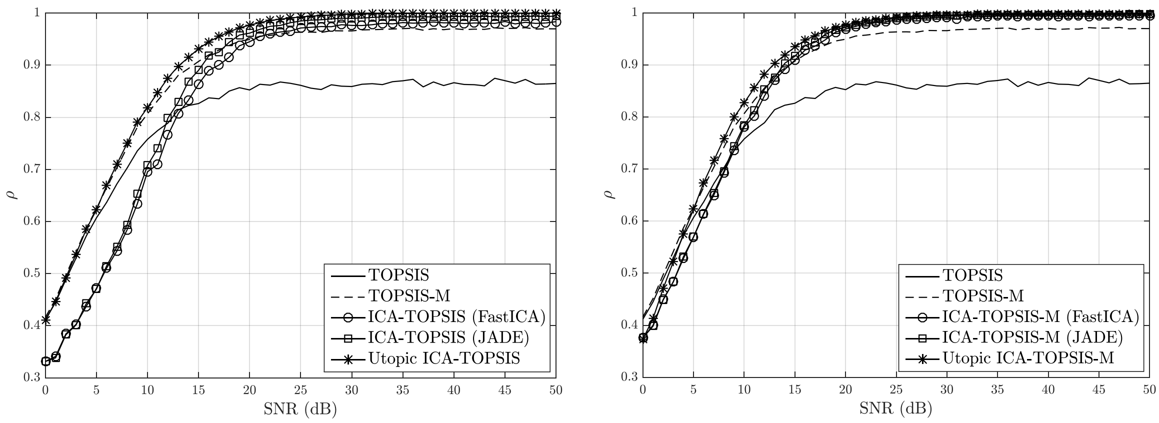
$$\rho\left(o_j^{\hat{\mathbf{L}}^{Adjc}}, o_j^{\mathbf{L}}\right) = \frac{\sum_{j=1}^{n}(o_j^{\hat{\mathbf{L}}^{Adjc}} - \bar{o}_j^{\hat{\mathbf{L}}^{Adjc}})(o_j^{\mathbf{L}} - \bar{o}_j^{\mathbf{L}})}{\sqrt{\sum_{j=1}^{n}(o_j^{\hat{\mathbf{L}}^{Adjc}} - \bar{o}_j^{\hat{\mathbf{L}}^{Adjc}})^2}\sqrt{\sum_{j=1}^{n}(o_j^{\mathbf{L}} - \bar{o}_j^{\mathbf{L}})^2}}, \tag{4.7}$$

where $\bar{o}_j^{\cdot} = (1/n)\sum_{j=1}^{n} o_j^{\cdot}$. Moreover, we also calculate the mean absolute error of the position of the first 20% of the total number (n) of alternatives (the first 20 alternatives, in this case). This measure is given by:

$$\varepsilon = \frac{1}{0.2n}\sum_{j=1}^{0.2n}|t_j - j|, \tag{4.8}$$

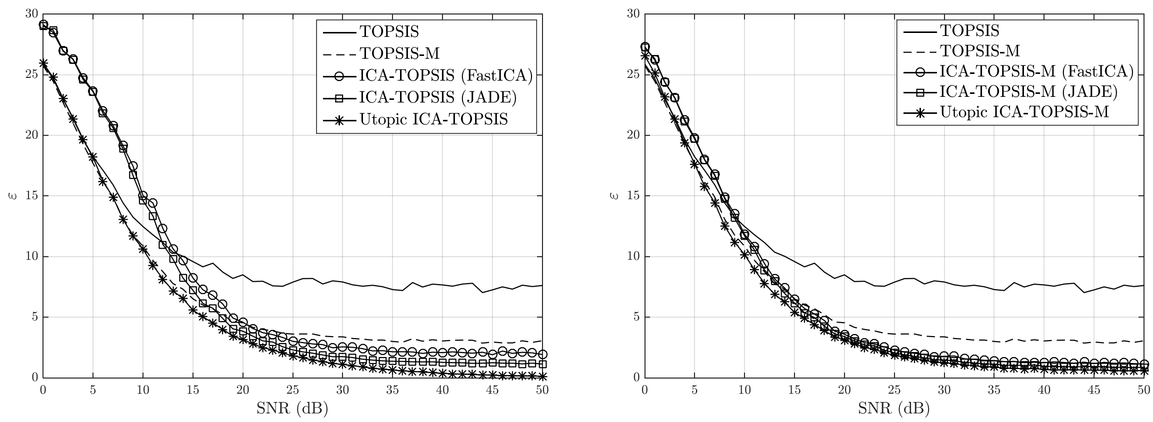
where t_j is the position, in the ranking derived by the considered approaches, of the j-th alternative in the correct ranking. Figures 19 and 20 present the obtained results.

Similarly as for the Kendall tau coefficient, one may note that both ICA-TOPSIS and ICA-TOPSIS-M approaches lead to better results, especially for SNR greater than 25 dB. One may also note that ICA-TOPSIS-M performed slightly better than ICA-TOPSIS.



(a) ICA-TOPSIS approach (b) ICA-TOPSIS-M approach

Figure 19 – Comparison of Pearson's correlation coefficients for different SNR values.



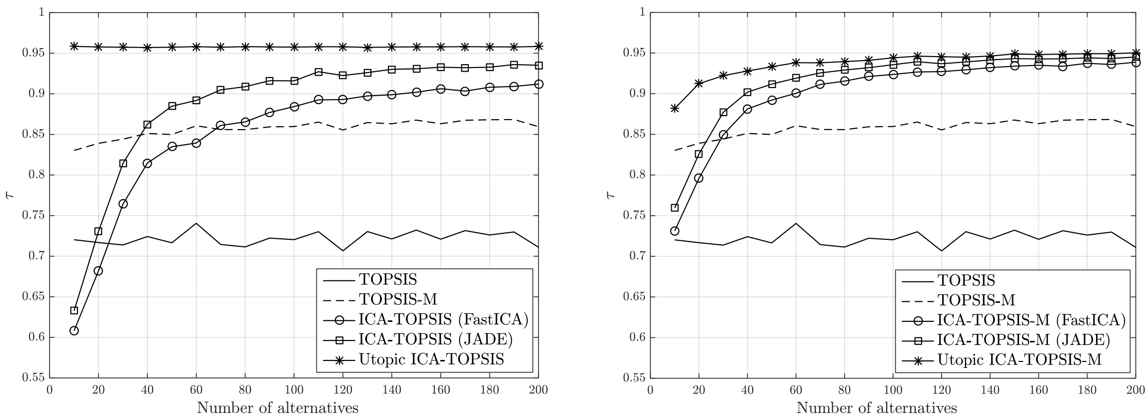
(a) ICA-TOPSIS approach (b) ICA-TOPSIS-M approach

Figure 20 – Comparison of mean absolute errors for different SNR values.

With respect to the ICA algorithms, the methods based on JADE achieved the highest values of the considered performance indices. This is directly related to the performance in estimating the mixing matrix and, consequently, the latent criteria. Since JADE provided a better estimation of $\mathbf{A}$ compared to the FastICA, it also provided a ranking of alternatives closer to the target one.

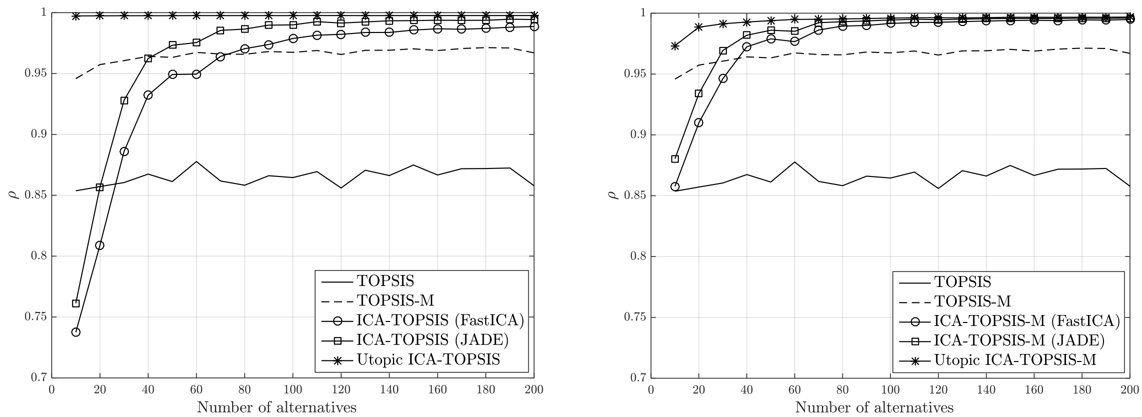
4.5.1.3 Comparison for different numbers of alternatives

In the previous experiment we compared the considered approaches by varying the level of noise in the mixing process and by fixing the number of alternatives. On the other hand, in this third experiment, we compare them by fixing $SNR = 30$ dB (a low noise level) and by varying the number of alternatives. The results (averaged over 500 simulations for each number of alternatives) for the Kendall tau coefficient, Pearson's correlation coefficient and mean absolute error are illustrated in Figures 21, 22 and 23, respectively[6].



(a) ICA-TOPSIS approach (b) ICA-TOPSIS-M approach

Figure 21 – Comparison of Kendall tau coefficients for different number of alternatives.

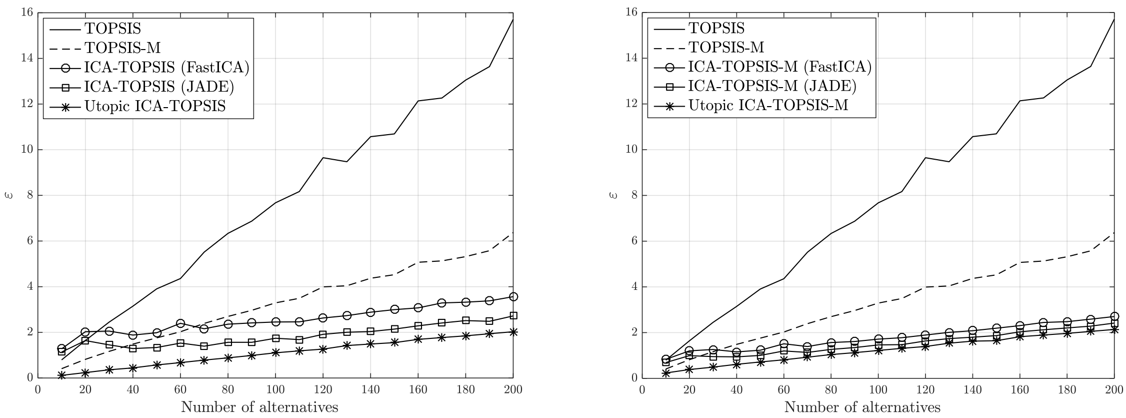


(a) ICA-TOPSIS approach (b) ICA-TOPSIS-M approach

Figure 22 – Comparison of Pearson's correlation coefficients for different number of alternatives.

Similarly as obtained in Section 4.5.1.2, both ICA-TOPSIS and ICA-TOPSIS-M approaches performed better in comparison with TOPSIS and TOPSIS-M, especially for $n \geq 30$. For lower values of n, the number of samples is not enough for the ICA technique to perform a good estimation of the mixing matrix. If we compare the performance of the ICA algorithms, JADE also provided a better estimation of the mixing matrix and, consequently, the ranking of alternatives.

[6] Note that the mean absolute error calculated in Equation (4.8) depends on the number of alternatives. Therefore, as can be seen in Figure 23, the higher is the number of alternatives, the higher is the mean absolute error.



(a) ICA-TOPSIS approach (b) ICA-TOPSIS-M approach

Figure 23 – Comparison of mean absolute error for different number of alternatives.

4.5.1.4 Numerical experiments with more than two criteria

In the aforementioned experiments, we considered MCDM problems with only two criteria. All the obtained results indicated that the ICA-TOPSIS-M approach based on JADE algorithm can better deal with the MCDM problem defined in Section 4.1. In order to verify the robustness of this approach, we compare it with TOPSIS and TOPSIS-M in decision problems with more than two criteria. For instance, we consider scenarios with different levels of noise ($SNR = 15$, $SNR = 30$ and $SNR = 45$ dB) and different numbers of alternatives ($n = 30$, $n = 100$ and $n = 170$). The average value and standard deviation σ (over 1000 simulations) of each performance index for 3, 4 and 5 decision criteria are shown in Tables 7, 8 and 9, respectively. The best result for each scenario and each performance index is highlighted in bold.

One may note that TOPSIS approach led to the worst result in all scenarios. By comparing TOPSIS-M and ICA-TOPSIS-M, the latter achieved the better performance for all scenarios whose level of noise and number of alternatives were, at least, 30 and 100, respectively. Moreover, for all scenarios with 170 alternatives, including the one with $SNR = 15$ dB, ICA-TOPSIS-M provided the better results. In most scenarios in which TOPSIS-M performed better in comparison with ICA-TOPSIS-M, the level of noise was 15 dB and the number of alternatives was 30. This is a consequence of the performance of the ICA algorithm, since the strong interference of noise and the small number of samples make the estimation process difficult to be conducted.

Table 7 – Results for a decision problem with $m = 3$ criteria.

Performance index	**Method**	**Scenarios:** $\{SNR, n\}$								
		$\{15, 30\}$	$\{15, 100\}$	$\{15, 170\}$	$\{30, 30\}$	$\{30, 100\}$	$\{30, 170\}$	$\{45, 30\}$	$\{45, 100\}$	$\{45, 170\}$
	TOPSIS	0.5088	0.5087	0.5043	0.5429	0.5365	0.5434	0.5372	0.5389	0.5228
		(0.1784)	(0.1671)	(0.1665)	(0.2026)	(0.1947)	(0.1932)	(0.2061)	(0.2052)	(0.1945)
τ	TOPSIS-M	**0.6402**	0.6561	0.6585	0.7235	0.7437	0.7458	0.7332	0.7455	0.7472
(σ_τ)		(0.1021)	(0.0761)	(0.0742)	(0.1299)	(0.1179)	(0.1133)	(0.1308)	(0.1239)	(0.1201)
	ICA-TOPSIS-M	0.5667	**0.6892**	**0.7167**	**0.7448**	**0.9089**	**0.9258**	**0.7808**	**0.9277**	**0.9478**
	(JADE)	(0.1979)	(0.0983)	(0.0691)	(0.2275)	(0.0337)	(0.0235)	(0.1842)	(0.0400)	(0.0269)
	TOPSIS	0.6900	0.6940	0.6901	0.7170	0.7149	0.7207	0.7105	0.7114	0.6985
		(0.1929)	(0.1817)	(0.1821)	(0.2036)	(0.1993)	(0.1959)	(0.2041)	(0.2070)	(0.1973)
ρ	TOPSIS-M	**0.8358**	0.8495	0.8519	**0.8933**	0.9055	0.9075	0.8971	0.9048	0.9064
(σ_ρ)		(0.0822)	(0.0629)	(0.0619)	(0.0914)	(0.0795)	(0.0767)	(0.0871)	(0.0822)	(0.0784)
	ICA-TOPSIS-M	0.7456	**0.8710**	**0.8957**	0.8787	**0.9893**	**0.9931**	**0.9092**	**0.9926**	**0.9963**
	(JADE)	(0.2300)	(0.0989)	(0.0662)	(0.2347)	(0.0086)	(0.0047)	(0.1714)	(0.0139)	(0.0037)
	TOPSIS	0.9157	3.0093	5.1098	0.8535	2.8094	4.6992	0.8838	2.8678	5.0050
		(0.5336)	(1.4848)	(2.4835)	(0.5579)	(1.6000)	(2.7342)	(0.5780)	(1.7304)	(2.7579)
ε	TOPSIS-M	**0.5943**	1.8161	3.0007	**0.4359**	1.2642	2.1041	0.4205	1.2642	2.1007
(σ_ε)		(0.3057)	(0.5962)	(0.9661)	(0.3138)	(0.7348)	(1.1600)	(0.2918)	(0.7806)	(1.2339)
	ICA-TOPSIS-M	0.7757	**1.6526**	**2.4284**	0.4428	**0.4003**	**0.5303**	**0.3749**	**0.2999**	**0.3584**
	(JADE)	(0.5646)	(0.8438)	(0.9990)	(0.5922)	(0.1910)	(0.1978)	(0.4669)	(0.2100)	(0.2016)

Table 8 – Results for a decision problem with $m = 4$ criteria.

Performance index	**Method**	**Scenarios:** $\{SNR, n\}$								
		$\{15, 30\}$	$\{15, 100\}$	$\{15, 170\}$	$\{30, 30\}$	$\{30, 100\}$	$\{30, 170\}$	$\{45, 30\}$	$\{45, 100\}$	$\{45, 170\}$
	TOPSIS	0.4257	0.4275	0.4238	0.4466	0.4427	0.4301	0.4409	0.4408	0.4319
		(0.1854)	(0.1690)	(0.1618)	(0.2011)	(0.1916)	(0.1913)	(0.2074)	(0.1884)	(0.1906)
τ	TOPSIS-M	**0.5690**	0.5829	0.5867	**0.6353**	0.6520	0.6478	**0.6253**	0.6489	0.6455
(σ_τ)		(0.1165)	(0.1000)	(0.0934)	(0.1435)	(0.1364)	(0.1542)	(0.1592)	(0.1378)	(0.1519)
	ICA-TOPSIS-M	0.4410	**0.5969**	**0.6488**	0.5961	**0.8793**	**0.9066**	0.5971	**0.9049**	**0.9317**
	(JADE)	(0.2060)	(0.1644)	(0.1247)	(0.2607)	(0.0740)	(0.0433)	(0.2597)	(0.0434)	(0.0282)
	TOPSIS	0.5984	0.6032	0.6003	0.6178	0.6148	0.6016	0.6081	0.6139	0.6038
		(0.2210)	(0.2039)	(0.1957)	(0.2306)	(0.2199)	(0.2244)	(0.2386)	(0.2147)	(0.2224)
ρ	TOPSIS-M	**0.7709**	**0.7825**	0.7863	**0.8245**	0.8373	0.8314	**0.8118**	0.8338	0.8280
(σ_ρ)		(0.1130)	(0.0982)	(0.0919)	(0.1251)	(0.1179)	(0.1507)	(0.1470)	(0.1178)	(0.1475)
	ICA-TOPSIS-M	0.6082	0.7824	**0.8356**	0.7515	**0.9774**	**0.9879**	0.7504	**0.9879**	**0.9940**
	(JADE)	(0.2609)	(0.1931)	(0.1374)	(0.2916)	(0.0632)	(0.0368)	(0.2844)	(0.0172)	(0.0048)
	TOPSIS	1.1511	3.7380	6.4295	1.1084	3.6643	6.3784	1.1105	3.6518	6.4229
		(0.6152)	(1.6822)	(2.6437)	(0.6111)	(1.7920)	(3.0382)	(0.6289)	(1.7573)	(3.0785)
ε	TOPSIS-M	**0.7628**	2.3556	3.9702	**0.6288**	1.8876	3.2587	**0.6345**	1.8933	3.2917
(σ_ε)		(0.3851)	(0.8902)	(1.3315)	(0.3859)	(1.0053)	(2.0816)	(0.4180)	(0.9950)	(2.0601)
	ICA-TOPSIS-M	1.1075	**2.3554**	**3.2790**	0.7673	**0.5652**	**0.7069**	0.7549	**0.4227**	**0.4938**
	(JADE)	(0.6696)	(1.5372)	(1.8270)	(0.7120)	(0.5634)	(0.5697)	(0.7086)	(0.2513)	(0.2239)

Table 9 – Results for a decision problem with $m = 5$ criteria.

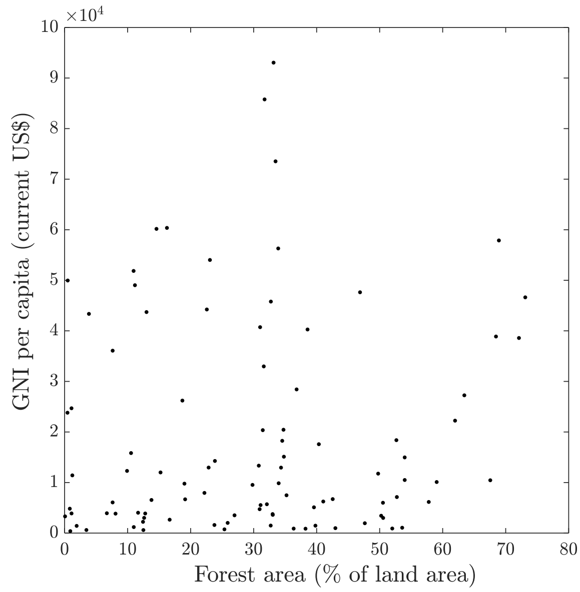
Performance index	**Method**	**Scenarios: $\{SNR, n\}$**								
		$\{15, 30\}$	$\{15, 100\}$	$\{15, 170\}$	$\{30, 30\}$	$\{30, 100\}$	$\{30, 170\}$	$\{45, 30\}$	$\{45, 100\}$	$\{45, 170\}$
τ (σ_τ)	TOPSIS	0.3659 (0.1848)	0.3623 (0.1529)	0.3606 (0.1584)	0.3710 (0.1895)	0.3671 (0.1738)	0.3720 (0.1656)	0.3661 (0.1932)	0.3726 (0.1755)	0.3682 (0.1661)
	TOPSIS-M	**0.4948** (0.1322)	**0.5238** (0.1087)	0.5206 (0.1164)	**0.5368** (0.1671)	0.5586 (0.1623)	0.5714 (0.1589)	**0.5293** (0.1786)	0.5647 (0.1776)	0.5662 (0.1817)
	ICA-TOPSIS-M (JADE)	0.3383 (0.2113)	0.4883 (0.1743)	**0.5567** (0.1725)	0.4322 (0.2494)	**0.8302** (0.1280)	**0.8802** (0.0718)	0.4215 (0.2627)	**0.8662** (0.0976)	**0.9181** (0.0376)
ρ (σ_ρ)	TOPSIS	0.5272 (0.2234)	0.5264 (0.1941)	0.5248 (0.2010)	0.5337 (0.2329)	0.5285 (0.2152)	0.5372 (0.2037)	0.5247 (0.2358)	0.5371 (0.2172)	0.5328 (0.2072)
	TOPSIS-M	**0.6882** (0.1449)	**0.7217** (0.1198)	0.7171 (0.1325)	**0.7291** (0.1752)	0.7474 (0.1740)	0.7599 (0.1714)	**0.7186** (0.1994)	0.7506 (0.1993)	0.7522 (0.2048)
	ICA-TOPSIS-M (JADE)	0.4805 (0.2824)	0.6690 (0.2136)	**0.7411** (0.2070)	0.5889 (0.3073)	**0.9488** (0.1158)	**0.9774** (0.0573)	0.5719 (0.3290)	**0.9688** (0.0777)	**0.9909** (0.0207)
ε (σ_ε)	TOPSIS	1.3198 (0.6289)	4.3508 (1.5793)	7.3898 (2.7033)	1.3007 (0.6360)	4.3489 (1.7759)	7.2475 (2.8046)	1.3135 (0.6422)	4.3116 (1.7707)	7.3241 (2.7918)
	TOPSIS-M	**0.9606** (0.4579)	**2.8516** (1.0423)	4.8531 (1.8217)	**0.8413** (0.4908)	2.6205 (1.4413)	4.2270 (2.3095)	**0.8696** (0.5239)	2.5992 (1.5956)	4.3285 (2.7424)
	ICA-TOPSIS-M (JADE)	1.4007 (0.7291)	3.2389 (1.7367)	**4.5328** (2.7336)	1.1393 (0.7576)	**0.8537** (0.9756)	**0.9379** (0.8476)	1.1933 (0.8199)	**0.6471** (0.6713)	**0.6097** (0.3811)

4.5.2 Experiment on real data

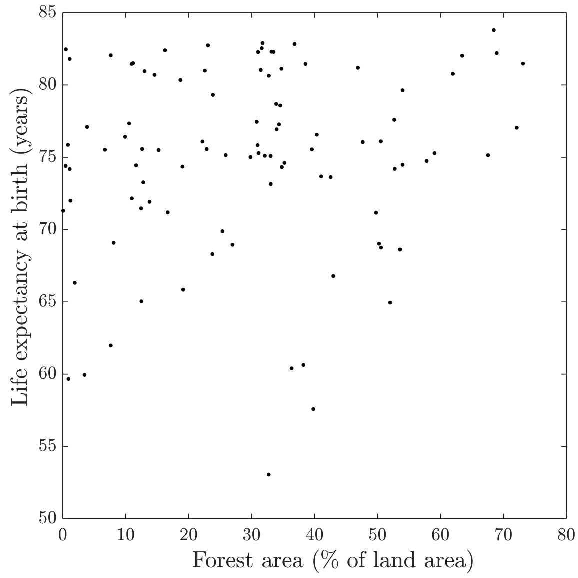
In the last experiment of this chapter, we apply the proposed approaches in a real dataset composed by $n = 96$ countries and $m = 3$ observed criteria: forest area (c_1, in % of land area), gross national income (GNI) per capita (c_2, in current US$) and life expectancy at birth (c_3, in years). This data refers to 2015 and was collected from the World Bank at http://www.worldbank.org/. It is worth mentioning that we adopted this set of criteria only to illustrate our proposal in a real data. Therefore, more criteria from different domains can be used to rank countries.

Figures 24a, 24b and 24c present the scatter plots of pairs of criteria. Since the life expectancy at birth is affected by several factors, such as social and economic ones, this criterion may be dependent on the GNI. The Pearson's correlation coefficient between these two criteria is 0.68. Therefore, even when equal importance is attributed to the decision criteria, the application of TOPSIS on the observed data shall amplify the importance of the latent factors that drive both GNI per capita and life expectancy at birth. So, in this scenario, it becomes interesting to find a representation in which the latent criteria are independent. These latent criteria may carry, for instance, information associated with environmental, economic and social aspects.

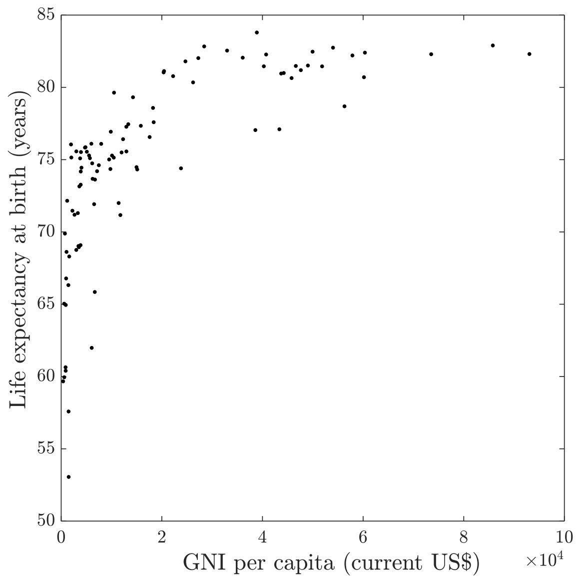
Table 10 presents a comparison with respect to the position of the first 10 alternatives in the ranking provided by each considered approach. With respect to the ICA algorithm, we adopted the JADE, since it led to the better results in the previous



(a) Scatter plot of criteria 1 and 2.



(b) Scatter plot of criteria 1 and 3.



(c) Scatter plot of criteria 2 and 3.

Figure 24 – Real data visualization.

experiments. By considering the TOPSIS, the alternatives with good evaluations in both GNI per capita and life expectancy at birth achieve a better position, even with a bad performance on the first criterion. For instance, the evaluations of the first alternative in the ranking, a_{66}, are equal to $(33.1613; 93050; 82.3049)$.

On the other hand, the application of the other methods favors alternatives with a good evaluation on forest area. For instance, by applying the TOPSIS-M, the first alternative in the ranking is a_{83}, whose evaluations are equal to $(68.9229; 57880; 82.2049)$. This can also be noted in the ranking provided by ICA-TOPSIS-M, in which the two first alternatives (a_{83} and a_{33}) have a good evaluation on the first criterion. However, the third,

Table 10 – Rankings obtained in experiments with real data.

Alternatives	Criteria evaluations			Position in the ranking			
	c_1	c_2	c_3	TOPSIS	TOPSIS-M	ICA-TOPSIS	ICA-TOPSIS-M
a_4	16.2388	60360	82.4000	10	15	59	15
a_5	46.8839	47630	82.4000	9	8	11	7
a_{13}	59.0488	10100	75.2840	28	24	8	30
a_{14}	72.1063	38590	77.0460	7	6	3	8
a_{33}	73.1072	46630	81.4805	5	4	2	2
a_{46}	68.4606	38880	83.7939	8	7	4	6
a_{49}	63.4386	27250	82.0244	14	10	5	9
a_{50}	53.9723	14970	74.4805	29	25	10	31
a_{52}	33.4749	73530	82.2927	4	5	25	5
a_{54}	67.5544	10450	75.1430	22	18	6	25
a_{66}	33.1613	93050	82.3049	1	2	21	4
a_{68}	57.7914	6160	74.7470	31	28	9	35
a_{79}	61.9662	22240	80.7756	18	12	7	14
a_{83}	68.9229	57880	82.2049	3	1	1	1
a_{84}	31.7340	85780	82.8976	2	3	23	3
a_{93}	33.8997	56300	78.6902	6	9	29	10

the fourth and the fifth alternatives in the ranking (a_{84}, a_{66} and a_{52}) perform well only in the GNI and life expectancy at birth. Therefore, although we addressed the problem of dependent criteria, we do not achieve a considerable improvement in the ranking provided by ICA-TOPSIS-M.

Better results are obtained by the application of ICA-TOPSIS. Most of the first 10 alternatives in the ranking have a good evaluation on the first criterion. Therefore, an alternative that performs well in forest area and in one of the other criteria is preferable compared to an alternative that performs well only in GNI per capita and life expectancy at birth, which are dependent criteria. Moreover, some alternatives well evaluated on these criteria that achieved good positions on the ranking provided by the other approaches were not well evaluated in ICA-TOPSIS. For instance, a_{66} was the first, second and fourth alternative in the ranking provided by TOPSIS, TOPSIS-M and ICA-TOPSIS-M, respectively. However, by applying ICA-TOPSIS, this alternative achieved the 21st position.

4.6 Conclusions

Dependent criteria are frequently observed in real situations formulated as multicriteria decision making problems. Since most MCDM methods do not consider this relation among criteria on the aggregation procedure, the ranking of alternatives may be biased towards an alternative that has good evaluations in dependent criteria. With the goal of mitigating this biased effect, this chapter proposed two different approaches, called ICA-TOPSIS and ICA-TOPSIS-M. In both approaches, the ICA technique can be seen as a preprocessing step whose aim is to extract information from the observed data in order to use as inputs of TOPSIS or TOPSIS-M methods.

By taking into account the experiments in synthetic data, in situations with two decision criteria, the obtained results indicate that both proposals performed better in comparison with TOPSIS and TOPSIS-M. By comparing ICA-TOPSIS and ICA-TOPSIS-M, the latter achieved the better results, especially with the application of JADE algorithm. The performance of ICA-TOPSIS-M based on JADE algorithm was also verified in scenarios with more than two decision criteria, which leads to the better results when we have moderate interference of noise and number of alternatives.

The application of the considered approaches on real data also provided interesting results. Since GNI and life expectancy at birth are dependent criteria, the use of TOPSIS directly on the observed decision data leads to a ranking in which the first alternatives are favored by good evaluations in these two criteria. TOPSIS-M, ICA-TOPSIS-M and ICA-TOPSIS improved the results of TOPSIS, leading to a ranking in which the importance of the other criteria was increased. An interesting aspect here, differently from the results on synthetic data, is that ICA-TOPSIS performed better in comparison with ICA-TOPSIS-M. A possible justification for this result is that ICA-TOPSIS could better exploit higher-order statistics from the collected data and, therefore, better capture the intercriteria relations. Recall that ICA-TOPSIS uses the estimated latent criteria in the TOPSIS method and ICA-TOPSIS-M only uses these data to extract both PIA and NIA.

It is important to highlight that the interpretation of the latent criteria depends on the problem that we are dealing with. For instance, in the problem of ranking students, we could associate the grades in calculus, physics and literature with competences in mathematics, natural sciences and linguistics. Moreover, in the experiment with real data conducted in the previous section, we could interpret the forest area, gross national income and life expectancy at birth as measures associated with environmental, economic and social aspects. However, we understand that this interpretation may not be straightforward in some applications. Therefore, it must be conducted case-by-case.

Recall that, in order to apply the proposed approaches, one should consider the assumptions described in Section 4.4. Therefore, a weakness of both ICA-TOPSIS and ICA-TOPSIS-M methods arises when these assumptions are not fulfilled, since the ambiguities provided by the ICA technique may not be mitigated. In that respect, future works may be conducted to exploit other signal processing techniques that can deal with these ambiguities in the context of MCDM problems. Moreover, in this chapter, we modeled and dealt with a linear dependence among criteria. Therefore, future perspectives also include the application of BSS methods that takes into account nonlinear relations among the decision data.

5 An unsupervised-LVA-based approach to adjust the WAM parameters

Let us recall that, as mentioned in Section 2.2.4, we should avoid redundancy among the decision data in order to associate each criterion to a different latent factor describing the alternatives. For example, in a situation in which we observe correlations among the decision data, this property is not achieved. As a consequence, one may have two or more criteria that measure the same latent factor.

Consider Example 2.3 (described in Section 2.4.1.2.1), which comprises the problem of ranking a set of students according to their grades in calculus, physics and literature. We argued that, since both grades in calculus and physics may be correlated, a student with a good performance in these two disciplines may achieve a better position in comparison with another one with more balanced grades. For instance, by using the weights $\mathbf{w} = [0.4, 0.4, 0.2]$, Student 1 (the one with the highest grades in both calculus and physics, but the worst in literature) achieved the first position.

In such situations, it is not convenient to remove either the calculus or the physics grades, since they bring relevant information for the decision problem. However, by neglecting the correlation between these criteria, we allow that the ranking is biased towards students with good academic performances in these two disciplines. Therefore, in order to keep the information carried by all criteria and, at the same time, to compensate compensate such a bias, it may be interesting to penalize the weights associated with both calculus and physics. For instance, if we decrease both w_1 and w_2 and increase w_3, leading to $\hat{\mathbf{w}} = [0.35, 0.35, 0.3]$, we obtain the results presented in Table 11. In this case, Student 3, who has the more balanced grades, achieves the first position.

Table 11 – Application of the WAM in Example 2.3, with $\hat{\mathbf{w}} = [0.35, 0.35, 0.3]$.

Students	$F_{WAM}(\cdot)$
Student 1	14.9
Student 2	13.1
Student 3	15.3

A remark in the aforementioned penalization is that the weights were redefined without any specific procedure. However, instead of randomly performing the adjustments in order to take into account the redundancies in the decision data, it may be interesting to adopt an automatic procedure to provide it. In this chapter, we propose an approach based on latent variable analysis. For this purpose, similarly as in the previous chapter, we also formulate the decision problem as a blind source separation one and investigate if the information obtained by a latent variable analysis technique (ICA and whitening, in this case) can be used to perform the adjustments. We detail our proposal in the sequel.

It is important to mention that preliminary results were presented in the 2nd conference paper described in Chapter 1.

5.1 Proposed approach

This section presents the proposed approach used to adjust the criteria weights. As mentioned in the previous section, we formulated the decision problem as a BSS problem. Therefore, we follow the same formulation discussed in the previous chapter (see Section 4.1). We assume that the redundancies that we observe in the decision data $\hat{\mathbf{M}}$ came from a mixture of latent variables $\mathbf{L}$, i.e., $\hat{\mathbf{M}} = \mathbf{AL}$. Since there is no redundancies among $\mathbf{L}$, we consider that these variables are the correct ones to apply the WAM. Therefore, the vector $\mathbf{r}$ which contains the overall scores for all alternatives can be obtained by

$$\mathbf{r} = \mathbf{w}^T\mathbf{L} = \mathbf{w}^T\mathbf{A}^{-1}\hat{\mathbf{M}} = \hat{\mathbf{w}}^T\hat{\mathbf{M}}, \tag{5.1}$$

where $\mathbf{w}$ is the initial weights vector provided by the decision maker[1] (through an elicitation method, for example) and $\hat{\mathbf{w}} = \left(\mathbf{w}^T\mathbf{A}^{-1}\right)^T$ is the adjusted weights vector which carries the information used to mitigate the bias introduced by correlations in the decision data. Therefore, by considering $\mathbf{w}$ in the overall values calculation, we allow the decision maker to express his preference over the set of criteria without taking into account redundancies among them. Moreover, remark that $\hat{\mathbf{w}}$ must be normalized in order to guarantee that $\sum_i^m \hat{w}_i = 1$.

Similarly as in the previous chapter, the goal here is to estimate the mixing matrix $\mathbf{A}$. Therefore, an ICA technique, which is based on higher-order statistics, is a good candidate for this task. However, we also intend to investigate if the information provided by second-order statistics is enough to adjust the weights and deal with redundancies. For this purpose, other than the separating matrix provided by an ICA technique, we also consider in our experiments the whitening matrix, which is described in Equation (3.17) (see Section 3.2.2.1). Note that, in this case, we do not solve the BSS problem. We only perform a preprocessing step and extract the information used in the adjustments. With respect to permutation and/or scaling ambiguities, we also adopted the procedure described in Section 4.4.1. Moreover, in order to avoid the scaling ambiguity provided by a positive factor, we divided all the elements in the same column by the associated

[1] Since in this thesis we consider that the number of latent variables is equal to the number of observed criteria, $\mathbf{w}$ may be elicited according to the latter. In the problem of ranking students, for example, $\mathbf{w}$ may reflect the decision maker opinion about the relative importance of each grade in the aggregation procedure.

diagonal element. By doing this procedure, all diagonal elements will be equal to 1 and all off-diagonal ones will be lower than 1 (in absolute value). For instance, if we consider the estimated mixing matrix achieve in Section 4.4.1, we will obtain

$$\hat{\mathbf{A}}^{Adj_c} = \begin{bmatrix} 1 & 0.7472 \\ -0.2273 & 1 \end{bmatrix}$$

5.2 Numerical experiments

In this section, we conduct numerical experiments in decision problems in order to verify the application of the proposed LVA-based-approach (for the ICA technique, we consider the JADE algorithm). We compare the obtained results with the techniques presented in Section 2.5.2.1, which also define the weights based on information extracted from the decision data.

5.2.1 On the use of ICA and whitening in a synthetic dataset

Aiming at comparing the use of ICA and whitening to extract the information used to adjust the weights, we applied both approaches in the synthetic dataset presented in Section 2.5.2.1 (see Figure 5), in which criteria 2 and 3 are positively correlated ($\rho_{2,3} \approx 0.7$). By assuming $\mathbf{w} = [1/3, 1/3, 1/3]$, we achieve the results presented in Table 12.

Table 12 – Weights adjusted by the proposed approaches and other objective methods.

Adjusted weights	**Method**				
	Entropy	Standard deviation	CRITIC	ICA	Whitening
$\hat{w}_1$	0.34	0.34	0.42	0.39	0.39
$\hat{w}_2$	0.33	0.33	0.29	0.32	0.30
$\hat{w}_3$	0.33	0.33	0.29	0.29	0.31

As already mentioned in Section 2.5.2.1, both entropy and standard deviation methods do not consider the correlation between pairs of criteria in the weights estimation. Therefore, these methods led to very similar values for all weights. On the other hand, by taking the CRITIC and our proposal, one may note that criteria 2 and 3 (the correlated ones) were penalized. However, if we compare the results provided by the application of ICA and whitening, we note that the latter led to a similar penalization for both redundant criteria. This can also be verified by the use of CRITIC method. A hypothesis for this difference is that, since ICA exploits higher-order statistics, this technique needs more data to achieve stable results. We investigate this issue in the next experiment.

5.2.2 Experiments varying the number of alternatives

Aiming at further investigating the difference that we may achieve by applying the approaches that take into account correlations in the decision data, we performed an experiment varying the number of alternatives from 50 to 1000 (the other parameters were the same as in the last experiment). Based on 1000 simulations, Figures 25, 26 and 27 present boxplots of the obtained weights[2].

It is evident that, as the number of alternatives increases, the results obtained by the application of ICA are more stable. However, even with a small dataset, the weights adjusted by using both CRITIC and whitening are stable. This suggests that the use of the latter methods, even if only a decorrelation procedure is taken into account, is enough to mitigate the redundancies. With respect to the adjusted weights in a stable scenario, we achieved very similar results, i.e., $\hat{\mathbf{w}}_{CRITIC} = [0.43, 0.29, 0.28]$, $\hat{\mathbf{w}}_{ICA} = [0.41, 0.29, 0.30]$ and $\hat{\mathbf{w}}_{White} = [0.41, 0.30, 0.29]$. In all cases, criteria 2 and 3 were penalized.

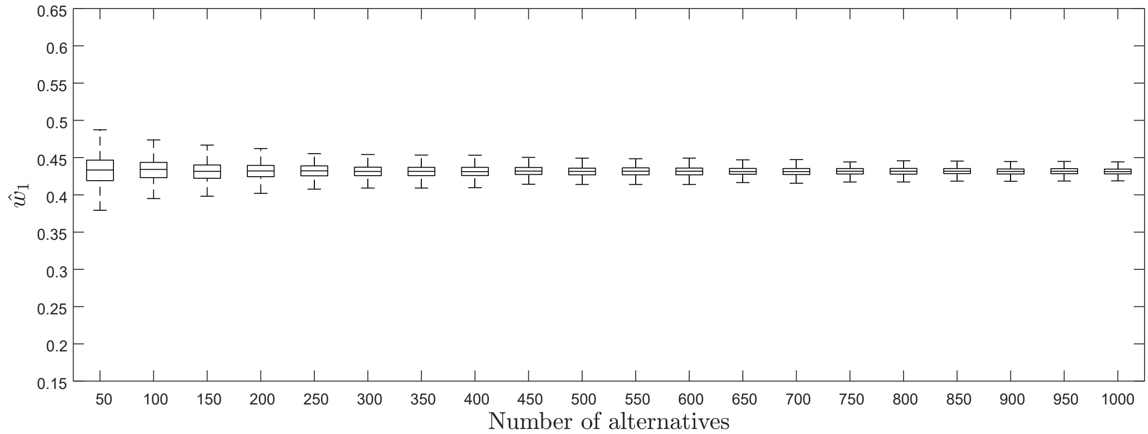
5.2.3 Experiments varying the degree of correlation

In the previous experiment, we varied the number of alternatives. On the other hand, in this section, we consider scenarios with different degrees of correlation between criteria 2 and 3. For this purpose, we generate decision problems with 500 alternatives and $\rho_{2,3}$ in the range $[-0.9, 0.9]$ (the other parameters were kept the same). Based on 1000 simulations, Figures 28, 29 and 30 present boxplots of the obtained weights.

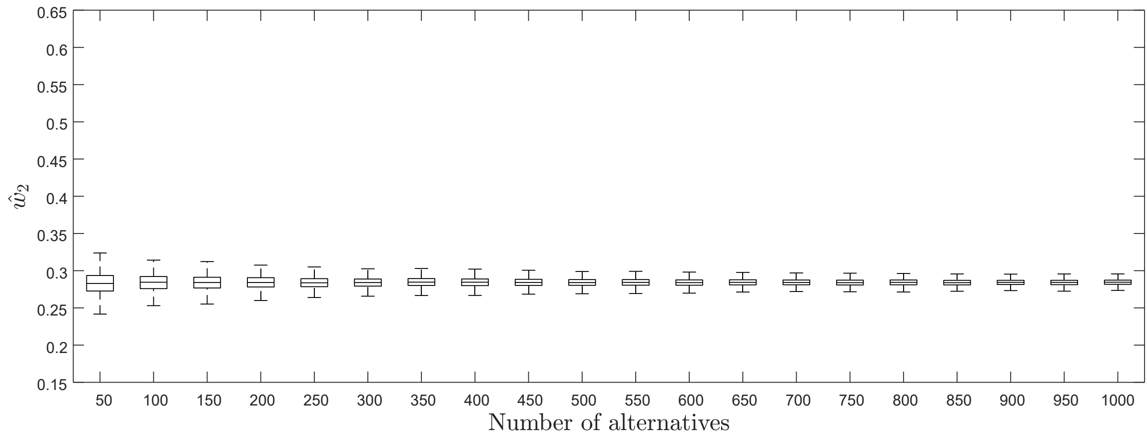
One may note that the higher is the correlation, the higher are the adjustments on the criteria weights. Moreover, the results are similar for all $\rho_{2,3} \geq 0$. However, if we analyze the scenarios in which the correlation coefficient is negative, the values of both $\hat{w}_2$ and $\hat{w}_3$ increase faster in our proposal in comparison with the CRITIC method (see a comparison in Figure 31). For instance, in the extreme case where $\rho_{2,3} \approx -0.9$, we achieve weights associated with criteria 2 and 3 close to 0.42. Therefore, our proposal favored negatively correlated criteria more in comparison with CRITIC method. Depending on the addressed situation, this may pose a problem since we associate a small weight to the independent criteria, which decreases its impact on the overall evaluations.

Another remark on the obtained results is that the weights adjusted by the ICA technique in scenarios with a strong correlation between criteria 2 and 3 ($\rho_{2,3} \approx -0.9$ and $\rho_{2,3} \approx 0.9$) are not stable even with 500 alternatives. Moreover, we achieved negative values. The explanation for such a result is that, since two criteria are practically the

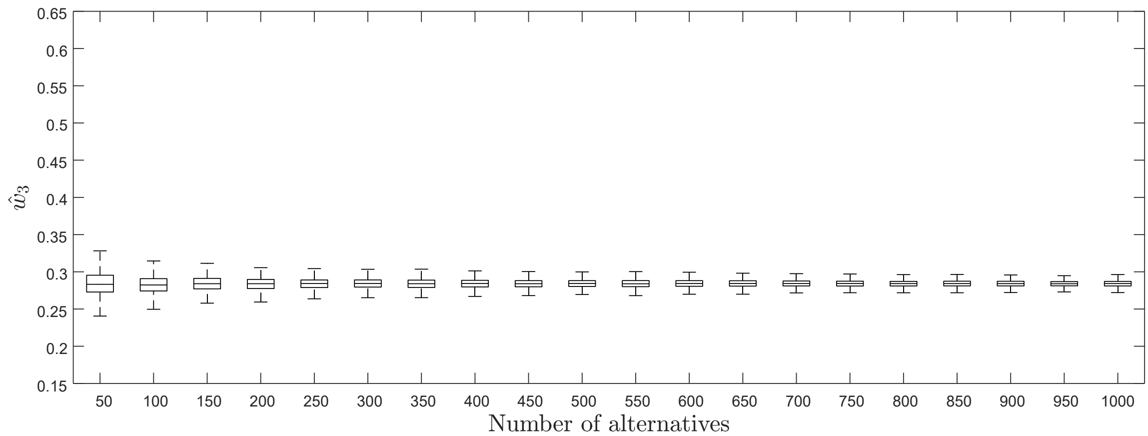
[2] In each box, the central mark, the top edge and the bottom edge represent the median (Q_2), the 75th percentile (Q_3) and the 25th percentile (Q_1), respectively. Moreover, the whiskers extend until the extremes points that lies in the range $[Q_3 - 1.5(Q_3 - Q_1), Q_1 + 1.5(Q_3 - Q_1)]$, which covers 99,3% of the points, approximately, if the data are normally distributed. The points outside this range are considered as outliers and were removed from the boxplot.



(a) Weight for criterion 1.



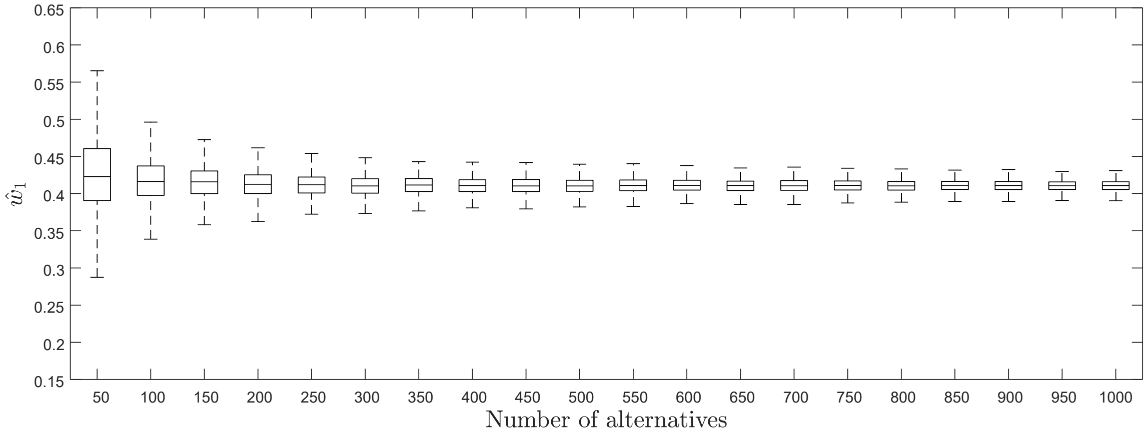
(b) Weight for criterion 2.



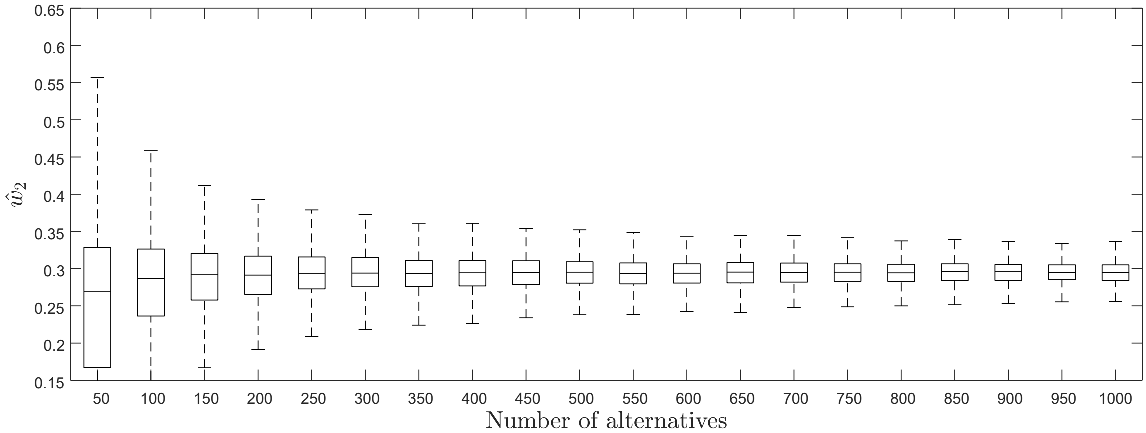
(c) Weight for criterion 3.

Figure 25 – Adjusted weights by using CRITIC method.

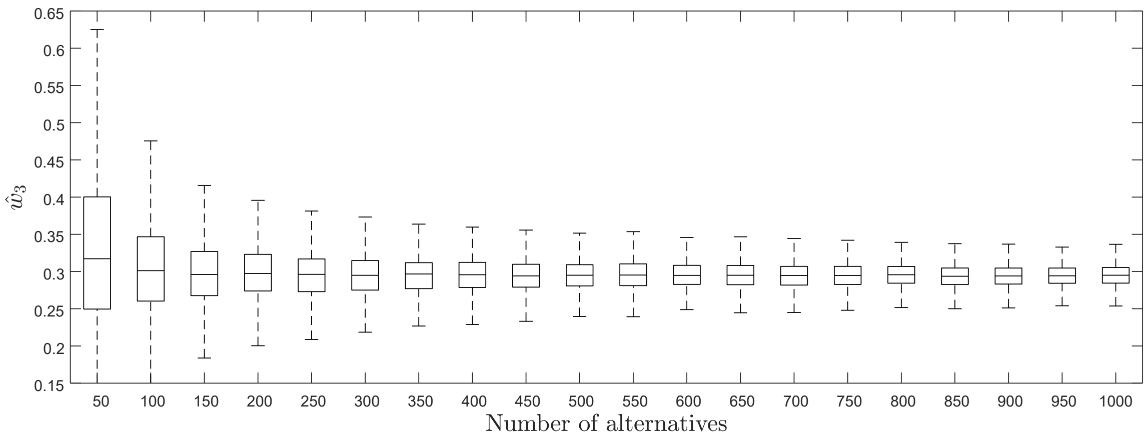
same (or the opposite, in a negative correlation), the estimated mixing matrix does not satisfy the assumptions used to mitigate the permutation and/or scaling ambiguities (see Section 4.4.1). As a consequence, a single latent factor will have a major influence in two observed criteria and, therefore, the weights will not be correctly adjusted.



(a) Weight for criterion 1.

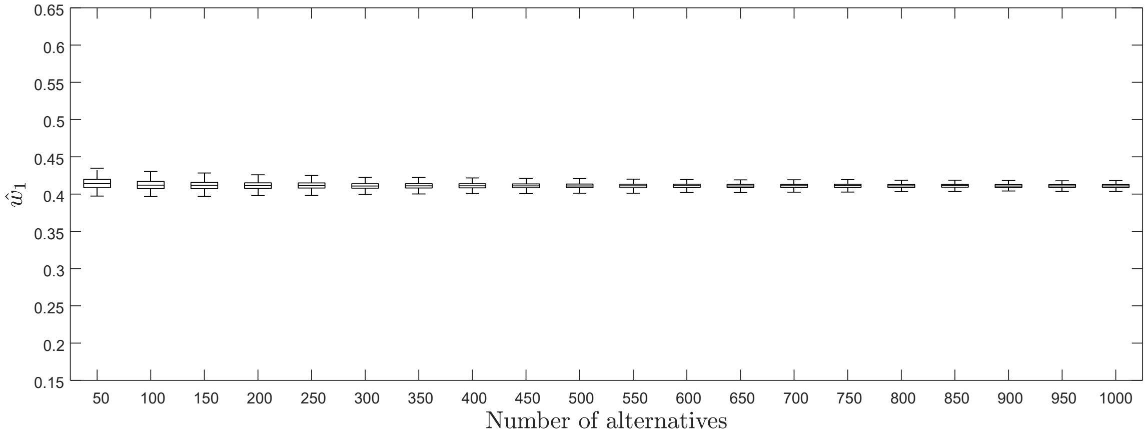


(b) Weight for criterion 2.

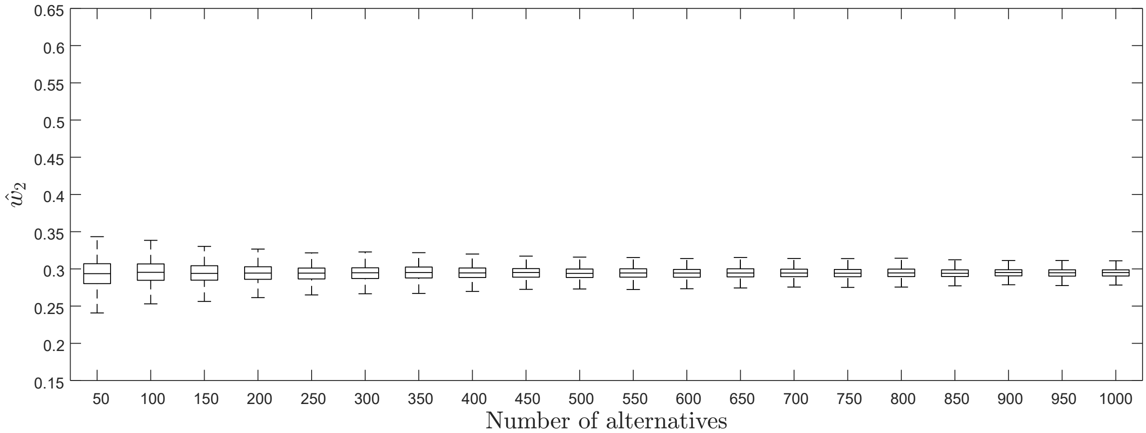


(c) Weight for criterion 3.

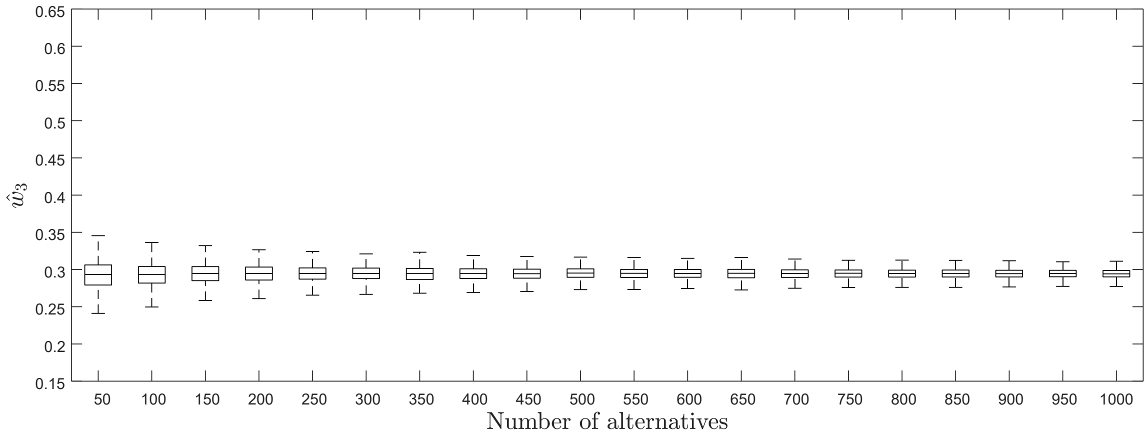
Figure 26 – Adjusted weights by using ICA.



(a) Weight for criterion 1.

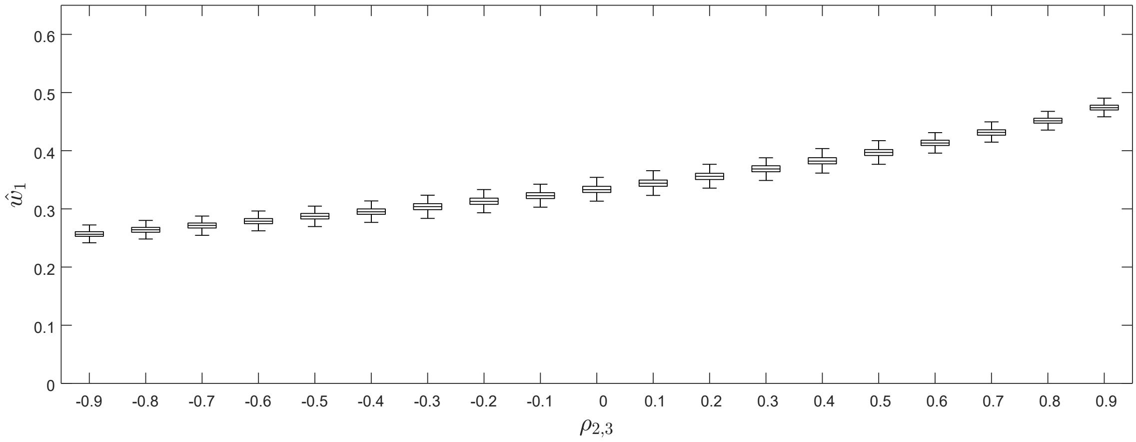


(b) Weight for criterion 2.

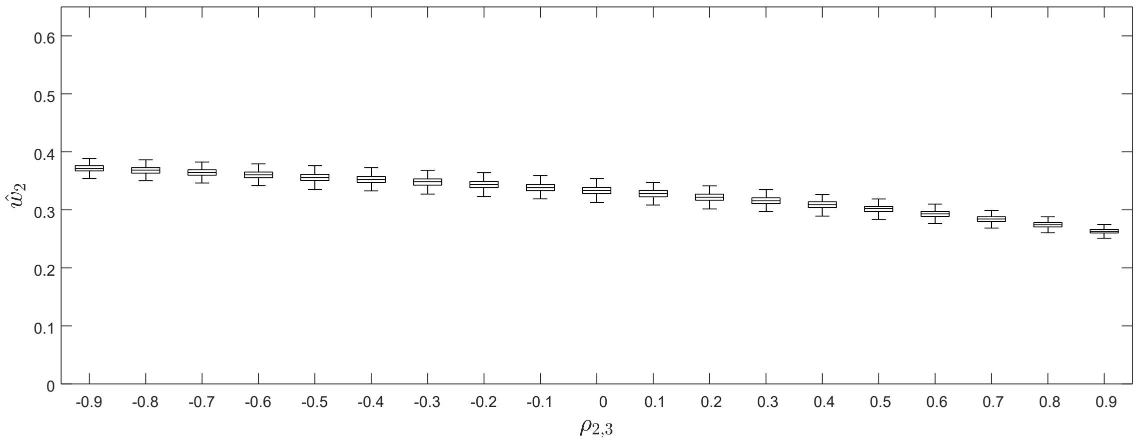


(c) Weight for criterion 3.

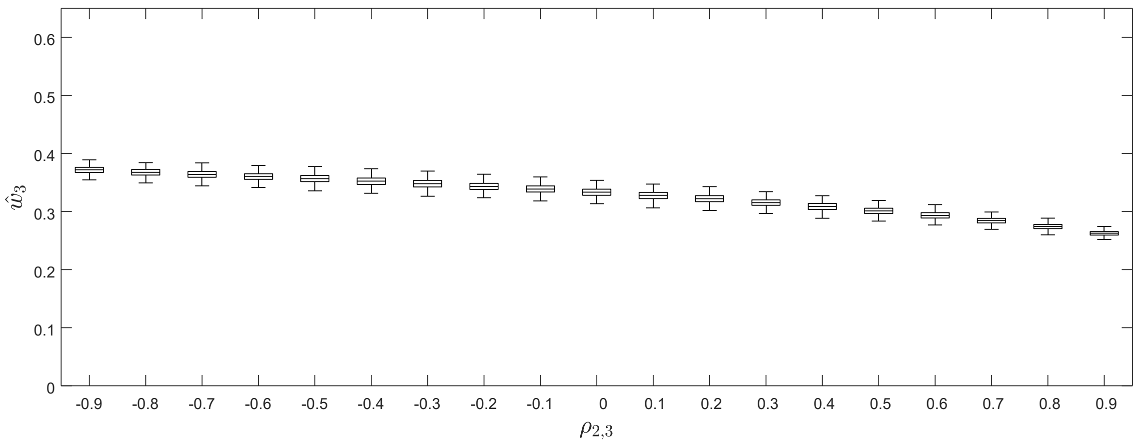
Figure 27 – Adjusted weights by using whitening.



(a) Weight for criterion 1.

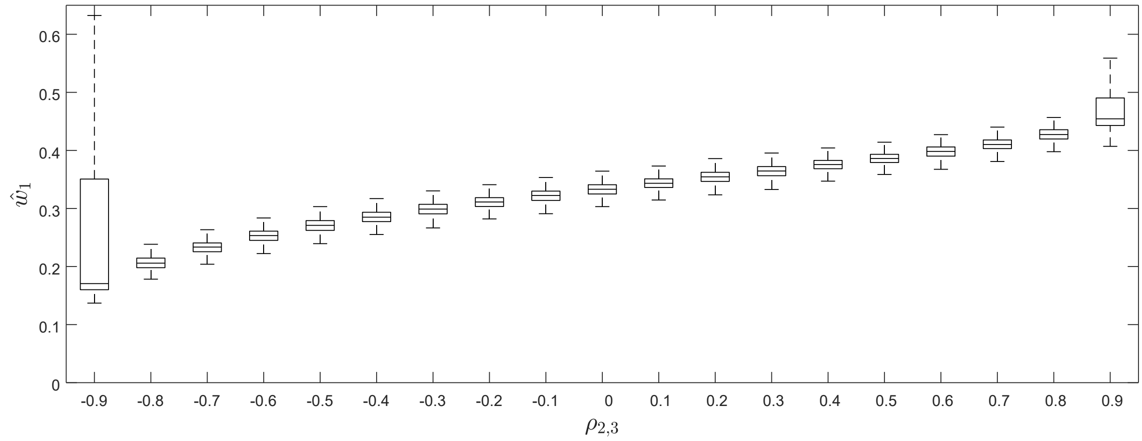


(b) Weight for criterion 2.

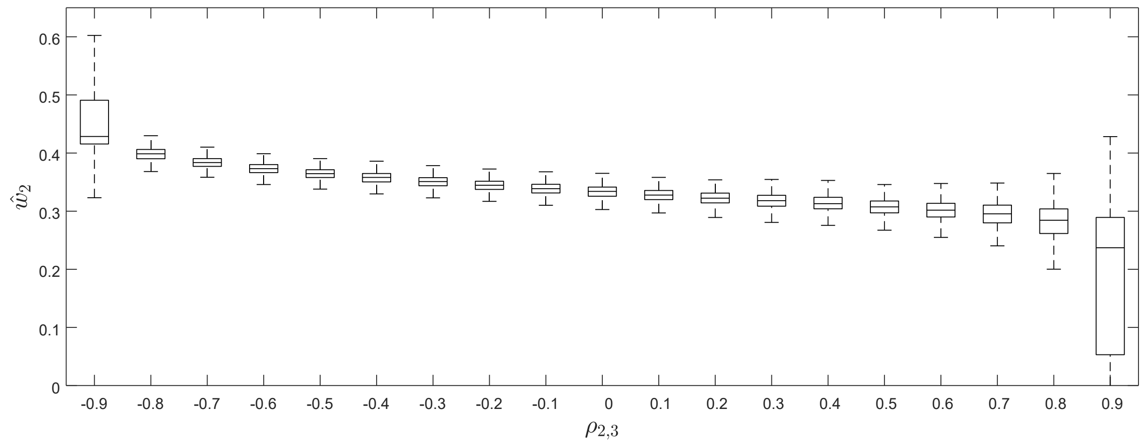


(c) Weight for criterion 3.

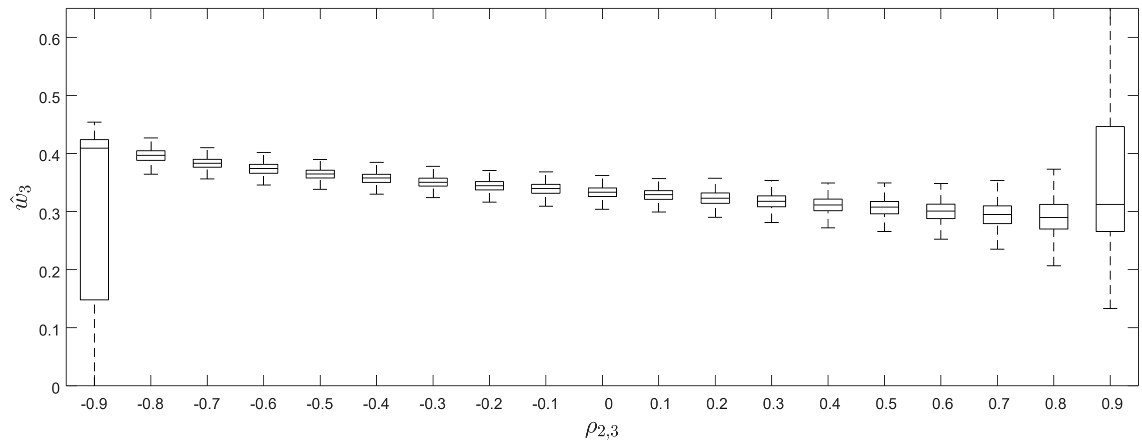
Figure 28 – Adjusted weights bu using CRITIC method.



(a) Weight for criterion 1.

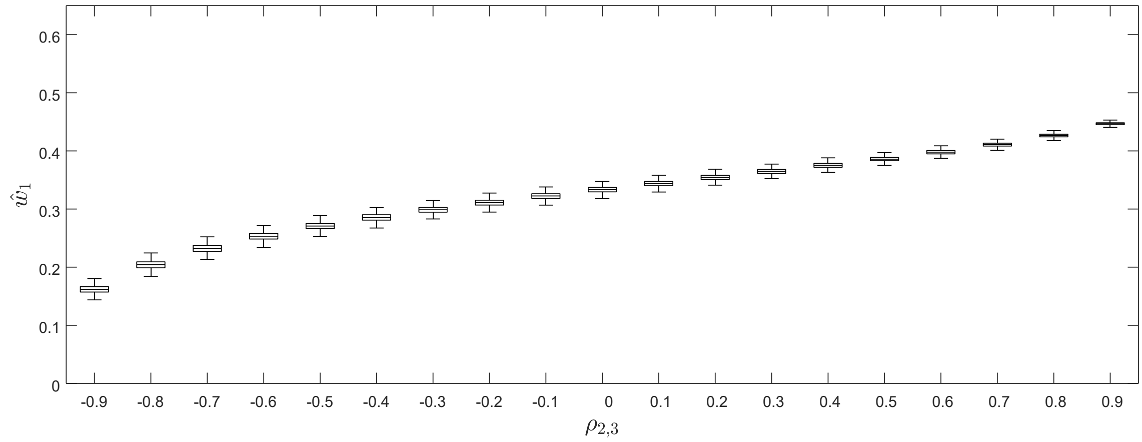


(b) Weight for criterion 2.

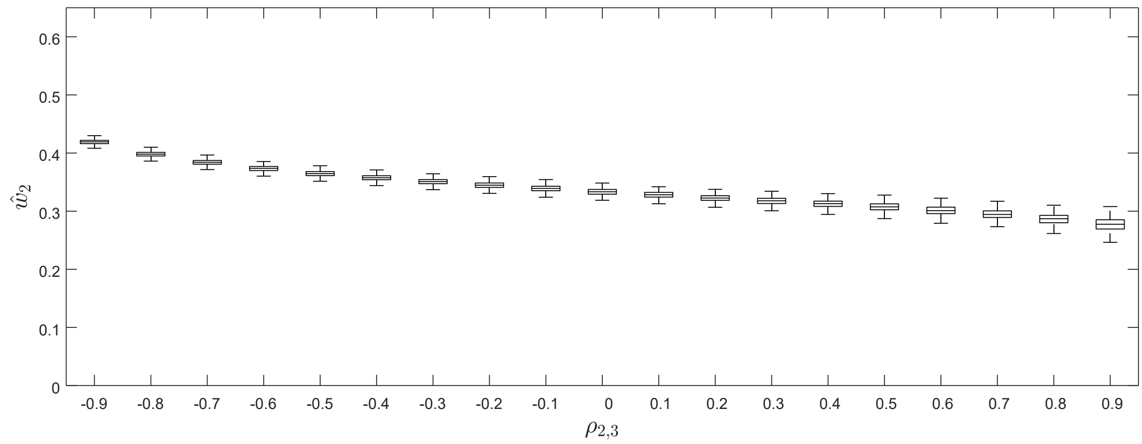


(c) Weight for criterion 3.

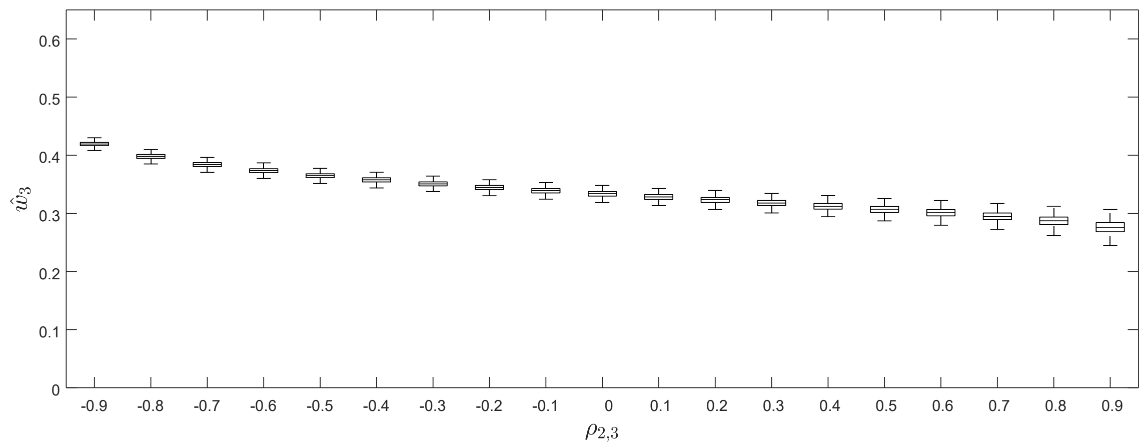
Figure 29 – Adjusted weights by using ICA.



(a) Weight for criterion 1.



(b) Weight for criterion 2.



(c) Weight for criterion 3.

Figure 30 – Adjusted weights by using whitening.

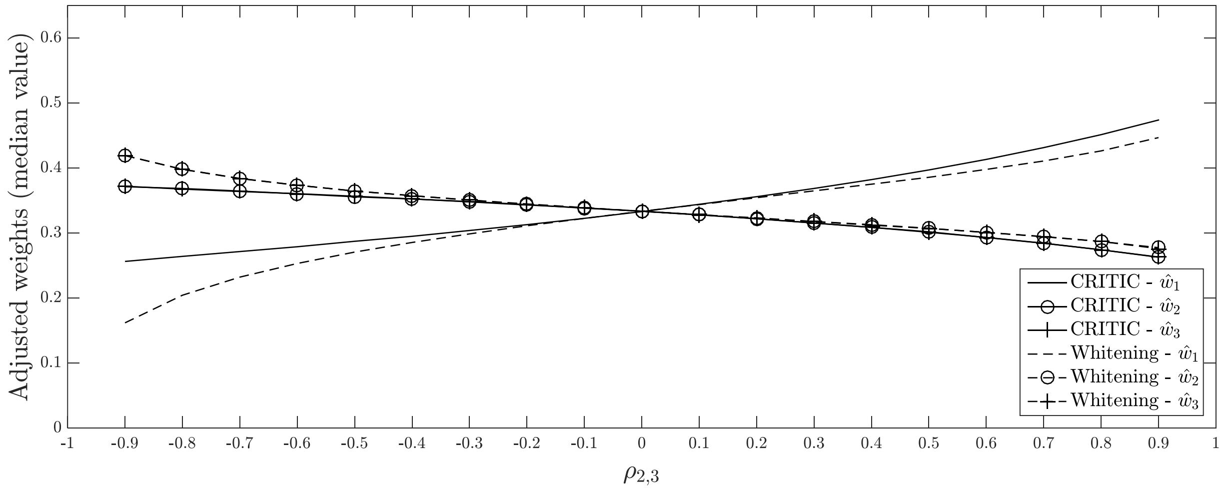


Figure 31 – Comparison between the CRITIC and whitening.

5.2.4 An extension to group MCDM

Different from the previous experiments, the aim in this section is to deal with correlations in the decision data in group MCDM problems. The group MCDM problem addressed here comprises the situation in which each decision maker d_k, $k = 1, \ldots, t$, provides his/her own criteria evaluations $u^k_{j,i}$ for all alternatives a_j, $j = 1, \ldots, n$ and all criteria c_i, $i = 1, \ldots, m$. In other words, each decision maker d_k provides his/her own decision matrix

$$\hat{\mathbf{M}}^k = \begin{bmatrix} u^k_{1,1} & u^k_{2,1} & \ldots & u^k_{n,1} \\ u^k_{1,2} & u^k_{2,2} & \ldots & u^k_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ u^k_{1,m} & u^k_{2,m} & \ldots & u^k_{n,m} \end{bmatrix} \tag{5.2}$$

Based on the collected information, one aggregates them according to predefined weights for criteria and DMs. Mathematically, the overall evaluation of alternative a_j is calculated by

$$r_j = \sum_{k=1}^{t} w^d_k \left(\sum_{i=1}^{m} w_i u^k_{j,i} \right) = \sum_{k=1}^{t} \sum_{i=1}^{m} \omega_{i,k} u^k_{j,i}, \tag{5.3}$$

where w^d_k ($w^d_k \geq 0$, $\forall k = 1, \ldots, t$, and $\sum_{k=1}^{t} w^d_k = 1$) is the weighting factor associated with d_k and $\omega_{i,k} = w^d_k w_i$.

In this scenario, our aim is to perform adjustments on the weights w_i and w^d_k in order to overcome bias introduced by correlations between criteria and between decision makers, respectively. Some group decision problems in the literature are solved by means of a consensus reaching process[3] (Dong et al., 2016), i.e., by finding a solution with a mutual agreement among the decision makers. In such situations, one may consider that

[3] In this thesis, we do not address the problem of reaching a consensus in group MCDM problem. Interested readers may refer to (Dong and Xu, 2016) for more details.

redundancies between decision makers may be neglected. However, one may have real scenarios in which there is a group of decision makers that, intentionally, combine their opinions in order to favor (or penalize) specific alternatives. Examples are the contests (e.g., sport and music), where a committee put scores on the performance of every participant, then these scores are averaged. If a subset of the committee members combine their scores, the achieved ranking may be biased towards the interests of such individuals. In this thesis, we consider that this should be avoided in order to achieve a fairest result.

With the purpose of extracting the statistics used in the considered approaches, we assume that, for each criterion, the evaluations provided by different DMs may be concatenated into a single vector. By doing this procedure for all criteria, one obtains the "extended criteria" matrix $\hat{\mathbf{M}}^{\mathcal{C}}$, defined by

$$\hat{\mathbf{M}}^{\mathcal{C}} = \left[\hat{\mathbf{M}}^{1}, \ldots, \hat{\mathbf{M}}^{t}\right]. \tag{5.4}$$

Similarly, we assume that, for each decision maker, the evaluations with respect to different criteria may also be concatenated into a single vector. In this case, one obtains the "extended decision makers" matrix $\hat{\mathbf{M}}^{\mathcal{D}}$, defined by

$$\hat{\mathbf{M}}^{\mathcal{D}} = \begin{bmatrix} \text{reshapeR}\left(\hat{\mathbf{M}}^{1}\right) \\ \vdots \\ \text{reshapeR}\left(\hat{\mathbf{M}}^{t}\right) \end{bmatrix}, \tag{5.5}$$

where the operator $\text{reshapeR}\left(\hat{\mathbf{M}}^{k}\right)$ converts the matrix $\hat{\mathbf{M}}^{k}$ into a row vector, i.e.,

$$\text{reshapeR}\left(\hat{\mathbf{M}}^{k}\right) = \left[u^{k}_{1,1}, u^{k}_{2,1}, \ldots, u^{k}_{n,1}, u^{k}_{1,2}, u^{k}_{2,2}, \ldots, u^{k}_{n,2}, \ldots, u^{k}_{1,m}, u^{k}_{2,m}, \ldots, u^{k}_{n,m}\right]. \tag{5.6}$$

Let us illustrate these assumptions in a simple example composed by 3 alternatives, 4 criteria and 3 decision makers, as presented in Figure 32.

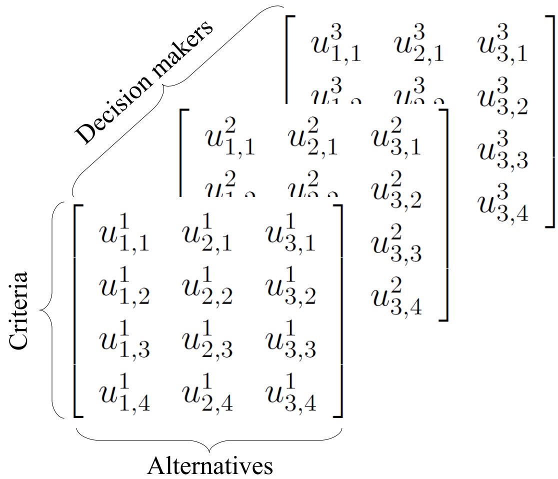


Figure 32 – Illustrative example of a dataset in group MCDM.

By concatenating these data, one obtains the following matrices:

$$\hat{\mathbf{M}}^{\mathcal{C}} = \begin{bmatrix} u_{1,1}^1 & u_{2,1}^1 & u_{3,1}^1 & u_{1,1}^2 & u_{2,1}^2 & u_{3,1}^2 & u_{1,1}^3 & u_{2,1}^3 & u_{3,1}^3 \\ u_{1,2}^1 & u_{2,2}^1 & u_{3,2}^1 & u_{1,2}^2 & u_{2,2}^2 & u_{3,2}^2 & u_{1,2}^3 & u_{2,2}^3 & u_{3,2}^3 \\ u_{1,3}^1 & u_{2,3}^1 & u_{3,3}^1 & u_{1,3}^2 & u_{2,3}^2 & u_{3,3}^2 & u_{1,3}^3 & u_{2,3}^3 & u_{3,3}^3 \\ u_{1,4}^1 & u_{2,4}^1 & u_{3,4}^1 & u_{1,4}^2 & u_{2,4}^2 & u_{3,4}^2 & u_{1,4}^3 & u_{2,4}^3 & u_{3,4}^3 \end{bmatrix} \tag{5.7}$$

and

$$\hat{\mathbf{M}}^{\mathcal{D}} = \begin{bmatrix} u_{1,1}^1 & u_{2,1}^1 & u_{3,1}^1 & u_{1,2}^1 & u_{2,2}^1 & u_{3,2}^1 & u_{1,3}^1 & u_{2,3}^1 & u_{3,3}^1 & u_{1,4}^1 & u_{2,4}^1 & u_{3,4}^1 \\ u_{1,1}^2 & u_{2,1}^2 & u_{3,1}^2 & u_{1,2}^2 & u_{2,2}^2 & u_{3,2}^2 & u_{1,3}^2 & u_{2,3}^2 & u_{3,3}^2 & u_{1,4}^2 & u_{2,4}^2 & u_{3,4}^2 \\ u_{1,1}^3 & u_{2,1}^3 & u_{3,1}^3 & u_{1,2}^3 & u_{2,2}^3 & u_{3,2}^3 & u_{1,3}^3 & u_{2,3}^3 & u_{3,3}^3 & u_{1,4}^3 & u_{2,4}^3 & u_{3,4}^3 \end{bmatrix}. \tag{5.8}$$

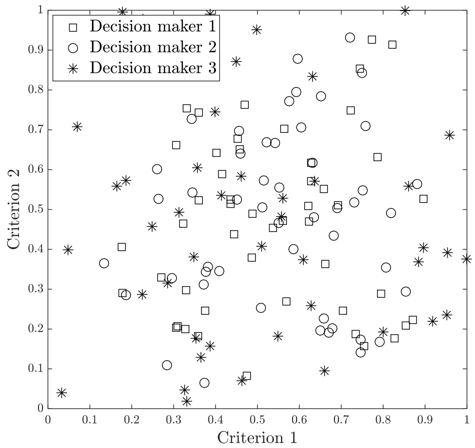
In order to attest our proposal in a group MCDM problem, we performed an experiment in a decision data comprised by 50 alternatives, 4 criteria and 3 decision makers. All data were randomly generated according to a uniform distribution in the range $[0, 1]$. Figures 33 and 34 present scatter plots of pairs of criteria and pairs of decision makers evaluations, respectively. One may note that criteria 2 and 3 and decision makers 1 and 2 are correlated ($\rho_{2,3} \approx 0.61$ and $\rho_{1,2}^d \approx 0.84$, respectively). In these figures, we also highlight the origin of each point in the concatenated vector.

Consider that, initially, we define $\mathbf{w} = [1/4, 1/4, 1/4, 1/4]$ and $\mathbf{w}^d = [1/3, 1/3, 1/3]$. With the application of the considered approaches on both $\hat{\mathbf{M}}^{\mathcal{C}}$ (to adjust the criteria weights w_i) and $\hat{\mathbf{M}}^{\mathcal{D}}$ (to adjust the decision makers weights w_k^d), one obtains the adjusted weights described in Table 13. Similarly as in the previous experiments, both CRITIC and the proposed approach based on the whitening procedure provided a penalization of correlated criteria and correlated decision makers.

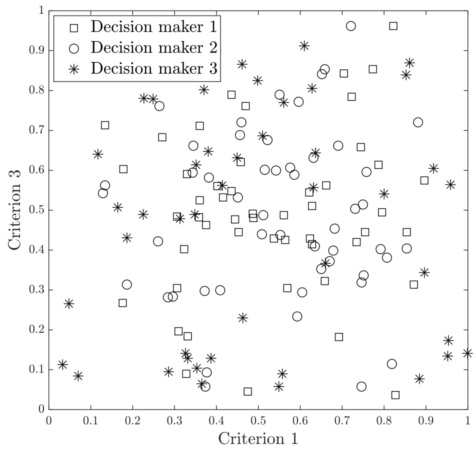
Table 13 – Criteria and decision makers weights adjusted by the considered approaches.

Methods	**Adjusted weights**						
	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	$\hat{w}_4$	$\hat{w}_1^d$	$\hat{w}_2^d$	$\hat{w}_3^d$
CRITIC	0.26	0.23	0.22	0.29	0.23	0.23	0.54
ICA	0.29	0.17	0.36	0.18	0.32	0.26	0.42
Whitening	0.29	0.21	0.20	0.30	0.29	0.28	0.43

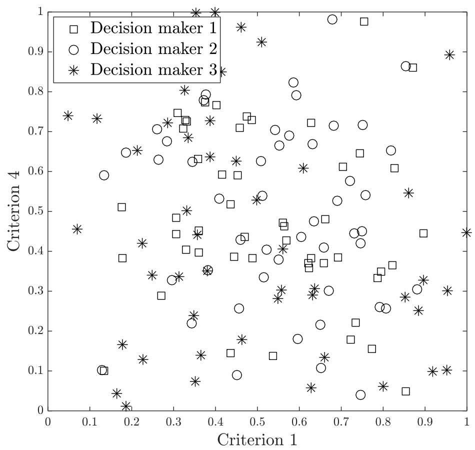
An important aspect that we would like to recall here is that the adjustments on the criteria and/or decision makers weights are associated with the degrees of redundancy between these entities. For instance, if the decision makers opinions are uncorrelated, our approach will not modify the DMs weights. Moreover, in a scenario in which all individuals share the same opinion (e.g., when we achieve consensus), the redundancies between different DMs (as well as the correlation between them) will be practically the same. Therefore, our approach will penalize all DMs with the same intensity and, after normalizing the weights, they will remain the same as they were initially defined.



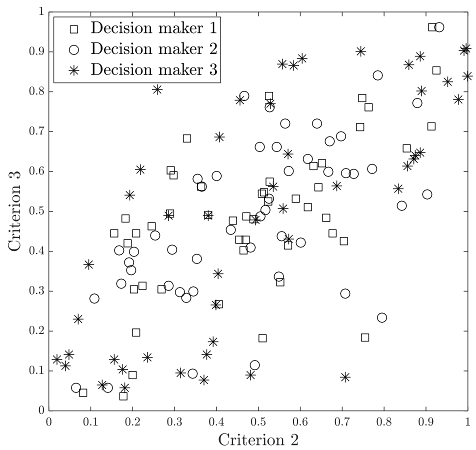
(a) Criteria 1 and 2.



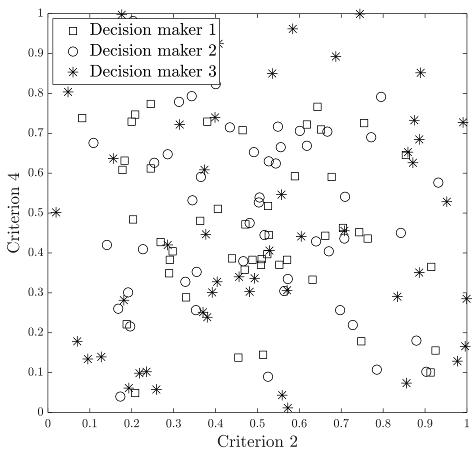
(b) Criteria 1 and 3.



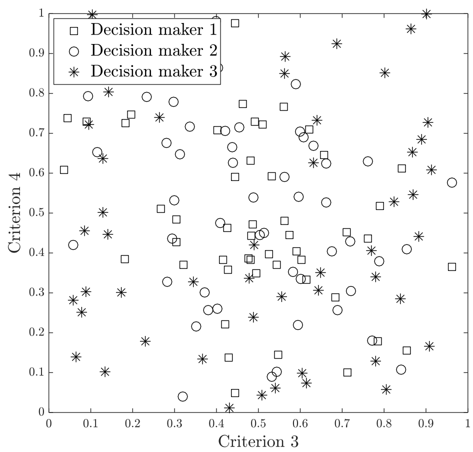
(c) Criteria 1 and 4.



(d) Criteria 2 and 3.



(e) Criteria 2 and 4.

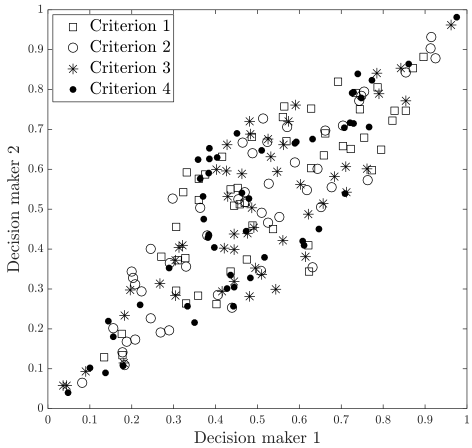


(f) Criteria 3 and 4.

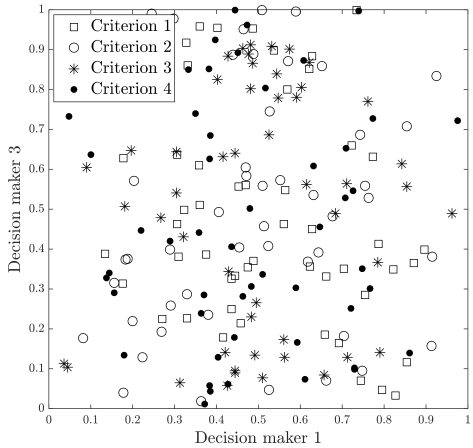
Figure 33 – Scatter plot of (concatenated) criteria.

5.3 Conclusions

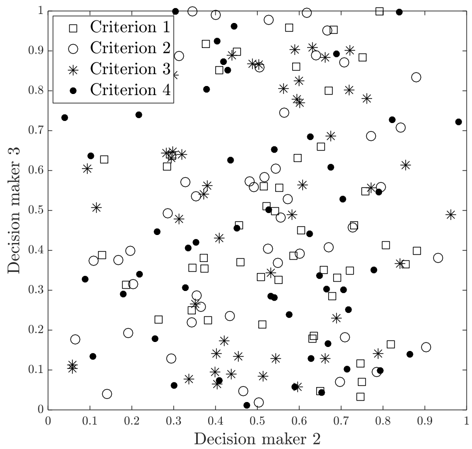
Similarly as in the previous contribution, this chapter proposed an approach that is able to deal with redundancies among criteria. However, we here combine our proposal with the weighted arithmetic mean. For instance, we applied both ICA and



(a) Decision makers 1 and 2.



(b) Decision makers 1 and 3.



(c) Decision makers 2 and 3.

Figure 34 – Scatter plot of (concatenated) decision makers evaluations.

whitening techniques to extract the information used to adjust the criteria weights. In situations in which two criteria are redundant, they are penalized and the weight associated with an independent criterion increases. Furthermore, in situations with no redundancies among criteria, the whitening matrix is diagonal and, therefore, no adjustments will be performed.

An interesting aspect in this contribution is that it can be easily explained by the decision maker. Although we use statistical techniques to perform the adjustments, the idea of penalizing redundant criteria according to their degree of correlation is clear. Therefore, it is convenient to the individuals that feel more comfortable to use a decision method that they better understand.

Other than ICA and whitening, we also considered in the experiments the CRITIC method. The results obtained by both CRITIC and whitening are similar for low to medium degrees of correlation. However, when this relation is strong, our approach penalizes (or favors, in the case of a strong negative correlation) the associated criteria more than the CRITIC method. Although both methods exploit second-order statistics, they use this information differently. CRITIC method perform the adjustments based on the linear correlation coefficients. In whitening, we conduct a decorrelation procedure followed by a power normalization. Therefore, we consider that whitening captures more information to adjust the criteria weights.

By comparing both whitening and ICA on synthetic data, we note that the former led to better results, specially in terms of the stability in the adjusted weights, in scenarios with small number of data. If we have a large number of data, they perform in a similar way. Remark that this result was achieved based on simulations in which the data was generated with a linear dependence between criteria 2 and 3. Therefore, whitening was sufficient for removing the redundancy. However, this conclusion may be different if one considers a dataset with nonlinear correlations. In such a scenario, new simulations can be conducted in order to verify if the ICA may lead to better results in comparison with both CRITIC and whitening. Moreover, one may also have different results comparing CRITIC and whitening.

We also extended our proposal to the group MCDM, in which the weights associated with decision makers can be adjusted in order to penalize the ill-intentioned ones who have combined their opinions. Therefore, in this scenario, we provided adjustments in both criteria and decision makers weights by applying our proposal twice. However, since the considered dataset can be seen as a fusion of multimodal data, novel works can be conducted to investigate the application of Joint Blind Source Separation methods (Chen et al., 2016), which can handle the similarities among both criteria and decision makers simultaneously.

Other future works may be conducted to exploit the case where the number of latent variables is different from the number of observed criteria. This scenario may be tricky for the decision maker, since he/she will need to either define the initial weights by assuming an additional information about the latent variables or, in the absence of such an information, set the same value for all weights. Another study that can be conducted consists in addressing the case in which the criteria are negatively correlated. As discussed by Roy (2009) and mentioned in Section 2.2.4, one firstly needs to verify if this redundancy deserves to be taken into account or may be neglected. If one decides that it is important to consider this correlation, a constraint may be introduced into the separating (or whitening) process in order to avoid that the adjustments lead to weights close to zero (or even negative). Moreover, we also may cite as a future perspective the investigation if it is possible to extend our proposal to estimate capacity coefficients. In this case, an idea consists in allowing the decision maker to provide the parameters associated with singletons and use the approach to estimate interaction indices of pairs of criteria.

6 The 2-additive multilinear model and supervised parameter identification

In Section 2.4.1.2, we defined the Choquet integral and the multilinear model. Moreover, in Section 2.4.1.2.2, we presented the 2-additive Choquet integral and argued that such an assumption drastically reduces the number of parameters used in this aggregation function. As commented in Section 2.5.1.2, this is an important issue, since it also reduces the computational effort of parameter identification algorithms.

A remark in the literature is that most of the works that consider a 2-additive capacity use such an assumption in the context of the Choquet integral. Similarly, the ones that deal with the problem of capacity identification also consider this aggregation function (Grabisch et al., 2008). For this purpose, most of them adopted a supervised approach (Grabisch, 1995; Marichal and Roubens, 2000; Miranda et al., 2003; Combarro and Miranda, 2006; Angilella et al., 2015) or a supervised approach with regularization (Anderson et al., 2014; Adeyeba et al., 2015; Oliveira et al., 2017). However, one does not find the aforementioned efforts in the context of the multilinear model.

Motivated by this lack in the literature, we propose in this thesis (i) to exploit the concept of a 2-additive capacity and (ii) to address the problem of capacity identification with respect to the multilinear model. For instance, a formalization of a 2-additive multilinear model, which is, as far as we know, unknown in the literature, is presented in Section 6.1. We also discuss the similarities between the obtained expression and the one in the context of the Choquet integral. Moreover, in Section 6.1.1, we present a graphical interpretation in a situation with 2 decision criteria.

With respect to the capacity identification problem, we extended the existing supervised approaches to the multilinear model case. The optimization models without and with a regularization term are presented in Sections 6.2.1 and 6.2.2, respectively. It is worth mentioning that the analysis conducted in this chapter as well as the obtained results were published in the 3rd conference paper (preliminary results) and in the 2nd journal article (a complete analysis) mentioned in Chapter 1.

6.1 The 2-additive multilinear model

In order to formalize the 2-additive multilinear model, let us consider an alternative representation of this function, given by (see (Grabisch, 2016), Chapter 6, for more details)

$$F_{ML}(u(\mathbf{x})) = \sum_{A \subseteq C} m^{\mu}(A) \prod_{i \in A} u_i, \tag{6.1}$$

where $m^{\mu}(A)$, defined by

$$m^{\mu}(A) = \sum_{B \subseteq A} (-1)^{|A \backslash B|} \mu(B), \; \forall A \subseteq C, \tag{6.2}$$

is the Möbius transform of μ (Rota, 1964). Since (6.2) comprises a linear transformation, one can easily invert it through the following expression:

$$\mu(A) = \sum_{B \subseteq A} m^{\mu}(B), \; \forall A \subseteq C. \tag{6.3}$$

Moreover, one may also express $m^{\mu}(A)$ in terms of $I^{\mathcal{B}}(B)$, i.e.,

$$m^{\mu}(A) = \sum_{B \supseteq A} \left(-\frac{1}{2}\right)^{|B|-|A|} I^{\mathcal{B}}(B), \; \forall A \subseteq C. \tag{6.4}$$

The definition of a 2-additive capacity (see Definition 2.11) can be extended to the Möbius transform. Indeed, a 2-additive capacity is such that $m^{\mu}(A) = 0$ for any $A \subseteq C$ and $|A| > 2$ (Grabisch, 1997a). By using this result and the aforementioned transformations, one achieves the following:

$$\begin{aligned} F_{ML}(u(\mathbf{x})) &= \sum_{\substack{A \subseteq C, \\ |A| \leq 2}} m^{\mu}(A) \prod_{i \in A} u_i \\ &= \sum_{\substack{A \subseteq C, \\ |A| \leq 2}} \sum_{\substack{B \supseteq A, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|-|A|} I^{\mathcal{B}}(B) \prod_{i \in A} u_i \end{aligned}$$

$$= \sum_{\substack{B, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) + \sum_{i} \sum_{\substack{B \ni i, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|-1} I^{\mathcal{B}}(B) u_i + \sum_{\{i,i'\}} \sum_{\substack{B \supseteq \{i,i'\}, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|-2} I^{\mathcal{B}}(B) u_i u_{i'}$$

$$= I^{\mathcal{B}}(\emptyset) - \frac{1}{2} \sum_{i} \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i,i'} I^{\mathcal{B}}_{i,i'} + \sum_{i} u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i'} I^{\mathcal{B}}_{i,i'} \right) + \sum_{i,i'} u_i u_{i'} I^{\mathcal{B}}_{i,i'}.$$

By assuming the 2-additive multilinear model, the axioms of a capacity can be defined by (for a detailed mathematical manipulation, see Appendix A):

- $\mu(\emptyset) = 0 \rightarrow I^{\mathcal{B}}(\emptyset) - \frac{1}{2} \sum\limits_{i} \phi_i^{\mathcal{B}} + \frac{1}{4} \sum\limits_{i,i'} I^{\mathcal{B}}_{i,i'} = 0;$
- $\mu(C) = 1 \rightarrow I^{\mathcal{B}}(\emptyset) + \frac{1}{2} \sum\limits_{i} \phi_i^{\mathcal{B}} + \frac{1}{4} \sum\limits_{i,i'} I^{\mathcal{B}}_{i,i'} = 1;$
- $\mu(\{A \cup i\}) - \mu(A) \geq 0 \rightarrow \phi_i^{\mathcal{B}} - \frac{1}{2} \sum\limits_{i' \neq i} \left| I^{\mathcal{B}}_{i,i'} \right| \geq 0, \; \forall i \in C.$

Based on these axioms and on the expression

$$\sum_{i'} I_{i,i'}^{\mathcal{B}} = \sum_{i'} \left|I_{i,i'}^{\mathcal{B}}\right| - 2 \sum_{\substack{i', \\ I_{i,i'}^{\mathcal{B}}<0}} \left|I_{i,i'}^{\mathcal{B}}\right|,$$

for every $i \in C$ (recall that $I_{i,i'}^{\mathcal{B}}$ may be positive or negative), the multilinear model can be written as

$$\begin{aligned}
F_{ML}(u(\mathbf{x})) &= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2}\sum_{i'} I_{i,i'}^{\mathcal{B}}\right) + \sum_{i,i'} u_i u_{i'} I_{i,i'}^{\mathcal{B}} \\
&= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2}\sum_{i'} \left|I_{i,i'}^{\mathcal{B}}\right|\right) + \sum_i u_i \sum_{i', I_{i,i'}^{\mathcal{B}}<0} \left|I_{i,i'}^{\mathcal{B}}\right| + \sum_{i,i'} u_i u_{i'} I_{i,i'}^{\mathcal{B}} \\
&= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2}\sum_{i'} \left|I_{i,i'}^{\mathcal{B}}\right|\right) + \sum_{I_{i,i'}^{\mathcal{B}}<0} (u_i + u_{i'}) \left|I_{i,i'}^{\mathcal{B}}\right| + \sum_{I_{i,i'}^{\mathcal{B}}>0} u_i u_{i'} I_{i,i'}^{\mathcal{B}} - \sum_{I_{i,i'}^{\mathcal{B}}<0} u_i u_{i'} \left|I_{i,i'}^{\mathcal{B}}\right| \\
&= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2}\sum_{i'} \left|I_{i,i'}^{\mathcal{B}}\right|\right) + \sum_{I_{i,i'}^{\mathcal{B}}<0} (u_i + u_{i'} - u_i u_{i'}) \left|I_{i,i'}^{\mathcal{B}}\right| + \sum_{I_{i,i'}^{\mathcal{B}}>0} u_i u_{i'} I_{i,i'}^{\mathcal{B}}.
\end{aligned}$$

An interesting aspect in the above expression is that it comprises an additive term, a disjunctive term $\mathbf{N_p}(u_i, u_{i'}) = u_i + u_{i'} - u_i u_{i'}$ and a conjunctive term $\mathbf{T_p}(u_i, u_{i'}) = u_i u_{i'}$. More specifically, the disjunctive term turns out to be the t-conorm usually called probabilistic sum of $(u_i, u_{i'})$ and the conjunctive term is the product t-norm of $(u_i, u_{i'})$ (Klement et al., 2000; Beliakov et al., 2007). Therefore, the 2-additive multilinear model may be defined by

$$F_{ML}(u(\mathbf{x})) = \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2}\sum_{i'} \left|I_{i,i'}^{\mathcal{B}}\right|\right) + \sum_{I_{i,i'}^{\mathcal{B}}<0} \mathbf{N_p}(u_i, u_{i'}) \left|I_{i,i'}^{\mathcal{B}}\right| + \sum_{I_{i,i'}^{\mathcal{B}}>0} \mathbf{T_p}(u_i, u_{i'}) I_{i,i'}^{\mathcal{B}}. \quad (6.5)$$

Another remarkable aspect is that Equation (6.5) is similar to the 2-additive expression of the Choquet integral (Grabisch, 1997b), which is given by

$$F_{CI}(u(\mathbf{x})) = \sum_i u_i \left(\phi_i^{\mathcal{S}} - \frac{1}{2}\sum_{i'} \left|I_{i,i'}^{\mathcal{S}}\right|\right) + \sum_{I_{i,i'}^{\mathcal{S}}<0} (u_i \vee u_{i'}) \left|I_{i,i'}^{\mathcal{S}}\right| + \sum_{I_{i,i'}^{\mathcal{S}}>0} (u_i \wedge u_{i'}) I_{i,i'}^{\mathcal{S}}, \quad (6.6)$$

where $\phi_i^{\mathcal{S}}$ and $I_{i,i'}^{\mathcal{S}}$ are the power and the interaction indices based on the Shapley value (Shapley, 1953; Grabisch, 1997a), respectively, $\vee$ represents the maximum operator (which is also a t-conorm) and $\wedge$ represents the minimum operator (which is also a t-norm). In that respect, since for a 2-additive capacity $\phi_i^{\mathcal{B}} = \phi_i^{\mathcal{S}} =: \phi_i$ and $I_{i,i'}^{\mathcal{B}} = I_{i,i'}^{\mathcal{S}} =: I_{i,i'}$ (Marichal and Roubens, 2000), one may generalize both aggregation functions by the following expression:

$$F_{\mathbf{N},\mathbf{T}}(u(\mathbf{x})) = \sum_i u_i \left(\phi_i - \frac{1}{2}\sum_{i'} |I_{i,i'}|\right) + \sum_{I_{i,i'}<0} \mathbf{N}(u_i, u_{i'}) |I_{i,i'}| + \sum_{I_{i,i'}>0} \mathbf{T}(u_i, u_{i'}) I_{i,i'}, \quad (6.7)$$

where $\mathbf{N}$ is a t-conorm and $\mathbf{T}$ is a t-norm

6.1.1 A graphical interpretation when $m = 2$

Similarly as presented in (Grabisch, 2000), we provide in this section a graphical interpretation of the 2-additive multilinear model. In a situation where $m = 2$, this aggregation function can be defined by

$$F_{ML}(u(\mathbf{x})) = \begin{cases} u_1\left(\phi_1^{\mathcal{B}} - \frac{1}{2}I_{1,2}^{\mathcal{B}}\right) + u_2\left(\phi_2^{\mathcal{B}} - \frac{1}{2}I_{1,2}^{\mathcal{B}}\right) + \mathbf{T_p}(u_1,u_2)I_{1,2}^{\mathcal{B}}, & I_{1,2}^{\mathcal{B}} \geq 0 \\ u_1\left(\phi_1^{\mathcal{B}} + \frac{1}{2}I_{1,2}^{\mathcal{B}}\right) + u_2\left(\phi_2^{\mathcal{B}} + \frac{1}{2}I_{1,2}^{\mathcal{B}}\right) + \mathbf{N_p}(u_1,u_2)\left|I_{1,2}^{\mathcal{B}}\right|, & I_{1,2}^{\mathcal{B}} \leq 0 \end{cases}. \tag{6.8}$$

With respect to the axioms of a capacity, the considered scenario leads to the following:

- $I^{\mathcal{B}}(\emptyset) - \frac{1}{2}\sum_i \phi_i^{\mathcal{B}} + \frac{1}{4}\sum_{i,i'} I_{i,i'}^{\mathcal{B}} = 0 \rightarrow I^{\mathcal{B}}(\emptyset) - \frac{1}{2}\left(\phi_1^{\mathcal{B}} + \phi_2^{\mathcal{B}}\right) + \frac{1}{4}I_{1,2}^{\mathcal{B}} = 0;$
- $I^{\mathcal{B}}(\emptyset) + \frac{1}{2}\sum_i \phi_i^{\mathcal{B}} + \frac{1}{4}\sum_{i,i'} I_{i,i'}^{\mathcal{B}} = 1 \rightarrow I^{\mathcal{B}}(\emptyset) + \frac{1}{2}\left(\phi_1^{\mathcal{B}} + \phi_2^{\mathcal{B}}\right) + \frac{1}{4}I_{1,2}^{\mathcal{B}} = 1;$
- $\phi_i^{\mathcal{B}} - \frac{1}{2}\sum_{i' \neq i}\left|I_{i,i'}^{\mathcal{B}}\right| \geq 0 \rightarrow \phi_i^{\mathcal{B}} \pm \frac{1}{2}I_{1,2}^{\mathcal{B}} \geq 0, \forall\, i \in \{1,2\}.$

One may note that these conditions involve four different parameters. However, by manipulating the first two, one eliminates $I^{\mathcal{B}}(\emptyset)$ and achieves that $\phi_1^{\mathcal{B}} + \phi_2^{\mathcal{B}} = 1$. Therefore, it remains that this scenario can be expressed in the $\left(\phi_1^{\mathcal{B}}, I_{1,2}^{\mathcal{B}}\right)$ coordinates (remark that, given $\phi_1^{\mathcal{B}}$, one obtains $\phi_2^{\mathcal{B}}$), which is illustrated in Figure 35.

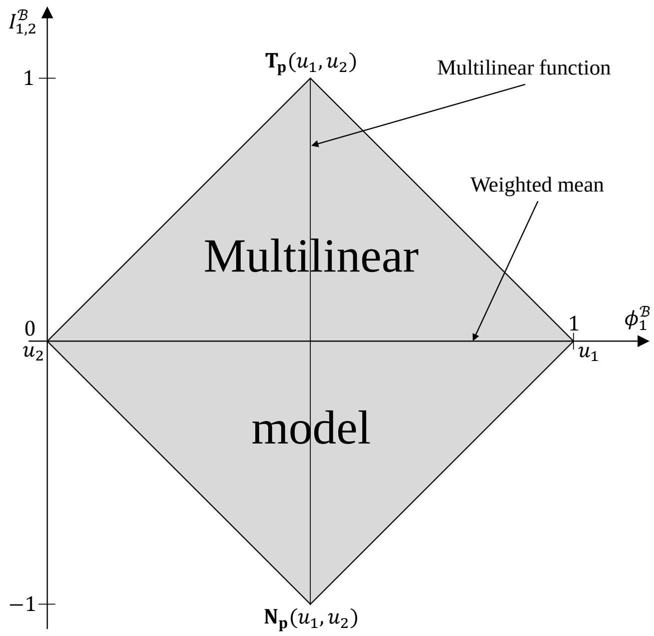


Figure 35 – Graphical interpretation of multilinear model when $m = 2$ criteria.

The horizontal axis comprises the situation in which $I_{1,2}^{\mathcal{B}} = 0$, i.e., there is no interaction between both criteria. Therefore, Equation (6.8) becomes

$$F_{ML}(u(\mathbf{x})) = u_1\phi_1^{\mathcal{B}} + u_2\phi_2^{\mathcal{B}}, \tag{6.9}$$

i.e., a weighted mean. Moreover, one may remark that, for the extreme cases of $\phi_1^{\mathcal{B}} = 0$ or $\phi_2^{\mathcal{B}} = 0$, one obtains $F_{ML}(u(\mathbf{x})) = u_2$ or $F_{ML}(u(\mathbf{x})) = u_1$, respectively.

On the other hand, the vertical axis comprises the situation in which the power indices of both criteria are the same, i.e., $\phi_1^{\mathcal{B}} = \phi_2^{\mathcal{B}} = 1/2$. In this scenario, Equation (6.8) becomes

$$F_{ML}(u(\mathbf{x})) = u_1 \left(\frac{1}{2} - \frac{1}{2} I_{1,2}^{\mathcal{B}} \right) + u_2 \left(\frac{1}{2} - \frac{1}{2} I_{1,2}^{\mathcal{B}} \right) + u_1 u_2 I_{1,2}^{\mathcal{B}}, \tag{6.10}$$

i.e., a bilinear function (Keeney and Raiffa, 1976) with weights $\left(\frac{1}{2} - \frac{1}{2} I_{1,2}^{\mathcal{B}} \right)$, $\left(\frac{1}{2} - \frac{1}{2} I_{1,2}^{\mathcal{B}} \right)$ and $I_{1,2}^{\mathcal{B}}$. One also may remark that, for the extreme cases of $I_{1,2}^{\mathcal{B}} = 1$ or $I_{1,2}^{\mathcal{B}} = -1$, one obtains $F_{ML}(u(\mathbf{x})) = \mathbf{T_p}(u_1, u_2) = u_1 u_2$ or $F_{ML}(u(\mathbf{x})) = \mathbf{N_p}(u_1, u_2) = u_1 + u_2 - u_1 u_2$, respectively.

6.2 The problem of capacity identification

The second main contribution of this chapter lies in the problem of capacity identification in the context of the multilinear model. Here, we considered a supervised approach, i.e., a learning algorithm that assumes the knowledge of both decision data $u_{j,i}$, for all $j = 1, \ldots, n$ and $i = 1, \ldots, m$, and the associated overall values $y(u(\mathbf{x}_j))$ over a subset $\mathcal{T}$ of all alternatives. For convenience of notation, let us assume that $\mathcal{T}$ contains n alternatives. Based on this information, we search for a capacity μ that leads to overall evaluations $F_{ML}(u(\mathbf{x}_j))$ as close as possible to $y(u(\mathbf{x}_j))$, for all $j = 1, \ldots, n$.

As will be detailed in the sequel, we considered two different supervised approaches. The main difference is that, in the second one, we included a regularization term in the optimization model.

6.2.1 Supervised approach

As mentioned in Section 2.5.1.2, in a typical supervised approach, the aim is to minimize the mean squared error between the obtained evaluations $F_{ML}(u(\mathbf{x}_j))$ and the desired ones $y(u(\mathbf{x}_j))$, which was provided by the decision maker. Mathematically, by using the multilinear model, this representation error is given by

$$E = \sum_{j=1}^{n} \left(F_{ML}(u(\mathbf{x}_j)) - y(u(\mathbf{x}_j)) \right)^2 . \tag{6.11}$$

Since the aggregation in multilinear model is linear with respect to the capacity coefficients[1], one may use vectors and matrices to represent (6.11). For this purpose, consider the vector of capacity coefficients $\mu = [\mu(\emptyset), \mu(\{1\}), \ldots, \mu(C)]^T$, the vector of

[1] Recall that the multilinear model is defined by $F_{ML}(u(\mathbf{x}_j)) = \sum_{A \subseteq C} \mu(A) \prod_{i \in A} u_{j,i} \prod_{i \in \overline{A}} (1 - u_{j,i})$, where $u_{j,i} \in [0,1]$ and $\overline{A}$ is the complement set of A.

overall evaluations $\mathbf{y} = [y(u(\mathbf{x}_1)), y(u(\mathbf{x}_2)), \ldots, y(u(\mathbf{x}_n))]^T$ and the matrix $\mathbf{R} \in \mathbb{R}^{2^m \times n}$, which contains the products of the multilinear model, given by

$$\mathbf{R} = \begin{bmatrix} \prod\limits_{i \in C} (1 - u_{1,i}) & \prod\limits_{i \in C} (1 - u_{2,i}) & \ldots & \prod\limits_{i \in C} (1 - u_{n,i}) \\ u_{1,1} \prod\limits_{i \in \overline{\{1\}}} (1 - u_{1,i}) & u_{2,1} \prod\limits_{i \in \overline{\{1\}}} (1 - u_{2,i}) & \ldots & u_{n,1} \prod\limits_{i \in \overline{\{1\}}} (1 - u_{n,i}) \\ \vdots & \vdots & \ddots & \vdots \\ \prod\limits_{i \in C} u_{1,i} & \prod\limits_{i \in C} u_{2,i} & \ldots & \prod\limits_{i \in C} u_{n,i} \end{bmatrix}.$$

Remark, for example, that the element in the second row and first column of $\mathbf{R}$ (which is associated with $A = \{1\}$ and the first learning data) is given by $\prod_{i \in \{1\}} u_{1,i} \prod_{i \in \overline{\{1\}}} (1 - u_{1,i}) = u_{1,1} \prod_{i \in \overline{\{1\}}} (1 - u_{1,i})$. Based on these elements, the representation error can be expressed by

$$E = \sum_{j=1}^{n} \left(F_{ML}(u(\mathbf{x}_j)) - y(u(\mathbf{x}_j))\right)^2 = \mu^T \mathbf{R}\mathbf{R}^T \mu - 2\mathbf{y}^T \mathbf{R}^T \mu + \mathbf{y}^T \mathbf{y}. \tag{6.12}$$

Moreover, the minimization of this cost function is equivalent to the minimization of $\mu^T \mathbf{R}\mathbf{R}^T \mu - 2\mathbf{y}^T \mathbf{R}^T \mu$, since $\mathbf{y}^T \mathbf{y}$ is a constant. Therefore, the capacity identification problem can be represented, in a quadratic form, by

$$\begin{aligned} \min_{\mu} \quad & \frac{1}{2} \mu^T \mathbf{Q} \mu + \mathbf{v}^T \mu \\ \text{s.t.} \quad & \mathbf{S}\mu = [0, 1]^T \\ & \mathbf{O}\mu \leq \mathbf{0} \end{aligned} \tag{6.13}$$

where $\mathbf{Q} = 2\mathbf{R}\mathbf{R}^T$, $\mathbf{v} = -2\mathbf{R}\mathbf{y}$ and the matrices $\mathbf{S}$ and $\mathbf{O}$ guarantee that the axioms of a capacity (normalization and monotonicity, respectively) are satisfied. In a scenario with $m = 2$ criteria, for example, we have

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{O} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

6.2.1.1 Example

In order to verify the performance of the quadratic model (6.13), let us consider synthetic decision problems with $m = 3$, $m = 4$ and $m = 5$ criteria and with a number of alternatives varying from 1 to 50. The evaluations $u(\mathbf{x}_j)$ were generated according to a uniform distribution in the range [0, 1]. With respect to the overall values $y(u(\mathbf{x}_j))$, they were obtained through the application of a capacity μ that was randomly generated according to the *Random-Node Generator* proposed by Havens and Pinar (2017). We adopted this generator since the analysis conducted by the authors pointed out that this procedure performs better in comparison with other existing methods (at least in the resulting distribution of the capacity coefficients).

Besides the multilinear model, we also considered the identification problem with respect to the Choquet integral. As a performance index, we adopted the squared error

$$e_\mu = \sum_{A \subseteq C, A \neq \emptyset, A \neq C} \left(\mu(A) - \hat{\mu}(A)\right)^2, \tag{6.14}$$

where μ is the capacity used to obtain the global evaluations $y(u(\mathbf{x}_j))$ and $\hat{\mu}$ is the retrieved one. Figure 36 presents the obtained results averaged over 1000 simulations.

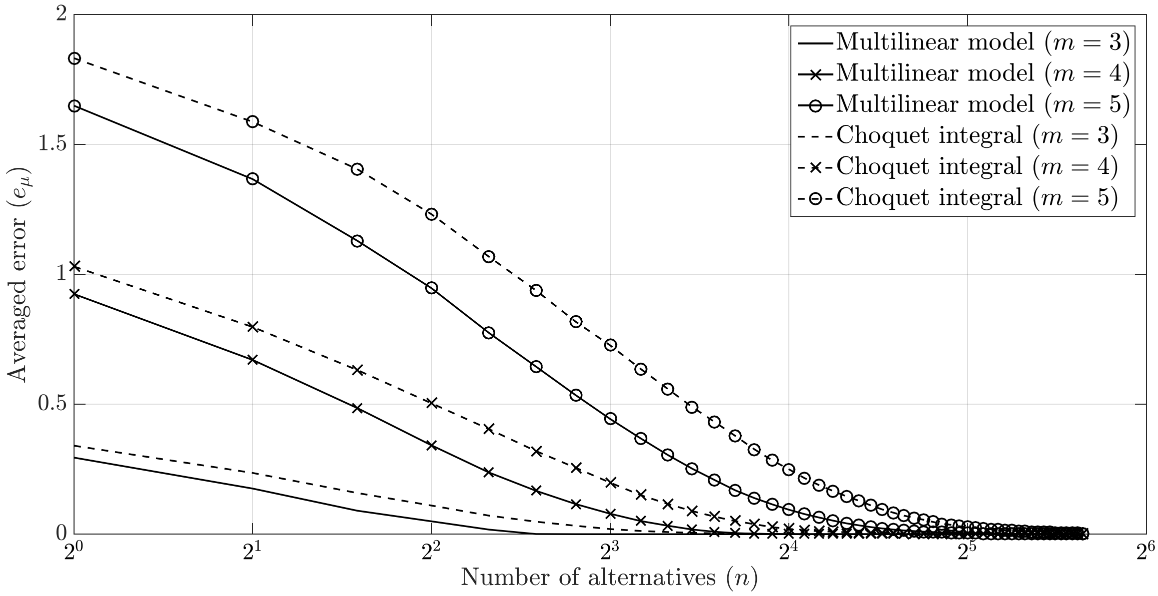


Figure 36 – Mean squared error for different numbers of alternatives.

The results with respect to the multilinear model indicate that, for $3, 4$ and 5 criteria, $e_\mu \approx 0$ when we have at least (2^3), (2^4) and (2^5) learning data, respectively. On the other hand, by taking the Choquet integral, one needs more samples in order to achieve the same performance. This result is in accordance with the properties that we highlighted in Section 2.4.1.2. The application of the Choquet integral only requires the use of a subset of all capacity coefficients, which depends on the order of the evaluations. As a consequence, the matrix $\mathbf{R}$ applied in the optimization model may have zero columns, which will not lead to the identification of the associated parameters. Aiming at achieving a similar performance as the one in the context of the multilinear model, one must guarantee that the considered learning data "activates" all the parameters that should be retrieved.

A common point of both aggregation functions is that, if one does not have enough learning data, the optimization problem (6.13) is ill-posed. This implies that the retrieved capacity coefficients are not necessarily the desired ones. As a consequence, the application of these parameters in a novel dataset will probably lead to distortions on the overall evaluations. Therefore, in such a scenario, one may include more information in the optimization model in order to retrieve a capacity with a high level of generalization.

This may be achieved by applying a regularization term, as will be discussed in the next section.

6.2.2 A supervised approach with regularization

As raised in the last section, one may consider a regularization term in order to deal with the ill-posed optimization problems. This approach has already been adopted in the literature in the context of the Choquet integral (Adeyeba et al., 2015). For instance, the authors considered the ℓ_1-norm of μ, expressed by[2]

$$\|\mu\|_1 = \sum_{A \subseteq C} \mu(A). \tag{6.15}$$

Based on this regularization term, the resolution of the optimization model

$$\begin{array}{ll} \min\limits_{\mu} & \frac{1}{2}\mu^T \mathbf{Q}\mu + \mathbf{v}^T\mu + \lambda \|\mu\|_1 \\ \text{s.t.} & \mathbf{L}\mu = [0,1]^T \\ & \mathbf{M}\mu \leq \mathbf{0} \end{array} \tag{6.16}$$

where λ is a constant, leads to a vector μ whose most part of the elements are close to zero. However, the interpretation of such an approach is not evident and, as pointed out in (Oliveira et al., 2017), the ℓ_1-norm regularization is more meaningful when applied to the interaction indices (recall that $I^{\mathcal{B}}(A) \approx 0$ means that the criteria in A acts independently).

By using the ℓ_1-norm in $\mathbf{I}^{\mathcal{B}} = \left[I^{\mathcal{B}}(\emptyset), I^{\mathcal{B}}(\{1\}), \ldots, I^{\mathcal{B}}(\{1,\ldots\}), \ldots, I^{\mathcal{B}}(C)\right]^T$, i.e., $\left\|\mathbf{I}^{\mathcal{B}}\right\|_1$, one needs to use the transformation described in Equation (2.26). The optimization problem is, therefore, given by:

$$\begin{array}{ll} \min\limits_{\mathbf{I}^{\mathcal{B}}} & \frac{1}{2}\left(\mathbf{I}^{\mathcal{B}}\right)^T \mathbf{Q}'\mathbf{I}^{\mathcal{B}} + \left(\mathbf{v}'\right)^T \mathbf{I}^{\mathcal{B}} + \lambda \left\|\mathbf{I}^{\mathcal{B}}\right\|_1 \\ \text{s.t.} & \mathbf{S}'\mathbf{I}^{\mathcal{B}} = [0,1]^T \\ & \mathbf{O}'\mathbf{I}^{\mathcal{B}} \leq \mathbf{0} \end{array} \tag{6.17}$$

where $\mathbf{Q}' = \mathbf{W}^T\mathbf{Q}\mathbf{W}$, with $\mathbf{W}$ being the transformation matrix from $\mathbf{I}^{\mathcal{B}}$ to μ (i.e. $\mu = \mathbf{W}\mathbf{I}^{\mathcal{B}}$), given by

$$\mathbf{W} = \begin{bmatrix} \left(\frac{1}{2}\right)^{|\emptyset|}(-1)^{|\emptyset\backslash\emptyset|} & \left(\frac{1}{2}\right)^{|\{1\}|}(-1)^{|\{1\}\backslash\emptyset|} & \ldots & \left(\frac{1}{2}\right)^{|C|}(-1)^{|C\backslash\emptyset|} \\ \left(\frac{1}{2}\right)^{|\emptyset|}(-1)^{|\emptyset\backslash\{1\}|} & \left(\frac{1}{2}\right)^{|\{1\}|}(-1)^{|\{1\}\backslash\{1\}|} & \ldots & \left(\frac{1}{2}\right)^{|C|}(-1)^{|C\backslash\{1\}|} \\ \vdots & \vdots & \ddots & \vdots \\ \left(\frac{1}{2}\right)^{|\emptyset|}(-1)^{|\emptyset\backslash C|} & \left(\frac{1}{2}\right)^{|\{1\}|}(-1)^{|\{1\}\backslash C|} & \ldots & \left(\frac{1}{2}\right)^{|C|}(-1)^{|C\backslash C|} \end{bmatrix},$$

$\mathbf{v}' = \mathbf{W}^T\mathbf{v}$, $\mathbf{S}' = \mathbf{S}\mathbf{W}$ and $\mathbf{O}' = \mathbf{O}\mathbf{W}$.

[2] The ℓ_1-norm of a b-dimensional vector $\mathbf{a} \in \mathbb{R}^b$ is given by $\|\mathbf{a}\|_1 = \sum_{l=1}^{b} |a_l|$. Therefore, in Equation (6.15), since $\mu(A) \geq 0$, $\forall A \subseteq C$, one may eliminate the absolute value.

One may remark that, if $\lambda = 0$ and $I^{\mathcal{B}}(A) = 0$ for all A such that $|A| \geq 3$, (6.17) leads to the optimization model with respect to a 2-additive capacity. Moreover, one also may note that (6.17) is not quadratic anymore. However, one may use auxiliary variables (Vanderbei, 2014) in order to turn it into a quadratic problem. For instance, consider a vector $\mathbf{z}_\gamma = [\mathbf{z}_\gamma(\emptyset), \mathbf{z}_\gamma(\{1\}), \ldots, \mathbf{z}_\gamma(\{1, \ldots\}), \ldots, \mathbf{z}_\gamma(C)]^T$, such that

$$\mathbf{z}_\gamma(A) = \begin{cases} 1, & \text{if } |A| > \gamma \\ 0, & \text{otherwise} \end{cases},$$

which controls the application of the ℓ_1-norm in a subset of all the interaction indices. In addition, consider that $\mathbf{I}^{\mathcal{B}} = \mathbf{I}^{\mathcal{B}+} - \mathbf{I}^{\mathcal{B}-}$ and $\left|\mathbf{I}^{\mathcal{B}}\right| = \mathbf{I}^{\mathcal{B}+} + \mathbf{I}^{\mathcal{B}-}$ (with both $\mathbf{I}^{\mathcal{B}+}, \mathbf{I}^{\mathcal{B}-} \geq \mathbf{0}$) and define

$$\tilde{\mathbf{I}}^{\mathcal{B}} = \begin{bmatrix} \mathbf{I}^{\mathcal{B}+} \\ \mathbf{I}^{\mathcal{B}-} \end{bmatrix}, \ \mathbf{Q}'' = \begin{bmatrix} \mathbf{Q}' & -\mathbf{Q}' \\ -\mathbf{Q}' & \mathbf{Q}' \end{bmatrix}, \ \mathbf{v}'' = \begin{bmatrix} \lambda \mathbf{z}_\gamma + \mathbf{v}' \\ \lambda \mathbf{z}_\gamma - \mathbf{v}' \end{bmatrix}, \ \mathbf{S}'' = \begin{bmatrix} \mathbf{S}' & -\mathbf{S}' \end{bmatrix} \text{ and } \mathbf{O}'' = \begin{bmatrix} \mathbf{O}' & -\mathbf{O}' \end{bmatrix}.$$

Based on the aforementioned matrices and vectors, the optimization problem described in (6.17) may be expressed, in a quadratic fashion, as follows:

$$\begin{aligned} \min_{\tilde{\mathbf{I}}^{\mathcal{B}}} \quad & \frac{1}{2} \left(\tilde{\mathbf{I}}^{\mathcal{B}}\right) \mathbf{Q}'' \tilde{\mathbf{I}}^{\mathcal{B}} + \left(\mathbf{v}''\right)^T \tilde{\mathbf{I}}^{\mathcal{B}} \\ \text{s.t.} \quad & \mathbf{S}'' \tilde{\mathbf{I}}^{\mathcal{B}} = [0, 1]^T \\ & \mathbf{O}'' \tilde{\mathbf{I}}^{\mathcal{B}} \leq \mathbf{0} \\ & \tilde{\mathbf{I}}^{\mathcal{B}} \geq \mathbf{0} \end{aligned} \tag{6.18}$$

As already mentioned, the use of the ℓ_1-norm regularization in the interaction index $I^{\mathcal{B}}(A)$ leads to a set parameters whose most part tends to be equal to zero. It is then possible to achieve a sparse solution for $\mathbf{I}^{\mathcal{B}}$, whose sparsity level depends on the λ value. Therefore, the rationality behind this approach is to achieve a simpler model in terms of reducing the parameters that assume a value different from zero. For example, if we consider $\gamma = 1$, the solution of (6.18) promotes a set of interaction indices in which $I^{\mathcal{B}}(A)$, $|A| \geq 2$, is close to zero. Conversely, if we consider $\gamma = 2$, the solution of (6.18) promotes a set of interaction indices in which $I^{\mathcal{B}}(A)$, $|A| \geq 3$, is close to zero. Therefore, one does not impose that the model must be a k-order additive one, but one reduces its flexibility to arbitrarily adjust the set of parameters in the identification problem. The price to be paid is that, by decreasing the level of flexibility, one may increase the representation error E between the obtained evaluations and the desired ones (expressed in Equation (6.11)).

As will be further discussed in the numerical experiments, although a simpler model may achieve a larger value of E in the training step, the retrieved capacity may be close to the correct one. Moreover, another benefit of such an approach is that one may find parameters that lead to a better generalization by applying the retrieved capacity in a dataset other than the one used in the training step.

6.3 Numerical experiments

In this section, we present the experiments conducted in the context of capacity identification. Besides the weighted arithmetic mean (WAM), we assume different optimization models, which will probably lead to different retrieved capacity coefficients. They are defined in the sequel, as well as the notation used along this section:

- WRE: Supervised approach without regularization, expressed in Equation (6.13);
- RE2: Supervised approach with regularization, expressed in Equation (6.18), with $\gamma = 2$;
- 2AD: Supervised approach by means of a 2-additive capacity, expressed in Equation (6.17), with $\lambda = 0$ and $I^{\mathcal{B}}(A) = 0$ for all A such that $|A| \geq 3$;
- RE1: Supervised approach with regularization, expressed in Equation (6.18), with $\gamma = 1$;

6.3.1 Experiments with synthetic data

The first experiment comprises the application of the considered methods with a set of synthetic data generated according to the procedure described in (Grabisch, 1995). We considered $n = 81$ learning data, with $m = 4$ criteria. For each evaluation $u_{j,i}$, we randomly assigned a value belonging to the set $\{0, 0.5, 1\}$. With respect to the global evaluations, they were obtained by the following expression:

$$y(u(\mathbf{x}_j)) = F_{ML}(u(\mathbf{x}_j)) + g, \tag{6.19}$$

where g represents an additive Gaussian noise with variance $\sigma^2 = 0.0125$.

Consider the weights $\lambda_{RE1} = 1$ and $\lambda_{RE2} = 0.015$ for RE1 and RE2 methods, respectively. It is important to mention that we experimentally defined both λ_{RE1} and λ_{RE2}. The application of WRE, RE2, 2AD, RE1 and WAM methods in the capacity identification problem leads to the representation errors $E_{WRE} = 0.0112$, $E_{RE2} = 0.0116$, $E_{2AD} = 0.0118$, $E_{RE1} = 0.0616$ and $E_{WAM} = 0.2254$, respectively. The retrieved capacities and interaction indices are described in Table 14. With respect to the WAM method, one obtains the set of weights $\mathbf{w} = [0.0825, 0.1850, 0.1581, 0.5744]$.

In Table 14, we also present the mean squared error between the correct capacity and the retrieved ones, given by

$$\epsilon_{\mu} = \frac{1}{2^m - 2} \sum_{A \subseteq C, A \neq \emptyset, A \neq C} \left(\mu(A) - \hat{\mu}(A)\right)^2, \tag{6.20}$$

Table 14 – Retrieved capacity and interaction indices (all synthetic learning data).

Capacity $\mu(A)$					
A	**Correct**	**WRE**	**RE2**	**2AD**	**RE1**
$\{\emptyset\}$	0	0	0	0	0
$\{1\}$	0.1	0.1029	0.0991	0.0994	0.0904
$\{2\}$	0.2105	0.2094	0.2066	0.2050	0.1988
$\{3\}$	0.2353	0.2441	0.2334	0.2298	0.1843
$\{4\}$	0.6667	0.6823	0.6749	0.6728	0.6316
$\{1,2\}$	0.3	0.2992	0.2994	0.2988	0.2891
$\{1,3\}$	0.3235	0.3212	0.3241	0.3222	0.2746
$\{1,4\}$	0.7333	0.7328	0.7438	0.7412	0.7016
$\{2,3\}$	0.4211	0.4216	0.4219	0.4225	0.3830
$\{2,4\}$	0.8070	0.7971	0.8053	0.8077	0.7773
$\{3,4\}$	0.8235	0.8152	0.8205	0.8217	0.7844
$\{1,2,3\}$	0.5	0.5010	0.5063	0.5093	0.4734
$\{1,2,4\}$	0.8667	0.8735	0.8692	0.8704	0.8473
$\{1,3,4\}$	0.8824	0.8833	0.8810	0.8831	0.8544
$\{2,3,4\}$	0.9474	0.9445	0.9444	0.9442	0.9300
$\{1,2,3,4\}$	1	1	1	1	1
Error ϵ_μ $(\times 10^{-4})$		0.4002	0.1911	0.2005	9.7187

Interaction index $I^{\mathcal{B}}(A)$					
A	**Correct**	**WRE**	**RE2**	**2AD**	**RE1**
$\{\emptyset\}$	0.5511	0.5518	0.5519	0.5518	0.5263
$\{1\}$	0.0743	0.0750	0.0770	0.0776	0.0802
$\{2\}$	0.1610	0.1581	0.1595	0.1610	0.1722
$\{3\}$	0.1811	0.1792	0.1792	0.1797	0.1685
$\{4\}$	0.5796	0.5786	0.5810	0.5818	0.5791
$\{1,2\}$	-0.0082	0.0007	-0.0056	-0.0056	≈ 0
$\{1,3\}$	-0.0093	-0.0099	-0.0083	-0.0070	≈ 0
$\{1,4\}$	-0.0297	-0.0247	-0.0295	-0.0311	-0.0204
$\{2,3\}$	-0.0201	-0.0145	-0.0123	-0.0123	≈ 0
$\{2,4\}$	-0.0644	-0.0654	-0.0698	-0.0702	-0.0531
$\{3,4\}$	-0.0725	-0.0798	-0.0820	-0.0809	-0.0315
$\{1,2,3\}$	0.0009	-0.0115	≈ 0	0	≈ 0
$\{1,2,4\}$	0.0033	0.0121	0.0014	0	≈ 0
$\{1,3,4\}$	0.0038	0.0165	≈ 0	0	≈ 0
$\{2,3,4\}$	0.0080	0.0195	0.0116	0	≈ 0
$\{1,2,3,4\}$	-0.0006	-0.0537	≈ 0	0	≈ 0
Error $\epsilon_{I^{\mathcal{B}}}$ $(\times 10^{-4})$		2.2202	0.1480	0.1805	2.1704

and the mean squared error between the correct interaction indices and the retrieved ones, given by

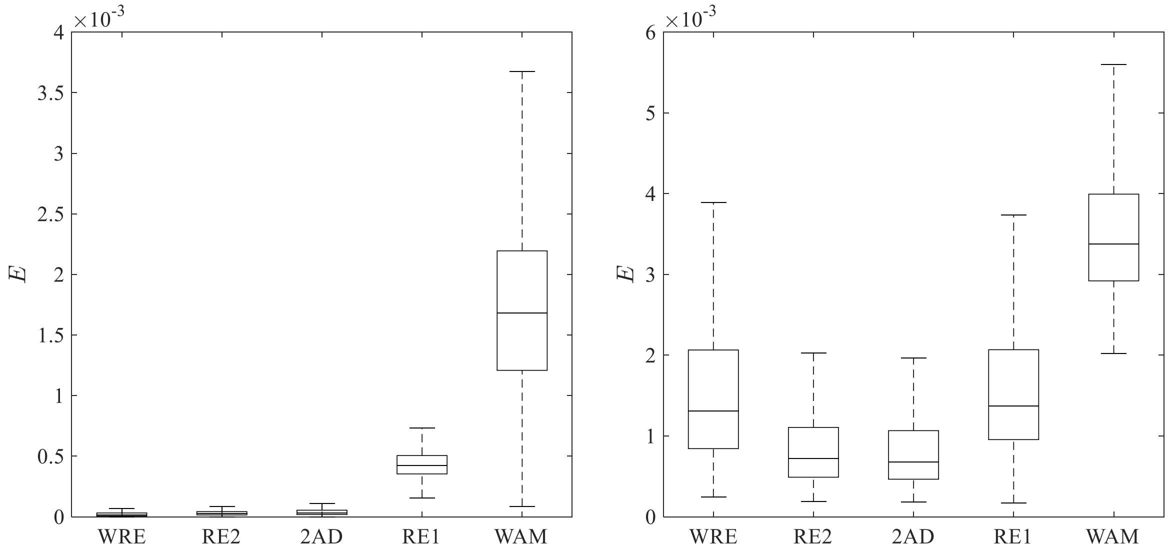
$$\epsilon_{I^{\mathcal{B}}} = \frac{1}{2^m} \sum_{A \in C} \left(I^{\mathcal{B}}(A) - \hat{I}^{\mathcal{B}}(A) \right)^2 . \tag{6.21}$$

As expected, the WRE method achieved the lower representation error E. However, if we compare the retrieved capacity coefficients, both RE2 and 2AD methods provided the better results (lower values of both ϵ_μ and $\epsilon_{I^{\mathcal{B}}}$).

Aiming at further exploiting this analysis, let us verify the generalization ability of the considered approaches. For this purpose, we divided the samples in two sets. The first one is used in the training step, which leads to the parameters identification. The second set is used in the test, which validates the retrieved capacity. Consider 12 samples for training (a subset of all samples less than $2^m - 2$ criteria) and the other 69 samples for test. Moreover, let us define $\lambda_{RE1} = 0.1$ and $\lambda_{RE2} = 0.0025$. Figures 37a and 37b present boxplots of the obtained representation error E (divided by the number of samples and over 10001 simulations[3]) in training and test steps, respectively. One may remark that, although the WRE method provides the lower E value in training step, the validation of the obtained capacity in the test does not lead to better results. In this case, we may say that WRE suffered from overfitting and the methods that achieved the higher levels of generalization were the RE2 and the 2AD.

The retrieved capacity and interaction indices are shown in Table 15 (based on the median representation error E in training step over the 10001 simulations). We also present the estimation errors for both adjusted capacity and interaction indices.

[3] We consider an odd number of simulations in order to have a median one without interpolating two results.



(a) Training. (b) Test.

Figure 37 – Representation error E based on a subset of the synthetic learning data.

The application of the WAM leads to $\mathbf{w} = [0.0997, 0.1860, 0.1870, 0.5273]$. Similarly as obtained by considering all the learning data in the training step, the application of the WRE does not lead to the lower values of ϵ_μ and $\epsilon_{I^\mathcal{B}}$.

Table 15 – Retrieved capacity and interaction indices (subset of the synthetic learning data).

Capacity $\mu(A)$					
A	**Correct**	**WRE**	**RE2**	**2AD**	**RE1**
$\{\emptyset\}$	0	0	0	0	0
$\{1\}$	0.1	0.0526	0.0926	0.0971	0.0984
$\{2\}$	0.2105	0.2038	0.2839	0.1969	0.1741
$\{3\}$	0.2353	0.3259	0.1465	0.1987	0.1092
$\{4\}$	0.6667	0.6597	0.6879	0.6982	0.6511
$\{1,2\}$	0.3	0.3859	0.3747	0.2941	0.2725
$\{1,3\}$	0.3235	0.3872	0.3178	0.2981	0.2076
$\{1,4\}$	0.7333	0.6755	0.6879	0.6982	0.7496
$\{2,3\}$	0.4211	0.3811	0.3492	0.4510	0.2831
$\{2,4\}$	0.8070	0.7115	0.8827	0.7598	0.7926
$\{3,4\}$	0.8235	0.8373	0.8057	0.8806	0.7603
$\{1,2,3\}$	0.5	0.4679	0.5188	0.5505	0.3815
$\{1,2,4\}$	0.8667	0.9085	0.8827	0.7598	0.8910
$\{1,3,4\}$	0.8824	0.8652	0.8863	0.8829	0.8587
$\{2,3,4\}$	0.9474	0.9500	0.9193	0.9977	0.9016
$\{1,2,3,4\}$	1	1	1	1	1
Error ϵ_μ $(\times 10^{-2})$		0.2809	0.2435	0.1951	0.5177

Interaction index $I^\mathcal{B}(A)$					
A	**Correct**	**WRE**	**RE2**	**2AD**	**RE1**
$\{\emptyset\}$	0.5511	0.5508	0.5523	0.5477	0.5082
$\{1\}$	0.0743	0.0842	0.0857	0.0497	0.0984
$\{2\}$	0.1610	0.1507	0.1983	0.1570	0.1577
$\{3\}$	0.1811	0.2021	0.1314	0.2194	0.1091
$\{4\}$	0.5796	0.5504	0.5836	0.5738	0.6348
$\{1,2\}$	-0.0082	0.0896	-0,0009	≈ 0	≈ 0
$\{1,3\}$	-0.0093	-0.0553	0.0797	0.0023	≈ 0
$\{1,4\}$	-0.0297	-0.0231	-0.0907	-0.0971	≈ 0
$\{2,3\}$	-0.0201	-0.1097	-0.0812	0.0555	-0.0002
$\{2,4\}$	-0.0644	-0.0352	-0.0882	-0.1353	-0.0326
$\{3,4\}$	-0.0725	-0.0556	-0.0277	-0.0163	≈ 0
$\{1,2,3\}$	0.0009	-0.1316	≈ 0	0	≈ 0
$\{1,2,4\}$	0.0033	0.0242	0.0018	0	≈ 0
$\{1,3,4\}$	0.0038	-0.0242	0.0019	0	≈ 0
$\{2,3,4\}$	0.0080	0.1819	0.0001	0	≈ 0
$\{1,2,3,4\}$	-0.0006	-0.0551	≈ 0	0	≈ 0
Error $\epsilon_{I^\mathcal{B}}$ $(\times 10^{-2})$		0.4650	0.1380	0.1304	0.1154

With respect to the computational effort[4] of each considered approach, we

[4] Computing device: Intel Core i7, 3.60 GHz, 32.00 GB RAM, software MATLAB 2018a.

calculated the average time (in seconds) spent in the optimization model. The results for WRE, RE2, 2AD, RE1 and WAM are 0.0045, 0.0047, 0.0044, 0.0047 and 0.0042, respectively. Therefore, although the approaches based on a regularization term led to the highest computational time, their difference with respect to the other ones are very small.

6.3.2 Experiments with real data

In this section, we attest our proposal in a real dataset collected from a mental workload experiment[5]. The mental workload is an important issue in ergonomics (Young et al., 2015) and can be measured through the well-known NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988). In the application of this procedure, after performing a task, the user provides subjective evaluations (in the range $[0, 100]$) based on six sources: mental demand, physical demand, temporal demand, performance, effort and frustration. Then, the user also provides 15 pairwise comparisons among the six sources, which indicate that a specific source contributes to the workload more than another one. Finally, based on these pairwise comparisons, one determines the "importance" of each source in the mental workload (the weight associated with the source) and obtains the global evaluation by means of a weighted arithmetic mean.

In view of the limitations of the WAM in modeling interactions among criteria, some works investigated the application of a capacity to provide the global mental workload evaluation (Raufaste et al., 2001; Grabisch et al., 2006). For instance, other than the six subjective evaluations, the users also provide a subjective global evaluation of the task (also in the range $[0, 100]$), which is used as a learning data to perform capacity identification. Therefore, a comparison between the WAM and a capacity-based aggregation function can be conducted in order to verify which one can better represent the information provided by the users.

The collected real dataset comprises the subjective evaluations (over the six sources and the global one) provided by a set of 143 users[6]. Similarly to the experiments with synthetic data, we firstly apply the considered approaches by taking into account all the 143 samples as the learning data. Before setting the regularization weights used in RE1 and RE2 methods, let us investigate the performance of the considered approaches for different values of λ. Figure 38 illustrates the obtained representation errors.

Clearly, in the case of WRE, 2AD and WAM, E is not affected by λ. However, as highlighted in Section 6.2.2, there is a trade-off between regularization and representation error when RE1 and RE2 methods are applied. In both cases, if one sets λ close to

[5] The authors would like to thank the *Institut National de Recherche et de Sécurité* (INRS) for allowing them to use, for the present work, the data collected in the framework of a previous collaboration (convention N° 5001053).

[6] Originally, there were 188 samples, however, 45 were eliminated due to inconsistencies.

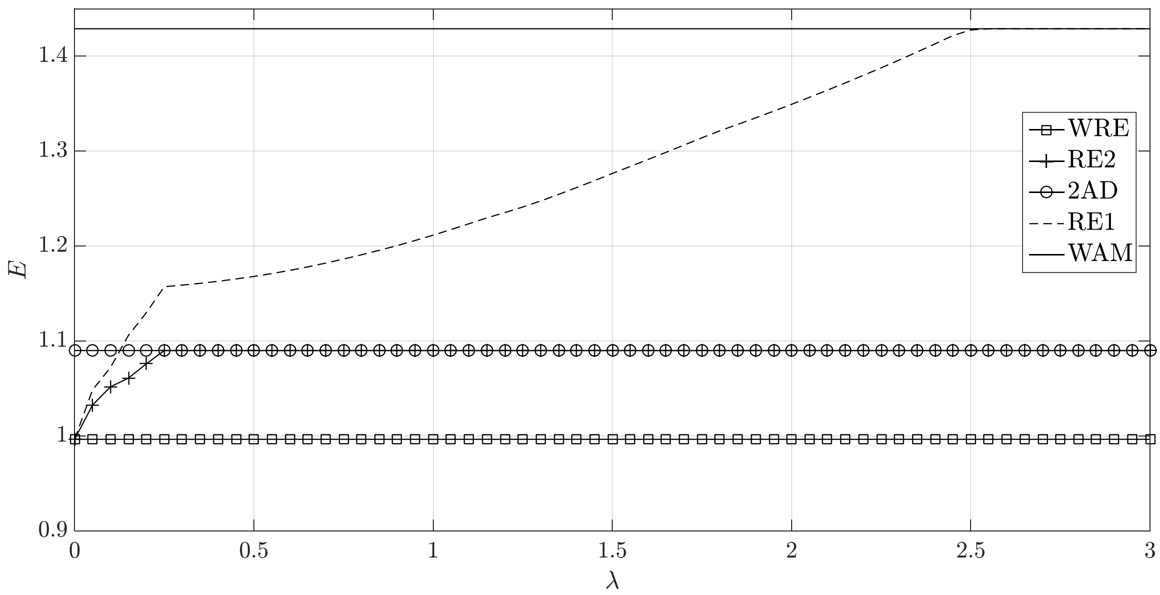


Figure 38 – Obtained representation error E for different values of λ.

zero, which practically eliminates the regularization term, the methods achieve the same performance as the WRE. However, if one adopts a large value of λ, which means that one considers that the minimization of the ℓ_1-norm is very important in the optimization model, the parameters associated with this regularization term will be approximately zero. Therefore, RE1 and RE2 will converge to the WAM and 2AD methods, respectively.

As an illustrative scenario, let us assume $\lambda_{RE1} = 1$ and $\lambda_{RE2} = 0.1$ for RE1 and RE2 methods, respectively. The application of WRE, RE2, 2AD, RE1 and WAM methods in the capacity identification problem leads to the representation errors $E_{WRE} = 0.9966$, $E_{RE2} = 1.0519$, $E_{2AD} = 1.0901$, $E_{RE1} = 1.2116$ and $E_{WAM} = 1.4290$, respectively. Note that the methods that take into account interactions among criteria achieved values of representation error E that are considerably lower compared to the application of the WAM. In other words, the additive aggregation provided by the WAM is not sufficient to satisfy the information provided by the users.

The retrieved capacities and interaction indices are described in Table 16. Since we have $m = 6$ criteria and, therefore, $2^6 = 64$ parameters, for the purpose of illustrating the obtained results, we omitted the ones such that $|A| > 2$. The application of the WAM leads to $\mathbf{w} = [0.5056, 0, 0.0622, 0.3056, 0.0854, 0.0412]$.

Similarly as in the experiments with synthetic data, we conducted here an analysis whose aim is to verify the generalization ability of the considered approaches. For instance, let us consider a training step and a test step with 100 and 43 samples, respectively. Moreover, let us define $\lambda_{RE1} = 0.1$ and $\lambda_{RE2} = 0.05$. Figures 39a and 39b present the boxplots of the obtained representation error E (divided by the number of samples and over 1001 simulations) in training and test steps, respectively. Although the

Table 16 – Retrieved capacity and interaction indices (all real learning data).

Capacity $\mu(A)$				
A	**WRE**	**RE2**	**2AD**	**RE1**
$\{1\}$	0.5580	0.5194	0.4953	0.4671
$\{2\}$	0.0001	0.1423	0.1821	0.0375
$\{3\}$	0.1344	0.1440	0.1310	0.1076
$\{4\}$	≈ 0	0.3032	0.4077	0.3179
$\{5\}$	0.0001	0.0044	0.0054	0.0824
$\{6\}$	0.0001	0.0168	0.0353	0.1171
$\{1,2\}$	0.5584	0.6275	0.5946	0.5046
$\{1,3\}$	0.6546	0.6536	0.6262	0.5715
$\{1,4\}$	0.6196	0.6388	0.7113	0.7756
$\{1,5\}$	0.5581	0.5238	0.5007	0.5495
$\{1,6\}$	0.6547	0.5943	0.6426	0.5842
$\{2,3\}$	0.3569	0.2864	0.3301	0.1451
$\{2,4\}$	0.4237	0.4456	0.4992	0.3555
$\{2,5\}$	0.0012	0.1467	0.1875	0.1200
$\{2,6\}$	0.0011	0.1591	0.2087	0.1501
$\{3,4\}$	0.4563	0.3279	0.4297	0.4255
$\{3,5\}$	0.1345	0.1483	0.1360	0.1900
$\{3,6\}$	0.1344	0.1483	0.1447	0.1422
$\{4,5\}$	00.6464	0.5193	0.5371	0.4004
$\{4,6\}$	0.4563	0.5115	0.4430	0.4351
$\{5,6\}$	0.0001	0.0169	0.0357	0.1694

Interaction index $I^{\mathcal{B}}(A)$				
A	**WRE**	**RE2**	**2AD**	**RE1**
$\{1\}$	0.4414	0.4256	0.4140	0.4608
$\{2\}$	0.0979	0.0982	0.0995	0.0352
$\{3\}$	0.1194	0.0726	0.0740	0.0648
$\{4\}$	0.2933	0.2977	0.2741	0.3132
$\{5\}$	0.1074	0.0787	0.0647	0.0674
$\{6\}$	0.0906	0.0840	0.0737	0.0586
$\{1,2\}$	-0.0133	-0.0360	-0.0828	≈ 0
$\{1,3\}$	-0.1098	-0.0107	≈ 0	-0.0031
$\{1,4\}$	-0.1879	-0.1989	-0.1917	-0.0094
$\{1,5\}$	-0.0953	-0.0151	≈ 0	≈ 0
$\{1,6\}$	0.1175	0.0555	0.1121	≈ 0
$\{2,3\}$	0.0838	0.0001	0.0169	≈ 0
$\{2,4\}$	-0.0514	-0.0523	-0.0906	≈ 0
$\{2,5\}$	0.0148	≈ 0	≈ 0	≈ 0
$\{2,6\}$	-0.0432	-0.0540	-0.0087	-0.0045
$\{3,4\}$	-0.1709	-0.1196	-0.1089	≈ 0
$\{3,5\}$	-0.0117	-0.0003	-0.0004	≈ 0
$\{3,6\}$	-0.0397	-0.0134	-0.0216	-0.0824
$\{4,5\}$	0.1745	0.1530	0.1240	≈ 0
$\{4,6\}$	0.0769	0.0958	≈ 0	≈ 0
$\{5,6\}$	-0.0643	-0.0477	-0.0050	-0.0302

WRE method provides the lower E value in training step, the validation of the obtained capacity in the test step does not lead to the better results. One also obtained higher generalization levels with the application of RE2 and 2AD methods.

In Table 17, we show the retrieved capacity and interaction indices (based on the median representation error E in training step over the 1001 simulations).By applying the WAM, one obtains $\mathbf{w} = [0.5382, 0.0287, 0.0799, 0.2476, 0.0537, 0.0519]$. With respect to the computational effort, the average time (in seconds) spent by WRE, RE2, 2AD, RE1 and WAM are 0.1999, 0.3030, 0.0315, 0.2987 and 0.0020, respectively. The results for both RE2 and RE1 are similar. However, since one includes more terms (variable and constraints) in the optimization model, they spend more time in comparison with the other approaches. Indeed, their average time are approximately 1.5 and 10 times higher than the WRE and 2AD approaches, respectively. The WAM has the lowest computational effort, which is justified by the small number of variables. Moreover, if one compares the average time achieved in the previous section, one notes that the experiment on real data led to higher values. This is explained by the number of capacity coefficients to be adjusted, which increased (by assuming the WRE model) from $2^4 - 2 = 14$ to $2^6 - 2 = 62$.

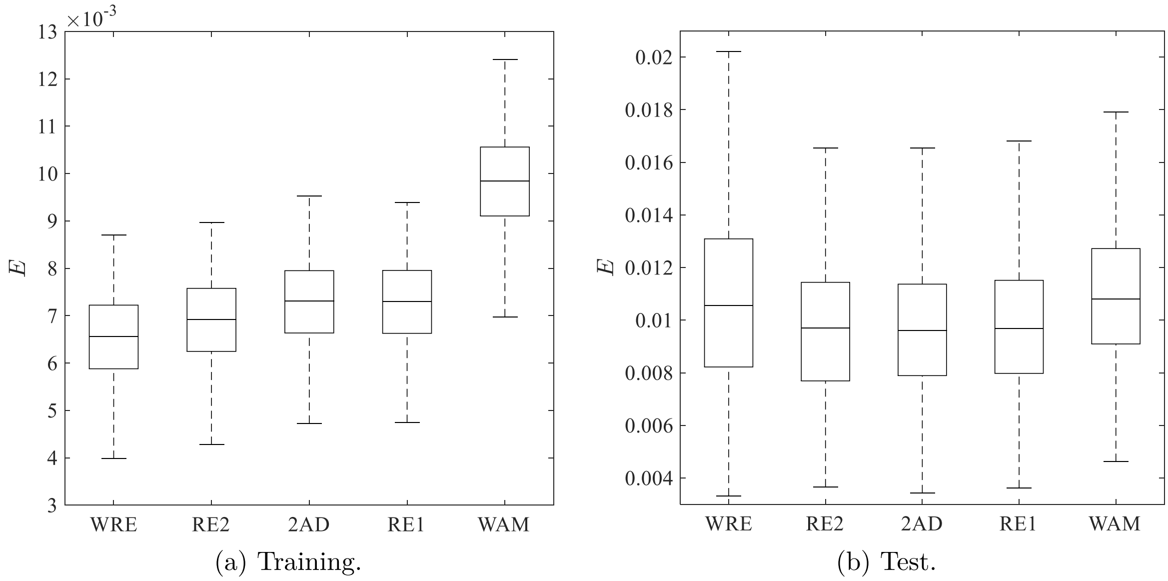


Figure 39 – Representation error E based on a subset of the real learning data.

6.4 Conclusions

Although both Choquet integral and multilinear model can be useful in several practical situations, few works in the literature exploit theoretical aspects or applications with respect to the latter one. Therefore, in this chapter, we address some issues in MCDM problems by means of the multilinear model and open the path for future researches on this subject.

As a first analysis, we exploited the concept of a 2-additive capacity in the multilinear model and provide an analytical expression. We noted that such an expression is similar to the one in the context of the Choquet integral and, therefore, we could generalize them. For instance, both functionals apply an additive, a disjunctive and a conjunctive term to aggregate the set of evaluations.

In a second contribution, we addressed the problem of capacity identification in a supervised fashion. A first result that we achieved indicated that, in order that the Choquet integral achieves the same performance of the multilinear model, one needs to ensure that the considered learning data will lead to the identification of all capacity coefficients. Otherwise, some parameters may not be retrieved by the optimization model. As a second experiment, we addressed the problem of capacity identification by using a regularization term. The resolution of the associated optimization problem can lead to a simpler model compared to the one obtained without the use of this additional term. Moreover, although there is a loss of performance with respect to the representation error, one can retrieve a capacity close to the correct one or, at least, with a higher level of generalization when applied in a new dataset.

Table 17 – Retrieved capacity and interaction indices (subset of the real learning data).

Capacity $\mu(A)$					Interaction index $I^{\mathcal{B}}(A)$				
A	**WRE**	**RE2**	**2AD**	**RE1**	A	**WRE**	**RE2**	**2AD**	**RE1**
$\{1\}$	0.0350	0.5282	0.4917	0.3968	$\{1\}$	0.5197	0.3470	0.4832	0.4459
$\{2\}$	0.0313	0.1544	0.2262	0.1503	$\{2\}$	0.1023	0.1277	0.1131	0.1222
$\{3\}$	0.1384	0.0286	0.2025	0.1737	$\{3\}$	0.1262	0.0750	0.1013	0.0868
$\{4\}$	≈ 0	0.4506	0.3009	0.3808	$\{4\}$	0.2488	0.3433	0.1987	0.2535
$\{5\}$	0.0313	≈ 0	≈ 0	0.0374	$\{5\}$	0.0549	0.0813	0.0266	0.0537
$\{6\}$	0.0313	0.0268	≈ 0	0.0374	$\{6\}$	0.1166	0.0761	0.0770	0.0791
$\{1,2\}$	0.5565	0.5365	0.7177	0.6295	$\{1,2\}$	-0.0009	-0.1463	-0.0001	≈ 0
$\{1,3\}$	0.6714	0.5552	0.5665	0.4953	$\{1,3\}$	-0.0944	-0.0017	-0.1276	-0.0751
$\{1,4\}$	0.7383	0.6400	0.7926	0.6778	$\{1,4\}$	0.0065	-0.3390	≈ 0	-0.0998
$\{1,5\}$	0.5565	0.5284	0.4917	0.5041	$\{1,5\}$	0.0589	≈ 0	≈ 0	0.0699
$\{1,6\}$	0.6714	0.6794	0.6024	0.6374	$\{1,6\}$	0.1475	0.1242	0.1108	0.1209
$\{2,3\}$	0.4463	0.3569	0.4287	0.3240	$\{2,3\}$	0.0305	0.0986	0.0001	≈ 0
$\{2,4\}$	0.3780	0.5972	0.3010	0.4748	$\{2,4\}$	-0.1510	-0.0842	-0.2261	-0.0564
$\{2,5\}$	0.0313	0.1572	0.2262	0.1878	$\{2,5\}$	-0.0596	-0.0001	≈ 0	≈ 0
$\{2,6\}$	0.0313	0.1830	0.2262	0.1878	$\{2,6\}$	0.0043	0.0008	≈ 0	-0.0824
$\{3,4\}$	0.4463	0.5248	0.4285	0.4560	$\{3,4\}$	-0.1642	-0.0520	-0.0748	-0.0986
$\{3,5\}$	0.1384	0.0287	0.2025	0.2111	$\{3,5\}$	-0.0593	-0.0233	≈ 0	≈ 0
$\{3,6\}$	0.1384	0.0286	0.2025	0.2111	$\{3,6\}$	-0.0936	-0.0268	≈ 0	≈ 0
$\{4,5\}$	0.3753	0.6363	0.3542	0.4183	$\{4,5\}$	0.0386	0.1610	0.0533	≈ 0
$\{4,6\}$	0.4463	0.4778	0.3442	0.4183	$\{4,6\}$	-0.0247	≈ 0	0.0433	≈ 0
$\{5,6\}$	0.0313	0.0269	≈ 0	0.0374	$\{5,6\}$	-0.0995	-0.0010	≈ 0	-0.0374

Note that the capacity identification models considered in this chapter adopt a least-squares-based cost function. We also assumed that the decision maker provides overall evaluations of a set of alternatives. As a future perspective, we would like to investigate the use of regularization terms in other optimization models, such as the ones based on linear programming (Marichal and Roubens, 2000). Moreover, instead of assuming that the decision maker provides the overall evaluations, we may consider that he/she expresses pairwise comparison between pairs of alternatives. In this case, the optimization model will lead to capacity coefficients that ensure that the partial preorder is satisfied.

In Section 6.3.1, we mentioned that the values of regularization weights were experimentally defined. Therefore, as another future work, one aims at investigating an automatic procedure to set these parameters. A remark on this issue is that, since the cost function comprises one part related to the global evaluations and another one associated to the interaction indices, one needs to take into account this difference in terms of the nature of each part.

7 On the unsupervised approaches for capacity identification

We discussed in Section 2.5.2.2 that some methods in the literature aim at dealing with the problem of capacity identification through an unsupervised approach (Kojadinovic, 2008; Rowley et al., 2015; Duarte, 2018). Let us recall that, by using such an approach, the parameters are defined based only on the information about the decision data. In other words, one does not have access to the overall evaluations as learning data[1]. Therefore, one needs to assume a characteristic about the decision problem that we would like to deal with and include it into the identification model. For instance, Duarte (2018) associates some interaction indices used in the Choquet integral to similarity measures of pairs of criteria in order to deal with the bias provided by correlations in the decision data. In fact, as discussed by Marichal (2000), the interaction indices can be used to deal with such a bias.

Although there are some works that deal with the unsupervised capacity identification problem, most of them were developed in the context of the Choquet integral. Since, to the best of our knowledge, no study on unsupervised capacity identification was conducted in the context of the multilinear model, the analysis presented in this chapter lies in this aggregation function.

We may divide our contributions into two approaches. The first one, presented in Section 7.1, consists in extending the approach proposed by Duarte (2018) to the multilinear model. For this purpose, we consider the 2-additive multilinear model formulation defined in the previous section (see Equation (6.5)) and improve the optimization model in order to turn it into a quadratic one (which is easier to be solved).

As a second contribution, described in Section 7.2, instead of using similarity measures, we deal with correlations by considering the Sobol' indices of coalitions of criteria. An interesting aspect of such indices is that they can be associated with the multilinear model (Grabisch and Labreuche, 2017). It is important to highlight that this analysis led to the 4th conference paper mentioned in Chapter 1.

7.1 An approach based on correlation coefficients

In order to apply the approach proposed by (Duarte, 2018) in the 2-additive multilinear model, one should take into account the axioms of a capacity associated with

[1] One may also argue that further information provided by the decision maker can be considered in the identification model. However, in this context, one should refer to it as a semi-supervised approach.

this aggregation function (described in Section 6.1). For instance, the monotonicity condition is the same as for the Choquet integral. For the normalization ones, by assuming that $\phi_i^{\mathcal{B}} = 1/m$, $i = 1, \ldots, m$, both conditions lead to $I^{\mathcal{B}}(\emptyset) + \frac{1}{4}\sum_{i,i'} I_{i,i'}^{\mathcal{B}} = \frac{1}{2}$, which is automatically satisfied. Therefore, one could use the same procedure in the context of the multilinear model.

However, in the optimization model (2.43), there are absolute values and a constraint that guarantees that $\hat{\Psi}$ be a positive semidefinite matrix. With such elements we cannot rewrite the considered optimization model as a linear or quadratic problem (which would facilitate the algorithm convergence). Therefore, in this section, we provide some modifications in the optimization problem in order to transform it into a quadratic one.

As a first aspect, instead of searching for a covariance matrix $\hat{\Psi}$ that is similar to the one obtained from the dataset, we only consider the correlation coefficients in the optimization model. For this purpose, since we expect that[2] $I_{i,i'}^{\mathcal{B}} \approx -\psi_{i,i'}$, the new cost function is defined by

$$\min_{\mathbf{I}^{\mathcal{B}}} \sum_{i,i'} \left(I_{i,i'}^{\mathcal{B}} + \psi_{i,i'} \right)^2. \tag{7.1}$$

Recall that $\psi_{i,i'}$ represents a similarity measure (e.g., the Pearson's or the Spearman's correlation coefficient) between criteria i and i'.

Aiming at avoiding the absolute values, one may consider all possible signals combinations of $I_{i,i'}^{\mathcal{B}}$. Therefore, the associated constraint can be defined by

$$\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i \neq i'} \pm I_{i,i'}^{\mathcal{B}} \geq 0, \ \forall i \in C. \tag{7.2}$$

The price to be paid is that, instead of m constraints for this axiom of a capacity, we will have $m(2^m)$.

Besides the aforementioned modifications, we also intend to verify the "general" optimization model, i.e., the one that does not assume $\phi_i^{\mathcal{B}} = 1/m$ for all $i = 1, \ldots, m$. Both models are presented in the next section.

7.1.1 Optimization models

The first optimization model that will be presented in this section does not assume any further condition about the interaction indices (and, as a consequence, in the capacity coefficients). On the other hand, the next one will impose the same assumption as the one adopted in (Duarte, 2018).

---

2 Recall that, as discussed in Section 2.5.2.2, we associate the interaction index $I_{i,i'}^{\mathcal{B}}$ to the negative of a similarity measure $\psi_{i,i'}$ in order to model a redundancy or a complementary effect between criteria i, i'. For instance, if criteria i, i' are positively correlated, (i.e., $\psi_{i,i'} > 0$), we model this by a redundancy between them (i.e., $I_{i,i'}^{\mathcal{B}} < 0$).

General model (EA1)

The general optimization problem used to estimate the interaction indices in the multilinear model is given by[3]

$$\begin{array}{ll} \min\limits_{\mathbf{I}^{\mathcal{B}}} & \sum_{i,i'} \left(I^{\mathcal{B}}_{i,i'} + \psi_{i,i'} \right)^2 \\ \text{s.t.} & \phi^{\mathcal{B}}_i - \frac{1}{2} \sum_{i' \neq i} \pm I^{\mathcal{B}}_{i,i'} \geq 0, \ \forall i \in C \\ & I^{\mathcal{B}}(\emptyset) + \frac{1}{4} \sum_{i,i'} I^{\mathcal{B}}_{i,i'} = \frac{1}{2} \\ & \sum_i \phi^{\mathcal{B}}_i = 1, \end{array} \tag{7.3}$$

or, in a quadratic form,

$$\begin{array}{ll} \min\limits_{\mathbf{I}^{\mathcal{B}}} & \frac{1}{2} (\mathbf{I}^{\mathcal{B}})^T \mathbf{H} \mathbf{I}^{\mathcal{B}} + \mathbf{v}^T \mathbf{I}^{\mathcal{B}} \\ \text{s.t.} & \mathbf{Q}' \mathbf{I}^{\mathcal{B}} \leq \mathbf{b}' \\ & \begin{bmatrix} \mathbf{Q}'' \\ \mathbf{Q}''' \end{bmatrix} \mathbf{I}^{\mathcal{B}} = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}, \end{array} \tag{7.4}$$

where $\mathbf{H} = (h_{f,l}) \in \mathbb{R}^{\frac{m(m+1)}{2}+1 \times \frac{m(m+1)}{2}+1}$ is such that

$$h_{f,l} = \begin{cases} 2, & f = l > m+1 \\ 0, & \text{otherwise} \end{cases}, \mathbf{v} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 2\psi_{1,2} \\ \vdots \\ 2\psi_{m-1,m} \end{bmatrix}$$

and $\mathbf{Q}'$, $\mathbf{Q}''$, $\mathbf{Q}'''$, $\mathbf{b}'$ are matrices and vectors used to ensure that the axioms of a capacity are satisfied. For instance, if we consider $m = 3$, we have

$$\mathbf{Q}' = \begin{bmatrix} 0 & -1 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & -1 & 0 & 0 & -1/2 & 1/2 & 0 \\ 0 & -1 & 0 & 0 & 1/2 & -1/2 & 0 \\ 0 & -1 & 0 & 0 & -1/2 & -1/2 & 0 \\ 0 & 0 & -1 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & -1 & 0 & -1/2 & 0 & 1/2 \\ 0 & 0 & -1 & 0 & 1/2 & 0 & -1/2 \\ 0 & 0 & -1 & 0 & -1/2 & 0 & -1/2 \\ 0 & 0 & 0 & -1 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & -1 & 0 & -1/2 & 1/2 \\ 0 & 0 & 0 & -1 & 0 & 1/2 & -1/2 \\ 0 & 0 & 0 & -1 & 0 & -1/2 & -1/2 \end{bmatrix}, \mathbf{Q}'' = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1/6 \\ 1/6 \\ 1/6 \end{bmatrix}^T, \mathbf{Q}''' = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T,$$

[3] It is worth mentioning that we will describe the optimization problems with respect to the multilinear model. However, they can be easily adapted to the Choquet integral. One only needs to change $I^{\mathcal{B}}(\emptyset) + \frac{1}{4} \sum_{i,i'} I^{\mathcal{B}}_{i,i'} = \frac{1}{2}$ by $I^{\mathcal{S}}(\emptyset) + \frac{1}{6} \sum_{i,i'} I^{\mathcal{S}}_{i,i'} = \frac{1}{2}$.

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 2\psi_{1,2} \\ 2\psi_{1,3} \\ 2\psi_{2,3} \end{bmatrix}, \mathbf{b}' = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Assuming $\phi_i^{\mathcal{B}} = 1/m$ for all i (EA2)

In this second optimization model, we assume that $\phi_i^{\mathcal{B}} = 1/m$ for all i. In this case, the optimization model (7.3) can be formulated as

$$\begin{array}{ll} \min\limits_{\mathbf{I}^{\mathcal{B}}} & \sum_{i,i'} \left(I_{i,i'}^{\mathcal{B}} + \psi_{i,i'} \right)^2 \\ \text{s.t.} & \sum_{i' \neq i} \pm I_{i,i'}^{\mathcal{B}} \leq \frac{2}{m}, \ \forall i \in C \\ & I^{\mathcal{B}}(\emptyset) + \frac{1}{4} \sum_{i,i'} I_{i,i'}^{\mathcal{B}} = \frac{1}{2} \\ & \phi_i^{\mathcal{B}} = \frac{1}{m}, \ \forall i \in C \end{array} \tag{7.5}$$

or, in a quadratic form,

$$\begin{array}{ll} \min\limits_{\mathbf{I}^{\mathcal{B}}} & \frac{1}{2} (\mathbf{I}^{\mathcal{B}})^T \mathbf{H} \mathbf{I}^{\mathcal{B}} + \mathbf{v}^T \mathbf{I}^{\mathcal{B}} \\ \text{s.t.} & \mathbf{Q}' \mathbf{I}^{\mathcal{B}} \leq \mathbf{b}' \\ & \begin{bmatrix} \mathbf{Q}'' \\ \mathbf{Q}''' \end{bmatrix} \mathbf{I}^{\mathcal{B}} = \begin{bmatrix} 1/2 \\ 1/m \\ \vdots \\ 1/m \end{bmatrix}, \end{array} \tag{7.6}$$

where $\mathbf{H}$ and $\mathbf{v}$ are the same as before, and $\mathbf{Q}'$, $\mathbf{Q}''$, $\mathbf{Q}'''$, $\mathbf{b}'$ are the adjusted matrices and vectors that ensure that the axioms of a capacity are satisfied. If we consider $m = 3$, we have

$$\mathbf{Q}' = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix}, \mathbf{Q}'' = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1/6 \\ 1/6 \\ 1/6 \end{bmatrix}^T, \mathbf{Q}''' = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 2\psi_{1,2} \\ 2\psi_{1,3} \\ 2\psi_{2,3} \end{bmatrix}, \mathbf{b}' = \begin{bmatrix} 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

It is important to highlight that, by assuming $\phi_i^{\mathcal{B}} = 1/m, \forall i \in C$ and, therefore, $\sum_{i \neq i'} \pm I_{i,i'}^{\mathcal{B}} \leq \frac{2}{m}$, one implies an upper bound to $\left| I_{i,i'}^{\mathcal{B}} \right|$. Moreover, the higher the number of criteria the lower the bound.

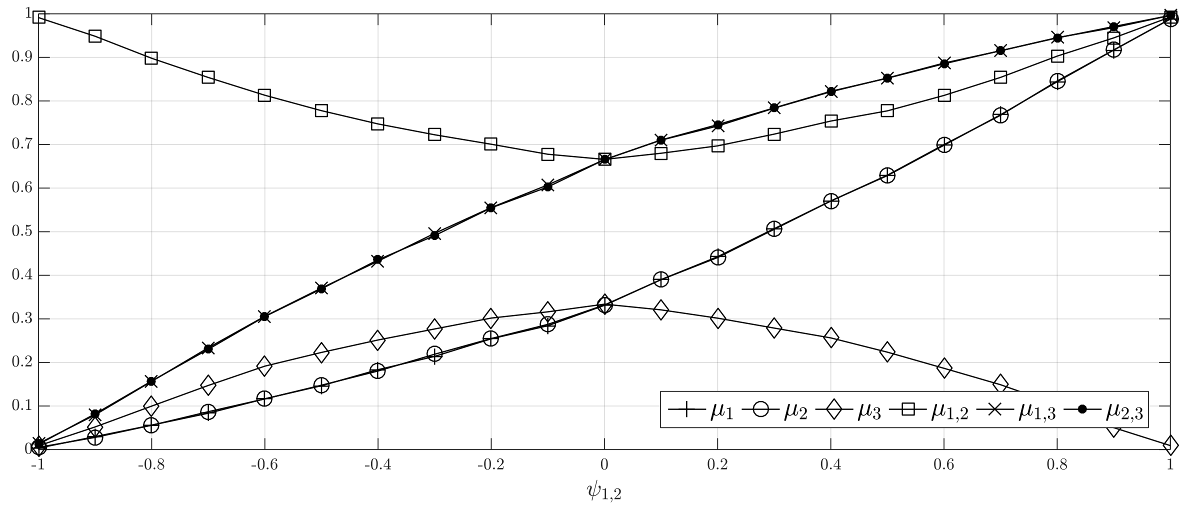
7.1.2 Numerical experiments on synthetic data

In order to verify the application of the aforementioned models in an unsupervised capacity identification problem, let us conduct a numerical experiment with 1000 alternatives and 3 criteria (synthetic random data generated according to a uniform distribution on $[0, 1]$). We vary the degree of correlation among criteria 1 and 2 in the range $[-1, 1]$. Based on 100 simulations, the results for models EA1 and EA2 are presented in Figures 40 and 41, respectively. We also conducted this experiment in the context of the Choquet integral and we obtained the same parameters values.

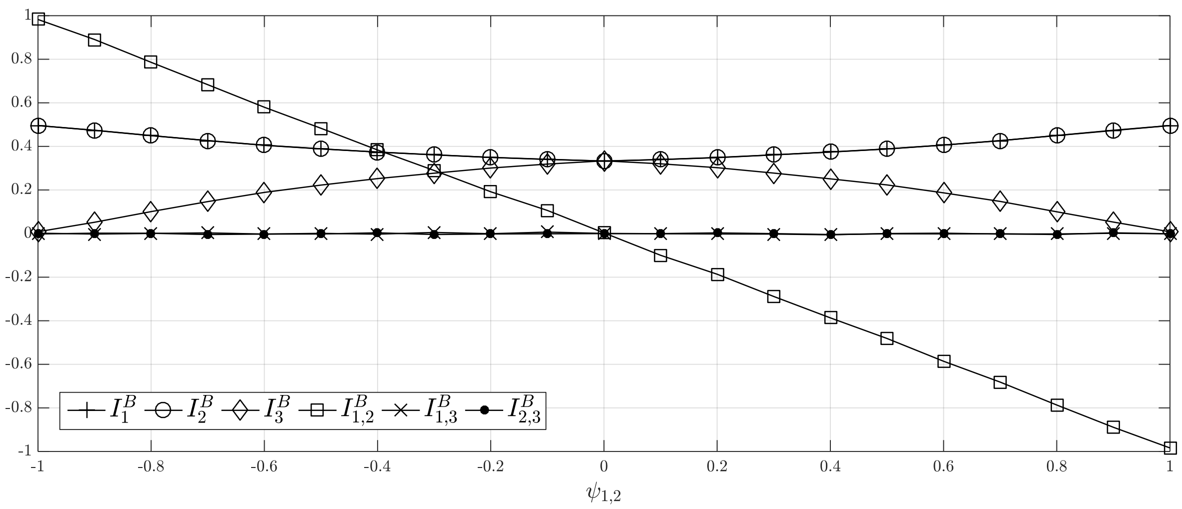
Based on these results, one may note that, by assuming the general model EA1, we have more flexibility in finding interaction indices that are similar to the correlation coefficients (even with values close to -1 or 1). Therefore, in order to satisfy the first constraint, if the correlation coefficient between criteria i and i' are high (in absolute value), both $\phi_i^{\mathcal{B}}$ and $\phi_{i'}^{\mathcal{B}}$ will assume high values. As a consequence, another $\phi_{i''}^{\mathcal{B}}$, whose criterion i'' is not correlated with any other one, may assume a value close to 0, which means that this criterion, alone, has no marginal contribution in the overall value. In real situations, this may be an inconvenience, so it would be interesting to avoid such a result. This could be avoided by assuming $\phi_i^{\mathcal{B}} = 1/m$ for all i, which corresponds to the model EA2. In this case, the obtained results also indicates that we can model the interaction between criteria 1 and 2, however, we have less flexibility in the adjustments of $I_{1,2'}^{\mathcal{B}}$. This is explained by the bounds imposed on this parameter.

7.1.3 Numerical experiments on real data

Aiming at applying the proposed models into a real dataset, let us consider a ranking problem composed by $n = 182$ countries evaluated according to $m = 3$ criteria: forest area (% of land area), life expectancy at birth (years) and gross domestic product (GDP) per capita (current US$). This data refers to 2014 and was collected from World



(a) Capacity coefficients.



(b) Interaction indices.

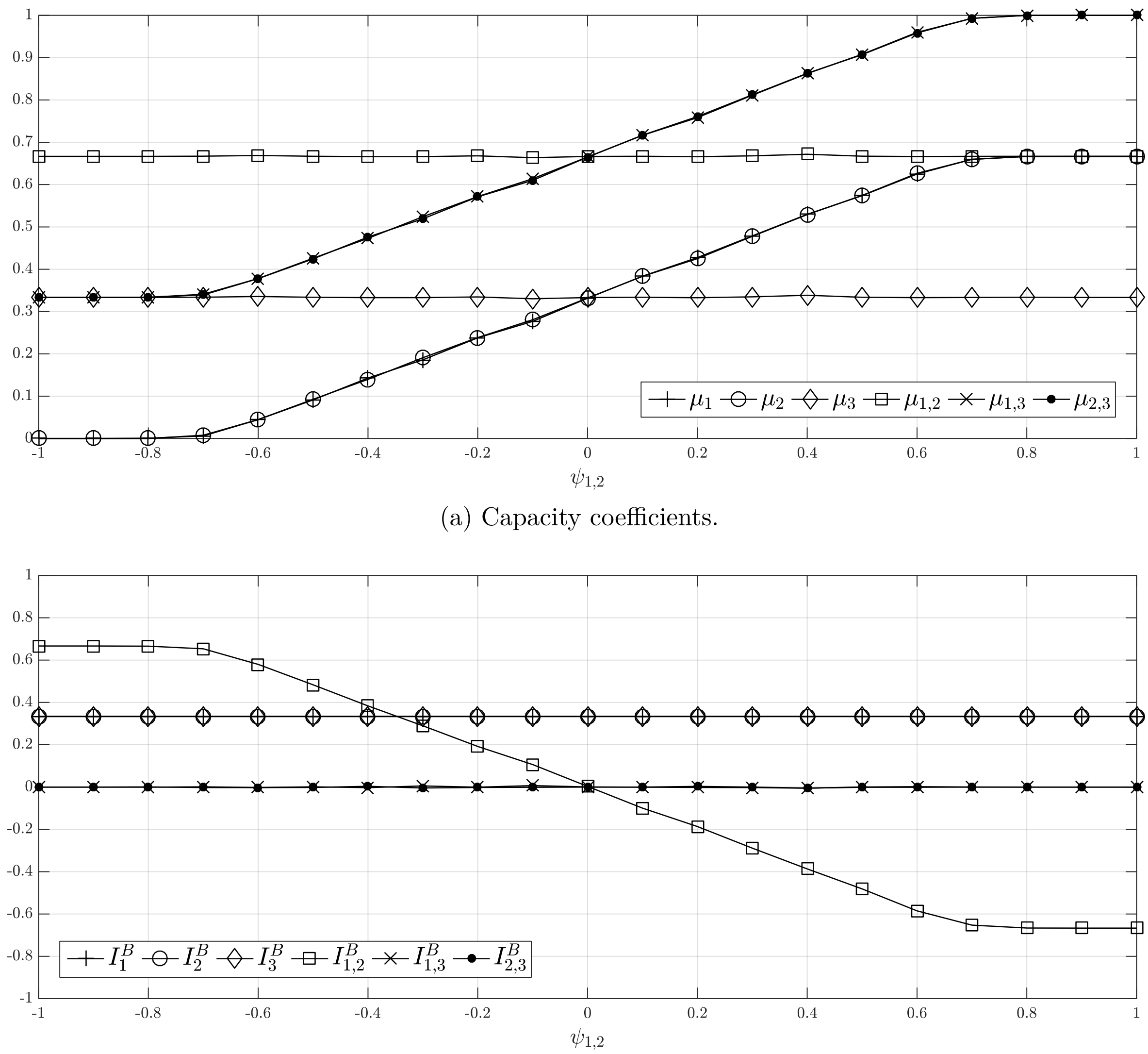
Figure 40 – Results for the general model.

Bank at http://www.worldbank.org/. We also normalized the decision data in the interval $[0, 1]$. The scatter plots of each pair of criteria are presented in Figure 42.

One may remark that life expectancy at birth and GDP per capita (criteria 2 and 3, respectively) have a certain degree of dependence. For instance, if we take the Pearson's correlation coefficient as a similarity measure, we have $\psi_{1,2} = 0.0064$, $\psi_{1,3} = -0.0292$ and $\psi_{2,3} = 0.5787$. Therefore, criteria 2 and 3 are positively correlated.

With the application of the proposed models, we obtain the parameters described in Table 18. We also compared our model with the method (DA) proposed by (Duarte, 2018).

With respect to the interaction indices, one may note that, in all cases, we achieved negative values for $I^{\mathcal{B}}_{2,3}$, which was expected since there is a positive correlation between criteria 2 and 3. Moreover, our proposed model EA2 and the approach pro-



(a) Capacity coefficients.

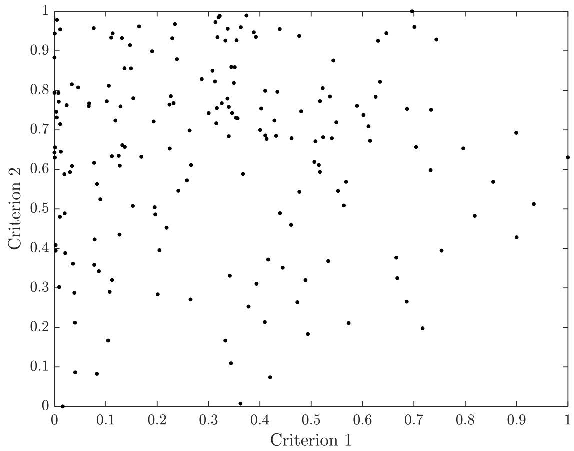
(b) Interaction indices.

Figure 41 – Results assuming $\phi_i^{\mathcal{B}} = 1/m$ for all i.

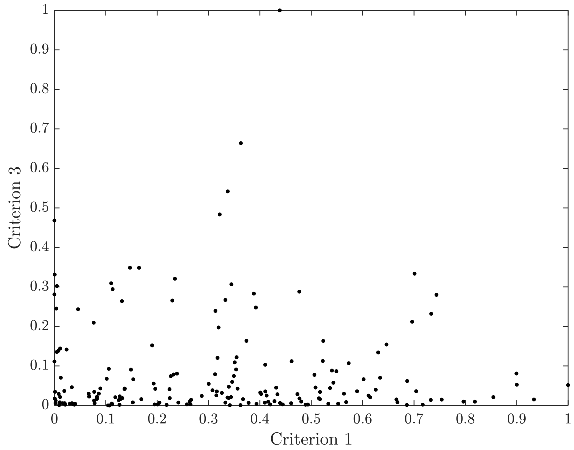
posed by (Duarte, 2018) led to the same results. This means that we achieved the same performance with an optimization model easier to be solved.

It is worth mentioning that we applied our approaches by using both multilinear model (ML) and Choquet integral (CI). For both aggregation functions, the obtained parameters were the same. However, as presented in Table 19, the derived rankings have some differences. Table 19 also shows the ranking provided by the application of the WAM (assuming $w_i = 1/m$, $i = 1, \ldots, m$).

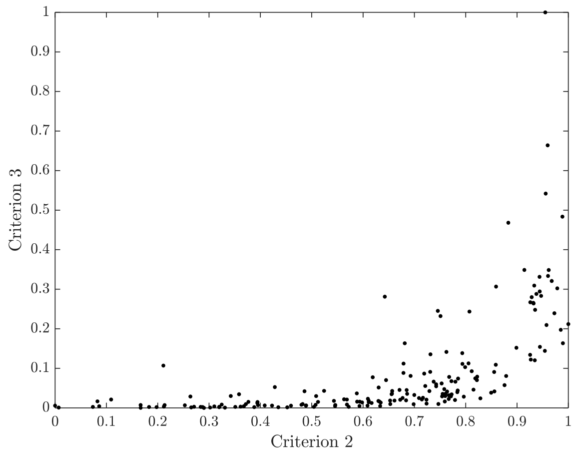
It is interesting to note that, when applying the WAM, the alternative $\mathbf{x}_{99}$ achieves the first position. However, in the EA2 model, this alternative has the worse performance. This can be explained by the fact that, since the WAM does not take into account the correlation among criteria 2 and 3, the aggregation is biased towards alternatives that have good evaluations in such criteria, because they measure the same hidden factor. If we apply the multilinear model (or the Choquet integral) with negative



(a) Scatter plot of criteria 1 and 2.



(b) Scatter plot of criteria 1 and 3.



(c) Scatter plot of criteria 2 and 3.

Figure 42 – Real data visualization.

values for $I^{\mathcal{B}}_{2,3}$ (or $I^{\mathcal{S}}_{2,3}$), this bias can be mitigated and the result could be fairer, since we avoid that the same latent factor is considered twice in the aggregation procedure. Therefore, alternatives like $\mathbf{x}_{85}$, $\mathbf{x}_{158}$ and $\mathbf{x}_{59}$ (the first three in the ranking provided by EA2 model), which do not have good evaluations on both criteria 2 and 3 but perform well in criterion 1, achieve good positions in the ranking.

On the other hand, the aforementioned explanation is not true for EA1. In this case, since the optimization model is less restrictive with respect to the interaction indices, it led to a small value for $\phi^{\mathcal{B}}_1$. Therefore, even with negative values for $I^{\mathcal{B}}_{2,3}$, the evaluation of criteria 1 has a weak impact on the aggregation and, as a consequence, alternatives that have good evaluations in both criteria 2 and 3 achieve better positions on the ranking.

In order to better exploit the application of the considered approaches, let us now include a fourth criterion, the inverse of CO2 emissions (metric tons per capita). It was also extracted from the same database. The Pearson's correlation coefficient between

Table 18 – Retrieved capacity and interaction indices (case with 3 criteria).

Capacity $\mu(A)$				**Interaction index** $I^{\mathcal{B}}(A)$			
A	DA	EA1	EA2	A	DA	EA1	EA2
$\{1\}$	0.3220	0.1482	0.3220	$\{1\}$	0.3333	0.1596	0.3333
$\{2\}$	0.6259	0.7170	0.6259	$\{2\}$	0.3333	0.4245	0.3333
$\{3\}$	0.6081	0.6907	0.6081	$\{3\}$	0.3333	0.4159	0.3333
$\{1,2\}$	0.9414	0.8589	0.9414	$\{1,2\}$	-0.0064	-0.0064	-0.0064
$\{1,3\}$	0.9592	0.8681	0.9592	$\{1,3\}$	0.0292	0.0292	0.0292
$\{2,3\}$	0.6553	0.8290	0.6553	$\{2,3\}$	-0.5787	-0.5787	-0.5787

Table 19 – Positions of the first 5 alternatives (case with 3 criteria).

Alternatives	Criteria evaluations			Position in the ranking				
	$u_1(\cdot)$	$u_2(\cdot)$	$u_3(\cdot)$	WAM	CI EA1	CI EA2	ML EA1	ML EA2
$\mathbf{x}_{59}$	0.7437	0.9286	0.2799	4	6	3	5	3
$\mathbf{x}_{85}$	0.6964	1	0.2119	5	2	1	2	1
$\mathbf{x}_{90}$	0.6464	0.9446	0.1542	8	8	5	8	4
$\mathbf{x}_{99}$	0.4387	0.9551	1	1	1	4	1	5
$\mathbf{x}_{101}$	0.3629	0.9597	0.6639	3	4	9	4	8
$\mathbf{x}_{158}$	0.7011	0.9604	0.3336	2	3	2	3	2
$\mathbf{x}_{159}$	0.3218	0.9884	0.4835	7	5	10	6	11

this criterion and the other ones are $\psi_{1,4} = -0.0819$, $\psi_{2,4} = -0.5752$ and $\psi_{3,4} = -0.2620$. Therefore, this criterion is negatively correlated with both criteria 2 and 3.

Table 20 presents a comparison in the obtained capacity coefficients and interaction indices. Moreover, Table 21 presents a comparison in the position of the first 5 alternatives. In this experiment, we do not compare with the approach proposed by (Duarte, 2018), since EA2 model leads to very similar estimations.

In terms of the parameters, we obtained the same conclusions as in the experiment with three criteria, i.e., the values of the estimated interaction indices $I^{\mathcal{B}}_{i,i'}$ were the negative of the Pearson's correlation coefficients between criteria i and i'. However, it is more evident that the lack of constraint restricting $\phi^{\mathcal{B}}_i$ may lead to values close to zero. This is the case of model EA1, where $\phi^{\mathcal{B}}_1 \approx 0$. Therefore, this model practically ignores the evaluation of criterion 1. In Table 21, one may note that, although some alternatives have bad performances on criterion 1, they achieve good positions in the ranking. This

Table 20 – Retrieved capacity and interaction indices (case with 4 criteria).

Capacity $\mu(A)$			**Interaction index** $I^{\mathcal{B}}(A)$		
A	EA1	EA2	A	EA1	EA2
$\{1\}$	0	0.2170	$\{1\}$	0	0.2500
$\{2\}$	0.4401	0.2640	$\{2\}$	0.4383	0.2500
$\{3\}$	0.4401	0.2640	$\{3\}$	0.2817	0.2500
$\{4\}$	0	0	$\{4\}$	0.2800	0.2500
$\{1,2\}$	0.4401	0.4810	$\{1,2\}$	0	0
$\{1,3\}$	0.4401	0.5000	$\{1,3\}$	0	0.0190
$\{1,4\}$	0	0.2640	$\{1,4\}$	0	0.0471
$\{2,3\}$	0.4401	0.2640	$\{2,3\}$	-0.4401	-0.2640
$\{2,4\}$	0.8766	0.5000	$\{2,4\}$	0.4366	0.2360
$\{3,4\}$	0.5634	0.4810	$\{3,4\}$	0.1234	0.2170
$\{1,2,3\}$	0.4401	0.5000	$\{1,2,3\}$	0	0
$\{1,2,4\}$	0.8766	0.7640	$\{1,2,4\}$	0	0
$\{1,3,4\}$	0.5634	0.7640	$\{1,3,4\}$	0	0
$\{2,3,4\}$	1	0.7170	$\{2,3,4\}$	0	0

is the case of alternatives $\mathbf{x}_{83}$, $\mathbf{x}_{153}$ and $\mathbf{x}_{159}$ (positions 5, 4 and 25, respectively, in the multilinear model).

By modeling redundancy between pairs of criteria, we may also note that alternatives such as $\mathbf{x}_{99}$, which have good evaluations on both correlated criteria 2 and 3, achieve good positions in the ranking provided by WAM but worse positions in the EA2 model. Moreover, by introducing synergy (or complementary effect) between criteria 2 and 4 and also 3 and 4, even if an alternative has a bad evaluation on criterion 4 but is well evaluated on criterion 2 and/or criterion 3, it can achieve a good position. This is the case, for example, of alternatives $\mathbf{x}_{59}$, $\mathbf{x}_{85}$ and $\mathbf{x}_{158}$ by assuming EA2 model.

7.2 An approach based on Sobol' indices

As mentioned in the beginning of this chapter, in this section we deal with the unsupervised capacity identification problem by using the Sobol' indices. Before describing our proposal, let us present an interesting result achieved by Grabisch and Labreuche (2017), which associates the Sobol' index to the multilinear model. Under the assumption that the input variables (criteria, in the addressed decision problem) are independent and that they follow a uniform distribution on $[0,1]$, the authors proved the following (see Appendix B for an example):

Table 21 – Positions of the first 5 alternatives (case with 4 criteria).

Alternatives	Criteria evaluations				Position in the ranking				
	$u_1(\cdot)$	$u_2(\cdot)$	$u_3(\cdot)$	$u_4(\cdot)$	WAM	CI EA1	CI EA2	ML EA1	ML EA2
$\mathbf{x}_{59}$	0.7437	0.9286	0.2799	0.0041	4	27	3	24	2
$\mathbf{x}_{83}$	0.3197	0.9852	0.1974	0.0075	24	6	26	5	27
$\mathbf{x}_{85}$	0.6964	1	0.2119	0.0037	5	3	1	3	1
$\mathbf{x}_{90}$	0.6464	0.9446	0.1542	0.0028	8	18	4	22	4
$\mathbf{x}_{99}$	0.4387	0.9551	1	0.0368	1	1	7	1	8
$\mathbf{x}_{101}$	0.3629	0.9597	0.6639	0.0016	3	14	23	8	18
$\mathbf{x}_{139}$	0.1951	0.5043	0.0025	0.5896	57	2	48	58	99
$\mathbf{x}_{153}$	0.3739	0.9894	0.1634	0.0079	19	5	17	4	16
$\mathbf{x}_{157}$	1	0.6304	0.0515	0.0114	12	122	5	111	5
$\mathbf{x}_{158}$	0.7011	0.9604	0.3336	0.0090	2	9	2	10	3
$\mathbf{x}_{159}$	0.3218	0.9884	0.4835	0.0094	7	4	20	2	22

Theorem 7.1. *(Grabisch and Labreuche (2017)) Consider the multilinear model F_{ML} of a capacity μ. The (nonnormalized) Sobol' index of a subset $\emptyset \neq A \subseteq C$ is given by*

$$Var[(F_{ML})_A] = \frac{1}{3^{|A|}} \left(\hat{\mu}(A)\right)^2, \tag{7.7}$$

where $\hat{\mu}$, defined by

$$\hat{\mu}(A) = \left(\frac{-1}{2}\right)^{|A|} I^{\mathcal{B}}(A), \tag{7.8}$$

is the Fourier transform[4] *of μ.*

Therefore, one may remark that the Sobol' index is associated with the Banzhaf interaction index through the equation

$$\mathrm{Var}[(F_{ML})_A] = \frac{1}{12^{|A|}} \left(I^{\mathcal{B}}(A)\right)^2. \tag{7.9}$$

As will be discussed in the next section, this relation will be used in our proposal to deal with the unsupervised capacity identification problem.

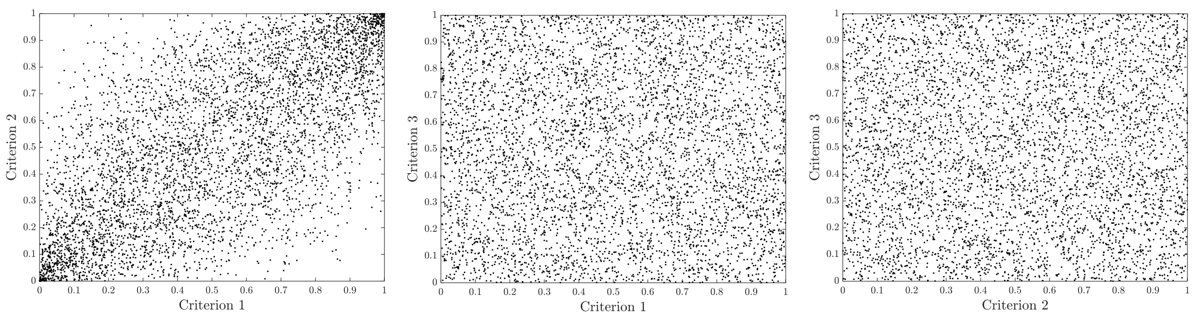
7.2.1 The proposed unsupervised approach for capacity identification

In order to present the motivations for our proposal, let us consider a synthetic example with $n = 5000$ alternatives and $m = 3$ criteria (generated according to a uniform distribution on $[0, 1]$). Figure 43 presents the scatter plot among pairs of criteria. One

[4] The Fourier transform is defined as the coordinates of a set function in the basis of the parity functions. For more details about this relation, see (Grabisch and Labreuche, 2017).

may note that there is a positive correlation between criteria 1 and 2, with a Pearson's correlation coefficient $\rho_{1,2} = 0.6768$.

Suppose that we want to analyze the impact that both criteria 1 and 3, alone, have in the output model. Let us assume that the considered capacity is additive, i.e., $\mu = [0, 1/3, 1/3, 1/3, 2/3, 2/3, 2/3, 1]$. Recall that this capacity leads to $\phi_1^{\mathcal{B}} = \phi_2^{\mathcal{B}} = \phi_3^{\mathcal{B}} = 1/3$. If we calculate the nonnormalized Sobol' index according to Equation (7.9), one obtains $\text{Var}[(F_{ML})_1] = \text{Var}[(F_{ML})_3] = (1/12)\left(\phi_i^{\mathcal{B}}\right)^2 \approx 0.0093$. On the other hand, if we consider Equation (3.34) (without normalization and with Y and Z_i being the multilinear model and criterion i, respectively), which takes into account statistics extracted from the considered dataset, one obtains $\text{Var}[\mathbb{E}[F_{ML}|\mathbf{u}_1]] \approx 0.0263$ and $\text{Var}[\mathbb{E}[F_{ML}|\mathbf{u}_3]] \approx 0.0091$, where $\mathbf{u}_i = [u_{1,i}, u_{2,i}, \ldots, u_{n,i}]$. Although both values for $\text{Var}[(F_{ML})_3]$ and $\text{Var}[\mathbb{E}[F_{ML}|\mathbf{u}_3]]$ are similar, one may clearly note that we achieved very different ones comparing $\text{Var}[(F_{ML})_1]$ and $\text{Var}[\mathbb{E}[F_{ML}|\mathbf{u}_1]]$. This difference is due to the existing correlation between criteria 1 and 2, which violates the hypothesis about the independence between the input variables assumed by Theorem 7.1. As a consequence, since these criteria are positively correlated, $\text{Var}[\mathbb{E}[F_{ML}|\mathbf{u}_1]]$ is also influenced by criterion 2, which increases the impact of criterion 1 in the output model. With respect to $\text{Var}[\mathbb{E}[F_{ML}|\mathbf{u}_3]]$, since criterion 3 has no correlation with any other criterion, its value remains independent on the other inputs.



(a) Criteria 1 and 2. (b) Criteria 1 and 3. (c) Criteria 2 and 3.

Figure 43 – Scatter plot of pairs of criteria.

In the aforementioned example, we adopted an additive capacity. Therefore, in terms of relative importance assigned to each criterion, we assumed that they all are the same (see, for instance, that $\phi_1^{\mathcal{B}} = \phi_2^{\mathcal{B}} = \phi_3^{\mathcal{B}} = 1/3$). However, due to the correlation between criteria 1 and 2, criterion 1 (and also criterion 2) have more impact on the overall values in comparison with criterion 3. As a consequence, the existing correlation may bias the obtained ranking by favoring alternatives with good evaluations only in criteria 1 and 2. In such a scenario, in which we have neither further information about the capacity coefficients nor overall evaluations provided by the decision maker, it may be interesting to adopt a capacity that is able to compensate this bias and lead to fairer overall evaluations for the alternatives.

In this context, our hypothesis is that the bias provided by correlations in the decision data can be mitigated by adopting a capacity whose coalitions of criteria with the same cardinality have similar impacts on the obtained overall evaluations. For instance, in order to deal with the correlation in the aforementioned example, one may define a capacity μ such that the impact of criterion 3 is the same as the impact of criteria 1 and 2. If one extends this idea to any $A \subseteq C$, one avoids disparities in the impacts of all possible coalitions of criteria with the same cardinality. Since these impacts can be measured by Equation (3.34), the aim here is to adjust a capacity μ that minimizes the difference between the Sobol' indices for subsets of criteria with the same cardinality. Mathematically, the optimization problem is given by

$$\min_{\mu} \sum_{\substack{A \subset C, \\ A \neq \emptyset}} \sum_{\substack{B \subset C, \\ |B|=|A|}} \left(S_A - S_B\right)^2 . \tag{7.10}$$

Recall that we must also satisfy the axioms of a capacity.

7.2.2 Numerical experiments

This section presents the numerical experiments and the obtained results.

7.2.2.1 Application of our proposal in an illustrative example

In order to verify the application of our proposal to a synthetic dataset, let us consider the example described in Section 7.2.1. Aiming at reducing the number of parameters to be estimated while keeping the flexibility to model interactions, we considered the 2-additive multilinear model. Moreover, in the absence of further information about the capacity coefficients, we adopted the same value for all $\mu(\{i\})$, i.e., $\mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = 1/3$.

Given the aforementioned assumptions, we only need to find the capacity coefficients associated with pairs of criteria. With respect to the constraints, other than the axioms of capacity, we must also guarantee that $I^{\mathcal{B}}(C) = 0$. This leads to the condition

$$\mu(\{1,2\}) + \mu(\{1,3\}) + \mu(\{2,3\}) = 2. \tag{7.11}$$

Moreover, aiming at achieving an aggregation function whose individual criteria have similar impacts on the overall evaluations, we only considered the minimization of the difference between first-order Sobol' indices, i.e., the subsets A in (7.10) are restricted to singletons i. The resulting optimization problem is the following:

$$\begin{array}{ll} \min\limits_{\mu(\{1,2\}),\mu(\{1,3\}),\mu(\{2,3\})} & \sum\limits_{i,i' \in C} \left(S_i - S_{i'}\right)^2 \\ \text{s.t.} & \mu(\{1,2\}) + \mu(\{1,3\}) + \mu(\{2,3\}) = 2, \\ & \mu(\{i\}) \leq \mu(\{i,i'\}),\ \forall i, i \in C, \\ & \mu(\emptyset) = 0, \\ & \mu(C) = 1 \end{array} \tag{7.12}$$

In order to solve (7.12), we adopted a simple iterative heuristic method based on the golden section search (Vajda, 1989). For instance, we started with the additive capacity $\mu = [0, 1/3, 1/3, 1/3, 2/3, 2/3, 2/3, 1]$ and selected at random a $\mu(\{i, i'\})$ to be fixed ($\mu(\{1, 3\}) = 2/3$, for example). Thereafter, we selected another $\mu(\{l, l'\})$ ($\mu(\{1, 2\})$, for example) to be optimized and applied the golden section search in the associated dimension to deal with the addressed optimization problem. In other words, by fixing $\mu(\{i, i'\})$, we perform a one-dimensional search on $\mu(\{l, l'\})$ that solves (7.10). It is important to note that, since we must satisfy the condition (7.11), when applying the golden section search on $\mu(\{l, l'\})$, the other capacity coefficient that was not selected so far is automatically adjusted ($\mu(\{2, 3\}) = 2 - 2/3 - \mu(\{1, 2\})$, in this case). This procedure is repeated until the convergence to the minimum (possibly a local one).

The application of the proposed approach in the illustrative example led to the capacity and the associated interaction indices presented in Table 22.

Table 22 – Obtained capacity and interaction indices.

	A							
	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{1, 3\}$	$\{2, 3\}$	C
$\mu(A)$	0	0.3333	0.3333	0.3333	0.4190	0.7778	0.8033	1
$I^{\mathcal{B}}(A)$	0.5000	0.2650	0.2778	0.4572	-0.2477	0.1111	0.1366	0

One may note that we achieved $I^{\mathcal{B}}_{1,2} < 0$, which was expected since both criteria 1 and 2 are correlated. Moreover, the obtained Banzhaf power index $\phi^{\mathcal{B}}_3$ was higher in comparison to the other ones, which also contributes to increase the impact of criterion 3 (the independent one) in the output model. With respect to the first-order Sobol' indices, we achieved $S_1 \approx S_2 \approx S_3 \approx 0.0173$.

7.2.2.2 Experiments varying the number of alternatives and the degree of the correlation

With the purpose of further investigating our proposal, we considered several scenarios with different numbers of alternatives and different degrees of the correlation between criteria 1 and 2. In all cases, we considered 3 decision criteria and generated the evaluations according to a uniform distribution on $[0, 1]$. Based on the same assumptions made in the last experiment, the obtained interaction indices (averaged over 100 simulations) for decision problems with $\rho_{1,2} \approx 0.75$, $\rho_{1,2} \approx 0$ and $\rho_{1,2} \approx -0.75$ are presented in Figures 44, 45 and 46, respectively.

One may note that, in all cases, the stability in the obtained capacity increases as the number of alternatives also increases. Since the Sobol' indices calculation requires statistical measures extracted from the decision data, the more data we have the better is

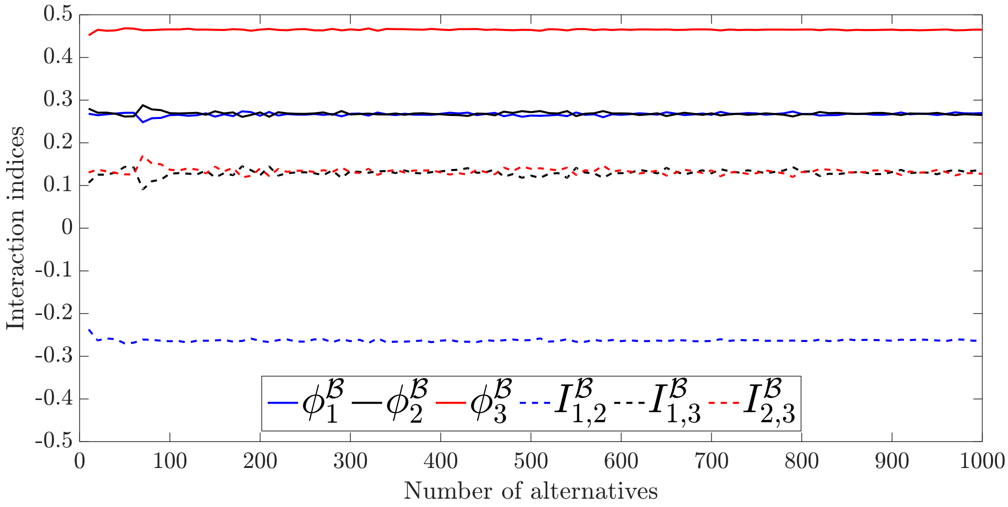


Figure 44 – Results for $\rho_{1,2} \approx 0.75$.

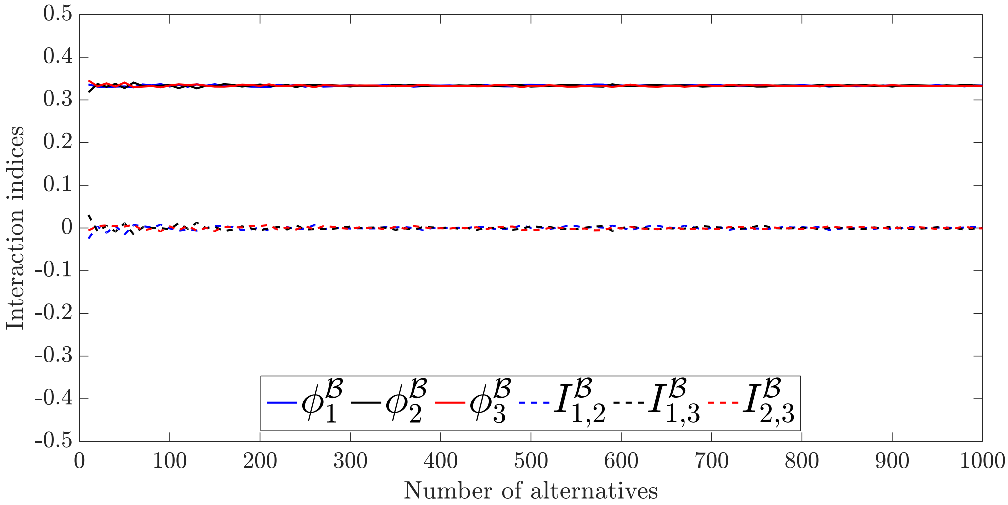


Figure 45 – Results for $\rho_{1,2} \approx 0$.

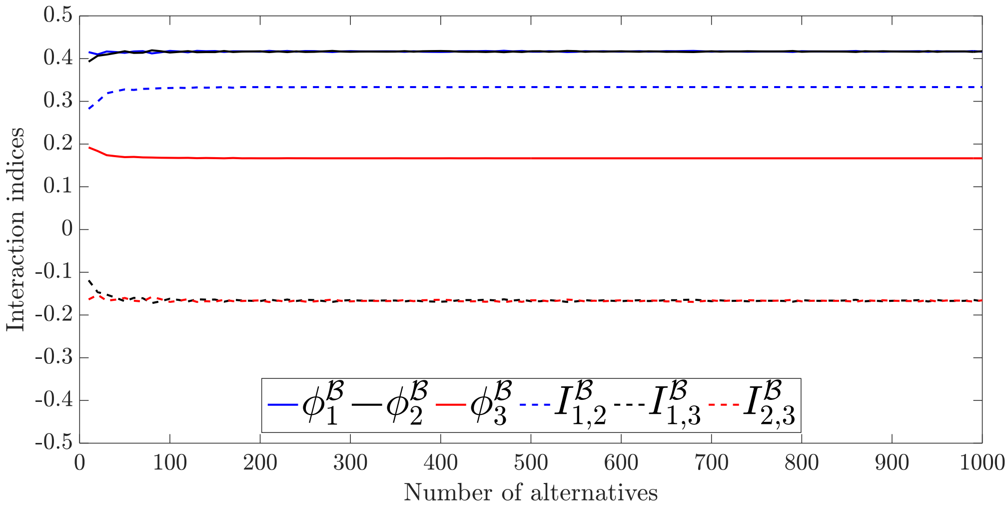


Figure 46 – Results for $\rho_{1,2} \approx -0.75$.

the estimation of the variances and conditional expectations and, as a consequence, the more stable are the estimated capacity coefficients.

With respect to the obtained interaction indices, we clearly see in Figure 44 that, in order to compensate the positive correlation between criteria 1 and 2, we achieved $I_{1,2}^{\mathcal{B}} < 0$ (a redundant effect for this correlated criteria). Moreover, both $I_{1,3}^{\mathcal{B}}$ and $I_{2,3}^{\mathcal{B}}$ values were slightly greater than zero and the marginal contribution of criterion 3 ($\phi_3^{\mathcal{B}}$) was greater than the other ones. Conversely, in Figure 46, the negative correlation between criteria 1 and 2 led to $I_{1,2}^{\mathcal{B}} > 0$ (a complementary effect for this correlated criteria), both $I_{1,3}^{\mathcal{B}}$ $I_{2,3}^{\mathcal{B}}$ values were slightly smaller than zero and the marginal contribution of criterion 3 ($\phi_3^{\mathcal{B}}$) was lower than the other ones.

In Figure 45, which contains the results when all criteria are independent, one may see that the obtained capacity is an additive one. Therefore, it was not needed to model interactions or increase marginal contributions to balance the Sobol' indices, since they were already similar.

7.2.2.3 Application to a real dataset

The experiments conducted in the two previous sections led to interesting results when dealing with correlated criteria. Remark that, in both cases, we generated all data based on a uniform distribution, which is in accordance with the assumption of Theorem 7.1. However, the hypotheses that the observed decision criteria follow a uniform distribution may not be true in practical situations. In order to illustrate this scenario, we conduct an experiment based on the same real dataset used in Section 7.1.3 (by considering as decision criteria the forest area, life expectancy at birth and GDP). The application of our proposal leads to the parameters and ranking presented in Tables 23 and 24, respectively.

Table 23 – Obtained capacity and interaction indices.

	A							
	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	C
$\mu(A)$	0	0.3333	0.3333	0.3333	0.5774	1	0.4226	1
$I^{\mathcal{B}}(A)$	0.5000	0.4554	0.1667	0.3779	-0.0892	0.3333	-0.2441	0

In Table 23, one may note that, although $I_{2,3}^{\mathcal{B}} = -0.2411$, we also achieved $I_{1,3}^{\mathcal{B}} = 0.3333$. The first result was expected since criteria 2 and 3 are positively correlated, however, we did not expect the latter one since there is no correlation between criteria 1 and 3. This counter-intuitive result can be explained by two facts: (i) the data distribution, which is different from uniform, and (ii) the number of learning data, which may not be enough to properly estimate the statistics used in the Sobol' indices calculation. As a

Table 24 – Positions of the first 5 alternatives.

Alternatives	Criteria evaluations			Position in the ranking	
	$u_1(\cdot)$	$u_2(\cdot)$	$u_3(\cdot)$	WAM	Proposal
$\mathbf{x}_{59}$	0.7437	0.9286	0.2799	4	3
$\mathbf{x}_{85}$	0.6964	1	0.2119	5	4
$\mathbf{x}_{99}$	0.4387	0.9551	1	1	1
$\mathbf{x}_{101}$	0.3629	0.9597	0.6639	3	5
$\mathbf{x}_{158}$	0.7011	0.9604	0.3336	2	2

consequence, as can be seen in Table 24, the obtained ranking was very similar to the one achieved by the WAM (or by using an additive capacity).

7.3 Conclusions

In this chapter, we addressed the problem of estimating the multilinear model parameters by means of an unsupervised approach. In this formulation, we assumed that the correlation among decision data may bias the obtained ranking. Therefore, we aim at defining a capacity that leads to fairer overall evaluations in the sense that this bias is mitigated.

As a first contribution, described in Section 7.1, we extended the optimization problem proposed in (Duarte, 2018) to be used in the multilinear model case. We associate the interaction indices used in the multilinear model to similarity measures between pairs of criteria. Moreover, we perform some adaptations in order to turn it into a quadratic problem. The results obtained by applying the proposed method in both synthetic and real dataset indicated that the estimated capacity can overcome the bias provided by correlations among decision data by penalizing (resp., favoring) criteria that are positively (resp., negatively) correlated. However, a drawback of our proposal lies in the bounds that we impose to the interaction indices, which may reduce the flexibility in scenarios with a large number of criteria.

The second analysis conducted in this chapter associates the Banzhaf interaction indices to the Sobol' indices. We assumed that all singletons should have the same impact on the output model and used the Sobol' indices as a means of comparison. The obtained results attested the application of the proposed approach in a scenario where the data distribution is uniform. By using the synthetic data, the results were similar to the ones obtained in the previous analysis. In situations with a positive (resp., negative) correlation between a pair of criteria, the achieved interaction index modeled a redundancy (resp., complementary) effect. Moreover, the power indices were also adjusted in order to balance the Sobol' indices.

However, the application on a real data, which does not follow a uniform distribution, led to unexpected results. The bias provided by correlations could not be avoided in the obtained ranking. Therefore, future perspectives consist in a further analysis on the impact that distributions different from the uniform may have in the Sobol' indices. Moreover, we would like to investigate different search algorithms as well as other assumptions about the capacities that one may consider to deal with the optimization problem (7.10).

8 Preliminary investigations on the use of capacities in group MCDM

The last chapter of this thesis investigates some issues in group MCDM problems by using capacity-based aggregation functions. We consider two different situations. In the first one, addressed in Section 8.1, we assume that we have a single observed decision matrix $\mathbf{M}$, which is independent of the decision makers. However, each DM defines his/her own capacity coefficients and aggregates the criteria evaluations. In this scenario, we investigate if it is possible to combine the capacities provided be the DMs and the one that aggregate their opinions into a single set of parameters.

As a second study, presented in Section 8.2, we assume the same scenario as described in Section 5.2.4. We consider that each decision maker provides his/her own criteria evaluations and that the capacities associated with criteria and with decision makers are predefined. In this situation, we investigate if one may face inconveniences with respect to the commutativity property. A preliminary study on both cases are described in the sequel.

8.1 Case 1: A single decision data and different capacities

As mentioned in the previous section, in the first group MCDM scenario considered in this chapter we have a single decision matrix $\mathbf{M} = (u_{j,i})_{n\times m}$, where j and i indicate the alternative j and the criterion i, respectively. Therefore, even if we have a set of t decision makers, the decision data will be the same for all of them. However, each one will aggregate the criteria according the his/her capacity μ^k $(k = 1, \ldots, t)$. In this case, the overall value of an alternative $\mathbf{x}$ (we skip the index j for convenience of notation) according to the decision maker k, by using the Choquet integral, is given by

$$F_{CI}^{k}(u(\mathbf{x})) = \sum_{i=1}^{m} \left(u_{(i)} - u_{(i-1)}\right) \mu^k(\{(i), \ldots, (m)\}). \tag{8.1}$$

Based on all $F_{CI}^{k}(u(\mathbf{x}))$, $k = 1, \ldots, t$, one may aggregate the decision makers opinions, also by using the Choquet integral, as follows:

$$F_{CI}(F_{CI}^{1}, \ldots, F_{CI}^{t}) = \sum_{k=1}^{t} \left(F_{CI}^{(k)} - F_{CI}^{(k-1)}\right) \mu^D(\{(k), \ldots, (t)\}), \tag{8.2}$$

where $F_{CI}^{(k)}$ indicates a permutation of the indices k such that $0 \leq F_{CI}^{(1)} \leq \ldots \leq F_{CI}^{(t)} \leq 1$ and μ^D is a predefined capacity used to aggregate the decision makers opinions.

One may note that Equation (8.2) comprises a multilevel Choquet integral. Our investigation in this situation is that the capacities μ^D and μ^k, $k = 1, \ldots, t$ can be combined into a single one. In order to address this issue, we used the result obtained by Murofushi (2004), which indicates that a multilevel Choquet integral can be represented by a 2-step aggregation procedure in which the first step comprises an additive aggregation and the second one comprises a 0-1 Choquet integral[1]. This conclusion is based on a result provided by Ovchinnikov (2002a) on piecewise linear functions (further details can also be seen in (Ovchinnikov, 2002b)). In the next section, we present this result on an illustrative example.

8.1.1 Illustrative example

Suppose a decision problem with 3 decision makers and 3 criteria, whose capacities are given by

$$\mu^1 = [0, 0.3, 0.2, 0.4, 0.5, 0.9, 0.6, 1],$$

$$\mu^2 = [0, 0.4, 0.4, 0.2, 0.9, 0.5, 0.7, 1],$$

$$\mu^3 = [0, 0.3, 0.5, 0.1, 0.8, 0.8, 0.6, 1]$$

and

$$\mu^D = [0, 0.3, 0.3, 0.3, 0.6, 0.6, 0.6, 1].$$

In order to achieve the aggregation procedure described by Murofushi (2004), one firstly needs to find the piecewise linear function and the associated domains. Therefore, we have to verify all possible order combinations among the evaluations u_i, $i = 1, \ldots, m$, and the overall values $F^k_{CI}(u(\mathbf{x}))$, $k = 1, \ldots, t$. For example, by assuming $u_2 \leq u_3 \leq u_1$, we achieve

$$F^1_{CI}(u(\mathbf{x})) = 0.3u_1 + 0.1u_2 + 0.6u_3,$$

$$F^2_{CI}(u(\mathbf{x})) = 0.4u_1 + 0.5u_2 + 0.1u_3$$

and

$$F^3_{CI}(u(\mathbf{x})) = 0.3u_1 + 0.2u_2 + 0.5u_3.$$

Moreover, by assuming $F^3_{CI}(u(\mathbf{x})) \leq F^2_{CI}(u(\mathbf{x})) \leq F^1_{CI}(u(\mathbf{x}))$, which leads to the conditions $0.4u_3 \leq 0.1u_1 + 0.3u_2$ and $0.1u_1 + 0.4u_2 \leq 0.5u_3$, the overall evaluation is given by

$$F_{CI}(F^1_{CI}, \ldots, F^t_{CI}) = 0.33u_1 + 0.26u_2 + 0.41u_3.$$

Since we have 3 criteria and 3 decision makers, we have $3! * 3! = 36$ possible combinations. However, some of them may be infeasible. Moreover, some combinations

[1] It is important to note that one may also apply the multilinear model to aggregate the criteria and the decision makers opinions. However, the result obtained by Murofushi is restricted to the use of Choquet integrals.

may also have the same additive weights, which may be merged into a single function (with the union of the domains). After verifying all combinations and performing the adjustments on the domains, one achieves the following piecewise linear function:

$$F_{PLF}(u(\mathbf{x})) = \begin{cases} 0.33u_1 + 0.38u_2 + 0.29u_3, & u_3 \leq u_2 \leq u_1 \\ 0.36u_1 + 0.35u_2 + 0.29u_3, & u_3 \leq u_1 \leq u_2 \\ 0.33u_1 + 0.26u_2 + 0.41u_3, & u_2 \leq u_3, 0.40u_3 \leq 0.10u_1 + 0.30u_2 \\ 0.34u_1 + 0.29u_2 + 0.37u_3, & u_2 \leq u_3 \leq u_1, 0.10u_1 + 0.30u_2 \leq 0.40u_3 \\ 0.52u_1 + 0.26u_2 + 0.22u_3, & u_2 \leq u_1, 0.40u_1 \leq 0.30u_2 + 0.10u_3 \\ 0.48u_1 + 0.29u_2 + 0.23u_3, & u_2 \leq u_1 \leq u_3, 0.30u_2 + 0.10u_3 \leq 0.40u_1 \\ 0.37u_1 + 0.41u_2 + 0.22u_3, & u_1 \leq u_2 \leq u_3 \\ 0.37u_1 + 0.35u_2 + 0.28u_3, & u_1 \leq u_3 \leq u_2 \end{cases} . \tag{8.3}$$

For simplicity of notation, let us represent each function in (8.3) by $f_v(u(\mathbf{x}))$, $v = 1, \ldots, V$, and the respective domain by Q_v. For example, $f_1(u(\mathbf{x})) = 0.33u_1 + 0.38u_2 + 0.29u_3$ and $Q_1 = \{u(\mathbf{x}) | u_3 \leq u_2 \leq u_1\}$. Given all $f_v(u(\mathbf{x}))$, $v = 1, \ldots, V$, the next task consists in determining the capacity μ^{0-1} of the 0-1 Choquet integral. For this purpose, we need to find the hyperplanes $H_{v,v'} = \{u(\mathbf{x}) | f_v(u(\mathbf{x})) = f_{v'}(u(\mathbf{x}))\}$ composed by non-empty adjacent regions Q_v and $Q_{v'}$ whose intersection is at most $(m-1)$-dimensional. Therefore, among these hyperplanes, we select the ones that satisfy[2] $F_{PLF}(u(\mathbf{x})) = f_v(u(\mathbf{x})) \vee f_{v'}(u(\mathbf{x}))$, where $\vee$ is the maximum operator. In the considered example, one obtains the hyperplanes $H_{1,3}$ and $H_{5,7}$.

The next step comprises the definition of a set of regions $T \in \mathbf{T}$ by dividing the whole space $\mathbb{R}^m$ by means of the aforementioned hyperplanes. In our example, we have (remark that $\cup T = \mathbb{R}^m$)

$$T_{1,3,5,7} = \{u(\mathbf{x}) | f_1(u(\mathbf{x})) \leq f_3(u(\mathbf{x})), f_5(u(\mathbf{x})) \leq f_7(u(\mathbf{x}))\},$$

$$T_{1,3,7,5} = \{u(\mathbf{x}) | f_1(u(\mathbf{x})) \leq f_3(u(\mathbf{x})), f_7(u(\mathbf{x})) \leq f_5(u(\mathbf{x}))\},$$

$$T_{3,1,5,7} = \{u(\mathbf{x}) | f_3(u(\mathbf{x})) \leq f_1(u(\mathbf{x})), f_5(u(\mathbf{x})) \leq f_7(u(\mathbf{x}))\}$$

and

$$T_{3,1,7,5} = \{u(\mathbf{x}) | f_3(u(\mathbf{x})) \leq f_1(u(\mathbf{x})), f_7(u(\mathbf{x})) \leq f_5(u(\mathbf{x}))\}.$$

For each $T \in \mathbf{T}$, one defines $S(T) = \{v \in \{1, \ldots, V\} \, | f_v(u(\mathbf{x})) \geq F_{PLF}(u(\mathbf{x})) \;\; \forall u(\mathbf{x}) \in T\}$, which are used to find the overall value of alternative $\mathbf{x}$ through the Max/Min operator

$$F(u(\mathbf{x})) = \vee_{T \in \mathbf{T}} \wedge_{v \in S(T)} f_v(u(\mathbf{x})), \tag{8.4}$$

[2] One may remark that $F_{PLF}(u(\mathbf{x})) = f_v(u(\mathbf{x})) \vee f_{v'}(u(\mathbf{x}))$ is satisfied when $f_v(u(\mathbf{x})) \leq f_{v'}(u(\mathbf{x}))$ for $u(\mathbf{x}) \in Q_v$ and $f_{v'}(u(\mathbf{x})) \leq f_v(u(\mathbf{x}))$ for $u(\mathbf{x}) \in Q_{v'}$.

where $\wedge$ is the minimum operator. Therefore, one may remark that Equation (8.4) comprises a 2-step aggregation procedure in which $f_v(u(\mathbf{x}))$, $v = 1, \ldots, V$, and the Max/Min operator represent the additive part and the 0-1 Choquet integral, respectively.

In our example, we have

$$S(T_{1,3,5,7}) = \{1, 2, 3, 4, 7, 8\},$$

$$S(T_{1,3,7,5}) = \{3, 4, 5, 6\},$$

$$S(T_{3,1,5,7}) = \{1, 2, 7, 8\}$$

and

$$S(T_{3,1,7,5}) = \{1, 2, 5, 6, 7, 8\},$$

which leads to

$$F(u(\mathbf{x})) = (f_1 \wedge f_2 \wedge f_3 \wedge f_4 \wedge f_7 \wedge f_8) \vee (f_3 \wedge f_4 \wedge f_5 \wedge f_6) \vee (f_1 \wedge f_2 \wedge f_7 \wedge f_8) \vee (f_1 \wedge f_2 \wedge f_5 \wedge f_6 \wedge f_7 \wedge f_8)$$

and, therefore,

$$F(u(\mathbf{x})) = (f_1 \wedge f_2 \wedge f_7 \wedge f_8) \vee (f_3 \wedge f_4 \wedge f_5 \wedge f_6), \tag{8.5}$$

whose capacity μ^{0-1} is given by

$$\mu^{0-1}(A) = \begin{cases} 1, & \text{if } \{1, 2, 7, 8\} \text{ or } \{3, 4, 5, 6\} \subseteq A \\ 0, & \text{otherwise} \end{cases}. \tag{8.6}$$

In order to attest this procedure, let us consider that the evaluations of an alternative $\mathbf{x}$ are $u(\mathbf{x}) = (0.8, 0.2, 0.5)$. By using Equation (8.5), one achieves

$$F(u(\mathbf{x})) = (0.485 \wedge 0.503 \wedge 0.488 \wedge 0.5060) \vee (0.521 \wedge 0.515 \wedge 0.578 \wedge 0.557) = 0.515.$$

With the application of Equations (8.1) and (8.2), one obtains $F_{CI}^1 = 0.56$, $F_{CI}^2 = 0.47$, $F_{CI}^3 = 0.53$ and, then, $F_{CI} = 0.515$, i.e., both procedures are equivalent.

8.2 Case 2: Different decision data and same capacity

This section considers that each decision maker d_k, $k = 1, \ldots, t$, provides his/her own criteria evaluations (i.e., his/her own $\mathbf{M}_k$) . However, we assume that the information about criteria is aggregated according to a single predefined capacity μ^C, which is independent of the DMs. We also assume a predefined capacity μ^D, which is associated with coalitions of decision makers.

Remark that this decision data is equivalent to the one described in Section 5.2.4. For instance, it can be seen as a 3-dimensional matrix $\mathcal{M} = (u_{j,i,k})_{n \times m \times t}$.

In this scenario, one may think in two different 2-step aggregation procedures. In the first one (called here C/D procedure), we aggregate the criteria evaluations, which leads to a $n \times t$ (alternatives $\times$ decision makers) matrix, and, then we aggregate the DMs. Therefore, for an alternative $\mathbf{x}$ (we also skip the index j for convenience of notation) and by using the Choquet integral, we have the following procedure:

$$F_{CI}^{k}(u^k(\mathbf{x})) = \sum_{i=1}^{m} \left(u_{(i)}^{k} - u_{(i-1)}^{k} \right) \mu^C(\{(i), \ldots, (m)\}),\ k = 1, \ldots, t, \tag{8.7}$$

where $u^k(\mathbf{x}) = (u_1^k, \ldots, u_m^k)$, and, given all $F_{CI}^k(u^k(\mathbf{x}))$,

$$F_{CI}^{C/D}(F_{CI}^1, \ldots, F_{CI}^t) = \sum_{k=1}^{t} \left(F_{CI}^{(k)} - F_{CI}^{(k-1)} \right) \mu^D(\{(k), \ldots, (t)\}). \tag{8.8}$$

In the second case (called here D/C procedure), we first aggregate the decision makers opinions, which leads to a $n \times m$ (alternatives $\times$ criteria) matrix, and, then we aggregate the merged criteria. Therefore, for an alternative $\mathbf{x}$ and by using the Choquet integral, we have

$$F_{CI}^{i}(u^i(\mathbf{x})) = \sum_{k=1}^{t} \left(u_{i}^{(k)} - u_{i}^{(k-1)} \right) \mu^D(\{(k), \ldots, (t)\}),\ i = 1, \ldots, m \tag{8.9}$$

where $u^i(\mathbf{x}) = (u_i^1, \ldots, u_i^t)$, and, given all $F_{CI}^i(u^i(\mathbf{x}))$,

$$F_{CI}^{D/C}(F_{CI}^1, \ldots, F_{CI}^m) = \sum_{i=1}^{m} \left(F_{CI}^{(i)} - F_{CI}^{(i-1)} \right) \mu^C(\{(k), \ldots, (t)\}). \tag{8.10}$$

An interesting aspect is that, since the Choquet integral depends on the order of its arguments, both Equations (8.8) and (8.10) may not lead to the same value for an alternative $\mathbf{x}$. In order words, these equations may not commute. For instance, consider that the decision makers 1, 2 and 3 provide the criteria evaluations $u^1(\mathbf{x}) = [4, 8, 3]$, $u^2(\mathbf{x}) = [7, 5, 9]$ and $u^3(\mathbf{x}) = [9, 8, 5]$, respectively. Moreover, assume the following capacities:

$$\mu^C = [0, 0.1, 0.5, 0.4, 0.7, 0.5, 0.8, 1] \text{ and } \mu^D = [0, 0.2, 0.2, 0.2, 0.6, 0.6, 0.6, 1].$$

If one aggregates the criteria and, then, the decision makers, one achieves $F_{CI}^{C/D} = 6.44$. However, if one aggregates the decision makers and, then, the criteria, one achieves $F_{CI}^{D/C} = 6.14$. Therefore, the obtained overall values are different. This may be inconvenient in applications in which there is no explanation for the use of one procedure over the other one. In this case, it should be used a set of parameters that ensures that the overall values calculated for all alternatives are independent on the order that the collected decision data are aggregated.

Motivated by the aforementioned problem, in this section, we conduct an analysis whose aim is to verify which particular capacities used in Choquet integral ensure the commutativity property. It is worth mentioning that this problem has been treated in the context of Sugeno integrals (Dubois et al., 2018, 2019).

8.2.1 Particular cases of Choquet integral and results on the commutativity property

Before discussing the commutativity property, let us define some particular cases of Choquet integral. They are presented as follows (see (Grabisch et al., 2009) for further details):

- The smallest normalized capacity: Defined by $\mu_s(A) := 0$, $\forall A \subset C$, this capacity leads to the minimum operator.
- The greatest normalized capacity: Defined by $\mu_g(A) := 1$, $\forall A \subseteq C$, $A \neq \emptyset$, this capacity leads to the maximum operator.
- The Dirac measure centered on i (P_i): Defined by

$$\mu(A) = \begin{cases} 1, & \text{if } i \in A \\ 0, & \text{otherwise} \end{cases}.$$

- The arithmetic mean: It consists of a capacity $\mu(A)$ such that $\mu(A) = |A|/m$.
- The weighted arithmetic mean: It consists of a capacity $\mu(A)$ such that $\mu(A \cup B) = \mu(A) + \mu(B)$, for any disjoint subsets $A, B \subseteq C$. Therefore, both the Dirac measure and the arithmetic mean are special cases of the weighted arithmetic mean.
- The threshold measure (OS_t): Defined by

$$\mu(A) = \begin{cases} 1, & \text{if } |A| \geq t \\ 0, & \text{otherwise} \end{cases}.$$

  One may remark that both minimum and maximum operators are special cases of the threshold measure.
- The symmetric capacity: It consists of a capacity $\mu(A)$ such that, for any $A, B \subseteq C$, $|A| = |B|$ implies that $\mu(A) = \mu(B)$. This capacity leads to the ordered weighted averaging (OWA) operator (Yager, 1988). Moreover, both arithmetic mean and threshold measure are special cases of the symmetric capacity.
- The 0-1 capacity (0-1c): It consists of a capacity $\mu(A)$ such that each coefficient is either equal to 0 or equal to 1. By using the Choquet integral, this capacity leads to the lattice polynomial function (Grabisch et al., 2009). Moreover, it includes both Dirac and threshold measures.
- The general Choquet integral: We denote as the general Choquet integral the case in which the capacity $\mu(A)$ only satisfies the axioms.

Figure 47 presents the aforementioned particular cases of Choquet integral and the relation among them.

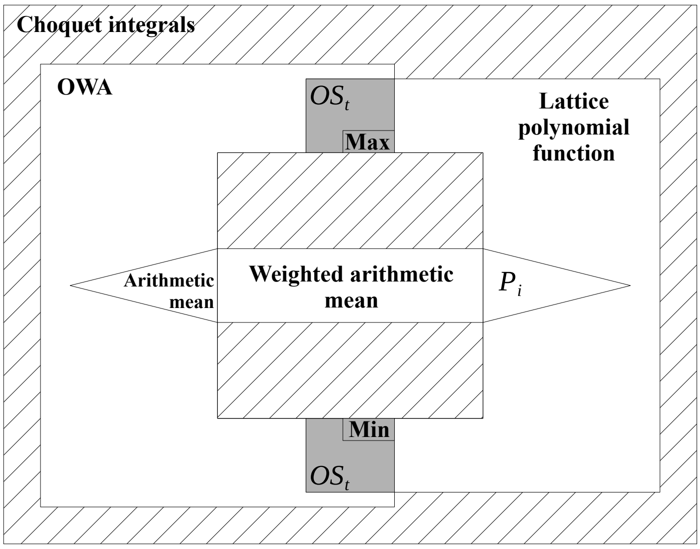


Figure 47 – Particular Choquet integrals (adapted from (Grabisch et al., 2009)).

In order to verify which combinations of particular Choquet integrals ensure commutativity in the 2-step aggregation procedure, let us start with the weighted arithmetic mean. Suppose that both μ^C and μ^D lead to this particular case. Therefore, we have

$$F_{CI}^{C/D} = \sum_{k=1}^{t} F_{CI}^{k} \mu^D(k) = \sum_{k=1}^{t} \sum_{i=1}^{m} u_i^k \mu^C(i) \mu^D(k) = \sum_{i=1}^{m} \sum_{k=1}^{t} u_i^k \mu^D(k) \mu^C(i) = \sum_{i=1}^{m} F_{CI}^{i} \mu^C(i) = F_{CI}^{D/C},$$

i.e., both procedures commute. Moreover, since both arithmetic mean and P_i capacities are special cases of WAM, they also commute. Therefore, any combination of weighted arithmetic mean, arithmetic mean and P_i ensure the commutativity property.

Let us now verify both smallest and greatest normalized capacities. If one assumes that μ^C and μ^D consist of the minimum and the maximum operators, respectively, we have

$$F_{CI}^{C/D} = \vee_{k=1,\ldots,t} \left(F_{CI}^{1}, \ldots, F_{CI}^{t} \right)$$
$$= \vee_{k=1,\ldots,t} \left(\wedge_{i=1,\ldots,m} \left(u_1^1, \ldots, u_m^1 \right), \ldots, \wedge_{i=1,\ldots,m} \left(u_1^t, \ldots, u_m^t \right) \right)$$

and

$$F_{CI}^{D/C} = \wedge_{i=1,\ldots,m} \left(F_{CI}^{1}, \ldots, F_{CI}^{m} \right)$$
$$= \wedge_{i=1,\ldots,m} \left(\vee_{k=1,\ldots,t} \left(u_1^1, \ldots, u_1^t \right), \ldots, \vee_{k=1,\ldots,t} \left(u_m^1, \ldots, u_m^t \right) \right),$$

which clearly does not commute. Therefore, one may not ensure that any combination of OS_t capacities and, consequently, 0-1c and OWA cases, satisfies the commutativity property. However, if we assume that both μ^C and μ^D consist either of the smallest or the greatest capacity, this property is guaranteed.

As a last analysis, consider the Dirac measure centered on i. By assuming that μ^C consists of a P_i capacity, we achieve

$$F_{CI}^{C/D} = \sum_{k=1}^{t} \left(F_{CI}^{(k)} - F_{CI}^{(k-1)} \right) \mu^D(\{(k), \ldots, (t)\})$$

$$= \sum_{k=1}^{t} \left(u_i^{(k)} - u_i^{(k-1)} \right) \mu^D(\{(k), \ldots, (t)\}) = F_{CI}^{i} = F_{CI}^{D/C}.$$

Independently of μ^D, we ensure commutativity. Therefore, this property is guaranteed for any combination of capacities in which at least one consists of a P_i capacity.

In summary, the aforementioned findings attest that if

- at least one capacity is a Dirac measure,
- or both capacities correspond to the weighted arithmetic mean (which includes the arithmetic mean),
- or both capacities correspond to the smallest normalized capacity or
- or both capacities correspond to the greatest normalized capacity,

the commutativity property is ensured. This highlights that in scenarios in which the criteria (and/or decision makers) interactions should be modeled, one will possibly violates the commutativity property.

8.3 Conclusions

In this chapter, we addressed two different scenarios in group MCDM and conducted preliminary investigations. In the first one, we consider that the set of decision makers have the access to a single decision matrix and that they provide their own capacity used to aggregate the criteria evaluations. In this context, we revisited a result in the literature which argue that the 2-level Choquet integral can be represented by a 2-step aggregation procedure in which the first and the second steps comprise an additive aggregation and a 0-1 Choquet integral, respectively. We follow the procedure described in (Murofushi, 2004) (written in Japanese) and attest this result with an illustrative example. However it only comprises an alternative representation of the 2-level Choquet integral. Moreover, the analysis of the piecewise linear functions and its domains can be expensive in scenarios with a high number of criteria and/or decision makers.

As a second analysis, we consider that each decision maker provides his/her criteria evaluations for all alternatives and that there are predefined capacities associated with criteria and with DMs. In this scenario, we investigated the commutativity property in the 2-step aggregation procedure. The obtained results indicated that, if one intents to model interactions between criteria and/or between decision makers, one may not ensure this property. Therefore, one may face inconveniences in situations where the evaluation of all alternatives should not depend on the order in which the collected decision data are aggregated. For future perspectives, we intend to investigate the commutativity property in the case of a 2-additive capacity.

9 Conclusions

The main purpose of this Ph.D. thesis was to deal with biased results provided by correlated criteria in multicriteria decision making problems. In such a situation, if one does not take into account criteria interactions, a single latent factor may be considered twice in the aggregation procedure. As a consequence, one may have unfair rankings, in which alternatives with good evaluations only in correlated criteria are favored in comparison to the other ones. Therefore, techniques which can explore the data structural information should be used in MCDM problems with redundant criteria.

In order to conduct this study, we started with the theoretical foundations of MCDM. In Chapter 2, we presented some definitions in decision analysis and discussed decision problems involving multiple criteria. With respect to the definition of this set of criteria, we highlighted the expected properties. We focused on the non-redundancy property, which is directly related to our concerns in this thesis. In Chapter 2, we also described some existing MCDM methods. Further attention was devoted to the TOPSIS-based methods and the capacity-based aggregation functions, which were adopted in our contributions. Moreover, we presented some techniques used to define (or estimate) the decisional parameters used in MCDM methods. We started presenting the subjective ones, which take into account the preference provided by the decision maker, and the objective approaches, which estimate the parameters based on the information extracted from the decision data.

Once the theory behind MCDM problems has been presented and the problem of redundant criteria has been raised, we presented in Chapter 3 the latent variable analysis techniques used in this thesis to deal with redundant criteria. In our contributions, they were applied to extract relevant information from the decision data that will be used by classical MCDM methods to mitigate biased results. We addressed signal processing methods, such as the principal and the independent component analysis, and the variance-based sensitivity analysis, which leads to the Sobol's indices.

The contributions of this thesis started in Chapter 4. In that chapter, we conducted a theoretical study on TOPSIS-based methods and proposed two approaches to deal with dependent criteria. These approaches, namely ICA-TOPSIS and ICA-TOPSIS-M, combined an ICA technique with TOPSIS-based methods. Since ICA could deal with dependencies in the decision data, the obtained results attested that our proposals can mitigate biased effects provided by redundant criteria.

In the next contribution, presented in Chapter 5, we tackled the problem of adjusting the weights used in the weighted arithmetic mean in order to take into account intercriteria relations. By using both ICA and PCA techniques, the application of the proposed approach led to a penalization on weights associated with positively correlated criteria. Therefore, we could reduce the contribution of such criteria in the aggregation procedure, which avoids that the same latent factor is considered twice in the determination of the ranking.

In Chapters 6 and 7, we addressed the use of the multilinear model in MCDM problems. As far as we know, this aggregation function has been rarely exploited in the literature to deal with such problems. Therefore, in Chapter 6, we conducted a theoretical study and provided a formulation of this function by means of a 2-additive capacity. Moreover, we also tackled the supervised capacity identification problem. For instance, we attested that the use of a regularization term in the optimization model can lead to a set of capacity coefficients with a better level of generalization. Still in the context of capacity identification, Chapter 7 addressed this problem by means of an unsupervised approach. For this purpose, we considered two different approaches. In the first one, we associated the interaction indices to similarity measures between pairs of criteria, assuming that the interaction index associated to a pair of criteria is negative if they are positively correlated. The other approach considered the use of Sobol' indices to determine the capacity coefficients. Since these measures indicate the impact that each criterion has in the overall evaluations, we assumed that, in scenarios with redundant criteria, the adopted capacity should be defined in order to lead to an aggregation procedure in which all criteria have the same impact. By doing this, we could compensate the biased effects provided by correlated criteria.

Finally, Chapter 8 addressed some issues in group MDCM problems by using the Choquet integral. We revisited an interesting result in the literature that argue that a multilevel Choquet integral can be represented by a 2-step procedure in which the first one comprises an additive aggregation and the second one consists in the application of a 0-1 capacity. Moreover, we also provided a preliminary discussion on the commutativity property in a 2-step Choquet integral. Indeed, if we would like to model interactions between criteria and/or between decision makers, this property may not be ensured.

Future perspectives

Based on the analysis conducted in this thesis and on the obtained results, we highlight in the sequel some perspectives that can be addressed by future works.

- We could note that redundant criteria may introduce bias in the achieved ranking of alternatives. Therefore, the development of new MCDM methods or the improvement of existing ones in order to deal with such an inconvenience and provide a fairer ranking is of importance in practical applications. For instance, we proposed an extension of TOPSIS-M method and an adjustment on the weighted arithmetic mean parameters. However, we only considered to deal with linear relations between criteria. Therefore, future works may be conducted in order to exploit nonlinear intercriteria relations.

- With respect to the multilinear model, we expect that our findings open the path for novel studies on this aggregation function. It can be used as an alternative to the Choquet integral in situations where the application of the weighted arithmetic mean is not sufficient to represent the preference provided by the decision maker.
- Some of the results achieved in our experiments indicate that our proposal can better deal with positively correlated criteria. Therefore, we also may consider as a future perspective the understanding if negative correlations may deserve attention when implementing a MCDM method. If it is the case, it should be verified how much this redundancy should be taken into account when defining the decisional parameters.
- In the context of the capacity identification problem, future works may be conducted in order to extend our ICA/PCA-based approaches in the adjustments of the capacity coefficients. Moreover, given our interesting findings in the context of unsupervised methods for capacity identification, other latent variable analysis methods that extract data structural characteristics could be tested in scenarios with redundant criteria.

Bibliography

Adeyeba, T. A., Anderson, D. T., and Havens, T. C. Insights and characterization of l_1-norm based sparsity learning of a lexicographically encoded capacity vector for the Choquet integral. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2015.

Anderson, D. T., Price, S. R., and Havens, T. C. Regularization-based learning of the Choquet integral. In *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 2519–2526, Beijing, China, 2014. IEEE.

Angilella, S., Corrente, S., and Greco, S. Stochastic multiobjective acceptability analysis for the Choquet integral preference model and the scale construction problem. *European Journal of Operational Research*, 240:172–182, 2015.

Antuchevičiene, J., Zavadskas, E. K., and Zakarevičius, A. Multiple criteria construction management decisions considering relations between criteria. *Technological and Economic Development of Economy*, 16(1):109–125, 2010.

Bai, C., Dhavale, D., and Sarkis, J. Integrating Fuzzy C-Means and TOPSIS for performance evaluation: An application and comparative analysis. *Expert Systems With Applications*, 41(.):4186–4196, 2014.

Bana e Costa, C. A. Les problématiques de l'aide à la décision: Vers l'enrichissement de la trilogie choix-tri-rangement. *RAIRO. Recherche opérationnelle*, 30(2):191–216, 1996.

Bana e Costa, C. A. and Vansnick, J.-C. A theoretical framework for Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH). In Clímaco, J., editor, *Multicriteria Analysis*, pages 15–24. Springer, Berlin, Heidelberg, 1997.

Bana e Costa, C. A., De Corte, J.-M., and Vansnick, J.-C. MACBETH. *International Journal of Information Technology & Decision Making*, 11(2):359–387, 2012.

Banzhaf, J. F. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19:317–343, 1965.

Batany, Y.-M., Duarte, L. T., Donno, D., and Romano, J. M. T. Adaptive multiple subtraction: Unification and comparison of matching filters based on the lq-norm and statistical independence. *Geophysics*, 81(1):V43–V54, 2016.

Behzadian, M., Otaghsara, S. K., Yazdani, M., and Ignatius, J. A state-of the-art survey of TOPSIS applications. *Expert Systems with Applications*, 39:13051–13069, 2012.

Beliakov, G., Pradera, A., and Calvo, T. *Aggregation functions: A guide for practitioners*. Springer, Germany, 2007.

Blackorby, C., Primont, D., and Russell, R. R. *Duality, separability, and functional structure: Theory and economic applications*. North-Holland, 1978.

Bouyssou, D. Outranking methods. In Floudas, C. A. and Pardalos, P. M., editors, *Encyclopedia of optimization*, volume 4, pages 249–255. Kluwer Academic Publishers, Dordrecht, 2001.

Bouyssou, D. and Pirlot, M. Preferences for multi-attributed alternatives: Traces, dominance, and numerical representations. *Journal of Mathematical Psychology*, 48:167–185, 2004.

Bouyssou, D. and Pirlot, M. Conjoint measurement tools for MCDM. In Greco, S., Ehrgott, M., and Figueira, J. R., editors, *Multiple Criteria Decision Analysis. International Series in Operations Research & Management Science*, volume 233, pages 97–151. 2016.

Bouyssou, D., Marchant, T., Pirlot, M., Tsoukiàs, A., and Vincke, P. *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*. Springer Science & Business Media, 2006.

Cables, E., García-Cascales, M. S., and Lamata, M. T. The LTOPSIS: An alternative to TOPSIS decision-making approach for linguistic variables. *Expert Systems with Applications*, 39:2119–2126, 2012.

Cardoso, J. and Souloumiac, A. Blind beamforming for non-gaussian signals. *IEE Proceedings F Radar and Signal Processing*, 140(6):362–370, 1993.

Chang, C.-H., Lin, J.-J., Lin, J.-H., and Chiang, M.-C. Domestic open-end equity mutual fund performance evaluation using extended TOPSIS method with different distance approaches. *Expert Systems with Applications*, 37:4642–4649, 2010.

Chen, P. Y. and Popovich, P. M. *Correlation: Parametric and nonparametric measures*. Sage Publications, Inc, Thousand Oaks, CA, 2002.

Chen, S.-Y. An improved fuzzy decision analysis framework with fuzzy Mahalanobis distances for individual investment effect appraisal. *Management Decision*, 55(5):935–956, 2017.

Chen, S.-Y. and Lu, C.-C. Assessing the competitiveness of insurance corporations using fuzzy correlation analysis and improved fuzzy modified TOPSIS. *Expert Systems*, 32 (3):392–404, 2015.

Chen, X., Wang, Z. J., and McKeown, M. J. Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods. *IEEE Signal Processing Magazine*, 33(3):86–107, 2016.

Cherry, E. C. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5):975–979, 1953.

Choquet, G. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1954.

Combarro, E. F. and Miranda, P. Identification of fuzzy measures from sample data with genetic algorithms. *Computers & Operations Research*, 33:3046–3066, 2006.

Comon, P. Independent component analysis. In Lacoume, J. L., editor, *Higher-Order Statistics*, pages 29–38. Elsevier Science Publishers B.V., 1992.

Comon, P. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, 1994.

Comon, P. and Jutten, C. *Handbook of blind source separation: Independent component analysis and applications*. Academic Press, Oxford, UK, 1 edition, 2010.

Cover, T. M. and Thomas, J. A. *Elements of information theory*. John Wiley & Sons, Inc., New York, NY, USA, 1991.

de Almeida, A. T., de Almeida, J. A., Costa, A. P. C. S., and de Almeida-Filho, A. T. A new method for elicitation of criteria weights in additive models: Flexible and interactive tradeoff. *European Journal of Operational Research*, 250:179–191, 2016.

De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L. The mahalanobis distance. *Chemometrics and intelligent laboratory systems*, 50:1–18, 2000.

Diakoulaki, D., Mavrotas, G., and Papayannakis, L. Determining objective weights in multiple criteria problems: The CRITIC method. *Computers and Operations Research*, 22(7):763–770, 1995.

Dias, L. C. and Tsoukiàs, A. On the constructive and other approaches in decision aiding. In *Proceedings of the 57th meeting of the EURO MCDA working group.*, 2003.

Dong, Y. and Xu, J. *Consensus in group decision making*. Springer Science+Business Media Singapore, Singapore, 2016.

Dong, Y., Zhang, H., and Herrera-Viedma, E. Integrating experts' weights generated dynamically into the consensus reaching process and its applications in managing non-cooperative behaviors. *Decision Support Systems*, 84:1–15, 2016.

Duarte, L. T. A novel multicriteria decision aiding method based on unsupervised aggregation via the Choquet integral. *IEEE Transactions on Engineering Management*, 65 (2):293–302, 2018.

Duarte, L. T., Moussaoui, S., and Jutten, C. Source separation in chemical analysis: recent achievements and perspectives. *IEEE Signal Processing Magazine*, 31(3):135–146, 2014.

Dubois, D., Fargier, H., and Rico, A. Sugeno integrals and the commutation problem. In Torra, V., Narukawa, Y., Aguiló, I., and González-Hidalgo, M., editors, *Modeling Decisions for Artificial Intelligence (MDAI 2018). Lecture Notes in Computer Science*, volume 11144, pages 48–63. Springer, Cham, 2018.

Dubois, D., Fargier, H., and Rico, A. Commuting double Sugeno integrals. *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, 27(Suppl.1): 1–38, 2019.

Edwards, W. How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-7(5):326–340, 1977.

Edwards, W. and Barron, F. H. SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, 60:306–325, 1994.

Figueira, J. R., Mousseau, V., and Roy, B. ELECTRE methods. In Greco, S., Ehrgott, M., and Figueira, J. R., editors, *Multiple criteria decision analysis: State of the art surveys. International Series in Operations Research & Management Science*, volume 233. Springer, New York, NY, USA, 2016.

Fishburn, P. C. *Utility theory for decision making.* John Wiley & Sons, Inc, New York, 1970.

García-Cascales, M. S. and Lamata, M. T. On rank reversal and TOPSIS method. *Mathematical and Computer Modelling*, 56:123–132, 2012.

Gilboa, I. *Theory of decision under uncertainty.* Cambridge University Press, New York, NY, 2010.

Golub, G. H. and Van Loan, C. F. *Matrix computations.* Johns Hopkins University Press, Baltimore, Maryland, 4 edition, 2013.

Govindan, K. and Jepsen, M. B. ELECTRE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, 250:1–29, 2016.

Grabisch, M. A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium*, pages 145–150, Yokohama, Japan, 1995.

Grabisch, M. The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89:445–456, 1996.

Grabisch, M. Alternative representations of discrete fuzzy measures for decision making. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 5(5): 587–607, 1997a.

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

Grabisch, M. A graphical interpretation of the Choquet integral. *IEEE Transactions on Fuzzy Systems*, 8(5):627–631, 2000.

Grabisch, M. *Set Functions, games and capacities in decision making.* Springer International Publishing, Switzerland, 2016.

Grabisch, M. and Labreuche, C. A note on the Sobol' indices and interactive criteria. *Fuzzy Sets and Systems*, 315:99–108, 2017.

Grabisch, M. and Roubens, M. An axiomatic approach to the concept of interaction among players in cooperative games. *International Journal of Game Theory*, 28:547–565, 1999.

Grabisch, M., Labreuche, C., and Vansnick, J.-C. On the extension of pseudo-Boolean functions for the aggregation of interacting criteria. *European Journal of Operational Research*, 148:28–47, 2003.

Grabisch, M., Prade, H., Raufaste, E., and Terrier, P. Application of the Choquet integral to subjective mental workload evaluation. *IFAC Proceedings Volumes*, 39(4):135–140, 2006.

Grabisch, M., Kojadinovic, I., and Meyer, P. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory: Applications of the Kappalab R package. *European Journal of Operational Research*, 186:766–785, 2008.

Grabisch, M., Marichal, J.-L., Mesiar, R., and Pap, E. *Aggregation functions.* Cambridge University Press, New York, NY, 2009.

Greco, S., Matarazzo, B., and Słowiński, R. Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *European Journal of Operational Research*, 158:271–292, 2004.

Greco, S., Ehrgott, M., and Figueira, J. R., editors. *Multiple criteria decision analysis: State of the art surveys.* Springer's International Series in Operations Research & Management Science, New York, NY, USA, 2 edition, 2016.

Hart, S. G. and Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Hancock, P. A. and Meshkati, N., editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

Havens, T. C. and Pinar, A. J. Generating random fuzzy (capacity) measures for data fusion simulations. In *2017 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2017)*, pages 1–8. IEEE, 2017.

Haykin, S. and Chen, Z. The cocktail party problem. *Neural Computation*, 17(9):1875–1902, 2005.

Hérault, J., Jutten, C., and Ans, B. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *10º Colloque sur le traitement du signal et des images, FRA, 1985*, volume 3, pages 1017–1022, Nice, France, 1985. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology*, 24(6):417–441, 1933.

Hu, J., Du, Y., Mo, H., Wei, D., and Deng, Y. A modified weighted TOPSIS to identify influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*, 444:73–85, 2016.

Hwang, C.-L. and Yoon, K. *Multiple attribute decision making: Methods and Applications.* Springer-Verlag Berlin Heidelberg, 1981.

Hyvärinen, A. A family of fixed-point algorithms for independent component analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, volume 5, pages 3917–3920, Munich, Germany, 1997. IEEE.

Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

Hyvärinen, A., Karhunen, J., and Oja, E. *Independent component analysis.* John Wiley & Sons, New York, NY, USA, 2001.

Jolliffe, I. T. *Principal component analysis.* Springer-Verlag, New York, 2 edition, 2002.

Jutten, C. *Calcul neuromimétique et traitement du signal: Analyse en composantes indépendantes.* PhD thesis, Grenoble INPG, 1987.

Jutten, C. and Taleb, A. Source separation: from dusk till dawn. In *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'2000)*, pages 15–26, Helsinki, Finland, 2000.

Kahneman, D. and Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

Kaya, T. and Kahraman, C. Multicriteria decision making in energy planning using a modified fuzzy TOPSIS methodology. *Expert Systems With Applications*, 38:6577–6585, 2011.

Keeney, R. L. *Value-focused thinking: A path to creative decisionmaking.* Harvard University Press, Cambridge, 1992.

Keeney, R. L. and Raiffa, H. *Decisions with multiple objectives.* Wiley, New York, 1976.

Kendall, M. G. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

Klement, E. P., Mesiar, R., and Pap, E. *Triangular norms.* Kluwer Academic Publishers, 2000.

Kojadinovic, I. Unsupervized aggregation of commensurate correlated attributes by means of the Choquet integral and entropy functionals. *International Journal of Intelligent Systems*, 23:128–154, 2008.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. *Foundations of measurement.* Academic Press, v. 1: Additive and Polynomial Representations. New York, NY, USA, 1971.

Kumar, S., Singh, B., Qadri, M. A., Kumar, Y. V. S., and Haleem, A. A framework for comparative evaluation of lean performance of firms using fuzzy TOPSIS. *International Journal of Productivity and Quality Management*, 11(4):371–392, 2013.

Li, G., Rosenthal, C., and Rabitz, H. High dimensional model representations. *Journal of Physical Chemistry A*, 105(33):7765–7777, 2001.

Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb, C. E., and Schoendorf, J. Global sensitivity analysis for systems with independent and/or correlated inputs. *Journal of Physical Chemistry A*, 114(19):6022–6032, 2010.

Lootsma, F. A. The French and the American school in multi-criteria decision analysis. *RAIRO - Operations Research*, 24(3):263–285, 1990.

Lovász, L. Submodular functions and convexity. In Bachem, A., Korte, B., and Grötschel, M., editors, *Mathematical Programming The State of the Art*, pages 235–257. Springer, Berlin, Heidelberg, 1983.

MacKay, D. J. C. *Information theory, inference, and learning algorithms.* Cambridge University Press, Cambridge, U.K., 2007.

Mahalanobis, P. C. On the generalized distance in statistics. In *Proceedings of the National Institute of Sciences of India*, volume 2 (1), pages 49–55, 1936.

Marichal, J.-L. An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems*, 8:800–807, 2000.

Marichal, J.-L. Aggregation of interacting criteria by means of the discrete Choquet integral. In Calvo, T., Mayor, G., and Mesiar, R., editors, *Aggregation Operators. Studies in Fuzziness and Soft Computing, v. 97*, pages 224–244. Physica, Heidelberg, 2002.

Marichal, J.-L. and Roubens, M. Determination of weights of interacting criteria from a reference set. *European Journal of Operational Research*, 124:641–650, 2000.

Martel, J.-M. and Roy, B. Analyse de la signifiance de diverses procédures d'agrégation multicritère. *INFOR: Information Systems and Operational Research*, 44(3):191–215, 2006.

Miranda, P. and Grabisch, M. Optimization issues for fuzzy measures. *International Journal of Uncertainty Fuzziness and Knowledge Based Systems*, 7(6):545–560, 1999.

Miranda, P., Grabisch, M., and Gil, P. Identification of non-additive measures from sample data. In Della Riccia, G., Dubois, D., Kruse, R., and Lenz, H. J., editors, *Planning Based on Decision Theory. International Centre for Mechanical Sciences (Courses and Lectures)*, pages 43–62, Vienna, 2003. Springer.

Moretti, S., Öztürk, M., and Tsoukiàs, A. Preference modelling. In Greco, S., Ehrgott, M., and Figueira, J. R., editors, *Multiple Criteria Decision Analysis. International Series in Operations Research & Management Science*, volume 233. Springer, New York, NY, 2016.

Mousseau, V. *Elicitation des préférences pour l'aide multicritère a la décision.* PhD thesis, Université Paris-Dauphine, 2003.

Munda, G. Multiple-criteria decision aid: Some epistemological considerations. *Journal of Multi-Criteria Decision Analysis*, 2:41–55, 1993.

Murofushi, T. Multilevel Choquet integral. *J. of Japan Soc. for Fuzzy Theory and Intelligent Informatics*, 16(4):319–327, 2004.

Oliveira, H. E., Romano, J. M. T., and Duarte, L. T. Identificação dos parâmetros da integral de Choquet via uma abordagem baseada em processamento de sinais esparsos

(Identification of Choquet integral parameters via an approach based on sparse signal processing). In *XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2017)*, pages 692–696, São Pedro, São Paulo, Brazil, 2017.

Ovchinnikov, S. Max-min representation of piecewise linear functions. *Contributions to Algebra and Geometry*, 43(1):297–302, 2002a.

Ovchinnikov, S. Piecewise linear aggregation functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(1):17–24, 2002b.

Owen, G. Multilinear extensions of games. *Management Science*, Part 2, 18(5):64–79, 1972.

Papoulis, A. and Pillai, S. U. *Probability, random variables and stochastic processes.* McGraw-Hill, New York, NY, USA, 4 edition, 2002.

Pomerol, J.-C. and Barba-Romero, S. *Multicriterion decision in management: Principles and practice.* Kluwer Academic Publishers, New York, 2000.

Poulton, E. C. *Behavioral decision theory: A new approach.* Cambridge University Press, 1994.

Pöyhönen, M. and Hämäläinen, R. P. On the convergence of multiattribute weighting methods. *European Journal of Operational Research*, 129:569–585, 2001.

Raufaste, E., Terrier, P., and Grabisch, M. Etude expérimentale de l'applicabilité de modèles d'agrégations flous à l'étude de la charge mentale (Experimental study of the applicability of fuzzy aggregation models to the study of mental workload). In *Journées d'études en Psychologie Ergonomique (EPIQUE'01)*, pages 171–176, Nantes, France, 2001.

Rieta, J. J., Castells, F., Sánchez, C., Zarzoso, V., and Millet, J. Atrial activity extraction for atrial fibrillation analysis using blind source separation. *IEEE Transactions on Biomedical Engineering*, 51(7):1176–1186, 2004.

Rota, G.-C. On the foundations of combinatorial theory I. Theory of Möbius functions. *Wahrscheinlichkeitstheorie verw Gebiete*, 2:340–368, 1964.

Roubens, M. Interaction between criteria through the use of fuzzy measures. In *44th Meeting of the European Working Group "Multicriteria Aid for Decisions'*, Brussels, Belgium, 1996.

Roubens, M. and Vincke, P. *Preference Modelling*, volume 250. Springer-Verlag, Berlin, Heidelberg, 1985.

Rowley, H. V., Geschke, A., and Lenzen, M. A practical approach for estimating weights of interacting criteria from profile sets. *Fuzzy Sets and Systems*, 272:70–88, 2015.

Roy, B. Classement et choix en présence de points de vue multiples. *Revue française d'informatique et de recherche opérationnelle*, 2(8):57–75, 1968.

Roy, B. Quelques remarques sur le concept d'indépendance dans l'aide à la décision multicritère. In *Cahier n° 51, Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision (LAMSADE)*. Université de Paris-Dauphine, 1983.

Roy, B. Decision-aid and decision-making. *European Journal of Operational Research*, 45:324–331, 1990a.

Roy, B. The outranking approach and the foundations of electre methods. In BANA E COSTA, C. A., editor, *Readings in multiple criteria decision aid*, pages 155–183. Springer-Verlag, Berlin, Heidelberg, 1990b.

Roy, B. *Multicriteria methodology for decision aiding*. Springer Science & Business Media, 1996.

Roy, B. À propos de la signification des dépendances entre critères: Quelle place et quels modes de prise en compte pour l'aide à la décision? *RAIRO - Operations Research*, 43: 255–275, 2009.

Roy, B. Paradigms and challenges. In Greco, S., Ehrgott, M., and Figueira, J. R., editors, *Multiple criteria decision analysis: State of the art surveys. International Series in Operations Research & Management Science*, volume 233, pages 19–39. Springer, New York, NY, USA, 2016.

Roy, B. and Bouyssou, D. *Aide multicritere à la décision: méthodes et cas*. Economica, Paris, France, 1993.

Roy, B. and Vanderpooten, D. The European school of MCDA: Emergence, basic features and current works. *Journal of Multi-Criteria Decision Analysis*, 5:22–38, 1996.

Saaty, T. L. The analytic hierarchy process: What it is and how it is used. *Mathematical Modelling*, 9(3-5):161–176, 1987.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. *Sensitivity analysis in practice: A guide toassessing scientific models*. John Wiley & Sons Ltd, Chichester, UK, 2004.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. *Global sensitivity analysis: The primer*. John Wiley & Sons Ltd, Chichester, UK, 2008.

Savage, L. J. *The foundations of statistics*. Wiley. Second revised edition, 1972, New York, 1954.

Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

Shapley, L. S. A value for n-person games. In Kuhn, W. and Tucker, A. W., editors, *Annals of mathematics studies: Vol. 28. Contributions to the theory of games, Vol. II*, pages 307–317. Princeton University Press, Princeton, 1953.

Shlens, J. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014a. URL `http://arxiv.org/abs/1404.1100`.

Shlens, J. A tutorial on independent component analysis. *arXiv preprint arXiv:1404.2986*, 2014b. URL `http://arxiv.org/abs/1404.2986`.

Sobol, I. M. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiment*, 1(4):407–414, 1990.

Sugeno, M. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.

Takemura, K. *Behavioral decision theory*. Springer Japan, 2014.

Triantaphyllou, E. Multi-criteria decision making methods: A comparative study. In *Applied Optimization*, volume 44. Springer, Boston, MA, 2000.

Tsoukiàs, A. On the concept of decision aiding process: An operational perspective. *Annals of Operations Research*, 154:3–27, 2007.

Tsoukiàs, A. From decision theory to decision aiding methodology. *European Journal of Operational Research*, 187:138–161, 2008.

Tversky, A. On the elicitation of preferences: Descriptive and prescriptive considerations. In Bell, D., Keeney, R. L., and Raiffa, H., editors, *Conflicting objectives in decisions*, pages 209–222, New York, 1977. John Wiley & Sons.

Vajda, S. *Fibonacci and Lucas numbers, and the golden section: Theory and applications*. Ellis Horword Limited, Chichester, UK, 1989.

Vanderbei, R. J. *Linear programming: Foundations and extensions*. Springer Science+Business Media, New York, 4 edition, 2014.

Vanderpooten, D. The construction of prescriptions in outranking methods. In BANA E COSTA, C. A., editor, *Readings in Multiple Criteria Decision Aid*, pages 184–215. Springer-Verlag, Berlin, Heidelberg, 1990.

Vega, A., Aguarón, J., García-Alcaraz, J., and Moreno-Jiménez, J. M. Notes on dependent attributes in TOPSIS. *Procedia Computer Science*, 31:308–317, 2014.

von Neumann, J. and Morgenstern, O. *Theory of games and economic behavior*. Princeton University Press. Second edition in 1947, third in 1954, Princeton, 1944.

von Winterfeldt, D. and Edwards, W. *Decision analysis and behavioural research*. Cambridge University Press, Cambridge, 1986.

Wakker, P. *Prospect theory for risk and ambiguity*. Cambridge University Press, New York, NY, 2010.

Wang, E. Benchmarking whole-building energy performance with multi-criteria technique for order preference by similarity to ideal solution using a selective objective-weighting approach. *Applied Energy*, 146:92–103, 2015.

Wang, J.-J., Jing, Y.-Y., Zhang, C.-F., and Zhao, J.-H. Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, 13:2263–2278, 2009.

Wang, Z.-X. and Wang, Y.-Y. Evaluation of the provincial competitiveness of the Chinese high-tech industry using an improved TOPSIS method. *Expert Systems with Applications*, 41:2824–2831, 2014.

Wang, Z.-X., Li, D.-D., and Zheng, H.-H. The external performance appraisal of China energy regulation: An empirical study using a TOPSIS method based on entropy weight and Mahalanobis distance. *International Journal of Environmental Research and Public Health*, 15(2):236, 2018.

Yager, R. R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(1):183–190, 1988.

Yoon, K. P. and Kim, W. K. The behavioral TOPSIS. *Expert Systems with Applications*, 89:266–272, 2017.

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. State of science: Mental workload in ergonomics. *Ergonomics*, 58(1):1–17, 2015.

Zardari, N. H., Ahmed, K., Shirazi, S. M., and Yusop, Z. B. *Weighting methods and their effects on multi-criteria decision making model outcomes in water resources management*. Springer, 2015.

Zavadskas, E. K., Mardani, A., Turskis, Z., Jusoh, A., and Nor, K. M. Development of TOPSIS method to solve complicated decision-making problems: An overview on developments from 2000 to 2015. *International Journal of Information Technology & Decision Making*, 15(3):645–682, 2016.

Zhu, F., Zhong, P.-A., Xu, B., Wu, Y.-N., and Zhang, Y. A multi-criteria decision-making model dealing with correlation among criteria for reservoir flood control operation. *Journal of Hydroinformatics*, 18(3):531–543, 2016.

APPENDIX A – Axioms of a capacity in a 2-additive multilinear model

If we consider the Banzhaf interaction index, the axioms of a capacity may be expressed by

- $\mu(\emptyset) = 0 \rightarrow \mu(\emptyset) = \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash \emptyset|} I^{\mathcal{B}}(B) = \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B|} I^{\mathcal{B}}(B)$
  $= \sum\limits_{B \subseteq C} \left(-\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = 0;$
- $\mu(C) = 1 \rightarrow \mu(C) = \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash C|} I^{\mathcal{B}}(B) = \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|0|} I^{\mathcal{B}}(B)$
  $= \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = 1;$
- $\mu(\{A \cup i\}) - \mu(A) \geq 0$
  $\rightarrow \mu(\{A \cup i\}) - \mu(A) = \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash \{A \cup i\}|} I^{\mathcal{B}}(B) - \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash A|} I^{\mathcal{B}}(B)$
  $= \sum\limits_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} \left[(-1)^{|B \backslash \{A \cup i\}|} - (-1)^{|B \backslash A|}\right] I^{\mathcal{B}}(B) \geq 0, \forall i \in C, \forall A \subseteq C \backslash i.$

If $i \notin B$, $(-1)^{|B \backslash \{A \cup i\}|} = (-1)^{|B \backslash A|}$ and, then $\left[(-1)^{|B \backslash \{A \cup i\}|} - (-1)^{|B \backslash A|}\right] = 0$. However, in the case where $B \ni i$, we have

$$\sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} \left[(-1)^{|B \backslash \{A \cup i\}|} - (-1)^{|B \backslash A|}\right] I^{\mathcal{B}}(B)$$

$$= \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} \left[(-1)^{|B \backslash A|} (-1)^{-1} - (-1)^{|B \backslash A|}\right] I^{\mathcal{B}}(B) = \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash A|} [(-1) - 1] I^{\mathcal{B}}(B)$$

$$= (-2) \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash A|} I^{\mathcal{B}}(B)$$

and, therefore,

$$\mu(\{A \cup i\}) - \mu(A) \geq 0 \rightarrow (-2) \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash A|} I^{\mathcal{B}}(B) \geq 0$$

$$\rightarrow \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \backslash A|} I^{\mathcal{B}}(B) \leq 0, \forall i \in C, \forall A \subseteq C \backslash i.$$

By assuming a 2-additive capacity, these axioms lead to the following conditions:

- $\mu(\emptyset) = 0 \rightarrow \sum\limits_{B \subseteq C, |B| \leq 2} \left(-\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = I^{\mathcal{B}}(\emptyset) - \frac{1}{2} \sum\limits_{i} \phi_i^{\mathcal{B}} + \frac{1}{4} \sum\limits_{i,i'} I_{i,i'}^{\mathcal{B}} = 0;$

- $\mu(C) = 1 \rightarrow \sum\limits_{B \subseteq C, |B| \leq 2} \left(\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = I^{\mathcal{B}}(\emptyset) + \frac{1}{2} \sum\limits_{i} \phi_i^{\mathcal{B}} + \frac{1}{4} \sum\limits_{i,i'} I_{i,i'}^{\mathcal{B}} = 1;$

- $\mu(\{A \cup i\}) - \mu(A) \geq 0$

  $\rightarrow \sum\limits_{\substack{B \subseteq C, \\ B \ni i, \\ |B| \leq 2}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus A|} I^{\mathcal{B}}(B) = -\frac{1}{2}\phi_i^{\mathcal{B}} + \frac{1}{4} \left(\sum\limits_{\substack{i' \neq i, \\ i' \notin A}} I_{i,i'}^{\mathcal{B}} - \sum\limits_{\substack{i' \neq i, \\ i' \in A}} I_{i,i'}^{\mathcal{B}} \right) \leq 0$

  $\rightarrow \phi_i^{\mathcal{B}} - \frac{1}{2} \left(\sum\limits_{\substack{i' \neq i, \\ i' \notin A}} I_{i,i'}^{\mathcal{B}} - \sum\limits_{\substack{i' \neq i, \\ i' \in A}} I_{i,i'}^{\mathcal{B}} \right) \geq 0, \, \forall i \in C, \forall A \subseteq C \backslash i$

  $\rightarrow \phi_i^{\mathcal{B}} - \frac{1}{2} \sum\limits_{i' \neq i} \left| I_{i,i'}^{\mathcal{B}} \right| \geq 0, \, \forall i \in C.$

APPENDIX B – The Banzhaf interaction index and the Sobol' index

Consider the multilinear model expressed in terms of the Banzhaf interaction indices, given by

$$F_{ML}(u(\mathbf{x})) = \sum_{A\subseteq C} I^{\mathcal{B}}(A) \sum_{B\subseteq A} \frac{-1}{2}^{|A|-|B|} \prod_{i\in B} u_i(x_i). \tag{B.1}$$

Suppose, without loss of generality, a decision problem with $m = 3$ criteria. Moreover, suppose that we want to estimate the impact of criterion 1 in the overall evaluation by means of the Sobol' index. The conditional expectation of F_{ML} given X_1 leads to

$$\mathbb{E}[F_{ML}|X_1] = X_1\phi_1^{\mathcal{B}} + \mathbb{E}[X_2|X_1]\phi_2^{\mathcal{B}} + \mathbb{E}[X_3|X_1]\phi_3^{\mathcal{B}} + \left(X_1\mathbb{E}[X_2|X_1] - \frac{1}{2}X_1 - \frac{1}{2}\mathbb{E}[X_2|X_1]\right) I_{1,2}^{\mathcal{B}}$$
$$+ \left(X_1\mathbb{E}[X_3|X_1] - \frac{1}{2}X_1 - \frac{1}{2}\mathbb{E}[X_3|X_1]\right) I_{1,3}^{\mathcal{B}} + \left(\mathbb{E}[X_2X_3|X_1] - \frac{1}{2}\mathbb{E}[X_2|X_1] - \frac{1}{2}\mathbb{E}[X_3|X_1]\right) I_{2,3}^{\mathcal{B}}$$
$$+ \left(X_1\mathbb{E}[X_2X_3|X_1] - \frac{1}{2}X_1\mathbb{E}[X_2|X_1] - \frac{1}{2}X_1\mathbb{E}[X_3|X_1] + \frac{1}{4}X_1 + \frac{1}{4}\mathbb{E}[X_2|X_1] + \frac{1}{4}\mathbb{E}[X_3|X_1]\right) I^{\mathcal{B}}(C).$$

If we assume that the variables are independent, we achieve

$$\mathbb{E}[F_{ML}|X_1] = X_1\phi_1^{\mathcal{B}} + \mathbb{E}[X_2]\phi_2^{\mathcal{B}} + \mathbb{E}[X_3]\phi_3^{\mathcal{B}} + \left(X_1\mathbb{E}[X_2] - \frac{1}{2}X_1 - \frac{1}{2}\mathbb{E}[X_2]\right) I_{1,2}^{\mathcal{B}}$$
$$+ \left(X_1\mathbb{E}[X_3] - \frac{1}{2}X_1 - \frac{1}{2}\mathbb{E}[X_3]\right) I_{1,3}^{\mathcal{B}} + \left(\mathbb{E}[X_2]\mathbb{E}[X_3] - \frac{1}{2}\mathbb{E}[X_2] - \frac{1}{2}\mathbb{E}[X_3]\right) I_{2,3}^{\mathcal{B}}$$
$$+ \left(X_1\mathbb{E}[X_2]\mathbb{E}[X_3] - \frac{1}{2}X_1\mathbb{E}[X_2] - \frac{1}{2}X_1\mathbb{E}[X_3] + \frac{1}{4}X_1 + \frac{1}{4}\mathbb{E}[X_2] + \frac{1}{4}\mathbb{E}[X_3]\right) I^{\mathcal{B}}(C).$$

Moreover, if we assume a uniform distribution on $[0, 1]$, we have

$$\mathbb{E}[F_{ML}|X_1] = X_1\phi_1^{\mathcal{B}} + \frac{1}{2}\phi_2^{\mathcal{B}} + \frac{1}{2}\phi_3^{\mathcal{B}} + \left(\frac{1}{2}X_1 - \frac{1}{2}X_1 - \frac{1}{4}\right) I_{1,2}^{\mathcal{B}}$$
$$+ \left(\frac{1}{2}X_1 - \frac{1}{2}X_1 - \frac{1}{4}\right) I_{1,3}^{\mathcal{B}} + \left(\frac{1}{4} - \frac{1}{4} - \frac{1}{4}\right) I_{2,3}^{\mathcal{B}} + \left(\frac{1}{4}X_1 - \frac{1}{4}X_1 - \frac{1}{4}X_1 + \frac{1}{4}X_1 + \frac{1}{8} + \frac{1}{8}\right) I^{\mathcal{B}}(C)$$

and, therefore,

$$\mathbb{E}[F_{ML}|X_1] = \phi_1^{\mathcal{B}}X_1 + \frac{1}{2}\phi_2^{\mathcal{B}} + \frac{1}{2}\phi_3^{\mathcal{B}} - \frac{1}{4}I_{1,2}^{\mathcal{B}} - \frac{1}{4}I_{1,3}^{\mathcal{B}} - \frac{1}{4}I_{2,3}^{\mathcal{B}} + \frac{1}{4}I^{\mathcal{B}}(C).$$

By taking the variance of $\mathbb{E}[F_{ML}|X_1]$, we achieve the nonnormalized Sobol' index of criterion 1, given by

$$\mathrm{Var}[\mathbb{E}[F_{ML}|X_1]] = \mathrm{Var}[\phi_1^{\mathcal{B}}X_1] = \frac{1}{12}(\phi_1^{\mathcal{B}})^2. \tag{B.2}$$

We recognize this expression in Equation (7.9). Therefore, we verify that, under the aforementioned conditions, there is a direct association between the Sobol' index and the Banzhaf interaction index.