



HAL
open science

Structural and algorithmic aspects of preference domain restrictions in collective decision making: contributions to the study of single-peaked and Euclidean preferences

Magdaléna Tydrichová

► To cite this version:

Magdaléna Tydrichová. Structural and algorithmic aspects of preference domain restrictions in collective decision making: contributions to the study of single-peaked and Euclidean preferences. Artificial Intelligence [cs.AI]. Sorbonne Université, 2023. English. NNT : 2023SORUS048 . tel-04143294

HAL Id: tel-04143294

<https://theses.hal.science/tel-04143294v1>

Submitted on 27 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



École doctorale n° 130 : Informatique, Télécommunications et Electronique
Sorbonne Université, Paris

Structural and algorithmic aspects of preference domain restrictions in collective decision making

Contributions to the study of single-peaked and Euclidean preferences

Thèse de doctorat en informatique
présentée publiquement par

Magdaléna TYDRICHOVÁ

le 21 mars 2023

Rapporteurs

Edith ELKIND
Jérôme LANG

Professor, University of Oxford, Royaume-Uni
Directeur de recherche CNRS
Université Paris-Dauphine, France

Examineurs

Jiehua CHEN
Jean-Paul DOIGNON
Nicolas MAUDET

Assistant professor, TU Wien, Autriche
Professeur émérite, Université Libre de Bruxelles, Belgique
Professeur, Sorbonne Université, France

Directeurs de thèse

Bruno ESCOFFIER
Olivier SPANJAARD

Professeur, Sorbonne Université, France
Maître de conférences, Sorbonne Université, France



Table of contents

- Acknowledgements** 9
- Introduction** 13
- I Opening** 19
- 1 Preliminaries** 21
 - 1.1 Preference profile 21
 - 1.2 Single-peaked preferences 24
 - 1.2.1 Axiomatic and algorithmic properties 26
 - 1.2.2 Recognition of single-peaked preferences 29
 - 1.3 Single-crossing preferences 31
 - 1.4 1-Euclidean preferences 33
- 2 Relations between single-peaked, single-crossing and 1-Euclidean preferences** 37
 - 2.1 Introduction 37
 - 2.2 Combinatorial structures behind restricted domains 38
 - 2.2.1 Permutohedron of order m 39
 - 2.2.2 Distributive lattice and weak Bruhat order 42
 - 2.3 The SPSC domain: state of art 44
 - 2.3.1 Characterization of the SPSC domain and permutohedron 45
 - 2.3.2 The SPSC domain and weak Bruhat orders 49
- Chapter appendices** 51
 - 2.A Graphs $SP_6(A)$ and $SP_7(A)$ 51
- II Extensions of single-peaked preferences** 53
- 3 Single-peaked preferences on an arbitrary graph** 57
 - 3.1 Introduction 57
 - 3.2 Problem definition 60
 - 3.3 ILP formulation and complexity 64

3.3.1	ILP Formulation	64
3.3.2	Minimizing the number of edges	66
3.3.3	Minimizing the Maximum Degree	69
3.4	Recognition of trees and paths	72
3.5	Recognition of pseudotrees	79
3.6	Experimental study	85
3.6.1	Numerical tests on real data	86
3.6.2	Experimental study on randomly generated data	88
3.7	Conclusion	93
Chapter appendices		95
3.A	Computation of the expected number of necessary edges	95
4	Nearly single-peaked preferences : forbidden triples	97
4.1	Introduction	97
4.2	Preliminaries	101
4.3	Positioning with respect to other metrics	103
4.4	Axiomatic Examination	108
4.4.1	Common points of nearly single-peakedness measures and comparison with single-peakedness	110
4.4.2	Differences between nearly single-peakedness measures	114
4.5	A Maximum likelihood estimator	120
4.6	Computational aspects	124
4.6.1	FT Single-peaked Evaluation and Consistency	124
4.6.2	An Integer Program to Determine $FT(\mathcal{P})$	127
4.7	Experimental study	129
4.7.1	Numerical tests on real data	131
4.7.2	Numerical tests on synthetic data	135
4.8	Conclusion	142
Chapter appendices		143
4.A	Detailed results on real data	143
III Multidimensional Euclidean preferences		147
5	An introduction to multidimensional preferences	149
5.1	Some words about multidimensional single-peaked preferences	149
5.2	Multidimensional Euclidean preferences	152
5.2.1	The State of the Art	152
5.2.2	Definitions	154
5.2.3	Weighted majority tournaments and Kemeny rankings with 2-Euclidean preferences	156

6	A Heuristic algorithm for recognizing 2-Euclidean preferences with respect to the norm ℓ_2	169
6.1	Introduction	169
6.2	Preliminaries	170
6.3	2-Euclidean recognition heuristic	171
6.3.1	NO-Certificates	172
6.3.2	The random generation of the representation f	173
6.3.3	The profile \mathcal{P}_f associated to a mapping f	176
6.3.4	Testing if the profile \mathcal{P} is a subprofile of $[\mathcal{P}_f]$	177
6.4	Experimental study	178
6.4.1	Experimental study on real-world data	179
6.4.2	Experimental study on synthetic data	181
6.4.3	Maximal 2-Euclidean profiles for 5 and 6 candidates	185
6.5	Conclusion	188
7	Euclidean preferences in the plane under ℓ_1, ℓ_2 and ℓ_∞ norms	189
7.1	Introduction	189
7.2	Hypersurfaces under ℓ_1 (resp. ℓ_∞) in the plane	191
7.2.1	Reminder on notations	191
7.2.2	Types of hypersurfaces	192
7.2.3	Classification of the hypersurfaces under ℓ_1 in the plane	197
7.2.4	Hypersurfaces under ℓ_∞ in the plane	198
7.3	Relations between different norms	199
7.3.1	Relation between norms ℓ_1 and ℓ_∞	200
7.3.2	Relation between norms ℓ_1 and ℓ_2	202
7.4	Properties of hypersurfaces under ℓ_1 in the plane	205
7.4.1	Intersection of two hypersurfaces	205
7.4.2	Intersection of three hypersurfaces	207
7.4.3	Some corollaries and further results	211
7.5	Euclidean profiles on 4 candidates in the plane	213
7.5.1	Maximum size of a Euclidean profile on 4 candidates	213
7.5.2	Characterization of ℓ_2 -Euclidean profiles	217
7.6	Euclidean profiles on $m \geq 5$ candidates in the plane	226
7.7	Conclusion	227
	Chapter appendices	229
7.A	Missing proofs of Section 7.4	229
7.A.1	Proof of Proposition 7.4.1	229
7.A.2	Missing proof of Property 7.4.2	231
7.B	Missing proofs of Section 7.5	234
7.B.1	Detailed proof of Lemma 7.5.1	234
7.B.2	Missing proof of Proposition 7.4.4	237
7.C	Missing proofs of Section 7.6	241
7.C.1	Missing part in the proof of Theorem 7.6.1	241

IV	Final thoughts	243
8	Conclusion	245
8.1	Looking back to Part I	246
8.1.1	1D-Euclidean preferences: towards a characterization	246
8.2	Looking back to Part II	253
8.2.1	Single-peaked preferences on an arbitrary graph: combining minimization criteria, the likelihood of solutions, and even more	254
8.2.2	Nearly single-peaked preferences on an arbitrary graph	259
8.3	Looking back to Part III	260
8.3.1	A heuristic algorithm for recognizing 2-Euclidean preferences with respect to the norm ℓ_1 : further studies	261
8.3.2	d -Euclidean preferences with respect to the norm ℓ_1 and ℓ_∞	264
	Bibliography	269

Alku aina hankala...

Acknowledgments



It is now time to come out of my shell and say out loud (a little part of) what I generally keep to myself.

To start, I would like to thank Edith Elkind and Jérôme Lang who accepted to review my thesis and take part in the committee. Thank you very much for the time you devoted to the lecture of my thesis, and for your feedback on my work which is very precious to me. I would also like to thank the others members of my committee - Jiehua Chen, Jean-Paul Doignon and Nicolas Maudet. During my thesis, I had many opportunities to come across your papers, and you have somehow become familiar to me thanks to your scientific results. Yet, I have never made an effort to really get in touch with you. I am therefore very pleased, and touched, that you have accepted to take part in my PhD defense committee and that I will finally have a chance to discuss with you.

The two persons I'm most grateful to are, obviously, my PhD supervisors, Olivier Spanjaard and Bruno Escoffier. Sometimes it happens that you meet someone and you feel like if you were on the same wavelength. It happened to me (and I wouldn't presume to say that it was mutual!) when I first met Olivier. It was in 2016 in his undergraduate algorithmic course. Impressed by teaching skills and human qualities of Olivier, I told myself that if one day I did a thesis, I would like it to be supervised by a person like him. And I had a chance - few years later, this wish came true. Thank you Olivier for your unlimited patience and kindness that you always had for me, for your huge involvement in the proofreading of my productions: if I compare my writing skills at the beginning and at the end of my thesis, I feel that I have progressed. And it is mainly thanks to you! Thank you finally for all your encouragements and advice - it is clearly thanks to your empathy that I became more confident, and able to do things I could never have done before (for example, pick up the phone and call the doctoral school).

When I first met Bruno, I did not have the impression that we were on the same wavelength. I think that we are not... which is actually a very good thing! Thank you Bruno for making my world richer by sharing with me your point of view, often different from mine, for providing me more practicality and organisation that I lack a bit, for your non-stop availability and involvement in every single of my (not only) scientific prob-

lems, for all your “t’es sûre que ça marche” questions which caused me many sleepless nights, and, last but not least, for your specific humour I appreciate so much. Thank you also for being probably one of the most intelligent persons I have ever known, and a model of success that motivates me to go further: I really hope that one day, I will be able to proudly answer “OUI !” to one of your “t’es sûre que ça marche” questions! For instance, would not we have a little conjecture to prove together?

Thank you both for your very strong confidence in me, for a very high level of freedom you always gave me for our research works, and for treating me as an equal. Thank you for following me down the dark paths plenty of permutohedra, combinatorics and other “black magic stuff”, sometimes (slightly) outside your domain of expertise. I feel that I have grown under your guidance full of kindness and good humour, and I just cannot thank you enough for all you have done for me, and for all the good time we spent together.

My thesis could never run well without a friendly environment. I am a shy tortoise who rarely dares to come out of her shell. And the COVID-19 pandemic as well as my two maternity leaves did not help me to socialize. Thank you Alexis for a huge dose of empathy that you had to make me come out of my shell and talk to other people. Thank you Martin and Alexis for all political discussions I was able to enjoy thanks to you, and, in consequence, for inbuing me with THE ideology to follow - I will remember you if one day I can vote in France! I would also like to thank other persons with whom I shared the office during some period of my thesis - Elena, Anja, and recently Maria Laura - for supporting my messy desk, my messy me, and for all the nice time we spent together. Finally, thank all the other people with whom I have a chance to meet for a lunch. If I don’t talk much (because of my shyness), I really enjoy being with you!

I was a teaching assistant during me thesis. If my teaching experience has gone well, it is also thanks to Pierre Fouilhoux and Nawal Benabbou - thank you for your confidence in me, and thank you for dealing with my (in)availability due to my two maternity leaves and lack of childcare solutions - sometimes it was quite complex. You were able to accompany me in my search for work-family balance, and working with you was a pleasure !

It is not always easy to be a PhD student, and several times I felt blue. I would like to thank to Pierre Fouilhoux and Meltem Öztürk who each crossed my path at one of those moments (one of them at the beginning of my thesis, another at the end), and who both cheered me up and helped me to completely change my mindset.

Finally, I would like to thank several people who introduced me to the world of research - it all started by my high-school teacher Dana Hilská who always supported me in my curiosity and thirst for knowledge. Thanks to her, I had a chance to do my several first research internships at the Charles University of Prague, supervised by

Jakub Staněk to whom I thank for his great patience, encouragements and education towards autonomy. Later, many other persons from the Charles University influenced me - among them, I would like to mention Jarik Nešetřil whose boundless enthusiasm made me understand that I was in love with discrete mathematics, and finally Rudolf Kryl who made me understand that I was perhaps able to follow my heart.

Talking about love, I finish these acknowledgements by mentioning some people who are very dear to my heart. I start by some very close friends that have been with me since my childhood, and of course also during my thesis. Honza, thank you for being THE person I can turn to when I need to be heard. I will never forget our crazy hiking trip in Romania, and all the happy time we spent together. Thank you also for consulting with me on the menu of my pot de thèse! Kiitos siitã, ettã olet kuka olet, sano hei Idalle ja nãhdããn pian Suomessa, Ranskassa tai missã tahansa! Aniãka, thank you for being one of the most ambitious and hardworking persons I have ever known - every time I get discouraged in my work, I think to you, and it helps me to continue. Pãja, thank you for reminding me, through your various little tricks, that there is always someone more clever than me - it's a good mindset for a PhD student I think! Dan, thank you for reminding me that one should follow his/her dreams, as crazy as they may seem. Týna, thank you to show me that is absolutely perfect not to be perfect. I think it is also thanks to you I knew to stop writing this document, even though there is still much to be improved in it.

In this last part, I would like to thank my family. My mother and father of course. I would never be who I am and where I am without them. Dãkuji za vaši vÝchovu a hodnoty, které jste mi pøedali, a díky kterým si pøipadám duševně bohatá. Dãkuji za vaši obãtavou podporu ve všem, co jsem si kdy usmyslela, ať už to bylo sebešilenější. A dãkuji, že se díky vám v tomto těžkém svãtã mũžu cítit bezpeãně. And then all my other family, especially my grandmothers and grandfathers, for a very joyful and care-free childhood and the home where I like to return so much. Finally, a very special thought to my brother František who will never read these lines, but who will be with me in my mind in every important moment of my life.

It was not easy to leave my country. I try hard to put down roots again, but it is too difficult, and I often feel lost and alone. I would like to thank my husband's family: Merci à Laurence et Patrick d'avoir toujours essayé de m'aider à surmonter ces sentiments de solitude, de m'avoir accueilli très chaleureusement et de me proposer une nouvelle maison. Ce n'est pas facile de refaire de nouvelles racines, mais si un jour j'y arrive, ça sera aussi grâce à vous. Merci bien sûr à mamie Monique - son amour immense et son soucis de savoir si je mange bien me donne vraiment une impression d'avoir une troisième grand-mère. Merci aussi à Denis et Sandra d'avoir compris et accepter qu'il me fallait beaucoup de temps avant de m'ouvrir aux autres, et que si souvent, je ne montrais pas beaucoup d'émotions, j'en ressentais plein. Enfin, je voudrais remercier à Marie-Laure, l'assistante maternelle de mes enfants, que je considère à présent un peu

comme un membre de la famille. Ce n'est pas facile de faire garder ses enfants pour aller travailler. Pour tout avouer, c'était même sacrément difficile pour moi. Merci Marie-Laure de m'avoir accompagné dans cette épreuve, et de t'avoir toujours occupé de mes enfants à la merveille. Quand je les vois tout heureux et épanouis chez toi, je pars travailler l'esprit tranquille... et ça, ça n'a pas de prix !

I would also like to thank my children. Thank you Esther for turning my life upside down three years ago, and for making me understand that the best things in life cannot be mastered, formalized and proved. Thank you Loïs for joining our family last year, and for making me understand that love was really infinite. Merci mes enfants de me faire voir le monde à travers de vos yeux, tellement plus parfaits que les miens.

Finally, I would like to thank my husband, Jean-François. And as I know his level of English, I will do it in French: Je te remercie, Jean-François, de m'accepter entièrement, avec mes nombreux défauts et mes petites habitudes et manies qui, auparavant perçues comme "ça fait ton charme", te font râler à présent. Merci pour ta générosité énorme et honnêteté absolue (bien que parfois dure à avaler sur le moment), merci de me donner de la force quand la mienne me fait défaut. Merci de savoir me ralentir quand j'ai tendance à faire trop de choses, et de savoir me faire comprendre que parfois, ça fait du bien de dormir, manger, rien faire.

Depuis la fin de ta thèse (et le début de la mienne), notre vie était tout sauf repos. Ce n'était pas toujours facile, mais je trouve que jusqu'à présent, on a su tout surmonter, et que nous en sommes ressortis plus soudés. Je sais que t'attends impatientement un moment de répit. Et je te promets que je t'en donnera - un jour, quand on sera à la maison de la retraite (si la retraite existera encore). Il n'y aura plus de MP 73 sur la ligne 6 (ça commence déjà), et tu ne rassemblera plus aux corbeaux que tu détestes tant puisque tes cheveux seront blancs (ça aussi, ça commence déjà). Mais il y aura probablement notre tortue Rachel, des hectolitres de la solution hydroalcoolique, et de l'amour qu'on a l'un pour l'autre. Que souhaiter de plus ?

Introduction

An autumn day in 2013, in the small town of Mělník, there was a group of high school friends that were all just 18, which in the Czech Republic means they were finally allowed to vote. And the parliamentary elections were just around the corner! Of course, our friends were very excited about it - they spent every little free time between two classes discussing about politics, and about who they were going to vote for. But one of them mostly kept out of these discussions. Artist in his body and soul, he was too much above the fray to get distracted by something as mundane as a parliamentary election. When the others finally asked him directly who he was going to vote for, he replied : “Well, I’m going for a walk around Mělník on Friday afternoon, and I’ll vote for whoever has the nicest posters.”

Social choice is a field which, among other things, studies, analysis or aggregates individual opinions, with the aim of reaching a common decisions which would be “the best one” in some sense. But actually, what does it mean to be “the best”? Should it be a solution preferred by the majority of voters? Or maybe a solution that is the best compromise? But in such a case, how “the degree of compromise” is measured? Should all voters be treated equally? And the candidates? How do we express our preferences? And what happens if some voter is not sincere? How do we ensure that the election is not manipulated?

Many questions arise if we think a bit about the election process (or, more generally, any decision making where a group of individuals have to choose together among a set of available alternatives), and we can have many requirements on the properties a “good” election system should fulfill. It is hence not a surprise that the first and most famous works in the field deal with identifying and formulating desirable properties (also called axioms) of electorates and voting rules. But of course, it is useless to come out with a theoretically perfect voting rule for which we are unable to compute the winner - and that is what *computational social choice* will deal with. More generally, this field, lying between social choice theory and theoretical computer science, studies computational problems arising from collective decision making (the winner determination, preference elicitation, manipulation issues etc.).

But back to looking for “an ideal” voting rule. Actually, social choice is full of para-

doxes and impossibility theorems. One of the oldest one, the so-called Condorcet paradox (or also Condorcet cycle), was observed at the end of the 18th century by Marquis de Condorcet. Let us consider the situation where a majority of voters prefer candidate c_1 to candidate c_2 , and candidate c_2 to candidate c_3 . One would then expect that a majority of voters would prefer c_1 to c_3 . However, Condorcet noticed that it is possible to have the majority of voters that prefer c_3 to c_1 .

Probably the most famous impossibility theorem is Arrow's impossibility theorem¹ (Arrow, 1951). Basically, Arrow assumes that each voter gives a complete ranking over the set of candidates, and a voting rule returns a single global societal ranking. He considers the following properties, reasonable for a "fair" voting rule:

- **Unanimity:** If candidate c_1 is preferred to candidate c_2 by each voter, then c_1 should be ranked better than c_2 in the global ranking.
- **Independence of irrelevant alternatives:** The preference between c_1 and c_2 in the global ranking should only depend on how c_1 and c_2 are ranked by the voters, regardless of the other candidates (in other words, if some voter change her preference for c_3 , it should not impact the global preference between c_1 and c_2).
- **Universality:** Each voter can give any possible complete ranking over the set of candidates.

Arrow proves that when there are at least 3 candidates, the only preference aggregation rule (returning a complete ranking over the set of candidates) satisfying these properties is a dictatorship. It should be emphasized that this does not mean that there is no possible "good" voting rule, but that any voting rule behaves "badly" from time to time. Despite that, the result may seem destabilizing, as each of the properties seems quite reasonable, while most of the readers will probably not find a dictatorship to be an appropriate voting rule.

Another famous result is the Gibbard–Satterthwaite theorem (Gibbard (1973), Satterthwaite (1975)). It states that, as soon as there are three or more candidates, any non-dictatorial voting rule can be manipulated - it basically means that it may be beneficial for some voter not to give her true opinion about candidates. Which is quite bad news...

... And that is where *structured preferences* come on the scene. They are natural to try to circumvent impossibility results and paradoxes, as an assumption of a common underlying preference structure shared by all voters relaxes for instance the universality condition of Arrow's impossibility theorem. We will present some popular preference structures in Chapter 1, but let us give here a brief overview.

¹What kind of thesis dealing with social choice this would be if Arrow's theorem was not mentioned...

One of the most popular preference structures is called *single-peaked domain*. This structure has been introduced by Black (1948) and independently by Arrow (1951). Basically, it assumes that the candidates can be ordered on an axis (for instance, we can imagine a left-right political axis). Each voter has her most preferred candidate somewhere on the axis - this position is called a peak. The more we move from the peak to the left or, independently, to the right, the worse the candidates are ranked by the voter. Under the assumption of single-peaked preferences, the Arrow's theorem does not hold, there is no Condorcet cycle and we can find some non-manipulable voting rules.

Another popular preference structure is the so-called *single-crossing domain*. The idea is similar to single-peaked domain, but this time, it is the set of voters that can be ordered on the axis, which can be seen as some scalable (sociological) "criterion" (as for instance the wealth, or the level of education). Then we suppose that for each pair of candidates c_1 and c_2 , there is a threshold value of the "criterion" beyond which each voter prefers c_1 to c_2 (resp. c_2 to c_1). On the other hand, all voters below the threshold value will prefer c_2 to c_1 (resp. c_1 to c_2). If the number of voters is odd, there is no Condorcet cycle.

Since good things always come in threes, we present here another preference structure. These are *one-dimensional Euclidean preferences* (or *1-Euclidean preferences*), first introduced by Coombs (1950). The idea is similar to single-peaked preferences. The candidates are ordered on an axis. Each voter is identified with a point of this axis called ideal point (note that this point is arbitrary, and there is in particular no reason it corresponds to any candidate position). The further the candidate is from this ideal point, the less it is appreciated by the voter. Note that, contrary to the notion of single-peakedness, the left and right axis part (separated by the most preferred candidate) are not treated independently. It is well-known that if the preferences are 1-Euclidean, they are both single-peaked and single-crossing (the reverse implication is false, as we will see later). Therefore, they benefit from all the good properties of these domains.

The interest of structured preferences goes further. Actually, some problems of computational social choice that are NP-hard in the general case can become polynomial-time solvable under some assumptions on the preference structure. For instance, as single-peaked preferences always admit a Condorcet winner, all Condorcet consistent voting rules become polynomial time computable under the assumption of single-peakedness. Last but not least, it is interesting to know the underlying preference structure of a given profile for better analysis, interpretation and understanding of data. Indeed, the knowledge of the structure can be used in recommender systems, voting advice applications, etc. For instance, if the preferences are known to be single-peaked and the recommender system knows the most preferred alternative A of a given voter, it will recommend her one of the alternatives lying next to A on the axis.

To sum up, structured preferences are undoubtedly worth attention, providing many appealing properties and interesting applications. Too good to be true? Unfortunately, the three aforementioned structures are nearly impossible to observe in real world collective decision making problems (at least, in the strict sense). Actually, it is not difficult to see why - to get the first insight, let us go back to the first paragraph of this introduction - even though the majority of voters might agree on the common structure (for instance, the left-right political axis), there will always be a person who decides in function of the beauty of the posters. More generally, real world applications are often complex, with many decision criteria to be taken into account, some of them being very subjective or non-scalable, some of them being relevant for only a part of the population... without forgetting the fact that some voters are not “rational”.

One of the goals of this thesis is to try to deal with these issues. The idea is simple: we relax the structure in order to make it less restrictive, and hence more likely to fit a real world data. Of course, the more the structure is relaxed, the more likely it will be observed, but on the other hand, many good properties, as well as structural information, will be lost. It is hence important to find a relaxation that achieves a good trade-off.

Part I is an introduction to the topic of structured preferences. In Chapter 1, we define the general framework of an election (or a decision making problem) that will be used in the whole thesis. We also introduce formally the three preference structures mentioned above, providing a state of the art and several examples. In Chapter 2, we discuss the relation between single-peaked, single-crossing and 1-Euclidean preferences. While most of the results are already known, we reinterpret and visualize them using graphs. The idea is to provide more insights on structured preferences, to illustrate the type of problems one can face while studying structured preferences, and to motivate the reader to dive deeper into the topic. In addition, some (marginal) new results are given in this chapter.

Part II focuses on single-peaked domain relaxations. In Chapter 3, we relax the axis structure by considering the notion of *single-peakedness on a graph*. A preference profile is single-peaked with respect to a given graph \mathcal{G} if each of its preferences is a traversal of \mathcal{G} . Trivially, any profile is single-peaked with respect to the complete graph. However, such a result is useless for data analysis purposes. Indeed, the sparser the graph is, the more structural information it provides. Two notions of sparsity are introduced in Chapter 3, and we are looking for solutions minimizing (one of) them. The works presented in this chapter has been published in Escoffier et al. (2020) (a full version is currently under evaluation in a journal).

In Chapter 4, a different approach is considered: this time, we keep the axis structure, but we allow the voters to derive a bit from it. We find then an axis which is the “nearest” to the set of preferences. This approach is known as *nearly single-peakedness*.

Obviously, we loose the good axiomatic properties of single-peaked preferences, but it is still interesting to capture the underlying structure of preferences in order to understand the data. Several metrics of single-peakedness have been considered in the literature. We introduce here a new metric, inspired directly by the definition of single-peakedness, we study its theoretical and computational properties, and we position it with respect to existing metrics of nearly single-peakedness. The works presented in this chapter has been published in Escoffier et al. (2021).

Another possible way to make the structure less restrictive is to add more dimensions. This is what we do in Part III. This time, we focus on multidimensional Euclidean preferences. We introduce the topic in Chapter 5. The works presented in the last section of this chapter has been published in Escoffier et al. (2022b).

In Chapter 6, we propose a heuristic algorithm for recognizing two-dimensional Euclidean preferences with respect to the norm ℓ_2 . In Chapter 7, we study two-dimensional Euclidean preferences with respect to the norm ℓ_1 . We mainly focus on geometrical and structural properties of this domain. The works presented in this chapter are available in Escoffier et al. (2022a), and currently under evaluation in a journal.

Finally, Part IV is devoted to the general conclusion where we discuss some possible research directions arising from our works.

How to read this document Let us terminate this introduction with some tips on how to read the document. My goal was to write it as “a book”, with the chapters following logically one another. However, it should be also possible to read the chapters independently. In this case, the reader should first read Section 1.1 of Chapter 1, where the common notations are introduced. It is also recommended (but not necessary) to read Section 1.2 of Chapter 1, devoted to the single-peaked preferences, if one is interested in Chapter 3 or Chapter 4.

Each chapter provides its own introduction; however, as in Part III both Chapters 6 and 7 deal with multidimensional Euclidean preferences, a common introduction is given in Section 5.2. Although the most important notations are quickly recalled at the beginning of each chapter, it is highly recommended to read this common introduction for better understanding. Also, each chapter provides a very brief conclusion whose main purpose is to summarize the principal contributions. Possible research directions are discussed in more details in the general conclusion of the thesis.

Another specificity of this document is that the reader is not alone - indeed, he or she is accompanied by a little tortoise named Mathilde. It is well-known that tortoises are quite slow animals. Hence, if you are in a hurry, you can skip the tortoise interventions - they are somehow “a bonus” to the text, and are not necessary for understanding. However, if you have the time and you want to go further, please do not hesitate to

read Mathilde's comments. She is there to point out some interesting details, cause you to ask questions, give you more intuition or, on the contrary, make you hesitate.

A Curious Tortoise Intervention



Hello reader, nice to meet you!

Finally, let me allow one more personal note. I make no secret on the fact that, as for the style of writing, I have been inspired by the book of Jiří Matoušek and Jaroslav Nešetřil entitled *Invitation to Discrete Mathematics*. I discovered this book when I was 14 years old, and I read it (nearly) in one go. It is clearly one of my favourite books that has a honour place in my library and to which I return when I need to be heartened. I always admired the ease of writing and the gentle humour this book is filled with. Everything is explained very clearly, with a great emphasis on intuition. It is affordable to a wide audience, but despite that, it is impeccably rigorous.

For me, intuition and visualization are extremely important. That is why my thesis is full of figures and examples, and I try to give as much motivation and intuition as I can. Of course, it would be extremely pretentious to want to do as good as Matoušek and Nešetřil. But I did my best...



Opening

Chapter 1

Preliminaries

1.1 Preference profile

Let us introduce basic notions and notations that will be used in the whole thesis. We consider a set $V = \{v_1, v_2, \dots, v_n\}$ of n voters. Each voter expresses her preference over the set of m candidates $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$. The preference of a voter v_i , denoted by $>_i$, is a total strict order on \mathcal{C} .

A Curious Tortoise Intervention



For convenience, let us recall what a *total strict order* is. It corresponds to a binary relation (denoted here by $>_i$) on the set \mathcal{C} that has the following properties:

- **irreflexive:** for any c_j , we do not have $c_j >_i c_j$ - in other words, no candidate can be strictly preferred to herself.
- **antisymmetric:** we cannot have at the same time $c_j >_i c_k$ and $c_k >_i c_j$ - if c_j is strictly preferred to c_k , then c_k cannot be strictly preferred to c_j .
- **transitive:** if $c_j >_i c_k$ and $c_k >_i c_l$, then $c_j >_i c_l$.
- **total:** any couple of candidates c_j and c_k can be compared - we have $c_j >_i c_k$ or $c_k >_i c_j$.

The multi-set¹ $\mathcal{P} = \{>_1, >_2, \dots, >_n\}$ is called a *preference profile*.

It is worth mentioning that other types of preferences exist in the literature. For instance, we can consider incomplete rankings where the voters may not give a complete ranking but only its upper part. We can also consider rankings (complete or not) with

¹Indeed, the elements of \mathcal{P} are not necessarily pairwise distinct, as several voters can give the same preference.

ties - in such a case, the voters have a possibility to express indifference between several candidates. However, only complete preferences without ties (corresponding to total strict orders) are considered in this thesis.

We write $c_1 >_i c_2$ if voter v_i prefers c_1 to c_2 . Moreover, given a set of candidates S , we write $S >_i c$ if for each $c' \in S$, we have $c' >_i c$. The preference $>_i$, called also the ranking of v_i , writes $c_j >_i c_k >_i \dots >_i c_l$. For brevity, we will often write it as (c_j, c_k, \dots, c_l) .

Let us define, for each voter v_i , her ranking application $r_i : \{1, \dots, m\} \rightarrow \mathcal{C}$, where $r_i(k)$ is the k -th most preferred candidate of v_i . This application is obviously a bijection between the candidates and the ranks, so we can define its inverse r_i^- , where $r_i^-(c_k)$ is the position of candidate c_k in the ranking $>_i$. Finally, we denote by c_i^* the most preferred candidate of v_i (i.e., the candidate $r_i(1)$).

Example 1.1.1. Let us consider a profile $\mathcal{P} = \{>_1, >_2, >_3, >_4\}$ of 4 preferences over 5 candidates with:

$$\begin{aligned} >_1 &: (c_2, c_3, c_1, c_4, c_5) \\ >_2 &: (c_3, c_4, c_5, c_2, c_1) \\ >_3 &: (c_4, c_5, c_3, c_2, c_1) \\ >_4 &: (c_2, c_3, c_1, c_4, c_5) \end{aligned}$$

We have, for instance, $c_1^* = c_2, r_1(2) = c_3$ and $r_1^-(c_4) = 4$. We also note that preferences of voters v_1 and v_4 are the same.

Given two preferences $>_i$ and $>_j$, it is natural to want to measure the “similarity” between them. A classical way to do so is to introduced the so called *Kendall tau distance* which counts the number of pairwise disagreements between two preferences. The greater this number is, the less $>_i$ and $>_j$ are similar. More formally, we have the following definition:

Definition 1.1.1: Kendall tau distance

Given two preferences $>_i$ and $>_j$, we define the Kendall-tau distance between $>_i$ and $>_j$, denoted $d_{KT}(>_i, >_j)$, as the number of pairs of candidates $\{c, c'\}$ such that $c >_i c'$ and $c' >_j c$, or vice versa:

$$d_{KT}(>_i, >_j) = |\{\{c, c'\} : c >_i c' \text{ and } c' >_j c, \text{ or vice versa}\}|$$

Example 1.1.2. To illustrate this notion, let us consider two preferences $>_i = (c_1, c_2, c_3, c_4)$ and $>_j = (c_3, c_1, c_2, c_4)$. The only pairs ranked in different order in $>_i$ and $>_j$ are $\{c_1, c_3\}$ and $\{c_2, c_3\}$. We have thus $d_{KT}(>_i, >_j) = 2$.

Given a preference profile \mathcal{P} , we call a *subprofile* of \mathcal{P} any preference profile that can be obtained from \mathcal{P} by removing some preferences. We call a *restriction* of \mathcal{P} on a set

$\mathcal{C}' \subset \mathcal{C}$ a preference profile \mathcal{P}' obtained from \mathcal{P} by removing from each preference $>_v$ all the candidates of $\mathcal{C} \setminus \mathcal{C}'$.

Definition 1.1.2: Subprofile

Let \mathcal{P} be a profile of n preferences over the set of m candidates $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$. A preference profile \mathcal{P}' over the same set \mathcal{C} of candidates is a *subprofile of \mathcal{P}* if, for each $>_v \in \mathcal{P}'$, we have $>_v \in \mathcal{P}$.

Definition 1.1.3: Restriction of profile \mathcal{P} on a set $\mathcal{C}' \subset \mathcal{C}$

Let $\mathcal{P} = \{>_1, >_2, \dots, >_n\}$ be a profile of n preferences over the set of m candidates $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$, and let $\mathcal{C}' \subset \mathcal{C}$. A preference profile $\mathcal{P}' = \{>'_1, >'_2, \dots, >'_n\}$ over the set of candidates $\mathcal{C}' \subset \mathcal{C}$ is a *restriction of \mathcal{P} on \mathcal{C}'* if for each $i \in \{1, \dots, n\}$, and for each pair of candidates (c_j, c_k) of \mathcal{C}' ,

$$c_j >'_i c_k \text{ iff } c_j >_i c_k.$$

With these notions, we can finally define a *minor of \mathcal{P}* as a subprofile of a restriction of \mathcal{P} . In other words, we remove some candidates from each preference of \mathcal{P} , as well as some preferences of \mathcal{P} .

Example 1.1.3. Let us consider the profile $\mathcal{P} = \{>_1, >_2, >_3, >_4\}$ of 4 preferences over the set of 5 candidates defined in Example 1.1.1. The profile $\mathcal{P}' = \{>_1, >_3\}$ of 2 preferences over the set of 5 candidates is a subprofile of \mathcal{P} .

Let us now consider a profile \mathcal{P}'' of 4 preferences over the set of candidates $\mathcal{C}'' = \{c_1, c_3, c_4\}$ with:

$$\begin{aligned} >''_1: (c_3, c_1, c_4) \\ >''_2: (c_3, c_4, c_1) \\ >''_3: (c_4, c_3, c_1) \\ >''_4: (c_3, c_1, c_4) \end{aligned}$$

We have $\mathcal{C}'' \subset \mathcal{C}$, and \mathcal{P}'' is a restriction of \mathcal{P} on \mathcal{C}'' .

Finally, $\mathcal{P}''' = \{>''_1, >''_3\}$ is a minor of \mathcal{P} .

We end this chapter by the definition of a *domain restriction*. We use in this thesis the definition given by Elkind et al. (2022), however, note that another definitions of this notions exist in the literature (see for instance Puppe (2018)).

Definition 1.1.4: Domain restriction (Elkind et al., 2022)

A domain restriction is a set of preference profiles.

As said in the introduction, this thesis deals with structured preferences. Each preference structure S can be associated to a domain restriction D_S in the following way: a preference profile \mathcal{P} is conform to the given structure S if and only if $\mathcal{P} \in D_S$. As emphasized by Elkind et al. (2022), a domain restriction can be viewed as a property of a profile. Typically, this property imposes some structure on the profile. In the rest of this chapter, we will present some of the most popular domain restrictions.

1.2 Single-peaked preferences

We will now formally introduce the notion of single-peakedness. This notion was originally introduced by Black (1948), and independently rediscovered by Arrow (1951). The underlying idea is simple: we suppose that voters decide according to a unique scalable criterion. The different alternatives can be ordered on an axis using the different values of the criterion. Let illustrate this idea on a classical example:

Example 1.2.1. *The decision to take is here very straightforward - we want to decide about the temperature to be set on a collective thermostat. Unless the voter is a tortoise or another poikilotherm animal, it seems very natural that she has her most preferred temperature somewhere on the scale, and that her preference will decrease as we lower (or, independently, increase) the temperature. For instance, if the ideal temperature of the voter is 18°C , it would be quite strange if her second most preferred temperature would be 24°C and the third one, let us say, 20°C . That is why we call this structure single-peaked - if we plot the graph of the level of preference in function of the attribute value, 18°C will be here a peak, and if we move to the left (or to the right), the preference will decrease. Therefore, there will be no other local maximum, i.e., peak. It is important to note that if two values are not on the same side of the peak, they can be ordered in both possible ways - for instance, we can have, in the above example, two preferences $>_v$ and $>_{v'}$ with the same peak 20°C such that $16^\circ\text{C} >_v 22^\circ\text{C}$, but $22^\circ\text{C} >_{v'} 16^\circ\text{C}$. However, for both $>_v$ and $>_{v'}$, we have $18^\circ\text{C} >_v 16^\circ\text{C}$.*

For a better visualisation, let us go back to the profile of Example 1.1.1. We will associate to each c_i a temperature, namely $c_1 = 16^\circ\text{C}$, $c_2 = 18^\circ\text{C}$, $c_3 = 20^\circ\text{C}$, $c_4 = 22^\circ\text{C}$ and $c_5 = 24^\circ\text{C}$. In Figure 1.1, we can see that the profile is indeed single-peaked with respect to the axis where the temperatures are ordered increasingly from left to right.

A Curious Tortoise Intervention



Another well-known example that should definitely be mentioned is the example of left-right political axis. Each candidate can be positioned somewhere on the left-to-right political spectrum. Each voter has her most preferred candidate, and her preference then decreases if we move towards the left or right extremity.

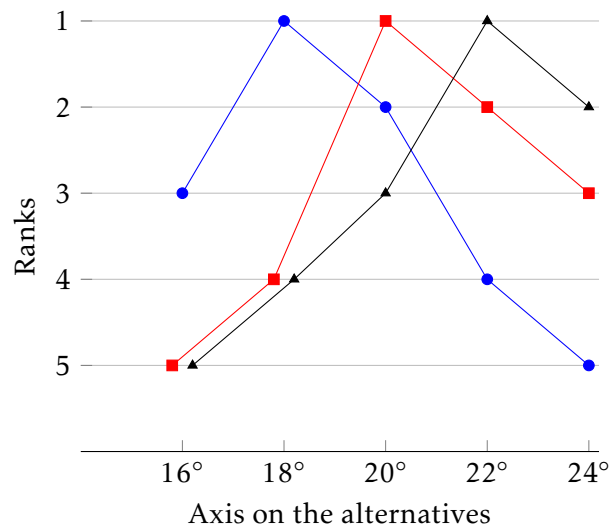


Figure 1.1: The profile from Example 1.1.1 is single-peaked with respect to the axis $c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_5$.



We did not choose it for the introductory example, as it is much more complex. Firstly, it is not always evident to place the candidates exactly on the axis. For instance, some candidates may be indistinguishable on the political axis, while opinions of other ones may vary from left to right, depending on the subject. Also, decisions of voters are typically very complex in a real-world election, and all decision criteria are not necessarily covered by (or in correlation with) the left-right position. Actually, it is often discussed whether the voter really decide according to the left-right axis and if so, to which extent. In particular, if we collect real-world election data, can we identify the underlying political axis? This point will be discussed in Chapter 4.

We are now ready to give the formal definition of single-peakedness. We start by defining an axis. Given a set of m candidates $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$, an axis A on the set \mathcal{C} is a total strict order \triangleleft_A on \mathcal{C} . We write $c_i \triangleleft_A c_j$ if c_i is on the left of c_j . Given an axis $A = c_i \triangleleft_A c_j \triangleleft_A \dots \triangleleft_A c_k$, its leftmost and rightmost candidates (here c_i and c_k) are called *the extremities* of the axis A . When no confusion is possible, we omit the index A in the notation - in other words, we write \triangleleft instead of \triangleleft_A .

We can now give the original definition of single-peakedness introduced by Black (1948):

Definition 1.2.1: Single-peaked preferences (Black, 1948)

Let $\mathcal{P} = \{>_1, >_2, \dots, >_n\}$ be a preference profile over the set of m candidates $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$, and let A be an axis on \mathcal{C} . The profile \mathcal{P} is *single-peaked with respect to A* if for any preference $>_i$ and any couple of candidates c_j, c_k such that $c_k \prec_A c_j \prec_A c_i^*$ or $c_i^* \prec_A c_j \prec_A c_k$, we have $c_j >_i c_k$.

We say that a preference profile is *single-peaked* if there exists an axis A such that the profile is single-peaked with respect to A .

As said earlier, this notion has been rediscovered by Arrow (1951) who formulated it in a slightly different way - instead of requiring the preferences to be decreasing while moving towards the extremities from the peak, he states that if a candidate c_j lies between candidates c_l and c_r on the axis, then she is never ranked worse than both c_l and c_r :

Definition 1.2.2: Single-peaked preferences (Arrow, 1951)

Let $\mathcal{P} = \{>_1, >_2, \dots, >_n\}$ be a preference profile over the set of m candidates $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$, and let A be an axis on \mathcal{C} . The profile \mathcal{P} is *single-peaked with respect to A* if for any preference $>_i$ and any triple of candidates $c_l \prec_A c_j \prec_A c_r$, we have $c_j >_i c_l$ or $c_j >_i c_r$ - in other words, c_j is never ranked last among these three candidates.

Again, a preference profile is *single-peaked* if there exists an axis A such that the profile is single-peaked with respect to A .

It is easy to see that both definitions are equivalent. Hence, we will use them interchangeably, depending on the context. We say that a preference is *compatible* with axis A if it is single-peaked with respect to A . Symmetrically, an axis A is *compatible* with a given preference (resp. profile) if this preference (resp. profile) is single-peaked w.r.t. A .

1.2.1 Axiomatic and algorithmic properties

Condorcet winner There are many desirable properties that hold under the assumption of single-peaked preferences. For instance, any single-peaked profile with an odd number of voters admits a Condorcet winner. A Condorcet winner is a candidate that beats all the other candidates in pairwise comparison. It is often advocated that if such a candidate exists, she should win the election. The notion of Condorcet winner was first proposed by Marquis de Condorcet at the end of the 18th century. He also noticed that some profiles do not admit a Condorcet winner: actually, there can be a cycle in

the majority relation, as shown in the following example.

Example 1.2.2. *Let us consider a profile of 3 preferences over 3 candidates:*

$$>_1: (c_1, c_2, c_3)$$

$$>_2: (c_2, c_3, c_1)$$

$$>_3: (c_3, c_1, c_2)$$

We have a majority of voters that prefer c_1 to c_2 , a majority of voters that prefer c_2 to c_3 and a majority of voters that prefer c_3 to c_1 . In other words, we get $c_1 > c_2 > c_3 > c_1$ - there is a cycle in the majority relation $>$. Note that we can easily check that this profile is not single-peaked.

Impossibility theorems We have already seen in the introduction that Arrow's impossibility theorem does not hold under the assumption of single-peakedness. We have also stated that the Gibbard–Satterthwaite theorem does not hold in such a case either - indeed, non-manipulable voting rules (called also strategyproof voting rules in literature) can be found if the preferences are known to be single-peaked. To give an example, let us consider the so called *median voter rule*. This rule returns the Condorcet winner (which is unique under the assumption of an odd number of voters): to do so, it finds an axis A w.r.t that the profile is single-peaked, and then it orders the voters according to their peaks (where the order on peaks is determined by A). It turns out that the Condorcet winner is the peak of the median voter (Black, 1948). Later, Moulin (1980) proved that this voting rule is non-manipulable.

Example 1.2.3. *To illustrate this rule, let us consider the subprofile of the profile \mathcal{P} of Example 1.1.1 containing its first three preferences $>_1, >_2$ and $>_3$.² We have seen that this profile is single-peaked w.r.t the axis $A = c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4$, so if we want to order the preferences according to their peaks, we first take all preferences with the peak c_1 , then all preferences with the peak c_2 , etc. Finally, the preferences with the peak c_5 are put to the end of this order. It turns out that the order $(>_1, >_2, >_3)$ respects already the order on peaks determined by A . The median voter (resp. median preference) is then v_2 (resp. $>_2$), so the Condorcet winner is the peak of $>_2$, i.e., c_3 . We can check that this candidate is a Condorcet winner - it is preferred to c_1 by all three voters, to c_2 by voters v_2 and v_3 , to c_4 by voters v_1 and v_2 , and to c_5 by voters v_1 and v_2 . In particular, we see that it is the unique Condorcet winner of this profile.*

A Curious Tortoise Intervention



It is quite easy to see why the median voter rule is non-manipulable. Let c^* be the Condorcet winner, and let us consider a voter v whose most preferred candidate c' is on the left of c^* on axis A . There are two possibilities

²For the simplicity, the fourth preference is omitted here to have an odd number of voters, which guarantees the uniqueness of Condorcet winner. However, note that any single-peaked profile admits a (weak) Condorcet winner.



for v to lie about her most preferred candidate:

- v declares that her most preferred candidate is c_l with $c_l \triangleleft_A c^*$. In such a case, the preference of v is still ordered on the left of the median voter - in other words, the number of preferences ranked on the left of the median voter does not change, hence the median voter remains the same
- v declare that her most preferred candidat is c_r with $c^* \triangleleft_A c_r$. In such a case, the preference of v moves from the left side to the right side of the (true) median voter in the ordering. Hence, the median voter either remains the same, or moves to the right - in other words, further from the true preference of v . As the preference of v is single-peaked w.r.t. A , the new Condorcet winner can only be worse for her (or, in the best case, it remains the same).

To sum up, v has nothing to gain by manipulating. We reason the same for any voter whose most preferred candidate lies on the right of the Condorcet winner. Finally, we note that a voter whose peak corresponds to the Condorcet winner has no motivation to lie.

Computational aspects From the computational point of view, it is known that single-peaked preferences help to simplify some NP-hard problems. For instance, there are many NP-hard voting rules (for instance, Kemeny rule (Brandt et al., 2015), Dodgson rule (Brandt et al., 2015) or Young rule (Peters and Lackner, 2020)) that become polynomial-time solvable for single-peaked preferences.

Preferences elicitation Another important topic of decision theory that we have not yet discussed, and that is actually not tackled in this thesis, is preference elicitation. Obviously, the preference profile is the key element of a social choice problem. However, it can be quite tricky to collect the preferences - especially with an important number of candidates, it can be a complex task for a voter to determine what is *exactly* her preference over *the whole* set of candidates. One of the interests of preference elicitation is to ask the voter *good* questions (called queries) about her preferences (for instance, one can ask if the voter prefers candidate c to candidate c') in order to get enough information to build voter's complete ranking. It has been shown (when no preference structure is assumed) by Conitzer and Sandholm (2005) that, in the worst case, it is necessary to question the voter about all her pairwise preferences to build her complete ranking of candidates. In contrast, Conitzer (2009) showed that if a preference is known to be single-peaked with respect to a known axis, it can be elicited using only a linear number (in terms of the number of candidates) of comparison queries.

1.2.2 Recognition of single-peaked preferences

In the case of the thermostat problem introduced in Example 1.2.1, the single-peaked axis is known a priori - different values of temperature are simply ordered increasingly. Also in the case of an election, the political left-right axis can be built a priori, as each candidate takes in general a position on the left-right spectrum. However, for some preference profiles, we can believe that they might be single-peaked, without knowing the underlying axis. Given a preference profile, it is thus natural to want to decide whether it is single-peaked, and, if so, to identify a compatible axis. This problem is called a *recognition problem*.

The question of the computational complexity of recognizing single-peaked preferences has been widely studied. Bartholdi III and Trick (1986) have proposed an $O(nm^2)$ algorithm to compute a compact representation of all axes on which a collection of n preferences over m candidates are single-peaked, or state that none exists. This complexity was then improved to $O(n^2 + nm)$ by Doignon and Falmagne (1994). It can be decreased to $O(nm)$ if one looks for only one possible axis (Escoffier et al., 2008).

Moreover, there exists an elegant characterization of the domain of single-peaked preferences using the so called *forbidden minors* - a given profile is single-peaked if and only if it does not contain any of these minors. This result has been found by Ballester and Haeringer (2011):

Theorem 1.2.1: Characterization of the single-peaked domain (Ballester and Haeringer, 2011)

A preference profile \mathcal{P} of n preferences over the set of m candidates \mathcal{C} is single-peaked if and only if:

- There does not exist any triple of candidates $\{c_{i_1}, c_{i_2}, c_{i_3}\}$ and any triple of voters $\{v_j, v_k, v_l\}$ such that

$$\{c_{i_1}, c_{i_2}\} >_j c_{i_3}, \{c_{i_1}, c_{i_3}\} >_k c_{i_2} \text{ and } \{c_{i_2}, c_{i_3}\} >_l c_{i_1}.$$

In other words, for any triples of candidates, at most two of them can be ranked last in the votes that make up the preference profile.

AND

- There does not exist any quadruple of candidates $\{c_{i_1}, c_{i_2}, c_{i_3}, c_{i_4}\}$ and any pair of voters $\{v_j, v_k\}$ such that

$$\{c_{i_1}, c_{i_4}\} >_j c_{i_2} >_j c_{i_3} \text{ and } \{c_{i_3}, c_{i_4}\} >_k c_{i_2} >_k c_{i_1}.$$

Strictly speaking, this theorem does not give explicitly the forbidden minors. However, they can be derived easily from it - namely, there are 5 *explicit* forbidden minors (up to renaming the candidates). For conciseness, we display each minor \mathcal{S}_i as a matrix, where preferences are given in columns:

$$\mathcal{S}_1 = \begin{pmatrix} c_1 & c_1 & c_3 \\ c_2 & c_3 & c_2 \\ c_3 & c_2 & c_1 \end{pmatrix}, \mathcal{S}_2 = \begin{pmatrix} c_1 & c_2 & c_3 \\ c_2 & c_3 & c_1 \\ c_3 & c_1 & c_2 \end{pmatrix}, \mathcal{S}_3 = \begin{pmatrix} c_4 & c_4 \\ c_1 & c_3 \\ c_2 & c_2 \\ c_3 & c_1 \end{pmatrix}, \mathcal{S}_4 = \begin{pmatrix} c_4 & c_3 \\ c_1 & c_4 \\ c_2 & c_2 \\ c_3 & c_1 \end{pmatrix}, \mathcal{S}_5 = \begin{pmatrix} c_1 & c_3 \\ c_4 & c_4 \\ c_2 & c_2 \\ c_3 & c_1 \end{pmatrix}$$

A Curious Tortoise Intervention



Without giving a proof of this theorem, let us visualize it a little bit - to get more insight, we will see why these subprofiles cannot be contained in a single-peaked profile. This is the necessary condition. We leave aside why this condition is sufficient.

At first, let us make an easy observation: if a candidate c is ranked last by at least one voter v_i , she must be an extremity of any compatible axis. Indeed, in the opposite case, there is at least one candidate c_l on the left and one candidate c_r on the right of c . We have $c_l \triangleleft c \triangleleft c_r$, but $c_l >_i c$ and $c_r >_i c$. This is in contradiction with Definition 1.2.2.

In particular, there are at most two candidates ranked last in any single-peaked profile. Note that this remains valid for any subprofile on any subset of candidates - in particular, in a subprofile on a triple of candidates, at most two of them can be ranked last. But this is nothing but the first two forbidden minors.

The second type of forbidden minors uses the same observation in a more complex way. Let us suppose that there are two voters v_i and v_j such that $\{c_1, c_4\} >_i c_2 >_i c_3$ and $\{c_3, c_4\} >_j c_2 >_j c_1$. As c_3 and c_1 are ranked last, we have necessarily (up to an axis reversal) $c_1 \triangleleft \{c_2, c_4\} \triangleleft c_3$. However, if $c_2 \triangleleft c_4$, the candidates c_1, c_2 and c_4 do not fulfill Definition 1.2.2 for $>_i$, and if $c_4 \triangleleft c_2$, the triple $\{c_4, c_2, c_3\}$ does not fulfill the definition for $>_j$.

However, as emphasized by Feld and Grofman (1986), single-peaked preferences are very difficult to observe in practice. Indeed, single-peakedness in the strictest sense requires that no individual preference deviates (even slightly) from the single-peakedness condition. Given an axis A , the number of rankings consistent with A is 2^{m-1} , over $m!$ possible rankings in total. The proportion of consistent rankings within all possible rankings thus quickly becomes tiny when m increases (e.g., for $m = 7$, $2^{m-1}/m! \approx 0.01$). Hence, a natural idea is to relax, hopefully just a little bit, the condition of single-peakedness, in order to get a structure that would be more likely to appear in practice, while conserving as many good properties as possible. This problem is studied in

Part II. In Chapter 3, we extend the notion of single-peakedness on an arbitrary graph. This idea is not completely new - several classes of structured preferences have been proposed to generalize the single-peaked domain with respect to an axis, i.e., a path, to more general graphs. For instance, Demange (1982) studied single-peakedness on a tree, Peters and Lackner (2017) on a circle. In Chapter 4, we study the so called nearly single-peakedness: we keep the structure of axis, but we allow a profile to derive from it if necessary. A distance measure needs then to be introduced to evaluate the distance of a profile from a given axis.

1.3 Single-crossing preferences

Another popular domain restriction is known as *single-crossing preferences*. The underlying idea of this structure is similar to single-peaked preferences. However, this time, it is the set of voters that is ordered on an axis. The motivation for doing that is the following: in some specific context, we can suppose that voters can be ordered according to some “ideological” aspect. For instance, if we go back to our thermostat example, the preferences could be governed by the level of environmental concern. Voters who care more about ecology will probably tend to prefer lower temperatures than voters who do not care about it. More generally, it is quite natural to assume that preferences somehow depend on the level of environmental concern - for each couple of candidates c_i and c_j , we can identify a level such that all voters situated above this level will prefer c_i to c_j (resp. c_j to c_i), and all voters below this level will prefer c_j to c_i (resp. c_i to c_j). Visually, suppose that we write the preferences as columns in the single-crossing ordering, and for each candidate, we connect her occurrences in each of the preferences by a polygonal chain. Then, for each couple of candidates, their corresponding chains will cross at most once - up to some ideological position, one of them will be preferred to the other by each voter, and then the preference will reverse (see Figure 1.2). Let us now formalize this idea:

Definition 1.3.1: Single-crossing preferences

A preference profile $\mathcal{P} = \{>_1, >_2, \dots, >_n\}$ over a set \mathcal{C} of m candidates is *single-crossing* if there exists an ordering $\pi = (>_{i_1}, >_{i_2}, \dots, >_{i_n})$ of the preferences such that, for each couple of candidates c and c' , the sets $\{i_k | c >_{i_k} c'\}$ and $\{i_k | c' >_{i_k} c\}$ form intervals of π .

The profile from Example 1.1.1 is single-crossing for the following order on voters: $(>_1, >_4, >_2, >_3)$. We can check that for each pair of candidates, the definition holds. For instance, we have $c_3 >_i c_5$ ³ for $i \in \{1, 2, 4\}$, and $c_5 >_i c_3$ for $i \in \{3\}$. These two sets both form an interval of the given ordering.

³We recall that in the example of the thermostat, we have associated to each c_i a temperature, namely $c_1 = 16^\circ\text{C}, c_2 = 18^\circ\text{C}, c_3 = 20^\circ\text{C}, c_4 = 22^\circ\text{C}$ and $c_5 = 24^\circ\text{C}$.

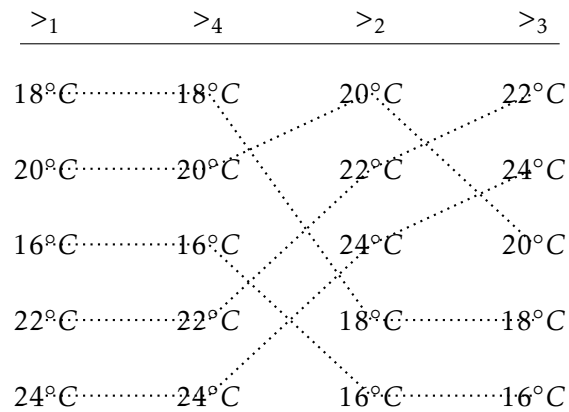


Figure 1.2: The profile from Example 1.1.1 is single-crossing for the following order on the voters: $(>_1, >_4, >_2, >_3)$.

Pioneered by Mirrlees (1971) and Roberts (1977), this domain restriction also guarantees some desirable properties. For instance, the majority relation is transitive for odd profiles, and, what is more, any (odd) profile contains a voter which coincides with this majority relation (note that this is not a case for single-peaked preferences) (Rothstein, 1991).

Similarly as for single-peaked preferences, it is natural to want to decide whether a given profile is single-crossing or not. We recall that this problem is called a *recognition problem*. It has been widely studied in the literature: Bredereck et al. (2013) give an $O(nm^2)$ recognition algorithm for single-crossing preferences by reducing the problem to the *consecutive ones problem* (see Booth and Lueker (1976) for more details on the consecutive ones problem). Elkind et al. (2022) show that this complexity can be refined to $O(nm \log(m))$. Contrary to single-peaked preferences, there is no known recognition algorithm that performs in $O(nm)$.

Similarly to the single-peaked domain, the single-crossing domain can be characterized by forbidden minors (Bredereck et al., 2013):

Theorem 1.3.1: Characterization of the single-crossing domain (Bredereck et al., 2013)

A preference profile \mathcal{P} of n preferences over a set \mathcal{C} of m candidates is single-crossing if and only if:

- There do not exist candidates c_{i_1}, \dots, c_{i_6} and voters v_i, v_j, v_k such that:

$$c_{i_2} >_i c_{i_1}, c_{i_1} >_j c_{i_2}, c_{i_1} >_k c_{i_2}$$

$$c_{i_3} >_i c_{i_4}, c_{i_4} >_j c_{i_3}, c_{i_3} >_k c_{i_4}$$

$$c_{i_5} >_i c_{i_6}, c_{i_5} >_j c_{i_6}, c_{i_6} >_k c_{i_5}$$

AND

- There do not exist candidates $c_{i_1}, c_{i_2}, c_{i_3}, c_{i_4}$ and voters v_i, v_j, v_k, v_l such that:

$$c_{i_1} >_i c_{i_2}, c_{i_2} >_j c_{i_1}, c_{i_1} >_k c_{i_2}, c_{i_2} >_l c_{i_1}$$

$$c_{i_3} >_i c_{i_4}, c_{i_3} >_j c_{i_4}, c_{i_4} >_k c_{i_3}, c_{i_4} >_l c_{i_3}$$

Again, the explicit list of forbidden minors can be inferred from the theorem (however, this time, there are 30 forbidden minors).

1.4 1-Euclidean preferences

The last domain restriction we present in this section is the so called *1-Euclidean domain*. This domain was introduced by Coombs (1964) under the name *unidimensional unfolding*. Each candidate and voter is associated to a position on the real line. The position of voter v is also called *the ideal point of v* . The idea here is that the nearer a voter v is to a candidate c , the higher c is ranked in the preference of v . For instance, imagine that you are in Paris, somewhere between Place d'Italie and Denfert-Rochereau. So basically, you walk along the metro line 6 which can be seen here as a straight line. If you want to take the metro (it is a cold, rainy afternoon and you are in a hurry), you will probably prefer to go to the nearest station. If by misfortune this station is closed, you would probably opt for the second nearest station. And so on.⁴ This is exactly the idea behind 1-Euclidean preferences. Let us now give a formal definition:

⁴Hopefully not the whole metro line 6 is closed... which would not surprise me!

Definition 1.4.1: 1-Euclidean preferences

A profile $\mathcal{P} = \{>_1, >_2, \dots, >_n\}$ of n preferences of a set V of n voters over a set \mathcal{C} of m candidates is *1-Euclidean* if there exists a mapping

$$f : V \cup \mathcal{C} \rightarrow \mathbb{R}$$

such that for each voter $v_i \in V$ and each couple of candidates $c_j, c_k \in \mathcal{C}$, we have

$$c_j >_i c_k \text{ iff } |f(v_i) - f(c_j)| < |f(v_i) - f(c_k)|.$$

The following example shows that the profile \mathcal{P} of Example 1.1.1 is 1-Euclidean.

Example 1.4.1. Figure 1.3 gives an example of a 1-Euclidean mapping of \mathcal{P} . For instance, for $>_1 = (c_2, c_3, c_3, c_4, c_5)$, we see that c_2 is the nearest candidate to v_1 . The second nearest candidate to v_1 is c_3 , followed by c_1 (note that the difference between the quantities $|f(c_3) - f(v_1)|$ and $|f(c_1) - f(v_1)|$ is very small). The fourth nearest candidate to v_1 is then c_4 , and the furthest candidate to v_1 is c_5 .

More precisely, we can give the values taken by f on \mathcal{C} and V : we have $f(c_1) = 0, f(c_2) = 5, f(c_3) = 9, f(c_4) = 11, f(c_5) = 12$, and $f(v_1) = 4.9, f(v_4) = 5.1, f(v_2) = 9.4$ and $f(v_3) = 10.9$.⁵ We can check numerically that this mapping satisfies Definition 1.4.1.

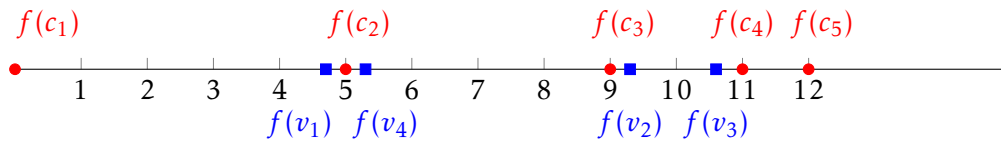


Figure 1.3: A 1-Euclidean mapping f of the profile \mathcal{P} from Example 1.1.1. The positions of candidates are drawn with red circles and the positions of voters are drawn with blue squares.

A Curious Tortoise Intervention

While for single-peaked preferences we are not able to state a preference between a candidate on the left-hand side of the peak and a candidate on the right-hand side of the peak, this is no more true for 1-Euclidean preferences. Contrary to the single-peaked axis which only gives an order on candidates, 1-Euclidean representation provides also information about the distance between candidates. For instance, in the above example, we

⁵Actually, it would be even easier to put $f(v_1) = f(v_5) = f(c_2)$, $f(v_2) = f(c_3)$ and $f(v_3) = f(c_4)$ - indeed, several voters or candidates can share the same position. However, we preferred to have all positions distinct to get a more readable plot of f in Figure 1.3.



have found a mapping f according to which c_3 is closer to c_4 than c_2 is to c_1 . If we go back to the thermostat example once again, the interpretation would be that the impact of changing from 20°C to 22°C is perceived as less important than the impact of changing from 16°C to 18°C .

However, this is only *one* possible mapping that fulfills the definition. Does it remain true for *any* convenient mapping, or can we find a mapping f' according to which c_2 will be closer to c_1 than c_3 to c_4 ? We will come back to this question in Chapter 2.

We have seen that the profile of Example 1.1.1 was also single-peaked and single-crossing. Actually, this is not a simple coincidence. It is well-known that any 1-Euclidean profile is both single-peaked and single-crossing (Grandmont, 1978) (note that the reverse is not true - we will detail this point in Chapter 2). This observation is actually used in all recognition algorithms for 1-Euclidean preferences that have been devised so far. Doignon and Falmagne (1994) combine the information that the profile must be single-peaked and single-crossing. They obtain an order on both candidates and voters which allows them to formulate a linear program to find the exact positions of voters/candidates. Knoblauch (2010) uses only the information that the profile needs to be single-peaked, and she then builds a linear program from a single-peaked order. Finally, Elkind and Faliszewski (2014) use the single-crossing order to formulate their linear program.

So far, no purely combinatorial algorithm (i.e., without using linear programming) has been found. As noticed by Peters (2017), this question seems closely related to whether there exists a *good* characterization of 1-Euclidean preferences. Chen et al. (2017) showed that contrary to single-peaked and single-crossing preferences, 1-Euclidean preferences cannot be characterized by finitely many forbidden minors - in other words, there are infinitely many forbidden minors. We say that a characterization (of a given domain by a, possibly infinite, set of forbidden minors) is *good* if there exists a polynomial-time algorithm to decide whether a given profile contains or not a forbidden minor. These questions highlight the fact that the 1-Euclidean domain is much less understood (from the structural point of view) than both single-peaked and single-crossing domains. We are going to discuss this issue in Chapter 2.

Chapter 2

Relations between single-peaked, single-crossing and 1-Euclidean preferences

2.1 Introduction

The aim of this short chapter is to deepen the state of the art concerning the domain restrictions introduced in Chapter 1. More precisely, we focus here on the relations between single-peaked, single-crossing and 1-Euclidean domains. Indeed, we have seen at the end of Chapter 1 that the 1-Euclidean domain is undoubtedly less understood, from the structural point of view, than the single-peaked or the single-crossing domain. For instance, contrary to both the single-peaked domain and the single-crossing domain, the 1-Euclidean domain cannot be characterized by finitely many forbidden minors (Chen et al., 2017). Moreover, no purely combinatorial (i.e., without using linear programming) recognition algorithm for 1-Euclidean preferences has been found so far.

We recall that it is well-known that any 1-Euclidean preference profile is both single-peaked and single-crossing (Grandmont, 1978). In other words, the 1-Euclidean domain is contained in the domain of profiles that are single-peaked and single-crossing. This domain will be denoted as the SPSC domain in which follows. Understanding the SPSC domain seems essential to better grasp the 1-Euclidean domain.

The SPSC domain is actually quite well understood. Recently, Elkind et al. (2020) gave a characterization of this domain - they proved that a preference profile is SPSC if and only if it can be obtained from a *minimally rich* single-crossing profile by deleting some voters (a profile is *minimally rich* if each candidate is ranked first by at least one voter). They also recall that the SPSC domain can be viewed as a “combinatorial” approximation of the 1-Euclidean domain. Actually, Chen et al. (2017) showed that a preference

profile over a set of at most 5 candidates is 1-Euclidean if and only if it is SPSC. They also provide a counterexample to show that this result no longer holds for $m \geq 6$.

Bredereck et al. (2013) explained how one can generate every possible single-crossing preference profile (up to renaming candidates). In fact, this result follows directly from a well-known connection between single-crossing preferences and the so called *weak Bruhat order* (see Definition 2.2.3 below). Indeed, several works (see for instance the works of Abello (1991) or Galambos and Reiner (2008)) come with a result that can be formulated as follows: in terminology of social choice theory, the problem of deciding whether a given preference profile is single-crossing can be reduced (in polynomial time) to the (shortest) path problem in a given, well-defined graph. This point will be detailed in Section 2.3.2. A direct combination of this result with results given by Elkind et al. (2020) yields an algorithm to generate all possible SPSC profiles (up to renaming candidates).

In this chapter, we reinterpret some of the known results by using a common framework, and we highlight the links between them. Our aim is to provide the reader a more global structural understanding of the relations between single-peaked, single-crossing and 1-Euclidean preferences. Actually, the existing works use several different terminologies of different fields as, for instance, combinatorics, group theory or social choice. Therefore, it is not always particularly easy to get a unified and global state of the art on the topic. The chapter is organised as follows: in Section 2.2, we introduce two main combinatorial structures, namely the *permutohedron* and the *weak Bruhat order*, that are often used to study restricted domains. Indeed, these structures help to grasp various (theoretical) properties of different domain restrictions and to understand the connexion between them. For the purposes of this chapter, we do not need to go deeply into the theory of permutohedra, or weak Bruhat orders. Instead, we will provide a quite high-level presentation of these notions, by focusing on visualisation and intuition issues.

We discuss then some known results on the SPSC domain in Section 2.3, and we explain how they can be reinterpreted using the common framework established in Section 2.2. Furthermore, we implemented the algorithm described by Bredereck et al. (2013) to enumerate all SPSC profiles up to 8 candidates. Using linear programming, we could then decide for each of these profiles whether or not it was 1-Euclidean. We provide the results of this numerical analysis to give some elements of answer to the question of how well the SPSC domain approximates the 1-Euclidean domain.

2.2 Combinatorial structures behind restricted domains

In this section, we introduce several notions that will be central in the remainder of the chapter. Let us start by introducing the notion of a *maximal preference profile*. In which follows, we consider that all preferences of a given profile are pairwise distinct. With

this assumption, we can define a maximal single-peaked (resp. single-crossing, SPSC, 1-Euclidean) profile as follows:

Definition 2.2.1: Maximal profile

A single-peaked (resp. single-crossing, SPSC, 1-Euclidean) profile is *maximal* if for any preference $\succ_i \notin \mathcal{P}$, the profile $\mathcal{P} \cup \{\succ_i\}$ is not single-peaked (resp. single-crossing, SPSC, 1-Euclidean) anymore.

2.2.1 Permutohedron of order m

As stated above, a common framework is set in this chapter in order to provide an overview of existing results. For this purpose, we use the notion of *permutohedron* (see Definition 2.2.2 below). This idea is not novel - the permutohedron is a classical structure in combinatorics and group theory. To the best of our knowledge, it was first used by Guilbaud and Rosenstiehl (1963) in the social choice context, and became quickly a useful¹ tool to study structural aspects of domain restrictions - for instance, see the excellent survey of Monjardet (2009) providing a detailed overview on the topic, or the more recent (and no less excellent!) work of Puppe and Slinko (2019) studying Condorcet domains.

Before introducing the notion of permutohedron, we need to define the notion of *swap*: given a permutation $\pi = (i_1, i_2, \dots, i_m)$ of the set $\{1, 2, \dots, m\}$, the operation of swap between two elements of π consists in exchanging their positions in π . For instance, applying the operation $swap(1, 4)$ on the permutation $(1, 2, 3, 4, 5)$ yields the permutation $(4, 2, 3, 1, 5)$. In the remainder of the chapter, only the swaps of adjacent elements will be allowed. For instance, given the permutation $(1, 2, 3, 4, 5)$, the only possible swaps are $swap(1, 2)$, $swap(2, 3)$, $swap(3, 4)$ and $swap(4, 5)$.

We are now ready to introduce the notion of permutohedron:

Definition 2.2.2: Permutohedron of order m

Let m be an integer. The permutohedron of order m is a graph $P_m = (V_m, E_m)$ where the set of vertices corresponds to the set of all permutations of the set $\{1, 2, \dots, m\}$ (in particular, $|V_m| = m!$), and we have $\{u, v\} \in E_m$ if and only if the permutation v can be obtained from the permutation u by making a unique swap of *adjacent* elements of u .

Example 2.2.1. We give here the permutohedron of order 3 and of order 4. For more readability, the vertices of P_4 are not labeled - the labels can be found by assuming, without loss of generality, that the *circle vertex* corresponds to the permutation $(1, 2, 3, 4)$.

¹but, in my humble opinion, perhaps not used enough by the social choice community.

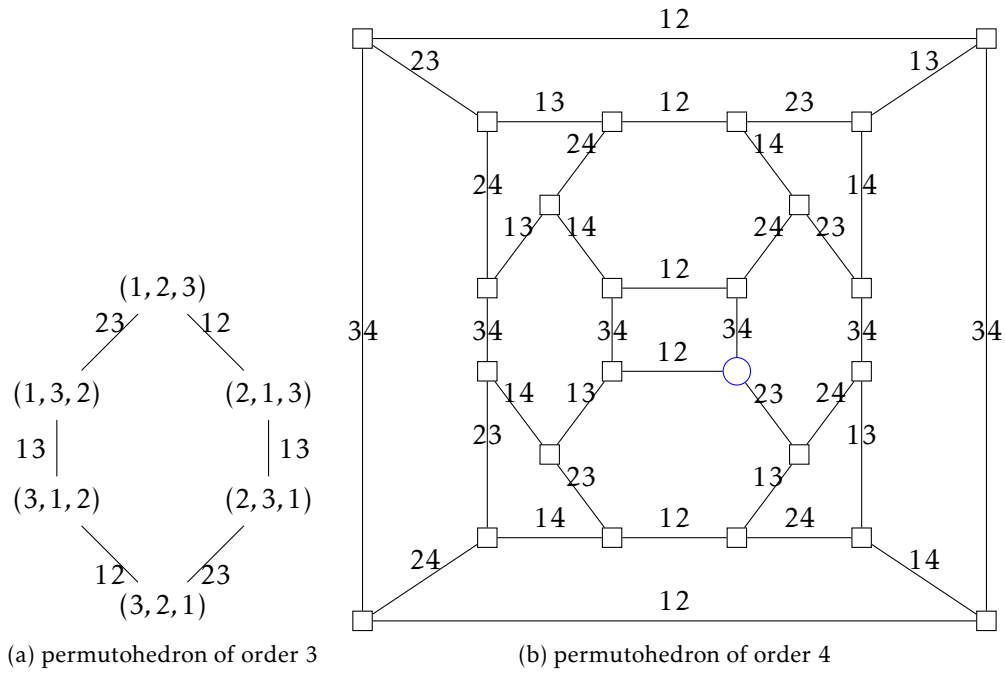



Figure 2.1: The permutohedra of order 3 and 4. The circle vertex corresponds to the permutation $(1, 2, 3, 4)$. The label ij on edge $\{u, v\}$ indicates that i and j are swapped between the permutations associated to the vertices u and v .

Each vertex of P_m can be identified with one possible preference over the set of candidates $\{c_1, c_2, \dots, c_m\}$, by simply associating the vertex corresponding to a permutation (i_1, i_2, \dots, i_m) to the preference $(c_{i_1}, c_{i_2}, \dots, c_{i_m})$. For conciseness and by abuse of language, we will use the terms *preference* $>_i$ and *vertex associated to preference* $>_i$ indifferently. Note also that given two preferences $>_i, >_j$, the Kendall-tau distance $d_{KT}(>_i, >_j)$ corresponds to the length of any shortest path between the vertices $>_i$ and $>_j$.

The single-peaked domain and the permutohedron Obviously, as P_m contains all possible permutations, each preference profile forms a subset of vertices of V_m . More precisely, each preference profile $\mathcal{P} = (V, E)$ can be seen as a subgraph of P_m induced by V . In particular, each domain restriction can be viewed as a set of subgraphs of P_m .

A Curious Tortoise Intervention



It should be noted, however, that other graphs have been proposed in the literature to represent a preference profile \mathcal{P} . For instance, Puppe and Slinko (2019) use the following definition: given two preferences $>, >'$ of \mathcal{P} , and denoting by $[>, >']$ the set of preferences that are between $>, >'$ (i.e., the set of preferences $>''$ such that $c_i >'' c_j$ whenever we have both $c_i > c_j$



and $c_i \succ' c_j$), they say that \succ and \succ' are *neighbours* if $[\succ, \succ'] \cap V = \{\succ, \succ'\}$, (i.e., there is no preference between \succ and \succ' in \mathcal{P}).

They then visualise \mathcal{P} by a graph $\Gamma_{\mathcal{P}} = (V, E_{\mathcal{P}})$ such that there is an edge $\{\succ, \succ'\}$ in $E_{\mathcal{P}}$ iff \succ and \succ' are neighbours in \mathcal{P} .

With this definition, it is obvious that the graph corresponding to \mathcal{P} may not be a subgraph of P_m , as there can be some edges in $E_{\mathcal{P}}$ that are not in E_m .

We denote by SP_m the set of subgraphs of P_m corresponding to all *maximal* single-peaked profiles. Indeed, the single-peaked domain can be characterized by the set of maximal single-peaked profiles, as any single-peaked profile is a subprofile of a maximal single-peaked profile. In terms of subgraphs of P_m , any (not necessarily maximal) single-peaked profile is then a subgraph of an element of SP_m .

Given an axis A , we denote by $SP_m(A)$ the subgraph of P_m corresponding to the (unique) maximal profile single-peaked with respect to A . If we denote by $\mathcal{A}_{\mathcal{C}}$ the set of all axes on the set of candidates \mathcal{C} , we have

$$SP_m = \{S_m(A) : A \in \mathcal{A}_{\mathcal{C}}\}$$

It is worth noting that, given two different axes A and A' , $SP_m(A')$ can be obtained from $SP_m(A)$ by renaming candidates. In other words, all elements of SP_m are the same, up to renaming of candidates. For this reason, we consider in which follows the axis $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$, and the graph $SP_m(A)$ corresponding to the maximal profile single-peaked w.r.t. A .

It is well-known that there are 2^{m-1} pairwise distinct preferences compatible with a given axis - therefore, $SP_m(A)$ has 2^{m-1} vertices. We also recall that any maximal single-peaked profile contains a unique pair of reverse preferences, corresponding to the order A and its reversal. These preferences will be denoted by \succ_A and $\succ_{\bar{A}}$ in which follows.

A straightforward way to build $SP_m(A)$ is to simply enumerate all preferences single-peaked with respect to A , which yields the set $V_m(A)$ of vertices. $SP_m(A)$ is then the subgraph of P_m induced by $V_m(A)$. Note that the single-peaked domain viewed as a subset of P_m (or a set of subsets of P_m) was studied by Nembua et al. (1989). For an illustration, we give a visualization of $SP_5(A)$ in Figure 2.2, and of $SP_6(A)$ and $SP_7(A)$ in Appendix 2.A.

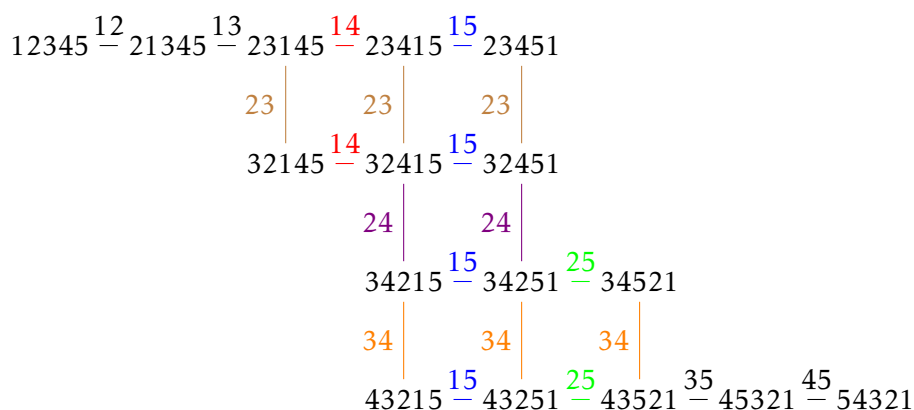


Figure 2.2: The visualisation of $SP_5(A)$

2.2.2 Distributive lattice and weak Bruhat order

Another mathematical structure that will be useful in the remainder of this chapter is the so called *weak Bruhat order*. Let us recall that a *partial (strict) order* is a transitive antisymmetric binary relation $<$ defined on the set S . In other words, for $x, y \in S$, we have either $x < y$, or $y < x$, or x and y are *incomparable*. The transitivity ensures that $x < y$ and $y < z$ imply $x < z$.

A *partially ordered set* (poset), denoted by $(S, <)$ is a set with a partial order. Let us consider two elements x and y of S such that $x < y$. x is called a *predecessor* of y and y is called a *successor* of x . An element of S without predecessors is called a *minimal element*. An element of S without successors is called a *maximal element*.

We will now introduce a partial order known as *weak Bruhat order*. It is an order defined on the set of all permutations of the set $\{1, 2, \dots, m\}$, denoted here by V_m (to refer to the set of vertices of P_m , which is nothing but the set of all permutations of the set $\{1, 2, \dots, m\}$). Let us denote by π_0 the identity permutation $(1, 2, \dots, m)$. The weak Bruhat order is then defined as follows:

Definition 2.2.3: The weak Bruhat order

We say that a permutation π_1 is between permutations π_2 and π_3 if for each pair $\{i, j\} \subset \{1, \dots, m\}$, if i is ranked better (resp. worse) than j in both π_2 and π_3 , it is also ranked better (resp. worse) than j in π_1 .

Let us denote by π_0 the identity permutation $(1, 2, \dots, m)$. The weak Bruhat order $<$ on the set V_m of all permutations of the set $\{1, 2, \dots, m\}$ is defined as follows: For each $\pi, \pi' \in V_m$, we have $\pi < \pi'$ if and only if π is between π_0 and π' .

This weak order can be seen as an orientation of the graph P_m : actually, an edge $\{>, >'\}$

is converted into an arc (\succ, \succ') if \succ lies on a shortest path between \succ_0 and \succ' in P_m (we recall that $\succ_0 = (1, 2, \dots, m)$). This high-level interpretation of the weak Bruhat order will be used in the remainder of this chapter; however, note that one can go further: Guilbaud and Rosenstiehl (1963) showed that the partially ordered set $(V_m, <)$ is a *distributive lattice*. Even though we do not need this information in which follows, for completeness, Mathilde will now present you the notion of *distributive lattice*.

A Curious Tortoise Intervention



A partially ordered set $(S, <)$ is a *lattice* if each pair of elements $\{x, y\} \subset S$ has a unique supremum (also called *join*), denoted by $x \vee y$, and a unique infimum (also called *meet*), denoted by $x \wedge y$.

Formally, for $\{x, y\} \subset S$, we define the join (\vee) and the meet (\wedge) binary operations on S as follows:

$$\begin{aligned} x \vee y &:= z \in S \text{ such that } x < z, y < z \\ &\text{and } z < z' \text{ for each } z' \in S \text{ such that } x < z', y < z'; \\ x \wedge y &:= z \in S \text{ such that } z < x, z < y \\ &\text{and } z' < z \text{ for each } z' \in S \text{ such that } z' < x, z' < y. \end{aligned}$$

With these notions, $(S, <)$ is a lattice if for each pair of elements $\{x, y\} \subset S$, $x \vee y$ and $x \wedge y$ exist. Note that in such a case, the values $x \vee y$ and $x \wedge y$ are unique. In particular, each lattice has a unique minimal element and a unique maximal element.

Finally, a partially ordered set $(S, <)$ is a *distributive lattice* if for each $x, y, z \in S$, we have:

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$$

We can check that the set V_m with the weak Bruhat order is actually a distributive lattice. To this purpose, let us denote by s the application which associates to each preference π the set of consecutive swaps that need to be done to transform π_0 into π . For instance, $s((2, 3, 1)) = \{\{1, 2\}, \{1, 3\}\}$. We can show that s is well-defined: even if there exist several “ways” to transform π_0 into π , they all yield the same set of consecutive swaps $s(\pi)$. Moreover, s is injective. If we denote by $im(s)$ the image of the application s , we can thus define an inverse application s^{-1} from $im(s)$ to V_m that associates to a given a set of consecutive swaps from $im(s)$ the permutation obtained when applying these consecutive swaps on



π_0 . For instance, we have $s^{-1}(\{1,2\}) = (2,1,3)$, as $(2,1,3)$ is obtained from $\pi_0 = (1,2,3)$ by performing the swap $\{1,2\}$ (we note that 1 and 2 are consecutive in $(1,2,3)$, so this swap is possible). Moreover, $s^{-1}(\{1,3\})$ is not defined, as $\{1,3\}$ is not in $im(s)$; indeed, there is no permutation that can be obtained from $(1,2,3)$ by performing the consecutive swap $(1,3)$, as 1 and 3 are not consecutive in $(1,2,3)$.

With these notions, we can check that following operations make from $(V_m, <)$ a distributive lattice:

$$\pi \vee \pi' = s^{-1}(s(\pi) \cup s(\pi'))$$

and

$$\pi \wedge \pi' = s^{-1}(s(\pi) \cap s(\pi')).$$

It is well known that the weak Bruhat order has the permutation $\pi_0 = (1, 2, \dots, m)$ as a (unique) minimal element and the permutation $\bar{\pi}_0 = (m, m-1, \dots, 1)$ as a (unique) maximal element. A list of elements $(\pi_1, \pi_2, \dots, \pi_k)$ is a *chain* of weak Bruhat order if for each $i < k$, $\pi_i < \pi_{i+1}$. It is well-known that each maximal chain in the weak Bruhat order (i.e., each longest path from π_0 to $\bar{\pi}_0$ in the graph of the weak Bruhat order) contains $\frac{m(m-1)}{2} + 1$ elements.

Let us finish this section with a concrete example of weak Bruhat order for $m = 3$. Its illustration is given in Figure 2.3a. Actually, it can be seen as an oriented graph on the set of vertices V_m with an arc from π to π' iff $\pi < \pi'$. We note that if we remove the transitivity arcs in such a graph, we obtain an oriented version of the graph P_m , where an edge $\{\pi, \pi'\}$ is converted into an arc $\{\pi, \pi'\}$ if π lies on a shortest path from π_0 to π' , as shown in Figure 2.3b.

The minimum element is $(1, 2, 3)$ and the maximum element is $(3, 2, 1)$. We have, for instance, $(2, 1, 3) < (2, 3, 1)$, as $(2, 1, 3)$ lies on a shortest path from $(1, 2, 3)$ to $(2, 3, 1)$. On the other hand, $(2, 1, 3)$ and $(3, 1, 2)$ are not comparable, as $(2, 1, 3)$ does not lie on any shortest path between $(1, 2, 3)$ and $(3, 1, 2)$, and vice versa. Finally, there are two maximal chains in the weak Bruhat order for $m = 3$, both of the length 3. These chains correspond to the longest paths (from π_0 to $\bar{\pi}_0$) in the graph of weak Bruhat order (see Figure 2.3a), or to the paths from π_0 to $\bar{\pi}_0$ in the oriented version of P_m (see Figure 2.3b).

2.3 The SPSC domain: state of art

After this long introduction, let us finally take a closer look at the SPSC domain. As said at the beginning of this chapter, one of the interests of the SPSC domain is that

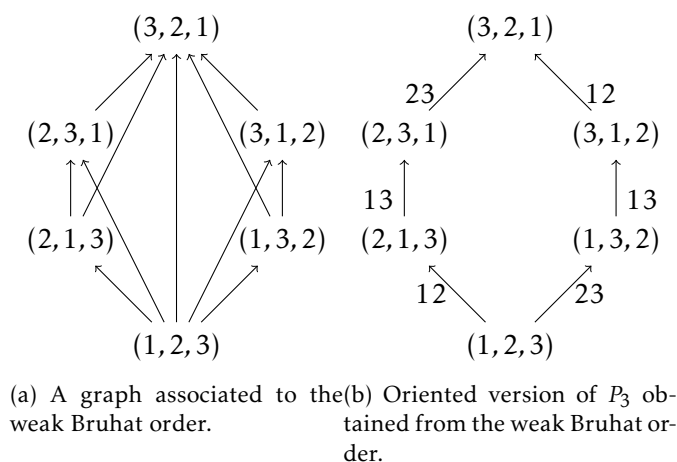


Figure 2.3: The graph on the left visualizes the weak Bruhat order for $m = 3$. There is an arc from π to π' iff $\pi < \pi'$. On the right side, we have the orientation of the graph P_3 obtained from the weak Bruhat order: there is a path from π to π' iff $\pi < \pi'$. In other words, the graph on the left is the transitive closure of the graph on the right.

it contains the 1-Euclidean domain. We recall here some important known results. To provide a more unified and visual overview, we reinterpret these results by using the notions of permutohedron and weak Bruhat orders.

2.3.1 Characterization of the SPSC domain and permutohedron

The SPSC domain is the intersection of the single-peaked domain and the single-crossing domain. As both these domains can be characterized by finitely many forbidden minors, this is thus also the case for the SPSC domain. However, as emphasized by Elkind et al. (2020), such a characterization is not particularly intuitive. Instead, Elkind et al. (2020) come with the following result:²

Proposition 2.3.1: Characterization of the SPSC domain (Elkind et al., 2020)

A preference profile is both single-peaked and single-crossing if and only if it can be obtained from a minimally rich single-crossing profile by deleting some voters.

Actually, Puppe (2018) (and implicitly Barberà and Moreno (2011)) already observed the right-left implication, i.e., that a minimally rich single-crossing profile was single-peaked with respect to the axis corresponding to the first preference in the single-crossing order. Mathilde is here to provide more insights on this observation.

²We recall that a preference profile is *minimally rich* if each candidate is ranked first by at least one of the voters.

A Curious Tortoise Intervention



To provide more insights on this latter observation, let us quickly recall the proof given by Elkind et al. (2020). The observation is proven by contradiction - we suppose for a contradiction that there is a minimally rich single-crossing profile that is not single-peaked.

Formally, consider a single-crossing preference profile \mathcal{P} where $>_A = (c_1, c_2, \dots, c_m)$ is the first preference in the single-crossing order. Assume (for contradiction) that:

1. \mathcal{P} is not single-peaked w.r.t. $A = c_1 \triangleleft_A c_2 \triangleleft_A \dots \triangleleft_A c_m$,
2. \mathcal{P} is minimally rich.

Put another way:

1. There exists a preference $>_v$ and indices $i < j < k$ (i.e., $c_i \triangleleft_A c_j \triangleleft_A c_k$) such that $c_i >_v c_j$ and $c_k >_v c_j$ (see Arrow's definition of single-peakedness - Definition 1.2.2).
2. There exists a preference $>_{v'}$ such that $c_j >_{v'} c_i$ and $c_j >_{v'} c_k$.

The pair of candidates $\{c_i, c_j\}$ implies that $>_v$ is between $>_A$ and $>_{v'}$ in any single-crossing order in which $>_A$ is ranked first: indeed, we have $c_i >_A c_j, c_i >_v c_j$, but $c_j >_{v'} c_i$.

On the other hand, the pair of candidates $\{c_j, c_k\}$ requires that $>_{v'}$ is between $>_A$ and $>_v$. It is a contradiction because the two betweenness relations cannot hold simultaneously. Consequently, any minimally rich single-crossing preference profile is single-peaked.

Obviously, the converse of this observation is not true - any SPSC profile remains SPSC if we remove any subset of its preferences. In particular, we can remove all preferences ranking a given candidate c first, to ensure that the profile is not minimally rich. So in other words, there are SPSC profiles that are not minimally rich.

However, Elkind et al. (2020) provide a constructive algorithm which allows to extend any SPSC profile to a minimally rich single-crossing profile. Their algorithm can be decomposed into two main steps:

1. Given a SPSC preference profile \mathcal{P} over the set $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ of candidates, we check whether the profile is single-peaked with respect to the first preference of the single-crossing order. If so, there is nothing to do in Step 1. Otherwise, Elkind et al. (2020) prove that there exists actually an axis A with respect to which the profile $\mathcal{P} \cup \{>_A\}$ remains SPSC. In other words, we can always ensure, by

adding a new preference to \mathcal{P} if necessary, that \mathcal{P} contains the preference $>_A$ corresponding to an axis A with respect to which \mathcal{P} is single-peaked. Obviously, $>_A$ will become the first preference in the single-crossing order.

A Curious Tortoise Intervention



Even though this step might seem as a “technical” step useful for the proof, it is actually more important than one could think at first sight: if a profile is single-peaked with respect to an axis A , the only preference which ranks the leftmost candidate of A first is $>_A$. Therefore, $>_A$ must necessarily be in the profile if we want it to be minimally rich. By the way, the same holds for $>_{\bar{A}}$.

2. In the second step, Elkind et al. (2020) explain how to add preferences to \mathcal{P} in order to make it minimally rich (without loosing the single-crossing property). Actually, they take the smallest $i \in \{1, \dots, m\}$ such that c_i is never ranked first in \mathcal{P} . We know that $i \neq 1$, as $>_A \in \mathcal{P}$.³ Let us denote by $>_{i-1}$ the last preference in the single-crossing order ranking c_{i-1} first. It is written $(c_{i-1}, \dots, c_{i-k}, c_i, \dots)$. Elkind et al. (2020) prove that the preference $>_i: (c_i, c_{i-1}, \dots, c_{i-k}, \dots)$ can be added to \mathcal{P} without loosing the SPSC property - note that $>_i$ is inserted just after $>_{i-1}$ in the single-crossing order.

Elkind et al. (2020) thus provide a polynomial time algorithm for constructing a minimally rich single-crossing profile from which a given SPSC profile can be obtained. In particular, a direct corollary of their result is that any maximal SPSC profile contains exactly $\frac{m(m-1)}{2} + 1$ pairwise distinct preferences and is minimally rich.

Note that these results can be visualised/reinterpreted by using the permutohedron P_m . Actually, the first step of the previous algorithm consists in identifying an axis A such that a given SPSC preference profile \mathcal{P} is a subgraph of $SP_m(A)$. For each preference $>$ of \mathcal{P} , we can then compute its Kendall tau distance from $>_A$, denoted by $d_{KT}(>_A, >)$. It is easy to see that a profile is SPSC if and only if these distances are pairwise distinct - in such a case, we obtain the single-crossing order by ordering the preferences increasingly with respect to the Kendall tau distances.

With this observation, any maximal SPSC profiles corresponds to a shortest path in $SP_m(A)$ between $>_A$ and $>_{\bar{A}}$ - we note that each such path contains exactly $\frac{m(m-1)}{2} + 1$ vertices, as $d_{KT}(>_A, >_{\bar{A}}) = \frac{m(m-1)}{2}$. In particular, given two preferences $>_i$ and $>_j$ that are consecutive in the single-crossing order of \mathcal{P} , there exists a (not necessarily unique) path between $>_i$ and $>_j$. Any vertex $>_k$ lying on this path can be added to \mathcal{P} without loosing the single-crossing property (note that $>_k$ is inserted between $>_i$ and $>_j$ in the single-crossing order). This idea is actually behind several lemmas used by Elkind et al. (2020) to prove the algorithm above.

³We recall that, up to renaming of candidates, we can suppose that $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$.

Example 2.3.1. Let us illustrate the ideas described above on the following profile \mathcal{P} of 3 preferences over 5 candidates:

$$\begin{aligned} >_1: (c_1, c_2, c_3, c_4, c_5) \\ >_2: (c_2, c_3, c_1, c_4, c_5) \\ >_3: (c_3, c_2, c_4, c_1, c_5) \end{aligned}$$

The profile is presented in a single-crossing order. It is also easy to check that \mathcal{P} is single-peaked with respect to the first preference of this order, i.e., the preference $>_1$. So there is nothing to do in Step 1.

A Curious Tortoise Intervention



What a cheat!

In particular, \mathcal{P} is contained in a graph $SP_5(A)$, with $A = c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4 \triangleleft c_5$, as shown in Figure 2.4 (the preferences of \mathcal{P} are framed).

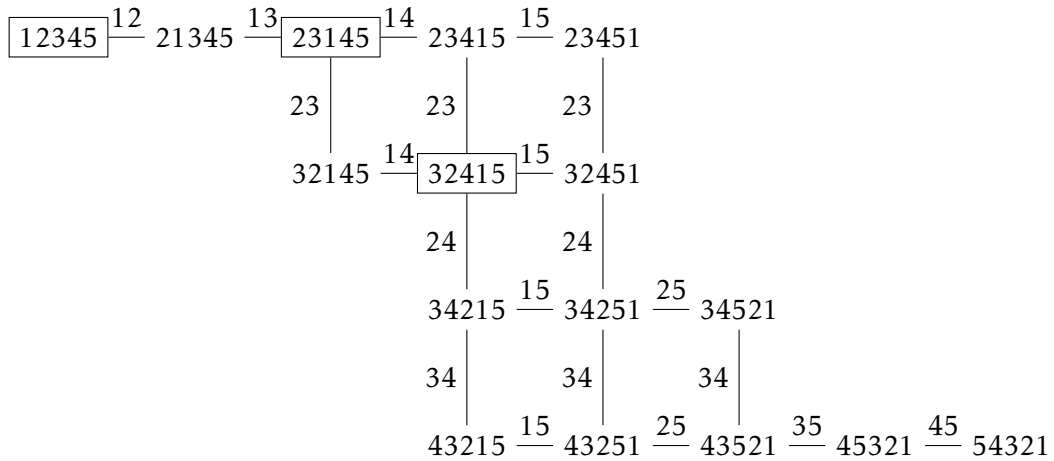


Figure 2.4

We have seen that any maximal SPSC profile is minimally rich. Moreover, any maximal SPSC profile on 5 candidates (and single-peaked with respect to A) corresponds to a shortest path in $SP_5(A)$ between $>_A$ and $>_{\bar{A}}$. In this example, we can therefore start by adding $>_{\bar{A}}: (c_5, c_4, c_3, c_2, c_1)$ to \mathcal{P} . There is a unique shortest path between $>_1$ and $>_2$, two paths between $>_2$ and $>_3$, and five paths between $>_3$ and $>_{\bar{A}}$. There are hence $1 \times 2 \times 5 = 10$ maximal SPSC profiles containing \mathcal{P} .

In practice, Elkind et al. (2020) only look for a minimally rich single-crossing profile, not a maximal minimally rich single-crossing profile. In this example, it is hence sufficient to

find a preference \succ_4 of most preferred candidate c_4 , so that $\mathcal{P} \cup \{\succ_4\} \cup \{\succ_{\bar{A}}\}$ is minimally rich. Obviously, any convenient preference \succ_4 must lie on a shortest path between \succ_3 and $\succ_{\bar{A}}$. But in practice, we do not even need to construct a path from \succ_3 to $\succ_{\bar{A}}$ - as discussed above, Elkind et al. (2020) explain how to construct \succ_4 : in this example, \succ_3 is the last preference of \mathcal{P} ranking c_3 first. It is written $(c_3, c_2, c_4, c_1, c_5)$. The result of Elkind et al. (2020) states that $\succ_4: (c_4, c_3, c_2, c_1, c_5)$ can be added to \mathcal{P} without losing the SPSC property. Note that \succ_4 is obtained from \succ_3 by simply placing c_4 to the first position of \succ_3 , i.e., by performing a swap between c_4 and c_2 and then a swap between c_4 and c_3 . Indeed, we see that all edges corresponding to these swaps come “after” the node corresponding to \succ_3 in $SP_5(A)$. And this is not a coincidence: one can actually rediscover the result of Elkind et al. (2020) by generalizing this last observation to $SP_m(A)$.

A Curious Tortoise Intervention



Note that although the approach using the permutohedron has an advantage to be visual, it is not to be used in practice. Indeed, its implementation requires the construction of the graph $SP_m(A)$ which has 2^{m-1} vertices! Step 2 of algorithm of Elkind et al. (2020) allows to find a minimally rich single-crossing profile containing \mathcal{P} in polynomial time, without the construction of $SP_m(A)$.

2.3.2 The SPSC domain and weak Bruhat orders

In the previous section, we have introduced the so called weak Bruhat orders. Actually, there is a well-known connection between weak Bruhat orders and the single-crossing domain. More precisely, there is a bijection between the maximum chains in the weak Bruhat order $(\sigma_m, <)$ and the maximum single-crossing profiles of preferences over the set of candidates $\{c_1, c_2, \dots, c_m\}$ containing the unique pair of reverse preferences (c_1, c_2, \dots, c_m) and $(c_m, c_{m-1}, \dots, c_1)$ - see for instance Abello (1991) or Galambos and Reiner (2008).⁴ As already pointed out by Bredereck et al. (2013), one can thus enumerate all (maximum) single-crossing profiles (up to renaming of candidates).

Example 2.3.2. Let us go back to the example of the weak Bruhat order for $m = 3$ (see Figure 2.3). It contains two maximum chains O_1, O_2 :

$$O_1 = ((1, 2, 3), (2, 1, 3), (2, 3, 1), (3, 2, 1)) \quad (2.1)$$

$$O_2 = ((1, 2, 3), (1, 3, 2), (3, 1, 2), (3, 2, 1)) \quad (2.2)$$

The corresponding maximal single-crossing profiles over the set of 3 candidates are:

$$\mathcal{P}_1 = \{(c_1, c_2, c_3), (c_2, c_1, c_3), (c_2, c_3, c_1), (c_3, c_2, c_1)\} \quad (2.3)$$

$$\mathcal{P}_2 = \{(c_1, c_2, c_3), (c_1, c_3, c_2), (c_3, c_1, c_2), (c_3, c_2, c_1)\} \quad (2.4)$$

Note that the profile \mathcal{P}_1 is minimally rich, so by the characterization of Elkind et al. (2020), it is single-peaked with respect to the axis $A = c_1 \triangleleft c_2 \triangleleft c_3$. The profile \mathcal{P}_2 is not minimally

⁴Note that this result is also mentioned, for instance, by Doignon and Falmagne (1994)

rich, and it is maximum, so it cannot be single-peaked, as any SPSC profile can be completed to a minimally rich single-crossing profile. Note indeed that \mathcal{P}_2 cannot be single-peaked, as there are more than two candidates ranked last at least once in it.

From here, it is just one step from enumerating all (maximum) SPSC profiles: one can enumerate all maximum chains in the weak Bruhat order, and keep only those chains that correspond to a minimally rich preference profile. In practice, these chains of the weak Bruhat order all correspond to paths (if we remove the orientation) in $SP_m(A)$. Therefore, enumerating all maximum SPSC profiles (up to renaming of candidates) is equivalent to listing all shortest paths from $>_A$ to $>_{\bar{A}}$ in $SP_m(A)$, as we have already seen at the end of Section 2.3.1.

We implemented an algorithm to put this idea into practice. In particular, we could enumerate all maximum SPSC profiles (up to renaming of candidates) up to 8 candidates. We have then determined (with the help of linear programming) how many of these SPSC profiles were actually 1-Euclidean. The results are summarised in Table 2.1.

m	4	5	6	7	8
SPSC	2	12	286	33 592	23 178 480
Euclidean	2	12	168 (59 %)	4680 (14%)	229 468 (1%)

Table 2.1: The number of maximum SPSC, resp. 1-Euclidean, profiles on m candidates (w.r.t. a given axis A).

According to these results, the SPSC domain seems to approximate the 1-Euclidean domain quite badly. A natural question is whether we can characterize the paths in $SP_m(A)$ leading to a 1-Euclidean profile. Note that solving this problem is equivalent to answering the open question of Chen et al. (2017) of whether or not there exists a good characterization of the 1-Euclidean domain. We provide some insights into this problem in the general conclusion.

Chapter appendix

2.A Graphs $SP_6(A)$ and $SP_7(A)$

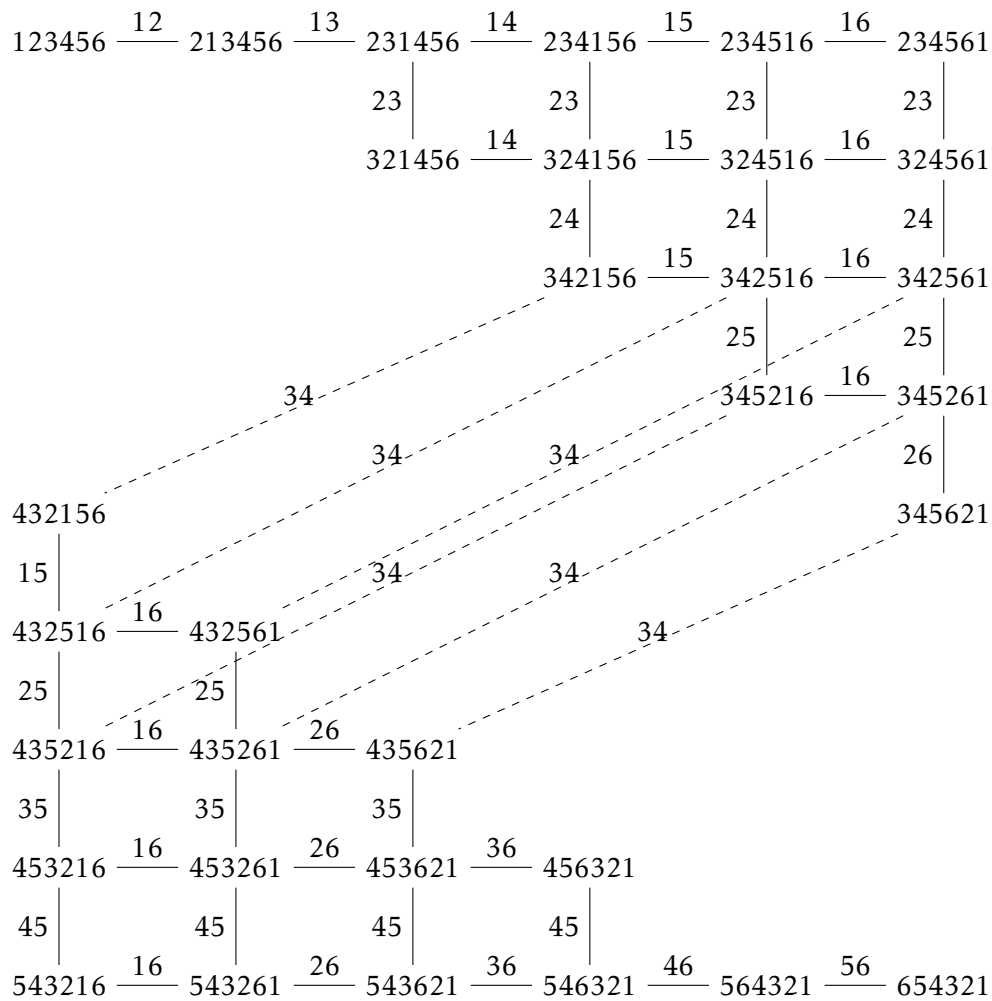


Figure 2.A.1: The visualisation of $SP_6(A)$. For more readability, the lower half of the graph is shifted to the left.

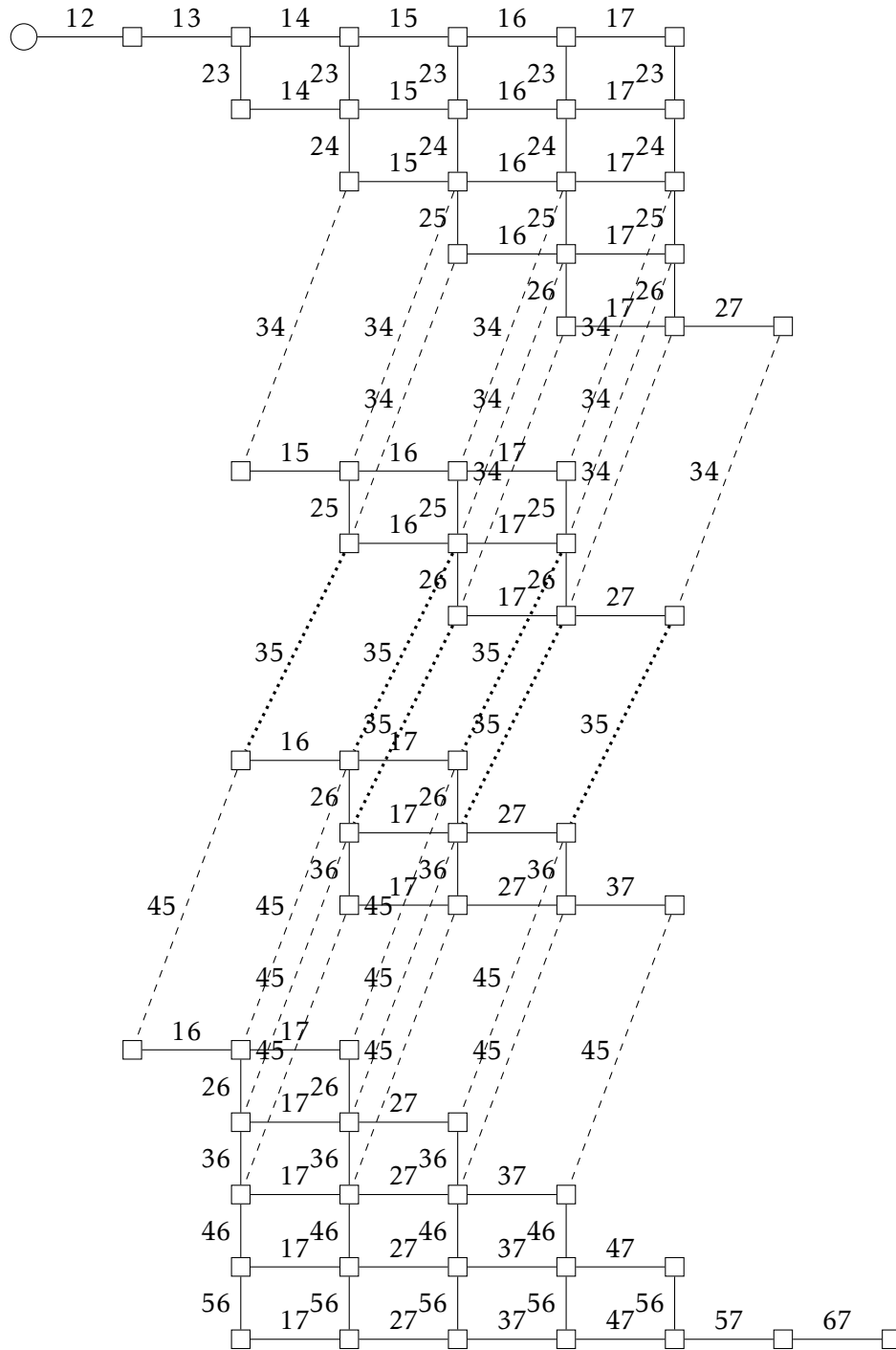


Figure 2.A.2: The visualisation of $SP_7(A)$. For more readability, the labels of nodes (corresponding to all possible preferences on the set of candidates $\{c_1, c_2, c_3, c_4, c_5, c_6, c_7\}$) are not displayed. The most top-left (circle) node corresponds to the preference (c_1, c_2, \dots, c_7)



Extensions of single-peaked preferences

Preamble of Part II

We have introduced in Chapter 1 the structure of single-peaked preferences as well as its theoretical and computational advantages. On the other hand, we have seen that single-peaked preferences were very unlikely to appear in real-world applications. Therefore, we would like to make the structure less restrictive while preserving as much of its good properties as possible. As we will see, unfortunately, the majority of these properties are lost if we relax the notion of single-peakedness. Nevertheless, it is still interesting to, given a preference profile, understand its underlying structure, in particular for data analysis purposes. This can be useful, for instance, for a recommendation system or voting advice applications, as already discussed in Chapter 1.

In this part, we propose two different approaches of structural relaxation:

1. In Chapter 3, it is the structure of the axis itself that is relaxed. A single-peaked axis is seen as a graph (more precisely a path), and we introduce the notion of the single-peakedness with respect to an arbitrary graph as follows: a preference profile is single-peaked with respect to a given graph \mathcal{G} if each of its preferences is a traversal of \mathcal{G} .

Trivially, any profile is single-peaked with respect to the complete graph. However, such a result is useless for data analysis purposes. Indeed, the sparser the graph is, the more structural information it provides. Two notions of sparsity are introduced in Chapter 3, and we look for solutions minimizing (one of) them.

2. In Chapter 4, we keep the axis structure, but we allow the preference to slightly deviate from it if necessary. Different distance measures have been proposed in the literature in order to quantify the distance of a profile from the an axis. Typically, we are looking for an axis which is the nearest possible to the profile in sense of a given distance measure. Several distance measures of (nearly) single-peakedness exist in literature. We propose here a new distance measure, we compare it with existing measures and study its theoretical and computational properties.

Chapter 3

Single-peaked preferences on an arbitrary graph

3.1 Introduction

As already said at the beginning of the part, this chapter is devoted to the extension of the notion of single-peakedness on an arbitrary graph. Indeed, a single-peaked axis can be seen as a graph (or, more precisely, a path). If a preference is not compatible with an axis, we allow to add more edges to the graph in order to make the preference compatible with it.

This idea is not completely new - several classes of graphs have been considered to generalize the classical single-peakedness with respect to an axis, i.e., a path. Given a set $\mathcal{C} = \{c_1, \dots, c_m\}$ of candidates, a preference order $>$ over \mathcal{C} is single-peaked on an undirected graph $\mathcal{G} = (\mathcal{C}, \mathcal{E})$ if it is a *traversal* of \mathcal{G} , i.e., for each $c_j \in \mathcal{C}$ the upper-contour set $\{c_i \in \mathcal{C} : c_i > c_j\}$ is connected. A preference profile is then single-peaked on \mathcal{G} if every preference is single-peaked on \mathcal{G} . Demange (1982) studied single-peakedness on a tree, Peters and Lackner (2017) on a circle.

Some good axiomatic properties remain valid when preferences are single-peaked on a tree: if the number of voters is odd, such profiles still admit a Condorcet winner (a candidate who is preferred over each other candidate by a majority of voters) (Demange, 1982), and returning this Condorcet winner is a strategyproof voting rule. On the contrary, every majority relation can be realized by a collection of preferences single-peaked on a circle (Peters and Lackner, 2017), hence single-peaked preferences on a circle do not inherit the good axiomatic properties of single-peakedness on an axis regarding voting rules that are based on the majority relation. In particular, the class of trees is somehow a maximal class of graphs keeping the most of good axiomatic properties - whenever a graph contains a cycle, there is no more the guarantee of Condorcet winner.

Some algorithmic advantages of single-peakedness on an axis also remain valid for different generalizations of single-peakedness, while others vanish. For instance, the winner determination problem for the Dodgson rule and the Young rule remain easy for single-peaked preferences on a tree with an odd number of voters, by Condorcet consistency (if a Condorcet winner exists, then it is the winner of the election for both rules), while determining an aggregate ranking with the Kemeny rule becomes NP-hard. Computing the winners for the Young rule is also easy if the preferences are single-peaked on a circle (Peters and Lackner, 2017). For single-peaked electorates on a tree, diverse variants of the proportional representation problem are in class FPT with respect to the number of internal vertices of the tree (Peters and Elkind, 2016), and in class XP with respect to the number of leaves (Yu et al., 2013). For single-peaked electorates on a circle, they are in class P because the corresponding integer linear program has a totally unimodular constraint matrix in this case, and therefore the integrality constraints can be relaxed (Peters, 2018; Peters and Lackner, 2017).

Our goal is to study the recognition problem for single-peaked preferences on arbitrary connected graphs. Although one cannot expect social choice theoretic guarantees from single-peakedness on arbitrary graphs (as it does not result in a domain restriction, any profile being single-peaked on the complete graph), a sparse graph gives some insights on the relations between candidates/items. This could be used, e.g., in recommendation systems: assume that one discovers that the preferences over movies $\{c_1, c_2, c_3, c_4, c_5\}$ are single-peaked w.r.t. axis $c_1 \prec c_2 \prec c_3 \prec c_4 \prec c_5$; if one knows that an agent likes movies c_3 and c_5 , then it is natural to recommend movie c_4 . More generally, one can take advantage of single-peakedness on a sparse graph to make recommendations in the neighbourhood of liked items. Note that, as already said, any preference profile is single-peaked on a complete graph, but it is obviously of no use in the example of a recommendation system.

Thereby, we focus here on determining a graph that minimizes the number of edges or the maximum degree of a vertex. This choice is motivated by the fact that these criteria are measures of sparsity of a graph (the sparser the graph is, the more informative), but also because they generalize known cases such as paths, cycles and trees. Let us indeed emphasize that the mathematical programming approach we propose to identify a graph generalizes the best known classes of the single-peaked recognition problem and provides a uniform treatment of them, leading to simple polynomial time algorithms. Last but not least, stating the recognition problem as an optimization problem is all the more interesting that the single-peakedness property on a path, a tree or a circle is a strong requirement that is rarely met in practice.

Our contribution We propose Integer Linear Programming formulations (ILP) of finding a compatible graph that minimize the number of edges (1) or the maximum degree of a vertex (2), and we show that both of them are NP-hard (Section 3.3). Nevertheless,

if the optimal value for problem (1) is $m-1$ (where m is the number of candidates), we prove the integrality of the optimal basis solution of the linear program (LP) obtained by relaxing the integrality constraint (Section 3.4). This provides an alternative polynomial time method, based on a simple LP solver, to recognize single-peakedness on a tree, as a connected graph with m vertices and $m-1$ edges is a tree. By adding some constraints on the maximum degree of a vertex, we obtain the same result for the case of paths. As a last theoretical result, we prove that single-peakedness on a *pseudotree* (a connected graph containing at most one cycle) is recognizable in polynomial time (Section 3.5). We also provide some experimental results on real-world and synthetic datasets, where we measure the density of the graphs depending on the diversity of preferences of voters (Section 3.6).

Related work We briefly describe here some previous contributions that have addressed the concept of single-peakedness on arbitrary graphs, the optimization view of the recognition problem and the use of ILP formulations for computational social choice problems related to structured preferences:

- Nehring and Puppe defined a general notion of single-peaked preferences based on abstract betweenness relations between candidates (Nehring and Puppe, 2007). In their setting, it is possible to define single-peaked preferences on a graph \mathcal{G} by considering the *graphic betweenness relation*: candidate c_j is between candidates c_i and c_k if and only if c_j lies on a shortest path between c_i and c_k in \mathcal{G} . A preference profile is then single-peaked on \mathcal{G} if for every preference \succ , if c^* is the most preferred candidate w.r.t. \succ and c_j is on a shortest path between c^* and c_k then $c_j \succ c_k$. This definition enables them to state general results regarding strategyproofness on restricted domains of preferences. Note that this definition of single-peakedness on a graph does not coincide with the one we use, as we will see in Section 3.2.
- Peters and Elkind showed how to compute in polynomial time a compact representation of *all* trees with respect to which a given profile is single-peaked (Peters and Elkind, 2016). This structure allows them to find in polynomial time trees that have, e.g., the minimum degree, diameter, or number of internal nodes among all trees with respect to which a given profile is single-peaked. On the contrary, they show that it is NP-hard to decide whether a given profile is single-peaked on a regular tree (where each vertex has degree either 1 or d), or if a profile is single-peaked on a tree isomorphic to a given tree. We provide here alternative proofs for some of the polynomial time results, based on linear programming arguments.
- Peters recently proposed ILP formulations for proportional representation problems, and showed that the binary constraint matrix is totally unimodular if preferences are single-peaked, because the matrix has then the consecutive ones property (Peters, 2018). We recall that the vertices of a polyhedron defined by a totally unimodular constraint matrix are all integer, thus solving the linear programming relaxation yields an optimal solution to the original ILP problem. We also rely on linear programming for proving the polynomial time complexity of some of the recognition problems we tackle here.

Although we do not rely on the consecutive ones property in our analysis.

Organization of the chapter The chapter is organized as follows: in Section 3.2, we start by introducing the notion of single-peakedness on an arbitrary graph. We note that any profile is single-peaked with respect to complete graph, and we motivate why we focus on solutions minimizing the number of edges (1) or the maximum degree of a vertex (2). We then provide in Section 3.3 an ILP formulation for the both problems (1) and (2), and we show that these problems are NP-hard in a general case. But it is well-known that they are polynomial-time solvable for some specific classes of graph - namely trees (i.e., connected graphs with exactly $m - 1$ edges) or cycles (i.e., connected graphs with m edges and the maximum degree of a vertex equals to 2). Indeed, we prove in Section 3.4 that if a profile is single-peaked on a tree, the ILP formulation can be relaxed, which provides an alternative recognition polynomial-time algorithm for preferences single-peaked on a tree. Similarly, we show that the ILP relaxation remains integral if the profile is single-peaked on a cycle or a path.

We study in Section 3.5 the recognition of pseudotrees (which are connected graphs containing at most one cycle), and we provide a polynomial-time recognition algorithm. Moreover, we show that, contrary to the case of trees, the continuous relaxation of the ILP formulation does not remain integral in the case of pseudotrees.

Finally, we perform numerical tests on both real and synthetic data in Section 3.6 to understand what type of graphs (in terms of a sparsity) we are supposed to obtain in practice.

3.2 Problem definition

As emphasized in the introduction, several definitions of single-peakedness on an arbitrary graph can be found in the literature. In our study, we are using the following one (Elkind et al., 2017):

Definition 3.2.1: Single-peakedness on an arbitrary graph (SP)

Let \mathcal{C} be a set of m candidates and \mathcal{P} a profile of preferences of n voters. Let $\mathcal{G} = (\mathcal{C}, E)$ be a connected undirected graph. We say that \mathcal{P} is *single-peaked (SP) on the graph \mathcal{G}* if every $\succ_i \in \mathcal{P}$ is a traversal of \mathcal{G} , i.e., for each $\succ_i \in \mathcal{P}$ and for each $k \in \{1, \dots, m\}$, the subgraph of \mathcal{G} induced by the vertices $\{r_i(1), \dots, r_i(k)\}$ is connected.

Example 3.2.1. To illustrate this notion, let us consider the following profile of only 2 preferences over 5 candidates:

$$\begin{aligned} \succ_1 &: (c_1, c_2, c_3, c_4, c_5) \\ \succ_2 &: (c_1, c_4, c_5, c_2, c_3) \end{aligned}$$

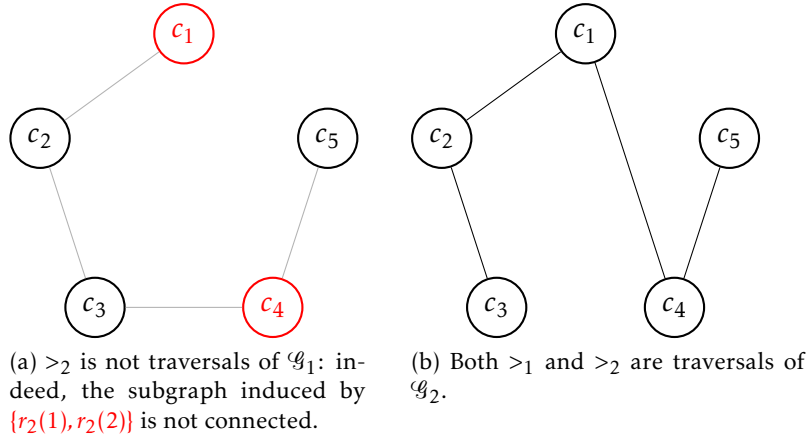


Figure 3.2.1: Illustration of Example 3.2.1: we recall that $>_1: (c_1, c_2, c_3, c_4, c_5)$ and $>_2: (c_1, c_4, c_5, c_2, c_3)$.

Let us check whether this profile is single-peaked on the graph $\mathcal{G}_1 = (\mathcal{C}, E_1)$ of Figure 3.2.1a, where $\{c_i, c_j\} \in E_1$ iff $j = i + 1$ - we note that \mathcal{G}_1 is nothing but the axis $c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4 \triangleleft c_5$. It is easy to see that $>_1$ fulfills the definition: for each $i \in \{1, 2, 3, 4, 5\}$, we have $r_1(i) = c_i$, and the subgraph of \mathcal{G}_1 induced by $\{c_1, c_2, \dots, c_i\}$ is connected.

However, the preference $>_2$ does not fulfill the definition - as the edge $\{c_1, c_4\}$ is not in E_1 , the subgraph of \mathcal{G}_1 induced by $\{r_2(1), r_2(2)\} = \{c_1, c_4\}$ is not connected. Therefore, the profile is not single-peaked w.r.t. graph \mathcal{G}_1 .

Nevertheless, we can see that the profile is single-peaked on the graph $\mathcal{G}_2 = (\mathcal{C}, E_2)$ of Figure 3.2.1b, where $E_2 = \{\{c_1, c_2\}, \{c_2, c_3\}, \{c_1, c_4\}, \{c_4, c_5\}\}$. This time, \mathcal{G}_2 is nothing but the axis $c_3 \triangleleft c_2 \triangleleft c_1 \triangleleft c_4 \triangleleft c_5$.

Example 3.2.1 illustrates that Definition 3.2.1 coincides with the standard definition of single-peakedness (Black, 1948) when \mathcal{G} is a path. Similarly, it coincides with the definition of preferences single-peaked on a cycle (Peters and Lackner, 2017) (resp. tree (Trick, 1989)) when \mathcal{G} is a cycle (resp. tree). Note that the definition based on shortest paths (Nehring and Puppe, 2007) mentioned earlier does *not* generalize single-peakedness on a circle as defined in (Peters and Lackner, 2017), as shown in the following example:

Example 3.2.2. Let us consider the cycle $\mathcal{G} = (V, E)$ with $V = \{c_1, c_2, c_3, c_4, c_5\}$ and $E = \{\{c_1, c_2\}, \{c_2, c_3\}, \{c_3, c_4\}, \{c_4, c_5\}, \{c_5, c_1\}\}$.

The preference $>: (c_1, c_2, c_3, c_4, c_5)$ is single-peaked w.r.t \mathcal{G} when we use Definition 3.2.1, but not when we use the definition of Nehring and Puppe (2007): we recall that according to their definition, a preference profile is single-peaked on \mathcal{G} if for every preference $>$ of peak c^* ,

we have $c_j > c_k$ whenever c_j is on a shortest path between c^* and c_k in \mathcal{G} . But here we have c_5 on a shortest path between c_1 and c_4 , but $c_4 > c_5$.

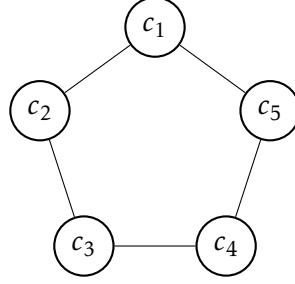


Figure 3.2.2: The preference $(c_1, c_2, c_3, c_4, c_5)$ is not single-peaked with respect to the cycle $\mathcal{G} = (V, E)$ with $V = \{c_1, c_2, c_3, c_4, c_5\}$ and $E = \{\{c_1, c_2\}, \{c_2, c_3\}, \{c_3, c_4\}, \{c_4, c_5\}, \{c_5, c_1\}\}$ in terms of the definition introduced by Nehring and Puppe (2007).

When a profile \mathcal{P} is single-peaked w.r.t. \mathcal{G} , for conciseness we say that \mathcal{P} is *compatible* with \mathcal{G} (or that \mathcal{G} is compatible with \mathcal{P}).

Example 3.2.3. Consider the following profile of 3 preferences over 5 candidates:

$$\begin{aligned} >_1: (c_1, c_2, c_3, c_4, c_5) \\ >_2: (c_1, c_4, c_5, c_2, c_3) \\ >_3: (c_2, c_3, c_4, c_5, c_1) \end{aligned}$$

We saw in Example 3.2.1 that preferences $>_1$ and $>_2$ are compatible with the graph $\mathcal{G}_2 = (\mathcal{C}, E_2)$ (where $E_2 = \{\{c_1, c_2\}, \{c_2, c_3\}, \{c_1, c_4\}, \{c_4, c_5\}\}$).¹ However, when adding $>_3$, this graph is no more compatible: indeed, there must be an edge between c_4 and c_2 or c_3 so that the subgraph induced by $\{c_2, c_3, c_4\} = \{r_3(1), r_3(2), r_3(3)\}$ is connected. In this case, for any other $i \in \{1, 2, 4, 5\}$, it is easy to check that the subgraph induced by $\{r_3(1), \dots, r_3(i)\}$ is connected. Therefore, the profile is single-peaked on $\mathcal{G}_3 = (\mathcal{C}, E_3)$ (resp. $\mathcal{G}_4 = (\mathcal{C}, E_4)$), where $E_3 = E_2 \cup \{c_2, c_4\}$ (resp. $E_4 = E_2 \cup \{c_3, c_4\}$).

We see that there are in general several graphs on which the preference profile is single-peaked. In this concrete example, the graphs \mathcal{G}_3 and \mathcal{G}_4 are minimal in the sense that removing any edge of \mathcal{G}_3 (resp. \mathcal{G}_4) would lead to a graph not compatible with the profile. Besides, we can find many others solutions by just adding more edges to \mathcal{G}_3 (resp. \mathcal{G}_4).

Example 3.2.3 points out the following issue: obviously, any profile is single-peaked w.r.t. the complete graph. However, this case is not interesting because it does not provide any information about the preference structure. That is why we are looking for a *minimal* graph on which the profile is single-peaked. The notion of minimality needs to be made more precise. In our study, we focus mainly on minimizing the number of graph edges. Another criterion we consider is the minimization of the maximum

¹By the way, one can easily show that this is the only path both these preferences are compatible with.

degree of vertices. Put another way, given a preference profile \mathcal{P} , we want to determine a graph \mathcal{G} on which the profile is SP, so as to minimize either the number of edges of \mathcal{G} , or its (maximum) degree. We emphasize the fact that:

- minimizing the number of edges allows to detect when the profile is compatible with a tree (this occurs iff the minimum number of edges is $m-1$, since \mathcal{G} is necessarily connected);
- minimizing the degree of \mathcal{G} allows to detect when the profile is compatible with a cycle (this occurs iff there exists a graph \mathcal{G} with maximum degree 2);
- combining the objective allows to detect when the profile is compatible with an axis: this occurs iff there is a graph \mathcal{G} with maximum degree 2 and $m-1$ edges.

So the tackled problems generalize the most well known (tractable) recognition problems of single-peakedness.

A Curious Tortoise Intervention



The argument “these minimization criteria are good because they generalize the most well known recognition problems of single-peakedness” may not satisfy a curious reader (at least, it does not satisfy me). So let us try to give some more intuition about them. It seems quite natural to minimize the number of edges. What about maximal degree minimization? Let us consider a profile single-peaked with respect of the axis $c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$: the corresponding graph is hence a path, each vertex (except the extremities) being of degree 2. In terms of single-peaked preferences, given a peak c_i , the subrankings on sets $\{c_1, c_2, \dots, c_{i-1}\}$ and $\{c_{i+1}, \dots, c_m\}$ can be merged in any way. In other words, given two candidates c_j and c_k with $j < i$ and $k > i$, we cannot decide whether c_j is preferred to c_k or c_k is preferred to c_j . In particular, a recommendation system which knows that the voter’s most preferred candidates is c_i will recommend her c_{i+1} and c_{i-1} , without any information on whether one of these two candidates could be more relevant - in terms of graph, there is a kind of incomparability among the neighbors of a given vertex.

The higher the vertex degree is, the more there are such “incomparable” candidates. Let us consider the following profile:

$$>_1: (c_3, c_1, c_4, c_2, c_5)$$

$$>_2: (c_2, c_1, c_4, c_3, c_5)$$

$$>_3: (c_3, c_1, c_5, c_4, c_2)$$

$$>_4: (c_1, c_2, c_5, c_3, c_4)$$

It is single-peaked on a special tree called *star graph* $\mathcal{G}_S = (\mathcal{C}, E_S)$, where $E_S = \{\{c_1, c_i\} \mid 1 < i \leq 5\}$ (see Figure 3.2.3a). One can check that this is the

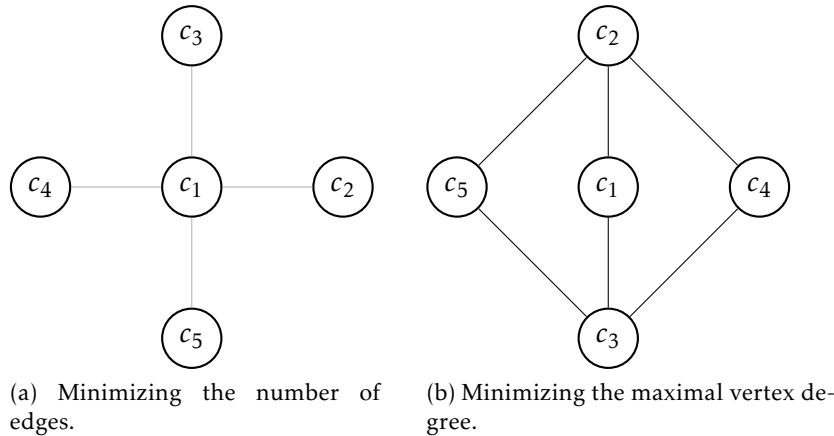


Figure 3.2.3: Different solutions in function of the minimization criterion (we consider here the profile $\{(c_3, c_1, c_4, c_2, c_5), (c_2, c_1, c_4, c_3, c_5), (c_3, c_1, c_5, c_4, c_2), (c_1, c_2, c_5, c_3, c_4)\}$).



only tree compatible with the profile - in terms of minimization of the number of edges, it is hence the best we can do. However, the vertex c_1 is of degree 4: the only information provided by \mathcal{G}_S is that c_1 is a very popular candidate, but we do not have any relation between the other candidates. While minimizing the maximal degree, we obtain the graph $\mathcal{G}_D = (\mathcal{C}, E_D)$ with $E_D = \{\{c_1, c_2\}, \{c_1, c_3\}, \{c_2, c_4\}, \{c_3, c_4\}, \{c_2, c_5\}, \{c_3, c_5\}\}$ (see Figure 3.2.3b). There are two more edges than in \mathcal{G}_S , but the maximal vertex degree is three. We note that more structural information can be read from this graph: for example, if the most preferred candidate of voter v_i is c_1 , then her second most preferred candidate must be c_2 or c_3 (and not c_4 nor c_5 as in the case of \mathcal{G}_S). We also note that c_2 and c_3 seem to be quite similar, as they have the same set of neighbours.

Besides motivating the minimization of maximum degree, we point out here another question: different solutions provide different information about the preference profile - if there are several solutions, how can we know which one fits the best our profile? Although this question is not really studied in this thesis, we will discuss it in the general conclusion.

3.3 ILP formulation and complexity

3.3.1 ILP Formulation

We now present an ILP formulation of the tackled problems. We are looking for a graph \mathcal{G} with m vertices (because there are m candidates). For each pair $\{k, l\} \subseteq \{1, \dots, m\}$, we

define a binary variable $x_{\{c_k, c_l\}}$ which is equal to 1 if the edge $\{c_k, c_l\}$ is present in graph \mathcal{G} , and 0 otherwise. By abuse of notation, we will often write $x_{\{k, l\}}$ instead of $x_{\{c_k, c_l\}}$. Hence, if we are minimizing the number of graph edges, the objective function $f(x)$ is

$$f(x) = \sum_{\{k, l\} \subseteq \{1, \dots, m\}} x_{\{k, l\}}.$$

If we are minimizing the maximum degree, then

$$f(x) = \max_{k \in \{1, \dots, m\}} \sum_{l=1, l \neq k}^m x_{\{k, l\}}.$$

In this latter case, the classical way of linearizing $f(x)$ is to minimize an auxiliary variable z with the constraints $\sum_{l=1, l \neq k}^m x_{\{k, l\}} \leq z$, for all $k \in \{1, \dots, m\}$.

Regardless of the objective function, the other constraints of the problem remain the same: for each $i \in \{1, \dots, n\}$, $\succ_i: (r_i(1), \dots, r_i(m))$ has to be a graph traversal. In other words, for each $k \in \{2, \dots, m\}$, there is an edge between $r_i(k)$ and at least one of the vertices $r_i(1), \dots, r_i(k-1)$. In terms of linear programming constraints, this yield for $i \in \{1, \dots, n\}$ and for $k \in \{2, \dots, m\}$

$$\sum_{j=1}^{k-1} x_{\{r_i(j), r_i(k)\}} \geq 1.$$

To sum up, the ILP formulation of the tackled problems is

$$\begin{aligned} & \min f(x) \\ \text{s.t. } & \begin{cases} \sum_{j=1}^{k-1} x_{\{r_i(j), r_i(k)\}} \geq 1 \quad \forall i \in \{1, \dots, n\}, \forall k \in \{2, \dots, m\} \\ x_{\{k, l\}} \in \{0, 1\} \quad \forall \{k, l\} \subseteq \{1, \dots, m\} \end{cases} \end{aligned}$$

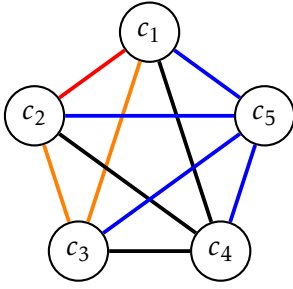
Example 3.3.1. Consider again the profile from Example 3.2.3:

$$\begin{aligned} & \succ_1: (c_1, c_2, c_3, c_4, c_5) \\ & \succ_2: (c_1, c_4, c_5, c_2, c_3) \\ & \succ_3: (c_2, c_3, c_4, c_5, c_1) \end{aligned}$$

For minimizing the number of edges, the objective function is:

$$f(x) = \min x_{\{1,2\}} + x_{\{1,3\}} + x_{\{1,4\}} + x_{\{1,5\}} + x_{\{2,3\}} + x_{\{2,4\}} + x_{\{2,5\}} + x_{\{3,4\}} + x_{\{3,5\}} + x_{\{4,5\}}$$

Figure 3.3.1 shows the constraints relatives to the preference \succ_1 . The constraints relative to preferences \succ_2 and \succ_3 can be obtained similarly.



$$\begin{cases} x_{\{1,2\}} & \geq 1 \\ x_{\{1,3\}} + x_{\{2,3\}} & \geq 1 \\ x_{\{1,4\}} + x_{\{2,4\}} + x_{\{3,4\}} & \geq 1 \\ x_{\{1,5\}} + x_{\{2,5\}} + x_{\{3,5\}} + x_{\{4,5\}} & \geq 1 \end{cases}$$

Figure 3.3.1: Constraints given by preference $>_1: (c_1, c_2, c_3, c_4, c_5)$. Any graph compatible with the profile must contain at least one edge of each color.

Note that for each voter $>_i$, we have necessarily $x_{r_i(1), r_i(2)} = 1$, as by Definition 3.2.1, the candidate $r_i(2)$ must be connected to at least one of her predecessors in $>_i$. In other words, there is the corresponding edge $\{r_i(1), r_i(2)\}$ in any graph \mathcal{G} with respect to which $>_1$ is single-peaked. More formally, we have the following definition:

Definition 3.3.1: Necessary edge

Let \mathcal{P} be a profile of n preferences over the set \mathcal{C} of m candidates. Let $\mathcal{G} = (V, E)$ with $V = \mathcal{C}$. An edge $\{c, c'\} \in E$ is called a *necessary edge* if there exists a voter $>_i \in \mathcal{P}$ such that $\{r_i(1), r_i(2)\} = \{c, c'\}$.

3.3.2 Minimizing the number of edges

In this section, we study the computational complexity of the problem of minimizing the number of edges of \mathcal{G} . A natural question while dealing with an ILP is to ask if its continuous relaxation can be used to solve the problem (here, it would consist in replacing the integrality constraints $x_{k,l} \in \{0,1\}$ by $x_{k,l} \in [0,1]$). We recall that if this can be done, the problem is solvable in polynomial time. Unfortunately, it would be too good to be true: indeed, the following example shows that the optimal solution (when minimizing the number of edges) of the continuous relaxation is not necessarily integer.

Example 3.3.2. Consider the profile of 3 preferences over 4 candidates:

$$\begin{aligned} >_1 &: (c_1, c_2, c_4, c_3) \\ >_2 &: (c_2, c_3, c_4, c_1) \\ >_3 &: (c_1, c_3, c_4, c_2) \end{aligned}$$

From the first two candidates of each voter, we see immediately that the edges $\{c_1, c_2\}$, $\{c_2, c_3\}$ and $\{c_1, c_3\}$ are necessarily present in the graph. Then, we observe that vertex c_4 needs to be connected to at least one of vertices c_1 and c_2 , at least one of vertices c_2 and c_3 and finally at least one of vertices c_1 and c_3 . Consequently any integer solution of the problem will be a graph with at least 5 edges. However, there exists a fractional solution of the continuous

relaxation with value 4.5: we set $x_{\{1,2\}} = x_{\{1,3\}} = x_{\{2,3\}} = 1$ and $x_{\{1,4\}} = x_{\{2,4\}} = x_{\{3,4\}} = 0.5$ (see Figure 3.3.2).

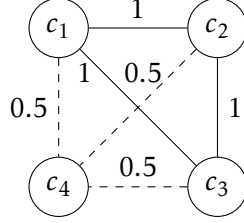


Figure 3.3.2: The continuous relaxation of the ILP for minimizing the number of edges can lead to fractional optima.

We now show that the problem is actually NP-hard.

Theorem 3.3.1

Given a preference profile \mathcal{P} , it is NP-hard to find a graph compatible with \mathcal{P} with a minimum number of edges.

Proof. We use a polynomial time reduction from the set cover problem, known to be NP-hard (Garey and Johnson, 1979), where given a finite set $\mathcal{U} = \{e_1, \dots, e_n\}$ of elements, a set $\mathcal{S} = \{S_1, \dots, S_m\}$ of subsets of \mathcal{U} and $k \in \mathbb{N}$, the question is to determine if there exists a subset $\mathcal{K} \subseteq \mathcal{S}$ of size k such that $\cup_{S \in \mathcal{K}} S = \mathcal{U}$.

From an instance of set cover, we define a preference profile \mathcal{P} as follows (see Figure 3.3.3 for better visualisation):

- (i) Let $\{S_1, \dots, S_m, z\}$ be a set of candidates.
- (ii) Let $\{v_1, \dots, v_n\}$ be a set of voters. Let S_{i_1}, \dots, S_{i_l} be the subsets in \mathcal{S} containing element $e_i \in \mathcal{U}$, and $S_{i_{l+1}}, \dots, S_{i_m}$ the remaining subsets in \mathcal{S} (which do not contain e_i). Then, the preference of voter v_i is defined as $>_{v_i}: (S_{i_1}, \dots, S_{i_l}, z, S_{i_{l+1}}, \dots, S_{i_m})$.

- (iii) We add $\frac{m \cdot (m-1)}{2}$ voters $v_{\{i,j\}}, \{i,j\} \subseteq \{1, \dots, m\}$ such that

$$>_{v_{\{i,j\}}}: (S_i, S_j, \underbrace{S_1, \dots, S_m, z}_{\text{except } S_i, S_j}).$$

We prove that there exists a set cover of size k if and only if there exists a graph \mathcal{G} compatible with \mathcal{P} that has $\frac{m \cdot (m-1)}{2} + k$ edges. Let \mathcal{K} be a set cover solution of size k . We build a graph \mathcal{G} compatible with \mathcal{P} in the following manner:

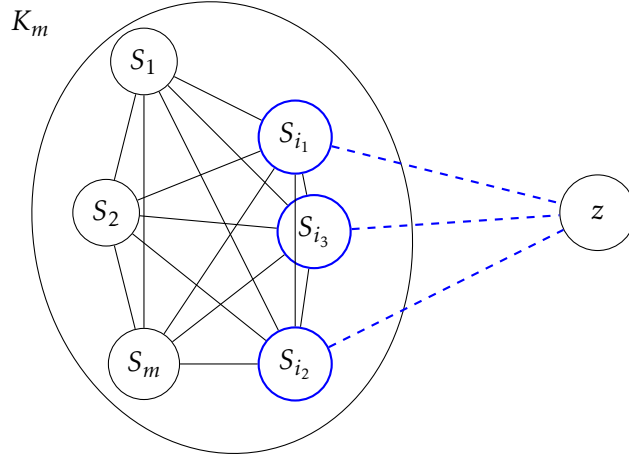


Figure 3.3.3: Any graph \mathcal{G} compatible with the reduction profile \mathcal{P} contains a complete subgraph on $\{S_1, \dots, S_m\}$ in order to be compatible with preferences $>_{v_{[i,j]}}$. To be compatible with preference $>_{v_i}$, there must be **at least one edge between z and a vertex corresponding to one of the sets containing e_i** .

- a) For each $\{i, j\} \in \{1, \dots, m\}^2, i \neq j$, the edge $\{S_i, S_j\}$ is in \mathcal{G} - this is necessary for the preferences of type (iii) above to be SP on \mathcal{G} .
- b) For each $i \in \{1, \dots, m\}$, the edge $\{S_i, z\}$ is in \mathcal{G} if and only if $S_i \in \mathcal{K}$.

Hence, the subgraph formed by vertices $\{S_1, \dots, S_m\}$ is a clique having $\frac{m \cdot (m-1)}{2}$ edges, and there are exactly k more edges adjacent to z - in total, \mathcal{G} has $\frac{m \cdot (m-1)}{2} + k$ edges. As $k > 0$, the graph is connected and all preferences of type (iii) are SP on \mathcal{G} . Let $>_{v_i}$ be one of the preferences of type (ii). We need to prove that z is connected to at least one of the vertices S_{i_1}, \dots, S_{i_l} . As the sets S_{i_1}, \dots, S_{i_l} are the only sets of \mathcal{S} containing the element e_i , and as \mathcal{K} is a solution of the set cover instance, this is true thanks to b). So, \mathcal{G} is a graph compatible with \mathcal{P} that has $\frac{m \cdot (m-1)}{2} + k$ edges. To prove the other implication, let \mathcal{G} be a graph compatible with \mathcal{P} that has $\frac{m \cdot (m-1)}{2} + k$ edges. As \mathcal{G} is compatible with \mathcal{P} , the subgraph induced by the set of vertices $\{S_1, \dots, S_m\}$ must be a clique so that the preferences $>_{v_{[i,j]}}$ of type (iii) are SP on \mathcal{G} . Hence, this subgraph contains $\frac{m \cdot (m-1)}{2}$ edges, and so, there are exactly k edges adjacent to z . Let us define \mathcal{K} containing S_i iff S_i is adjacent to z in \mathcal{G} . As \mathcal{G} is compatible with \mathcal{P} , each preference $>_{v_i}$ of type (ii) is SP on \mathcal{G} . It means that at least one of S_{i_1}, \dots, S_{i_l} is adjacent to z , so is in \mathcal{K} . As all these sets contains e_i , there is an element of \mathcal{K} that covers e_i . The subset $\mathcal{K} \subseteq \mathcal{S}$ is thus a solution of size k of the set cover instance. \square

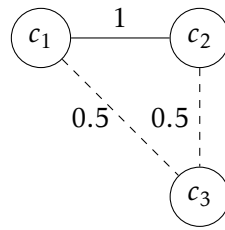


Figure 3.3.4: The continuous relaxation of the ILP for minimizing the maximum degree can lead to fractional optima.

3.3.3 Minimizing the Maximum Degree

We now consider our second objective function, namely the maximum degree of a vertex in the graph (to be minimized). We come up with similar results.

First, the same as for the minimization of the number of edges, the ILP formulation we have proposed in Section 3.3.1 is not integral, as we can see in the following example.

Example 3.3.3. Consider a profile with 3 candidates and one voter with ranking (c_1, c_2, c_3) . The ILP formulation of the problem of determining a graph \mathcal{G} of minimum max-degree compatible with \mathcal{P} is:

$$\begin{array}{l} \min z \\ \text{s.t.} \left\{ \begin{array}{l} x_{\{1,2\}} + x_{\{1,3\}} \leq z \\ x_{\{1,2\}} + x_{\{2,3\}} \leq z \\ x_{\{1,3\}} + x_{\{2,3\}} \leq z \\ x_{\{1,2\}} \geq 1 \\ x_{\{1,3\}} + x_{\{2,3\}} \geq 1 \\ x_{\{k,l\}} \in \{0, 1\} \quad \forall \{k, l\} \subset \{1, 2, 3\} \end{array} \right. \end{array}$$

The value of an optimal integer solution is $z=2$, but there exists a fractional solution of the continuous relaxation of value 1.5 ($x_{\{1,2\}} = 1, x_{\{1,3\}} = x_{\{2,3\}} = 0.5$ - see Figure 3.3.4).

Here again we show that the problem of minimizing the degree of \mathcal{G} is NP-hard, by using a similar reduction.

Theorem 3.3.2

Given a preference profile \mathcal{P} , it is NP-hard to find a graph compatible with \mathcal{P} with a minimum max-degree.

Proof. Let $\mathcal{U} = \{e_1, \dots, e_n\}$, $\mathcal{S} = \{S_1, \dots, S_m\}$, $k \in \mathbb{N}$ be an instance of the set cover problem. Consider the profile \mathcal{P} defined as follows:

- (i) Let $\{S_1, \dots, S_m, z, t_1, \dots, t_m\}$ be a set of candidates.
- (ii) Let $\{v_1, \dots, v_n\}$ be a set of voters. Let S_{i_1}, \dots, S_{i_i} be the subsets in \mathcal{S} containing the element $e_i \in \mathcal{U}$, and $S_{i_{i+1}}, \dots, S_{i_m}$ the other subsets in \mathcal{S} . Then, the preference of voter v_i is defined as

$$>_{v_i}: (S_{i_1}, \dots, S_{i_i}, z, S_{i_{i+1}}, \dots, S_{i_m}, t_1, \dots, t_m).$$

- (iii) We add $\frac{m(m-1)}{2}$ voters $v_{\{i,j\}}$, $\{i,j\} \subseteq \{1, \dots, m\}$ such that

$$>_{v_{\{i,j\}}}: (S_i, S_j, \underbrace{S_1, \dots, S_m}_{\text{except } S_i, S_j}, z, t_1, \dots, t_m).$$

- (iv) We add m voters v_{t_1}, \dots, v_{t_m} where the preference of v_{t_i} is defined as

$$>_{v_{t_i}}: (z, t_i, \underbrace{t_1, \dots, t_m}_{\text{except } t_i}, S_1, \dots, S_m),$$

and a voter w with preference $(t_1, S_1, S_2, \dots, S_m, z, t_2, \dots, t_m)$.

Figure 3.3.5 illustrates that in any graph \mathcal{G} compatible with the profile:

- a) the vertices S_i form a clique (edge $\{S_i, S_j\}$ is enforced by voter $v_{\{i,j\}}$),
- b) z is adjacent to all t_i (due to voter v_{t_i}),
- c) $\{S_1, t_1\}$ is in the graph (due to w).

Actually, these edges are all *necessary edges* (see Definition 3.3.1). We claim that there exists a set cover of size (at most) k iff there is a graph compatible with the profile with degree at most $m + k$.

Suppose that there is a set cover \mathcal{K} of size at most k . Then, beyond the necessary edges mentioned above, we put an edge $\{S_i, z\}$ iff S_i is in \mathcal{K} . Then the vertex with maximum degree is z , with degree $m+k$. The graph \mathcal{G} is compatible with each preference $>_{v_i}$ because \mathcal{K} is a set cover. It is compatible with preference $>_{v_{\{i,j\}}}$ thanks to the necessary edges (and $k > 0$, so z is connected as well). It is compatible with $>_{v_{t_i}}$ and $>_w$ thanks to the necessary edges.

Now suppose that there is a solution \mathcal{G} with degree at most $m + k$. In particular, z has degree at most $m + k$, hence is adjacent to at most k vertices S_i . The preference of voter v_i imposes that z is adjacent to some S_i which contains e_i . In other words, the set \mathcal{K} of these (at most) k sets S_i is a set cover of size at most k . \square

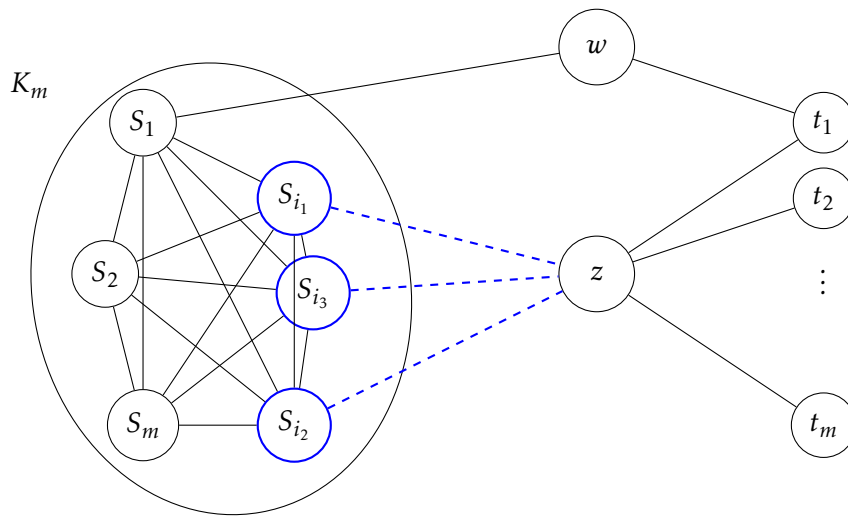


Figure 3.3.5: Any graph \mathcal{G} compatible with the reduction profile \mathcal{P} contains a complete subgraph on $\{S_1, \dots, S_m\}$ in order to be compatible with preferences $>_{v_{\{i,j\}}}$. To be compatible with preference $>_{v_i}$, there must be **at least one edge between z and a vertex corresponding to one of the sets containing e_i** . The set of vertices t_i ensures that the degree of z will be greater or equal to the degree of any other vertex. Finally, w is a “technical” vertex to establish the connection between the clique K_m and the set of vertices t_i .

3.4 Recognition of trees and paths

In this section, we focus on the recognition of trees and paths - given a profile \mathcal{P} , we are looking for a tree (or a path) on which the profile is SP.

Recognizing single-peaked preferences on a tree can be done using the combinatorial algorithm proposed by Trick (Trick, 1989). As an alternative proof of this result, we show in this section that the continuous relaxation of the ILP formulation given in Section 3.3.1 can be used to solve this recognition problem in polynomial time: in fact, all (optimal) extremal solutions are integral (Theorem 3.4.1). We show in Theorem 3.4.2 that a similar result holds for the recognition of profiles SP on a path.

We start by recalling Trick's procedure (Trick, 1989), as we will use it in the proof of the two previously mentioned theorems.

Recognition of profiles SP on a tree (Trick, 1989) Let $\mathcal{P} = \{>_1, \dots, >_n\}$ be a profile of n preferences over m candidates. Let c_l be a candidate ranked last by at least one voter. Trick shows that, if preferences are SP on a tree, then c_l must necessarily be a leaf. More formally, for each $i \in \{1, \dots, n\}$, let us denote by $A(c_l)_i$ the set of candidates ranked better than c_l by voter v_i if c_l is not ranked first by v_i ; if c_l is ranked first by v_i , then $A(c_l)_i$ is the singleton containing the second most-preferred candidate of v_i . From $A(c_l) = \bigcap_{i=1}^n A(c_l)_i$, the following conclusions can be drawn:

- if $A(c_l) = \emptyset$, there does not exist a tree solution.
- Otherwise, $A(c_l)$ is the set of vertices to which the leaf c_l can be connected.

In the latter case, the algorithm of Trick deletes c_l from all preferences, and repeats this process on the modified profile with preferences over $m-1$ candidates.

A Curious Tortoise Intervention



Obviously, we do not give here the proof of Trick's procedure (and we invite the reader to consult the original result of Trick (1989)). However, we think it can be useful for which follows to give some informal intuition of it. Here are some questions (and answer) that can arise while reading the previous paragraph:

Question 1: Why any candidate c_l ranked last at least once must be a leaf?

Answer: Suppose there is an optimal solution (i.e., a tree) in which c_l is not a leaf. Then it has at least two neighbours, let say c_i and c_j . We note that there is a path from c_i to c_j going through c_l . But c_l is ranked last by a voter v . In other words, v ranks both c_i and c_j better than c_l . Without loss of generality, let us assume $c_i >_v c_j$. There is then a path from c_i to c_j that does not go through c_l . Put together, there is a cycle in the solution, which is a contradiction with the optimality of the solution.



We highlight the parallel with the classical notion of single-peakedness (on an axis) - in any single-peaked profile, there are at most two candidates ranked last. Indeed, any path graph has only two leaves.

Question 2: Why is there no tree solution if $A(c_l) = \emptyset$?

Answer: Well, we have just seen that c_l must be a leaf. But if $A(c_l) = \emptyset$, there is no candidate ranked better than c_l in all preferences, so c_l need to be connected to at least two vertices to satisfy the definition of single-peakedness on a graph. So it cannot be a leaf.

Question 3: Why can we remove c_k at the end of the iteration?

Answer: As c_k is ranked last at least once, it will never be in $A(c_i)$ for any other candidate c_i . So removing it does not change anything.

Example 3.4.1. Consider the profile $\mathcal{P} = \{>_1, >_2, >_3\}$ defined by:

$$>_1: (c_1, c_2, c_3, c_4)$$

$$>_2: (c_2, c_1, c_3, c_4)$$

$$>_3: (c_4, c_1, c_2, c_3)$$

1. The candidate c_4 is ranked last by at least one voter - we will determine the set $A(c_4)$:

$$A(c_4) = \{c_1, c_2, c_3\} \cap \{c_1, c_2, c_3\} \cap \{c_1\} = \{c_1\}$$

The candidate c_4 is then deleted from all preferences and we continue next iteration with the subprofile $>_1^1: (c_1, c_2, c_3)$, $>_2^1: (c_2, c_1, c_3)$ and $>_3^1: (c_1, c_2, c_3)$.

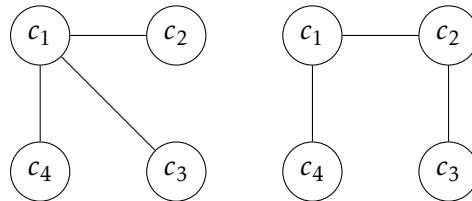
2. The only candidate ranked last by at least one voter is now c_3 . We see that

$$A(c_3) = \{c_1, c_2\}.$$

We continue with the subprofile $>_1^2: (c_1, c_2)$, $>_2^2: (c_2, c_1)$, $>_3^2: (c_1, c_2)$.

3. We get $A(c_2) = \{c_1\}$, and the algorithm stops as we obtain a subprofile involving only one candidate.

To sum up, we have obtained $A(c_4) = \{c_1\}$ (first iteration), $A(c_3) = \{c_1, c_2\}$ (second iteration) and $A(c_2) = \{c_1\}$ (third iteration). Consequently, in any tree compatible with \mathcal{P} , vertex 2 and vertex 4 have to be connected to vertex 1, and vertex 3 has to be connected to vertex 1 or 2. Hence, there exist two trees² on which profile \mathcal{P} is single-peaked, and these are:



²one of them being actually a path

Using LP to recognize SP preferences on a tree or a path Let us consider the following continuous relaxation LP-SP (linear program for single-peakedness) of the ILP introduced in Section 3.3.1:

$$(LP-SP) \quad \begin{aligned} & \min \sum_{\{k,l\} \subseteq \{1,\dots,m\}} x_{\{k,l\}} \\ & \text{s.t.} \quad \begin{cases} \sum_{j=1}^{k-1} x_{\{r_i(j), r_i(k)\}} \geq 1 \quad \forall i \in \{1, \dots, n\}, k \in \{2, \dots, m\} \\ x_{\{k,l\}} \in [0, 1] \quad \forall \{k,l\} \subseteq \{1, \dots, m\} \end{cases} \end{aligned}$$

We show in Theorem 3.4.1 that we can use LP-SP to solve in polynomial time the problem to determine, given a profile, whether or not there exists a tree compatible with it.

Theorem 3.4.1

If a profile \mathcal{P} is compatible with a tree, then any extremal optimal solution x of LP-SP is integral, i.e., $x_{\{k,l\}} \in \{0, 1\}$ for any $\{k,l\} \subseteq \{1, \dots, m\}$.

Proof of Theorem 3.4.1. The proof is based on two properties of optimal solutions of LP-SP when the profile is compatible with a tree. These two properties allow to come up with a reformulation of the problem as a maximum flow problem, where there is a bijection between the solutions of LP-SP of value $m - 1$ and the (optimal) flows of value $m - 1$. The result then comes from the fact that any extremal solution of the flow problem (with integral capacity) is integral (Ahuja et al., 1988).

The first property states that all constraints of LP-SP are tight in a solution of value $m - 1$.

Property 1

If the optimal value of LP-SP is $m - 1$, then all constraints are tight in an optimal solution x^* : $\sum_{j=1}^{k-1} x_{\{r_i(j), r_i(k)\}}^* = 1$.

Proof of Property 1 Let v_i be a voter. There are $m - 1$ constraints associated with v_i , and each variable $x_{\{k,l\}}$ appears in exactly one of these constraints. Since on the one hand the sum of all variables is $m - 1$ (objective function), and on the other hand the sum of variables in each of these $m - 1$ constraints is at least one, each constraint must be tight. This concludes the proof of Property 1. \square

Let us now consider that the profile is single-peaked with respect to a tree. The recognition procedure recalled above starts by identifying a candidate, say c_m , ranked last in at least one preference and such that $A(c_m) \neq \emptyset$. We recall that c_m is then removed from all preferences. This procedure is applied recursively, till there is only one candidate left. For simplicity, let us assume that the first removed (identified) candidate is c_m , the second c_{m-1} , and so on. Let us now focus on the step when candidate c_k is identified as a

leaf (and then removed from the profile). To avoid confusion, we denote by $B(c_k)$ the set $A(c_k)$ at this step, i.e., when considering the profile restricted to the first k candidates.

Property 2

If the profile is SP on a tree, then in an optimal solution of LP-SP, for any candidate $c_k \geq 2$ we have $\sum_{c_j \in B(c_k)} x_{\{c_j, c_k\}} = 1$, and $x_{\{c_j, c_k\}} = 0$ for any $c_j \in \{c_1, \dots, c_{k-1}\} \setminus B(c_k)$.

Proof of Property 2. Let us consider some candidate c_k (with $k \geq 2$) and some optimal solution X of LP-SP. The proof is organized as follows :

- **Step 1:** first, we consider the restriction of the problem to the candidates c_1, c_2, \dots, c_k and we show that the restricted solution X_k is a feasible and optimal solution for it.
- **Step 2:** We prove then the statement for the problem restriction.

Step 1: Let us define LP-SP(k) as the linear program corresponding to the problem restricted to the candidates c_1, c_2, \dots, c_k . We first show that the optimal solution X restricted to the first k candidates, let us call it X_k , is feasible and optimal for LP-SP(k).

A Curious Tortoise Intervention



Actually, it is not obvious why X_k should be feasible: let us consider the preference $>_1: (c_2, c_1, c_5, c_4, c_3)$. Any graph compatible with this preference must contain one of the edges $\{c_2, c_4\}$, $\{c_1, c_4\}$ or $\{c_5, c_4\}$. In terms of LP constraint, one gets:

$$x_{\{2,4\}} + x_{\{1,4\}} + x_{\{5,4\}} \geq 1$$

In LP-SP(4), the candidate c_5 has been removed, and the above constraint becomes:

$$x_{\{2,4\}} + x_{\{1,4\}} \geq 1$$

Imagine that in some optimal solution X of LP-SP, it is necessary to set $x_{\{2,4\}} = x_{\{1,4\}} = 0$ and $x_{\{5,4\}} = 1$. In such a case, X_4 would not be a solution of LP-SP(4) (which is the profile restricted to first 4 candidates)! So why can we assume that the restricted solution X_4 is still feasible? That is the goal of the first part of this proof...

If $k = m$, $X_k = X$, so there is nothing to do and we can directly continue to step 2 proving the property.

If $k < m$, assume that the properties are true for any $k' > k$. We will show that they are also true for k . To do this, let us consider a constraint of LP-SP(k) for connecting candidate c_j (with $j < k$) for some voter i . As all candidates $c_{k'}$ with $k' > k$ have been

removed, this constraint only involves the variables $x_{l,j}$ with $l \leq k$. However, the corresponding constraint in the initial program LP-SP possibly contains some other variables $x_{k',j}$ with $k' > k$. Let us suppose that there is such a variable $x_{k',j}$. It means that $c_{k'}$ is ranked better than c_j by v_j .

Suppose that $x_{k',j} = 1$ in the solution X . As we assumed the property true for $k' > k$, we have $x_{k',j} \in B(c_{k'})$ - in other words, c_j is ranked better than $c_{k'}$ by all voters - but this is a contradiction, as we have just said above that $c_{k'}$ is ranked better than c_j by the voter v_j . Therefore, $x_{k',j} = 0$ in X . More generally, all “removed variables” in the constraint was set to 0 in X - the only variables responsible for the feasibility of X are still present in X_k . Hence X_k is feasible for LP-SP(k). We can now easily see that it is optimal: each time a candidate $c_{k'}$ with $k' > k$ has been removed, $\sum_{c_j \in B(c_{k'})} x_{c_j, c_{k'}} = 1$ so the total weights of (remaining) variables reduce by 1. Thus X_k is a feasible solution of LP-SP(k) of value $k - 1$.

Step 2: Now we can focus on X_k on LP-SP(k).³ Note that the profile is trivially SP on the first k candidates (as it is SP on the whole set of candidates). Candidate c_k is ranked

in last position by some voter v_i , so we have $\sum_{j=1}^{k-1} x_{\{c_j, c_k\}} \geq 1$ (constraint of connecting c_k

for voter v_i), and by Property 1 we have $\sum_{j=1}^{k-1} x_{\{c_j, c_k\}} = 1$. If all candidates c_1, c_2, \dots, c_{k-1}

are in $B(k)$ then we are done. Otherwise, consider a candidate $c_\ell \notin B(c_k)$. Then c_ℓ is ranked after c_k by some other voter v_j (and, for this voter v_j , c_k and c_ℓ are not the best two candidates, as we have supposed $k \geq 2$). Then, if $x_{\{c_\ell, c_k\}} > 0$ we get $\sum_{c_j \notin \{c_\ell, c_k\}} x_{\{c_j, c_k\}} =$

$1 - x_{\{c_\ell, c_k\}} < 1$, and the constraint associated to v_j for connecting c_k to its predecessors is violated. So $x_{\{c_\ell, c_k\}} = 0$ for any $c_\ell \notin B(c_k)$, and consequently $\sum_{c_j \in B(c_k)} x_{\{c_j, c_k\}} = 1$. This

concludes the proof of Property 2. □

Back to the proof of Theorem 3.4.1.

We are now done with the proof of Property 2 and we return back to the proof of the theorem. We reformulate the problem as a flow problem. From \mathcal{P} , we build a network (directed graph) R with:

- A source s , a destination t , and for each candidate k two vertices ℓ_k and r_k .
- We have an arc from s to each ℓ_k with capacity 1, and an arc from each r_k to t with capacity ∞ .
- For each candidate k , we have an arc (ℓ_k, r_j) for each $j < k$. The capacity of this arc is 1 if $j \in B(k)$, and 0 otherwise.

The reader can consult Figure 3.4.1 for a better visualisation of the network R . Let us

³Let us note that the proof of this step also applies to the special case $k = m$. That justifies why we could assume in the previous step that the property was true for any $k' > k$.

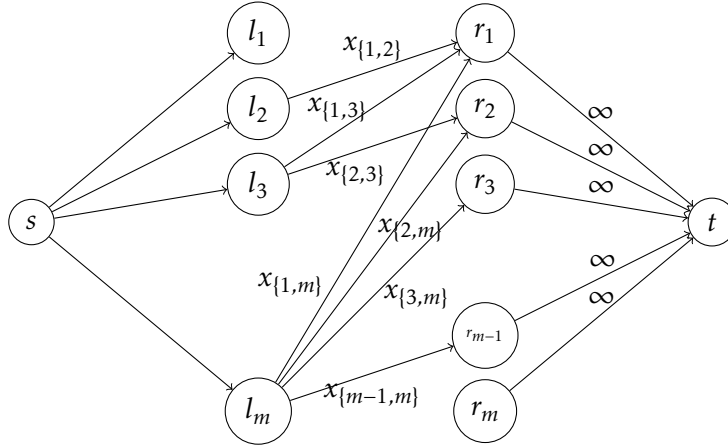


Figure 3.4.1: Reformulation of LP-SP as a maximum flow problem: the capacity of an arc $\{l_i, r_j\}$ corresponds to the value of LP-SP variable $x_{\{i,j\}}$; the capacities of arcs from s to l_i equal 1.

denote by ϕ a flow on this network, with $\phi(e)$ the flow on edge e . Note that ℓ_1 has no outgoing edge, so the optimal flow is at most $m-1$.

We show that the correspondence $x_{\{k,j\}} = \phi(k,j)$ (for each $j < k$) is a bijection between solutions of value $m-1$ of LP-SP and (optimal) flows of value $m-1$ in R .

Let ϕ be a flow of value $m-1$. As there is no flow through ℓ_1 , there is a flow of value 1 through each $\ell_k, k > 1$. Since arc (k,j) has capacity 0 if $j \notin B(k)$, by flow conservation we have $\sum_{j \in B(k)} \phi(k,j) = 1$, which means that $\sum_{j \in B(k)} x_{\{k,j\}} = 1$. Now consider a voter v_i for which k is not ranked first. By the procedure of Trick, when k is identified as a leaf, all candidates in $B(k)$ are ranked before k , and the corresponding constraint is satisfied. This is true for all candidates and voters, so x is a feasible solution of LP-SP, of value $m-1$.

Conversely, let x be a feasible solution of LP-SP of value $m-1$. From Property 2, we have $\sum_{j \in B(k)} x_{\{j,k\}} = 1$ for each candidate $k \geq 2$. This immediately gives a flow of value $m-1$.

By integrality of extremal flows (any non integral optimal flow is a convex combination of integral flows), any extremal optimal solution of LP-SP is integral (when there exists a tree compatible with \mathcal{P}). \square

Let us now turn to the recognition of profiles SP on a path. A (connected) graph is a path iff it is a tree with degree at most 2. Hence, we consider the following ILP formulation where we minimize the number of edges and add constraints on the vertex

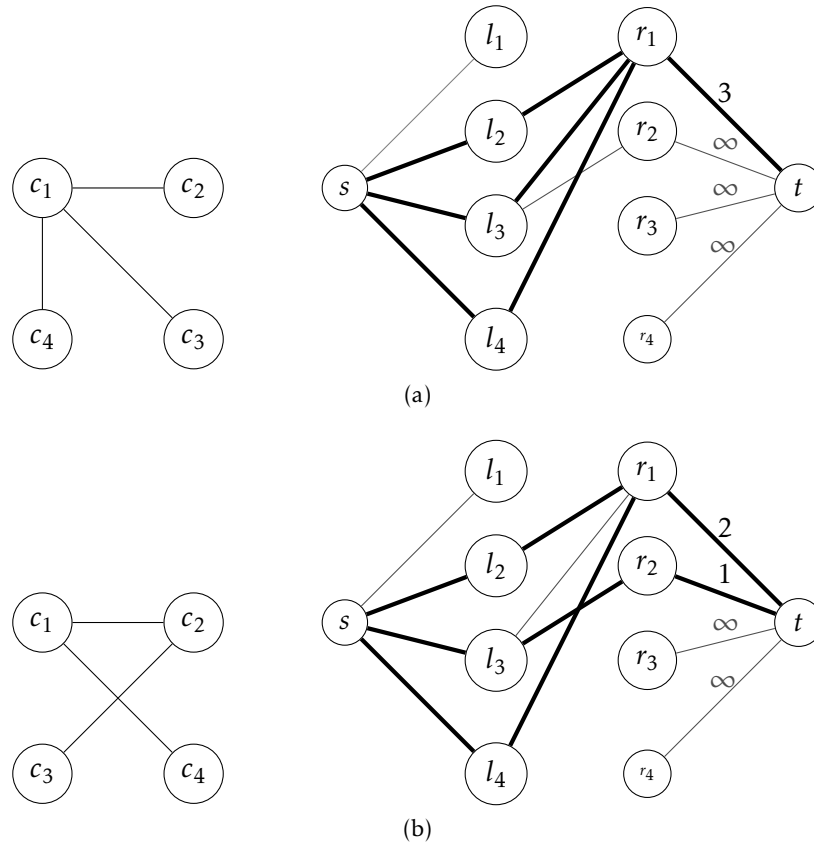


Figure 3.4.2: Two tree solutions of the profile \mathcal{P} of Example 3.4.1 and their associated flow solutions.

degrees:

$$\begin{aligned} \min \quad & \sum_{\{k,l\} \subseteq \{1,\dots,m\}} x_{\{k,l\}} \\ \text{s.t.} \quad & \begin{cases} \sum_{j=1}^{k-1} x_{\{r_i(j), r_i(k)\}} \geq 1 \quad \forall i \in \{1, \dots, n\}, k \in \{2, \dots, m\} \\ \sum_{l=1, l \neq k}^m x_{\{k,l\}} \leq 2 \quad \forall k \in \{1, \dots, m\} \\ x_{\{k,l\}} \in \{0, 1\} \quad \forall \{k,l\} \subseteq \{1, \dots, m\} \end{cases} \end{aligned}$$

Clearly, a profile is compatible with a path iff the optimal value of the previous ILP is $m-1$. Let us call LP-SP2 the continuous relaxation. As well as LP-SP, we can reformulate the problem as a flow problem, using a very similar network: from \mathcal{P} , we build a network (directed graph) R with:

- A source s , a destination t , and for each candidate k two vertices ℓ_k and r_k .
- We have an arc from s to each ℓ_k with capacity 1, and an arc from each r_k to t with capacity 2 (this is the only difference with the flow reformulation of LP-SP where the capacity of arcs (r_k, t) was unlimited).
- For each candidate k , we have an arc (ℓ_k, r_j) for each $j < k$. The capacity of this arc is 1 if $j \in B(k)$, and 0 otherwise.

If we suppose in Property 2 that the profile is compatible with a path, the proof remains true if we replace the LP-SP flow reformulation by the LP-SP2 one. We get then the following result:

Theorem 3.4.2

If a profile \mathcal{P} is compatible with a path, then any extremal optimal solution of LP-SP2 is integral, i.e., $x_{\{k,l\}} \in \{0, 1\}$ for any $\{k,l\} \subseteq \{1, \dots, m\}$.

3.5 Recognition of pseudotrees

So far, we have seen that our minimization problem was NP-hard in the general case, but polynomially solvable in the case where the optimal solution is a tree. As a natural extension, we consider the problem to recognize profiles that are single-peaked with respect to a graph with $m - 1 + k$ edges, for some fixed k , thus allowing k more edges than in a tree. In this section, we consider the case $k = 1$. A graph on m vertices with m edges is called a pseudotree. We show that recognizing if there exists a pseudotree compatible with a given profile can be done in polynomial time. We leave as open question the parameterized complexity of the problem when k is the parameter: would the problem be in XP? Or even in FPT?

Let us now deal with the case of pseudotree. Hence, the set of solutions we want to

recognize is the class of connected graphs having (at most) m edges. To solve the problem in polynomial time, we devise an algorithm that first identifies the leaves of the pseudotree and then the cycle on the remaining vertices. The second step (cycle recognition) is done using the polynomiality of recognizing single-peakedness on a cycle (Peters and Lackner, 2017). For the first step, we need to modify the Trick's procedure recalled in Section 3.4. This procedure was able to correctly identify leaves when the profile was compatible with a tree, but it fails to correctly identify leaves when the underlying structure is a pseudotree. With a slight modification though, we obtain in Proposition 3.5.1 a necessary and sufficient condition for a candidate to be a leaf in a pseudotree. This is the stepping stone leading to the polynomiality of detecting whether or not a given profile is compatible with a pseudotree, stated in Theorem 3.5.1.

Example 3.5.1. *Let us consider the following profile with 4 voters and 5 candidates:*

$$\begin{aligned} >_1: (c_1, c_2, c_3, c_4, c_5) \\ >_2: (c_1, c_3, c_4, c_2, c_5) \\ >_3: (c_2, c_5, c_3, c_4, c_1) \\ >_4: (c_3, c_5, c_4, c_2, c_1) \end{aligned}$$

Regarding the first two candidates of each voter, we see that any solution contains the edges $\{c_1, c_2\}$, $\{c_1, c_3\}$, $\{c_2, c_5\}$ and $\{c_3, c_5\}$ - i.e., the cycle $(c_1, c_2, c_5, c_3, c_1)$. From here it is easy to see that there is a unique pseudotree compatible with the profile. Trick's procedure for recognizing single-peaked preferences on a tree focuses on candidates ranked last at least once in order to detect leaves. Here, that would be c_1 and c_5 . However, we have $A(c_1) = A(c_5) = \emptyset$. Note also that the whole profile is not compatible with a cycle, so we need somehow to first detect c_4 as a leaf, and then detect that the remaining candidates are single-peaked on a cycle.

The central property that allows to recognize profiles compatible with a pseudotree is given in the following proposition.

Proposition 3.5.1

Let \mathcal{P} be a preference profile, and suppose that a candidate c_i is such that $A(c_i) \neq \emptyset$. Then \mathcal{P} is compatible with a pseudotree if and only if it is compatible with a pseudotree where c_i is a leaf.

Proof. Let \mathcal{G} be a pseudotree compatible with \mathcal{P} where c_i is not a leaf. We transform \mathcal{G} into a pseudo-tree \mathcal{G}' compatible with \mathcal{P} where c_i is a leaf. Let $c_j \in A(c_i)$.

Case 1: $\{c_i, c_j\} \in \mathcal{G}$.

Let us first consider an easy case, where $\{c_i, c_j\} \in \mathcal{G}$. Then we build \mathcal{G}' from \mathcal{G} by simply replacing each edge $\{c_i, c_k\}$ (with $k \neq j$) by the edge $\{c_j, c_k\}$. We need to justify why for each voter v , $>_v$ is still a traversal of \mathcal{G}' :

1. Let c_l be a vertex not adjacent to c_i in \mathcal{G} . Then the modifications of \mathcal{G} do not have any impact on c_l , and c_l can still be connected (in the same manner it was done in \mathcal{G}) to one of the candidates ranked better than her in $>_v$.

2. Similarly, the modifications have no impact on the candidate c_j , as we did not remove any of its neighbours while creating \mathcal{G}' .
3. Let us consider the vertex corresponding to the candidate c_i , and its neighbour in \mathcal{G} called here c_k . In \mathcal{G}' , we have replaced the edge $\{c_i, c_k\}$ by $\{c_j, c_k\}$. We need to show that both c_i and c_k can still be connected to one of their predecessors in \mathcal{G}' . There are two cases to distinguish:
 - (a) Suppose that c_k is ranked better than c_i by the voter v : then the edge $\{c_i, c_k\}$ was used in \mathcal{G} to connect the candidate c_i to one of its predecessors - i.e. c_k . But we have $c_j \in A(c_i)$, so c_j is ranked better than c_i by v^4 , so c_i can be connected to one of its predecessors, i.e. c_j , in \mathcal{G}' by the edge $\{c_i, c_j\}$.
 - (b) Suppose that c_i is ranked better than c_k by the voter v : then $\{c_i, c_k\}$ was used in \mathcal{G} to connect c_k to its predecessor c_i . But as $c_j \in A(c_i)$, c_j is also a predecessor of c_k , so we are done.

We have shown by the case distinction that the graph \mathcal{G}' is compatible with all the preferences. Note that \mathcal{G}' has (at most) as many edges as \mathcal{G} , so it is a pseudotree (or a tree, and we can add any edge to create a pseudotree).

Case 2: $\{i, j\} \notin \mathcal{G}$.

Let us now consider the case where $\{c_i, c_j\} \notin \mathcal{G}$. Note that then c_j is ranked before i in all preferences (otherwise c_i is first and c_j is second, which leads $A(c_i) = \{c_i\}$, and the edge $\{c_i, c_j\}$ is forced to be in any compatible graph - a contradiction). Then we first transform \mathcal{G} into a graph \mathcal{G}' which is a pseudotree solution containing the edge $\{c_i, c_j\}$, and then we apply Case 1 to \mathcal{G}' .

As \mathcal{G} contains exactly one cycle, there are at most two (simple) paths between each couple of vertices. This gives us two subcases to distinguish:

Case 2a:) Let us suppose that there is a unique simple path from c_j to c_i in \mathcal{G} . Let us denote by c_u the predecessor of c_i on this path. Then we create \mathcal{G}' by replacing the edge $\{c_u, c_i\}$ by the edge $\{c_i, c_j\}$. Consider a voter v . The only candidates possibly touched by transforming \mathcal{G} to \mathcal{G}' are c_i and c_u . But c_i is connected to its predecessor c_j by the new edge $\{c_i, c_j\}$. Regarding c_u , the modification does not affect this vertex - indeed, it does not wee the edge $\{c_u, c_i\}$ in \mathcal{G} because it is ranked better than c_i in $>_v$ (the subgraph induced by c_i and the candidates ranked before her by v is connected and contains c_i and c_j , so it contains the whole path connecting these edges, in particular c_u).

Case 2b. In the other case, in \mathcal{G} there are two simple paths from c_j to c_i . If on both of them, the predecessor of c_i is the same vertex c_u , the same reasoning as in the previous subcase applies.

Suppose that the predecessor of c_i is c_{u_1} in the first path and c_{u_2} in the second one (with $c_{u_1} \neq c_{u_2}$). We build \mathcal{G}' from \mathcal{G} by deleting edges $\{c_{u_1}, c_i\}$ and $\{c_{u_2}, c_i\}$, and adding edges $\{c_i, c_j\}$ and $\{c_{u_1}, c_{u_2}\}$. See Figure 3.5.1.

⁴The only way to have c_i ranked better than c_j is to rank them first and second in v . But this would be a contradiction with $c_k >_v c_i$.

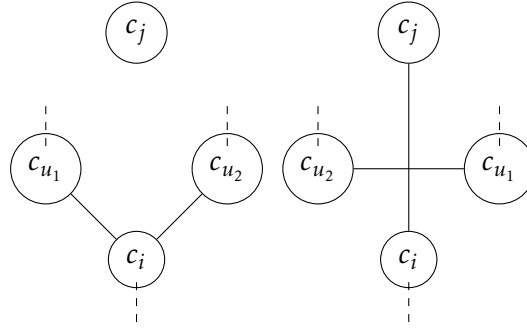


Figure 3.5.1: Edge swap.

Consider a voter v . Since v prefers c_j to c_i and the subgraph of \mathcal{G} induced by the candidates up to c_i in the ranking $>_v$ is connected, then c_{u_1} or c_{u_2} is ranked before c_i by v , say c_{u_1} (we assume wlog that c_{u_1} is preferred to c_{u_2} by v). Then we see that \mathcal{G}' is compatible with the preference of v : indeed, when considering candidates one by one in the order of v , the only modification holds for c_{u_2} , which is now connected to c_{u_1} (ranked before her), and for c_i , which is now connected to c_j (ranked before her). \square

Note that c_i can be connected to any vertex $c_j \in A(c_i)$. Before giving the procedure that recognizes preferences compatible with a pseudotree, we need to establish another property regarding such preferences.

Proposition 3.5.2

If a profile \mathcal{P} is compatible with a pseudotree, then either there exists a candidate c_i such that $A(c_i) \neq \emptyset$, or \mathcal{P} is compatible with a cycle.

Proof. Let \mathcal{G} be a pseudotree compatible with \mathcal{P} . If \mathcal{G} is not a cycle, then there exists a candidate c_i which is a leaf in \mathcal{G} . Let c_j be the unique neighbour of c_i in \mathcal{G} .

Suppose for contradiction that $c_j \notin A(c_i)$. There exists a voter v such that:

- either c_i is not the most preferred candidate of v and $c_i >_v c_j$: The subgraph induced by the candidates up to c_i (in the ranking of v) is not connected.
- or c_i is the most preferred candidate of v and c_j is not ranked second: in this case, the subgraph induced by the first two candidates (in the ranking of v) is not connected.

We get a contradiction in both cases, so c_j is necessarily in $A(c_i)$. \square

Consider now the following procedure `DETECT_PSEUDOTREE`:

Algorithm 1 DETECT_PSEUDOTREE

Input : a preference profile \mathcal{P}
Output: A pseudotree compatible with \mathcal{P} if it exists, NO otherwise.
Set $E' = \emptyset$
while there are at least 4 candidates, and a candidate c_i such that $A(c_i) \neq \emptyset$ **do**
 $E' \leftarrow E' \cup \{c_i, c_j\}$ for some (arbitrary) $c_j \in A(c_i)$
 remove c_i from the profile
end while

if the remaining profile is compatible with a cycle C **then**
 return $E' \cup C$
else
 return NO
end if

Theorem 3.5.1

Given a preference profile \mathcal{P} on at least 3 candidates, the procedure DETECT_PSEUDOTREE is polynomial-time and returns a pseudotree compatible with \mathcal{P} if some exists, or returns NO otherwise.

Proof. DETECT_PSEUDOTREE obviously runs in polynomial time. We proceed by induction on the number of candidates. If there are three candidates the procedure outputs a cycle on these 3 candidates. Now suppose that the result is true up to $m - 1$ candidates, and consider a profile \mathcal{P} on $m \geq 4$ candidates.

Suppose that \mathcal{P} is compatible with a pseudotree \mathcal{G} .

- If there exists a candidate c_i with $A(c_i) \neq \emptyset$, then by Proposition 3.5.1, there exists a pseudotree \mathcal{G}' compatible with \mathcal{P} where c_i is a leaf. Then the profile obtained from \mathcal{P} by removing c_i is compatible with a pseudotree $(\mathcal{G}' \setminus c_i)$, and adding the edge $\{c_j, c_i\}$ as done by DETECT_PSEUDOTREE gives a pseudotree compatible with \mathcal{P} .
- Otherwise, by Proposition 3.5.2, \mathcal{P} is compatible with a cycle, which is found by DETECT_PSEUDOTREE (Step 3).

Suppose now that DETECT_PSEUDOTREE does not output NO. If there were no candidate i with $A(c_i) \neq \emptyset$, then \mathcal{P} is compatible with the cycle C . Otherwise, let c_i be the candidate in the first iteration of the loop in Step 2 ($A(c_i) \neq \emptyset$). Then, on the profile \mathcal{P} where i is removed, DETECT_PSEUDOTREE outputs a pseudotree, compatible with this profile without i by induction. Since $c_j \in A(c_i)$, adding edge $\{c_j, c_i\}$ makes the pseudotree compatible with \mathcal{P} . \square

A Curious Tortoise Intervention



An attentive reader may have noticed that we do not need the hypothesis of pseudotree in the proof of Proposition 3.5.1. Indeed, this proposition can be reformulated in the following more general way: Let \mathcal{P} be a preference profile and c_i a candidate such that $A(c_i) \neq \emptyset$. Then there exists a graph compatible with \mathcal{P} in which c_i is a leaf.

In other words, we can always start by identifying leaves. When there is no more candidate c_i such that $A(c_i) \neq \emptyset$, the procedure `DETECT_PSEUDOTREE` checks if the restricted profile is single-peaked on a cycle. But maybe we could modify the procedure in a way that it checks the compatibility with a (no-leaf) graph with $m + k$ edges. This is discussed in the following, and also in the last paragraph of this section.

The generalization of this polynomiality result to connected graphs with $(m-1+k)$ edges seems to require new techniques (even for fixed k , i.e. to show that the problem is in XP when parameterized by k). Indeed, an enumeration of all subsets of k edges does not allow to reduce the problem to trees. Procedure `DETECT_PSEUDOTREE` does not seem to generalize either, as it specifically relies on the decomposition of the solution into one cycle and leaves.

Using LP to recognize SP preferences on a pseudotree A natural question is whether the continuous relaxation of the ILP introduced in Section 3.3.1 can be used to recognize pseudo-trees. Unfortunately, this is not possible: we will now show that, contrary to the case of trees, there may exist some non integrap solution of LP-SP in the case of pseudo-trees.

To this purpose, let us consider the following profile \mathcal{P} of 4 preferences over 6 candidates:

$$>_1: (c_1, c_2, c_f, c_3, c_4, c_5)$$

$$>_2: (c_2, c_3, c_4, c_f, c_1, c_5)$$

$$>_3: (c_4, c_3, c_5, c_f, c_2, c_1)$$

$$>_4: (c_5, c_3, c_2, c_f, c_4, c_5)$$

These preferences enforce 4 necessary edges : $\{c_1, c_2\}, \{c_2, c_3\}, \{c_3, c_4\}$ and $\{c_3, c_5\}$. The connectivity constraints of c_f induced by $>_1$ and $>_3$ implies that :

$$x_{c_1, c_f} + x_{c_2, c_f} \geq 1$$

$$x_{c_3, c_f} + x_{c_4, c_f} + x_{c_5, c_f} \geq 1$$

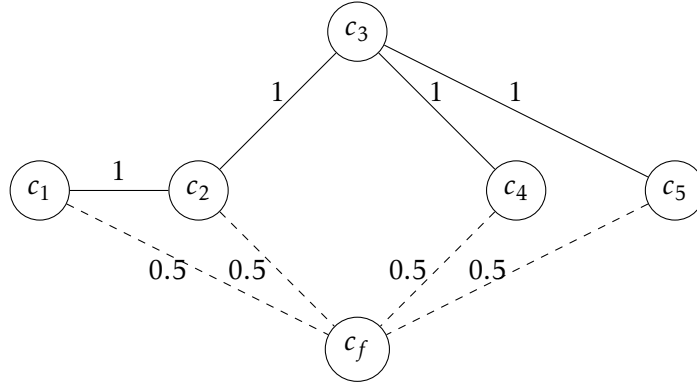


Figure 3.5.2: Example of a non-integral extremal solution of LP-SP

As $\{x_{c_1, c_f}, x_{c_2, c_f}\}$ and $\{x_{c_3, c_f}, x_{c_4, c_f}, x_{c_5, c_f}\}$ are disjoint variable sets, the optimum solution value (of both ILP and LP-SP) is at least 6. We obtain $f(x) = 6$ for the feasible vector x such that :

$$\begin{aligned} x_{c_1, c_2} &= x_{c_2, c_3} = x_{c_3, c_4} = x_{c_3, c_5} = 1 \\ x_{c_1, c_f} &= x_{c_2, c_f} = x_{c_4, c_f} = x_{c_5, c_f} = \frac{1}{2} \\ x_{c, c'} &= 0 \text{ otherwise} \end{aligned}$$

(See also Figure 3.5.2 for an illustration). Hence x is an optimal solution of LP-SP. What's more, x is an extremal solution : the only integral solution of value 6 x' such that $x'_{c_1, c_f} \neq 0$ is :

$$\begin{aligned} x'_{c_1, c_2} &= x'_{c_2, c_3} = x'_{c_3, c_4} = x'_{c_3, c_5} = x'_{c_1, c_f} = x'_{c_3, c_f} = 1 \\ x'_{c, c'} &= 0 \text{ otherwise} \end{aligned}$$

Thus, if x is a convex combination of optimal integral solutions, x' should be among them. But $x_{c_3, c_f} = 0$. Hence, x is not a convex combination of optimal integral solution, and consequently there exists some extremal optimal solution which is not integral.

3.6 Experimental study

We carried out numerical experiments⁵ on real and randomly generated instances of the problems tackled in the paper. In the case of real data, we compare the optimal solution of the ILP to that of its continuous relaxation. We also focus on the ability to detect structure in voters' preferences depending on the election context.

⁵All tests were performed on an Intel Core i7-1065G7 CPU with 8 GB of RAM under Windows OS. We used the IBM Cplex solver for the solution of ILPs.

To go further, we use randomly generated instances to study structural aspects of solutions; we notably study the graph density depending on the number of voters and on the dispersion of their opinions.

3.6.1 Numerical tests on real data

We used PrefLib data sets available on www.preflib.org (Mattei and Walsh, 2013) to perform our numerical tests on real data. While this database offers four different types of data, only the ED (Election Data) type is relevant for our study. Among the ED data sets, we used the complete strict order lists (which correspond to files with `.soc` extension). At the time we carried out these experiments, 315 data files of this type were available in PrefLib, however, many of them were not adapted to our study for several reasons. The first one is that many elections dealt with only 3 or 4 candidates and a great number of voters, hence the obtained graph was, unsurprisingly, always complete. We also met the opposite problem when there were very few voters, typically 4, so there was no point in looking for some general structure. Thus, in practice, there were 25 real data files usable for our purposes, namely:

- 20 files from the ED-00006 data set, which contains figure skating rankings from various competitions during the 1998 season including the World Juniors, World Championships, and the Olympics.
- 2 files from the ED-00009 data set, which contains the results of surveying students at AGH University of Science and Technology (Krakow, Poland) about their course preferences.
- 1 file from the ED-00012 data set, where individuals ranked T-shirt designs.
- 1 file from the ED-00014 data set, which contains preferences about various kinds of sushi (survey conducted by Toshihiro Kamishima).
- 1 file from the ED-00032 data set, which contains the results of surveying students in the Faculty of Informatics, Instituto Superior Politécnico José Antonio Echeverría (Cujae, Havana, Cuba), about the most important criteria affecting their performances as students.

We now present the results obtained for these 25 instances. The tackled optimization problem was to determine a graph with a *minimal number of edges*. For all instances considered here, an optimal graph has been obtained in about 40 milliseconds for the ILP formulation and 20 milliseconds for its LP relaxation. In fact, the linear programming formulation always returned an integer solution. Table 3.6.1 summarizes the obtained results.

ED-00006 data set. The number of candidates (skaters) varies from 14 to 30, and the number of voters (judges) from 7 to 9. For the 20 instances considered, a tree has been obtained 14 times, a pseudotree 5 times, and a solution with 23 edges for 19 candidates (vertices) has been obtained once. The possible interpretation of these results is that, even though the rankings are based on subjective opinions of the judges, there is something like a “true ranking” behind as some skaters are objectively better than other ones. Thus, the rankings given by the judges can be viewed as biased observations of the true ranking, so that they are quite close. In addition, it should be noted that even when the solution was not a tree, the LP continuous relaxation gave an integer solution (identical to the one of the ILP). Finally, let us mention that we also checked compatibility with an axis, and no profile was single-peaked with respect to an axis.

ED-00009 data set. Each student provided a rank ordering over all the courses with no missing elements. There were 9 courses to choose from in 2003 and 7 in 2004, and about 150 students. For both years, the optimal solution was a tree, more specifically a star. This is easily explained from the fact that, in both years, there was one course which was the most preferred for every student.

ED-00012 and ED-00014 data sets. The optimal solution for the preferences over the T-shirt designs had 25 edges, which is quite a lot regarding the number of candidates (11) and voters (30). However, it is consistent with the intuition that there is probably no structure behind T-shirt designs. The same can be said for the preferences over the kinds of sushi, where 5000 voters were asked for their preferences about 10 kinds of sushi (the optimal solution is a complete graph in this case).

ED-00032 data set In the single instance with no tie nor missing element, there were 15 students that ranked the 6 criteria affecting their performances. In both the ILP and LP formulations, a solution with 8 edges has been returned.

Set	File	#candidates	#votvoters	#edges
ED-00006	3	14	9	13 (tree)
ED-00006	4	14	9	13 (tree)
ED-00006	7	23	9	22 (tree)
ED-00006	8	23	9	22 (tree)
ED-00006	11	20	9	20 (pseudotree)
ED-00006	12	20	9	20 (pseudotree)
ED-00006	18	24	9	23 (tree)
ED-00006	21	18	7	17 (tree)
ED-00006	22	18	7	17 (tree)
ED-00006	28	24	9	23 (tree)
ED-00006	29	19	9	23
ED-00006	32	23	9	23 (pseudotree)
ED-00006	33	23	9	22 (tree)
ED-00006	34	23	9	22 (tree)
ED-00006	35	18	9	17 (tree)
ED-00006	36	18	9	17 (tree)
ED-00006	37	19	9	18 (tree)
ED-00006	44	20	9	19 (tree)
ED-00006	46	30	9	30 (pseudotree)
ED-00006	48	24	9	23 (tree)
ED-00009	1	9	146	8 (tree-star)
ED-00009	2	7	153	6 (tree-star)
ED-00012	1	11	30	25
ED-00014	1	10	5000	45 (clique)
ED-00032	2	6	15	7

Figure 3.6.1: Minimal number of edges (fifth/last column) on real data sets from PreLib. Specific structures are indicated in parentheses.

3.6.2 Experimental study on randomly generated data

The experimental study on real data revealed some interesting information. Nevertheless, it is limited by the small amount of data available. Here, we conduct experiments on random data in order to study the structure of solutions. As mentioned above, in some contexts we can assume that the voter's preferences are biased observations of a "true" ranking. This idea can be modeled using the Mallows distribution on rankings. In this model, the "true" ranking is called *central permutation* and its probability is the highest one. The probability of other permutations decreases with the Kendall-tau distance from the central permutation. Formally, let \succ_0 be the central permutation (preference). The probability of a permutation (preference) \succ is $P(\succ) = \frac{\exp(-\theta d_{KT}(\succ, \succ_0))}{\psi(\theta)}$, where $d_{KT}(\cdot, \cdot)$ is the Kendall-tau distance (see Definition 1.1.1), $\theta \geq 0$ is a dispersion

parameter modeling the opinion heterogeneity, and $\psi(\theta)$ is a normalisation constant. If $\theta = 0$, the uniform distribution is obtained. The greater the value of θ , the more the voters agree on the central permutation.

We used the PerMallows R package⁶ for generating the random data according to the Mallows model. The number of candidates was set to $m = 20$, the value of θ varied from 0 to 1 by step of 0.1. The number of voters n varied from 20 to 100 by step of 10. For each pair (θ_0, n_0) of parameter values, the results are averaged over 1000 randomly drawn preference profiles. The curves in Figure 3.6.2 show the evolution of the graph density according to these parameter values.

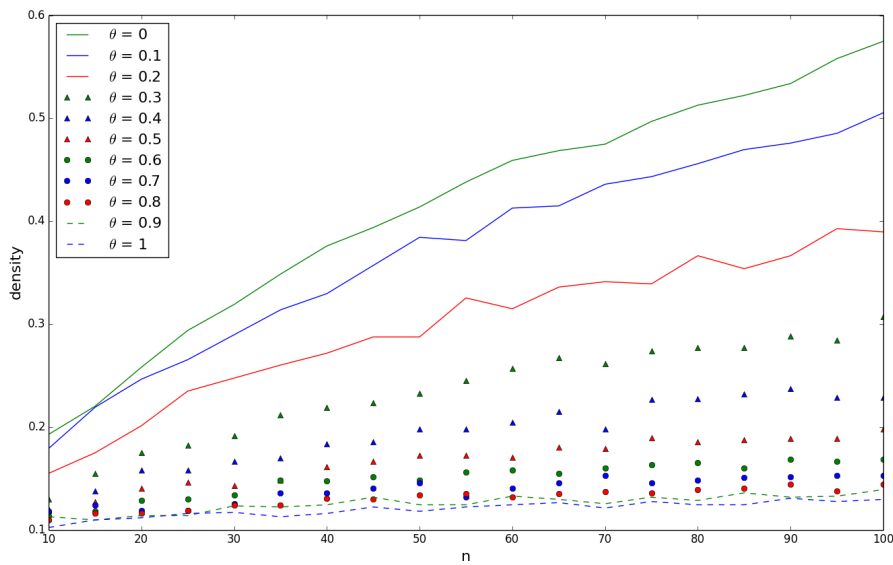


Figure 3.6.2: Density of the graph according to parameters θ and n (with $m = 20$).

In the best case, the obtained solution is a tree, hence, the density is $(m-1)/\frac{m(m-1)}{2} = \frac{2}{m}$. As we set $m = 20$, this corresponds to a density of 0.1. The function representing the graph density seems indeed to converge to the constant function of value 0.1 while the value of θ increases and the preferences in the profile become similar (the curves get closer and closer to the x-axis). Put another way, the density captures the similarity of voters' preferences, as clearly the higher θ the lower the curve. On the contrary, the graph density becomes of course higher when the number n of voters increases. Nevertheless, note that, even for 100 voters, the graph is still quite far from being complete. Besides, the slope of the curve decreases with n . During our experiments, we plotted functions $1 - \log(\text{density})$ and obtained a set of (approximate) straight lines, thus indicating that the convergence towards density 1 (complete graphs) is of the form $1 - e^{-\lambda_{\theta} n}$,

⁶<https://cran.r-project.org/web/packages/PerMallows/index.html>

where $\lambda_\theta > 0$ is a parameter decreasing with θ .

We now give some theoretical arguments that support this observation. Let us recall that if a voter ranks c_j first and c_k second (or the opposite), then edge $\{c_j, c_k\}$ must be present in the graph and is called *necessary edge*. Assuming that the preferences in the profile are generated with the Mallows model, let us now estimate the number of necessary edges in the graph for n voters and m candidates, which gives us an under-estimation and hopefully good approximation of the total number of edges. Let θ be the model parameter and $>_0$ the central permutation. The probability that a preference induces the necessary edge $\{c_j, c_k\}$ is

$$P(\{c_j, c_k\}) = \frac{1}{\psi(\theta)} \sum_{> \in \mathcal{R}_{\{j,k\}}} \exp(-\theta d_{KT}(>, >_0)) \quad (3.1)$$

where $\mathcal{R}_{\{j,k\}}$ is the set of permutations of $\{c_1, \dots, c_m\}$ that ranks c_j and c_k in the first two positions. In a profile with n voters, the probability that no preference induces the necessary edge $\{c_j, c_k\}$ is then written $(1 - P(\{c_j, c_k\}))^n$.

Hence, by switching to the complement, the probability that $\{c_j, c_k\}$ is a necessary edge is

$$P_n(\{c_j, c_k\}) = 1 - (1 - P(\{c_j, c_k\}))^n.$$

Finally, we obtain the expected value of the number of edges as

$$\sum_{\{c_j, c_k\} \subseteq \{c_1, \dots, c_m\}} P_n(\{c_j, c_k\}) = \binom{m}{2} - \sum_{\{c_j, c_k\} \subseteq \{c_1, \dots, c_m\}} (1 - P(\{c_j, c_k\}))^n. \quad (3.2)$$

For $\theta = 0$, as the distribution is uniform, we get that $P(\{j, k\}) = 1/\binom{m}{2}$. Then, we directly obtain that the average number of necessary edges is $\binom{m}{2}(1 - e^{-\alpha n})$ with $\alpha = -\ln(1 - 1/\binom{m}{2})$, thus contributing for $(1 - e^{-\alpha n})$ in the density, in accordance with the experiments. The curves in Figure 3.6.3 shows the evolution of the expected contribution of necessary edges in the graph density according to the values of θ and n . The result extends to any value of θ but requires a dedicated algorithm to compute efficiently $P(\{j, k\})$ in Equation 3.2 (see Appendix 3.A). As expected, we can see that the shapes of the curves coincide in Figures 3.6.2 and 3.6.3. Note, however, that the scale of the y-axis in Figure 3.6.3 slightly differs from the one in Figure 3.6.2 (the curves in Figure 3.6.3 indeed only account for *necessary* edges, thus the analytical values are smaller than the experimental ones).

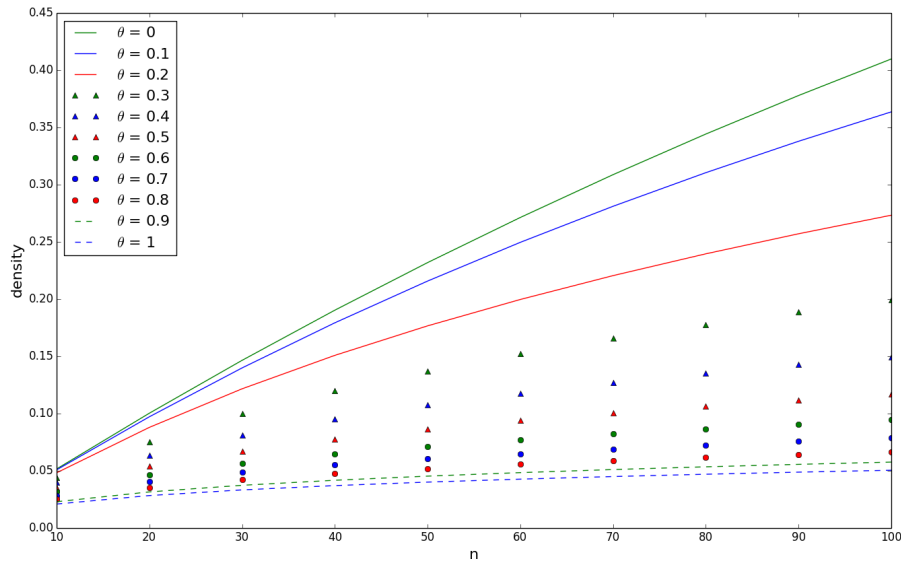


Figure 3.6.3: Expected number of necessary edges according to parameters θ and n (with $m = 20$).

By using the formula in Equation 3.2, we can have an idea of the evolution of the number of necessary edges in the graph for up to 1000 voters whose preferences follows the Mallows model. The obtained curves for various values of θ are shown in Figure 3.6.4. For instance, if one assumes that all votes are equally likely (*impartial culture assumption*, corresponding to $\theta = 0$), then the graph becomes complete for a thousand voters, while only around 45% of the edges are present if one sets $\theta = 0.3$ (i.e., a lower preference heterogeneity).

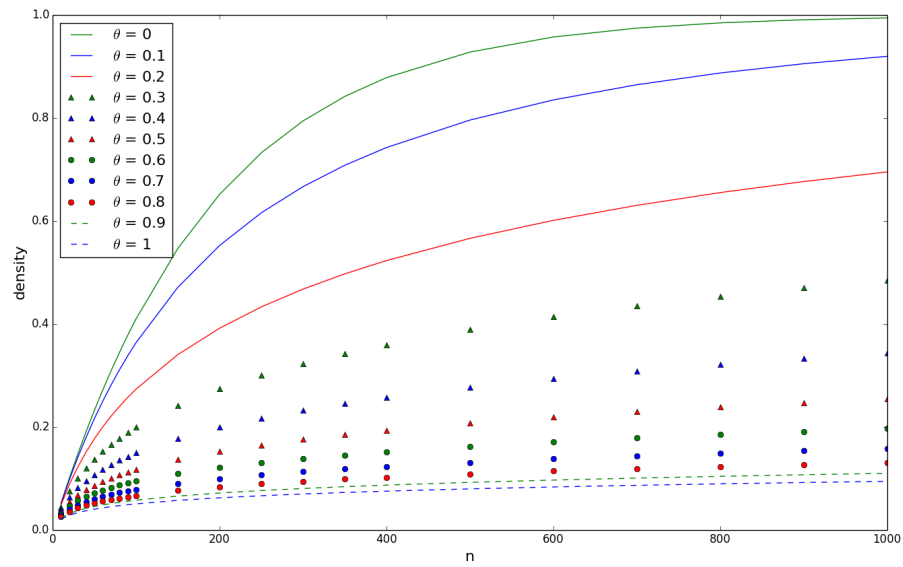


Figure 3.6.4: Expected number of necessary edges for up to $n = 1000$ voters following the Mallows model of parameter θ (with $m = 20$).

3.7 Conclusion

In this section, we have studied the single-peakedness on an arbitrary graph. In Section 3.2, we have introduced this notion, as well as two different measures of graph sparsity, namely the number of edges and the maximum vertex degree. In Section 3.3, we have proposed an ILP formulation for the both problems of minimizing the number of edges and of minimizing the maximum degree of a vertex. We have shown that these problems are NP-hard in a general case.

In Section 3.4, if the optimal value of the number of edges minimization equals $m - 1$, we prove the integrality of the continuous relaxation of the ILP. This provides an alternative polynomial time method to recognize single-peakedness on a tree. By adding some constraints on the maximum degree of a vertex, we obtain the same result for the case of paths.

In Section 3.5, we have provided a polynomial time recognition algorithm of preferences single-peaked on a pseudotree. Moreover, we have shown that the linear relaxation of the ILP formulation does not remain integral in the case of pseudotrees, and cannot thus be used for recognition of the single-peakedness of this class of graph.

Finally, we have performed numerical tests on both real-world and synthetic data in Section 3.6. The results are quite promising - the obtained solutions are generally quite sparse (in terms of the number of edges), and the theoretical analysis providing a bound on the number of edges seems to support this observation. Moreover, it is interesting to note that even when using the ILP continuous relaxation, there were almost no fractional edges the most of the time.

Several research directions, concerning mostly the interpretability of solutions and the impact of different sparsity measures to the solution, are discussed in the global conclusion.

Chapter appendix

3.A Computation of the expected number of necessary edges

In practice, it is computationally cumbersome to enumerate all permutations to compute $P(\{c_j, c_k\})$ according to Equation 3.1 (page 90). We present here another approach to compute efficiently this value. Note that the maximal value Kendall-tau distance between two permutations of length m is $m(m-1)/2$. For each value $\delta \in \{1, \dots, m(m-1)/2\}$, let $N^m(\delta)$ denote the number of permutations $>$ such that $d_{KT}(>, >_0) = \delta$, and $N_{\{c_j, c_k\}}^m(\delta)$ the number of permutations $>_i$ such that $\{r_i(1), r_i(2)\} = \{c_j, c_k\}$ and $d_{KT}(>_i, >_0) = \delta$. Then, $P(\{c_j, c_k\})$ can be computed as

$$P(\{c_j, c_k\}) = \frac{\sum_{\delta=0}^{m(m-1)/2} N_{\{c_j, c_k\}}^m(\delta) \exp(-\theta \delta)}{\psi(\theta)}. \quad (3.3)$$

The value $N_{\{j,k\}}^m(\delta)$ can be computed as follows:

- Firstly, we define the permutation $>_1$ such that $r_1(1) = c_j, r_1(2) = c_k$ and for each pair c_i, c_l of candidates different from c_j and c_k , we have

$$r_1^-(c_i) < r_1^-(c_l) \text{ if and only if } r_0^-(c_i) < r_0^-(c_l).$$

Similarly, we define the permutation $>_2$ such that $r_2(1) = c_k, r_2(2) = c_j$ and for each pair c_i, c_l of candidates different from c_j and c_k , we have

$$r_2^-(c_i) < r_2^-(c_l) \text{ if and only if } r_0^-(c_i) < r_0^-(c_l).$$

We denote by δ_1 (resp. δ_2) the Kendall-tau distance between $>_0$ and $>_1$ (resp. $>_2$).

- As $>_1$ (resp. $>_2$) is between⁷ $>_0$ and any permutation $>_i$ such that $r_i(1) = c_j$ and $r_i(2) = c_k$ (resp. $r_i(1) = c_k$ and $r_i(2) = c_j$), we have $d_{KT}(>_0, >_i) = d_{KT}(>_0, >_1) + d_{KT}(>_1, >_i)$ (resp. $d_{KT}(>_0, >_i) = d_{KT}(>_0, >_2) + d_{KT}(>_2, >_i)$) because the Kendall-tau distance

⁷A ranking $>_i$ is *between* rankings $>_j$ and $>_k$ if, for any pair c, c' of candidates, $r_j^-(c) < r_j^-(c')$ and $r_k^-(c) < r_k^-(c')$ implies that $r_i^-(c) < r_i^-(c')$.

satisfies the betweenness condition⁸ (Kemeny, 1959). Consequently, the number of permutations \succ inducing the necessary edge $\{c_j, c_k\}$ and such that $d_{KT}(\succ_0, \succ) = \delta$ can be computed as

$$N_{\{c_j, c_k\}}^m(\delta) = N^{m-2}(\delta - \delta_1) + N^{m-2}(\delta - \delta_2)$$

because $d(\succ_0, \succ_1) = \delta_1$ and $d(\succ_0, \succ_2) = \delta_2$ by definition of δ_1 and δ_2 .

Note that this equation is well defined because $N^m(\delta)$ is fully characterized by m and δ . The problem consists now in determining the values $N^{m-2}(\delta - \delta_1)$ and $N^{m-2}(\delta - \delta_2)$. In this purpose, for any value m and distance δ , $N^m(\delta)$ can be computed thanks to the following recursion principle:

- Regardless of the length m of the permutations considered, $N^m(0) = 1$ as there is only one permutation at distance 0 of \succ_0 - it is \succ_0 itself.
- Let $m > 1$ and $\delta > 0$. Let \succ be an arbitrary permutation of length m at distance δ from \succ_0 . The distance between \succ_0 and \succ can be calculated as the number δ' of swap operations needed to move $r_0(1)$ to the first position in \succ , to which one adds the distance between the restrictions of \succ_0 and \succ to their $(m-1)$ last elements. We have $0 \leq \delta' \leq m-1$, and so $N^m(\delta) = \sum_{\delta'=0}^{\min(\delta, m-1)} N^{m-1}(\delta - \delta')$.

Overall, after a single preprocessing step in $O(m^4)$, each probability $P(\{c_j, c_k\})$ can be computed in $O(m^2)$. The preprocessing step consists in determining $N^i(\delta_i)$ for each $i \in \{1, \dots, m\}$ and $\delta_i \in \{1, \dots, i(i-1)/2\}$. Hence, there are $N(m) \in O(m^3)$ values to compute, and each of them is obtained in $O(m)$. Once these values are computed, Equation 3.3 allows us to compute $P(\{c_j, c_k\})$ in $O(m^2)$. This is a significant improvement compared to the brute force implementation in $O(m!)$ of Equation 3.1.

⁸A distance d satisfies the betweenness condition if for all \succ, \succ', \succ'' such that \succ' is between \succ and \succ'' we have $d(\succ, \succ'') = d(\succ, \succ') + d(\succ', \succ'')$.

Chapter 4

Nearly single-peaked preferences : forbidden triples

4.1 Introduction

Although the approach presented in Chapter 3, consisting in generalization of single-peaked preferences on an arbitrary graph, seems quite promising (judging according to experiences held in Section 3.6), it still has some weak points. The main problem consists in the fact that each preference must be *perfectly* compatible with the graph. Therefore, the voters with a very specific preferences may force the presence of edges that are not “useful” for the majority of population. We could consider a weighted version of a graph, where each edge would be labeled with the number of voters for which the edge is necessary to ensure their compatibility with the graph. In such a case, we would probably observe that some edges seem more “relevant” than others, in the sense that their removing would make the structure incompatible for a great proportion of voters. An idea could be to remove a subset of edges in order to obtain a sparser graph. But how do we decide which subset should be removed? Actually, we need to measure the impact of this action - in other words, the distance of the profile from such a modified graph. This leads us to the notion of nearle single-peakedness which will be studied in this chapter.

Actually, removing some edges can be justified as follows: we assume that there is some exact underlying theoretical structure of preferences, which is in practice “noised” by real preferences of voters. In fact, the voters can more or less perceive the underlying structure, or can also more or less agree on it (typically, a voter can globally agree with the structure, with some minor deviations). In this chapter, we assume that the underlying structure is an axis. We will then study the question how to find the axis which is “nearer” to a given preference profile. Let us now set the framework of this topic in a more formal way:

As emphasized by Feld and Grofman (1986), the assumption that preferences are per-

fectly single-peaked is indeed very strong if the alternatives are candidates in an election (the case of numerical alternatives, such as tax levels, is obviously different). Single-peakedness in the strictest sense thus requires that no individual preference deviates (even slightly) from the single-peakedness condition. Given an axis A , the number of rankings consistent with A (i.e., such that condition 2 holds) is 2^{m-1} , over $m!$ possible rankings in total, where m is the number of alternatives. The proportion of consistent rankings within all possible rankings thus quickly becomes tiny when m increases ($2^{m-1}/m! \approx 0.01$ for $m=7$), as well as the likelihood that no voter deviates from this subset of preferences. This observation is also corroborated by the numerical tests carried out by Sui et al. (2013) on 2002 Irish General Election data in Dublin West and Dublin North, where the best axes explain only 2.9% and 0.4% of voters' preferences.

Conitzer (2009) distinguishes between two interpretations of nearly single-peakedness (see e.g. Erdélyi et al. (2013) for a systematic study of nearly single-peaked electorates): an interpretation where preferences are said nearly single-peaked if only a few voters' preferences deviates from a given axis A and the other voters' preferences are perfectly single-peaked w.r.t. A (the numerical tests reported above corresponds to this interpretation); another interpretation where one allows all voters' preferences to deviate to some extent from a given axis A . The distance measure we propose in this chapter falls under the second interpretation, which has been less studied and tested than the first one.

Given an axis A on the candidates and a set \mathcal{P} of preferences, the idea is to measure how far from single-peakedness w.r.t. A each individual preference is. Put another way, each preference in the electorate partially fits with the axis (according to a non-binary measure), and one sums up the degrees of fitness of preferences in \mathcal{P} to obtain the "degree of single-peakedness" of \mathcal{P} w.r.t. A . More precisely, one defines a *distance* to single-peakedness, i.e. the degree is 0 if \mathcal{P} is single-peaked w.r.t. A . We are thus seeking a procedure that returns both a degree of single-peakedness of a profile and an axis that witnesses the obtained value. This outputs allow the analysis of a political landscape, by answering the questions: How close to single-peakedness is an electorate? How the voters perceive the ideological proximities between candidates?

Related work While recognizing perfectly single-peaked preferences is a polynomial time problem (Bartholdi III and Trick, 1986; Doignon and Falmagne, 1994), determining the distance to single-peakedness (according to various measures) is often NP-hard. Various notions of nearly single-peakedness are present in the literature. We briefly review here notions that do not relax the assumption of a one-dimensional axis on all the candidates. Most of them have been introduced and/or studied by Faliszewski et al. (2011), Erdélyi et al. (2013) and Elkind and Lackner (2014). Faliszewski et al. (2011) studied *k-voter deletion single-peakedness*, also known as *partial single-peakedness* in economics (Niemi, 1969). One says that an electorate is *k-voter deletion single-peaked* consistent if all but k of the voters preferences ("maverick" voters) are consistent with a common axis on the candidates. The smallest number k such that there exists an

axis w.r.t. which the electorate is k -voter deletion single-peaked can be viewed as a distance to single-peakedness. Erdélyi et al. (2013) as well as Bredereck et al. (2016) have proved that determining this distance is NP-hard. Elkind and Lackner (2014) have proposed a polynomial time 2-approximation algorithm for this distance, and have established fixed-parameter tractability results (complexity $\mathcal{O}^*(1.28^k)$ if $k < n/2$, and $\mathcal{O}^*(2.08^k)$ if $k \geq n/2$, where n is the number of voters).

Erdélyi et al. (2013) introduced *k-local candidate deletion single-peakedness*. They first defined single-peaked consistency of a partial preference (linear order on a subset of candidates) w.r.t. an axis A on all candidates: a partial preference is single-peaked w.r.t. A if it is single-peaked w.r.t. the axis obtained from A by removing the missing candidates. Then they say that an electorate is k -local candidate deletion single-peaked consistent if, by removing at most k candidates from each preference, one obtains a set of partial preferences that are single-peaked with respect to a common axis. As above, the smallest k for which the property holds can be viewed as a distance. Here again, the authors have proved that determining this distance is an NP-hard problem.

The class of distance measures that is the closest to our work is that of swap distances. Erdélyi et al. (2013) introduced *k-global swaps single-peakedness*, where k is the number of swaps of consecutive candidates that need to be performed in the preferences to make the election single-peaked. Following Faliszewski et al. (2011), they also considered a “local budget” for swaps, i.e., they allow up to k swaps *per vote*. They call *k-local swaps* this notion of nearly single-peakedness. For both notions, Erdélyi et al. (2013) have proved that computing the smallest k enabling to make the election single-peaked is NP-hard. Finally, let us mention the notion of *PerceptionFlip_k single-peakedness* (Faliszewski et al., 2011). An electorate is *PerceptionFlip_k single-peaked* if there exists an axis A such that, for each voter, the axis A can be transformed into an axis A' by at most k swaps of consecutive candidates in A so that the voter’s preference is single-peaked with respect to A' . Erdélyi et al. (2013) have proved that k -local swaps single-peakedness and *PerceptionFlip_k single-peakedness* are equivalent, in the sense that an electorate is k -local swaps single-peaked iff it is *PerceptionFlip_k single-peaked*.

Our contribution The originality of the distance measure we introduce here is that it directly follows from the very definition of Black’s single-peakedness condition. For a given axis on the candidates, it consists in counting the number of violations of the single-peakedness condition in the preferences. We position this new distance with respect to existing measures of nearly single-peakedness. We provide an axiomatic study in order to identify theoretical properties guaranteed (or not guaranteed) by this measure. In particular, we provide some insights on the differences between this measure, k -voter deletion single-peakedness and k -global swap single-peakedness, and we also point out the differences with the exact single-peakedness. Moreover, we show that under a very specific assumptions, the minimization of the distance from single-peakedness in sense of *FT* can be interpreted as a maximum likelihood estimation problem. We then tackle computational complexity: we prove that, as for most of the proposed measures in the literature, computing an axis at minimum distance to a

given preference profile is NP-hard for our measure. We nevertheless propose an exact method to compute such an axis, that turns out to be efficient in practice. Then we present the results of numerical tests on both real and synthetic election data, to evaluate the relevance of the returned axes on the candidates, providing also comparisons with other notions of nearly single-peakedness.

Organization of the chapter The chapter is organized as follows. We introduce the notion of *k-forbidden triples single-peakedness* (*FT* in which follows) in Section 4.2. As said above, this notion consists in counting the number of triples of candidates that violate Black’s single-peakedness condition. In Section 4.3, we give a brief overview of existing measures of nearly single-peakedness, and we compare the *k-forbidden triples single-peakedness* with them.

In which follows, we only focus on nearly single-peakedness measures resulting in a (complete) axis. Namely, we consider the measures of *k-voter deletion single-peakedness*, *k-global swaps single-peakedness*, *k-local swaps single-peakedness* and, of course, *k-forbidden triples single-peakedness*. In Section 4.4, we highlight some common points and differences between *k-forbidden triples single-peakedness* and other notions resulting in a complete axis. In the first time, we provide a set of properties that hold, resp. do not hold, for any considered measure of single-peakedness. Moreover, all these properties hold for a perfectly single-peaked preferences. This gives a reader a first idea of what is lost by relaxing the notion of single-peakedness. We provide then a second set of properties such that each of them holds (or does not hold) for each measure except one. This gives a reader an idea of differences in behaviour of considered measures.

We show in Section 4.5 that under some very specific assumptions, the axis minimizing the number of forbidden triples can be interpreted as a maximum likelihood estimation of a “true” hidden axis A , where each vote is viewed as a noisy estimate of a preference single-peaked w.r.t. A .

In Section 4.6, we study so-called evaluation and consistency problems. Given an axis A , a preference profile \mathcal{P} and a nearly single-peakedness measure X , the evaluation problem consists in determining the distance (in sense of the measure X) of \mathcal{P} from A . The consistency problem decides, given a profile \mathcal{P} , a measure X and an integer k , whether there exists an axis A such that the distance of \mathcal{P} from A is at most k . We show that the evaluation problem is polynomial-time solvable for *FT*, while the consistency problem is NP-complete. Nevertheless, we provide an ILP formulation of the consistency problem which reveals quite efficient in practice, as we will see in Section 4.7.

4.2 Preliminaries

We start by recalling the definition of single-peakedness, as this notion will be central in this chapter.

Definition 4.2.1: Single-peakedness

Let an *axis* A be a total order \triangleleft_A over a set $\mathcal{C} = \{c_1, \dots, c_m\}$ of candidates. Let $>_v$ denote the preference relation (total order) of a voter v over \mathcal{C} . Let c_v^* denote the most preferred candidate of v (also called the *peak* of v), i.e., $c_v^* >_v c$ for all $c \neq c^*$. The preference $>_v$ is *single-peaked with respect to* A if for any $c_i, c_j \in \mathcal{C}$, such that $c_j \triangleleft_A c_i \triangleleft_A c_v^*$ or $c_v^* \triangleleft_A c_i \triangleleft_A c_j$, we have $c_v^* >_v c_i >_v c_j$. A profile \mathcal{P} is said to be *single-peaked with respect to* A if every vote is single-peaked with respect to A .

The subscript A in \triangleleft_A emphasizes that this relation is defined *with respect to an axis* A . We recall that, by abuse of language, the relation \triangleleft_A will be simply denoted by \triangleleft if no confusion is possible.

Another very important notion of this chapter is the so-called *betweenness relation*:

Definition 4.2.2: Betweenness relation

The *betweenness relation induced by an axis* A is the ternary relation R_A defined by:

$$R_A = \{(c_i, c_j, c_k) \in \mathcal{C}^3 : c_i \triangleleft_A c_j \triangleleft_A c_k \text{ or } c_k \triangleleft_A c_j \triangleleft_A c_i\}.$$

Put another way, $(c_i, c_j, c_k) \in R_A$ means that c_j is between c_i and c_k on the axis A (note that c_i, c_j and c_k do not need to be consecutive on A). By definition of R_A , $(c_i, c_j, c_k) \in R_A$ implies $(c_k, c_j, c_i) \in R_A$, and conversely.¹ The notion of *A-forbidden triple* that we introduce below will make it possible to measure the consistency of a profile with an axis:

Definition 4.2.3: A-forbidden triple

Let c_v^* be the peak of a voter v . If $c_v^* >_v c_i >_v c_j$ and $(c_v^*, c_j, c_i) \in R_A$, then the triple $T = (c_v^*, c_i, c_j)$ is called *A-forbidden in* $>_v$.

Counting the number of *A-forbidden triples* in a preference profile \mathcal{P} (i.e., the number of *A-forbidden triples* in $>_v$ for $>_v \in \mathcal{P}$) amounts to counting the number of violations of the definition of single-peakedness with respect to A . We denote by $FT(>_v, A)$ the

¹Actually, we may notice here a (small) abus of language: while we talk about *triples* (i.e., (unordered) sets of 3 elements), the relation R_A is defined by *3-tuples* (i.e., (ordered) sequences of 3 elements). However, we keep the notion of *triple* in which follows, as we are only interested in the middle position of each tuple in practice. Indeed, as we will see, the tuples (c_i, c_j, c_k) and (c_k, c_j, c_i) are counted as one in the notion of *k-forbidden triples* single-peakedness.

number of A -forbidden triples in the preference $>_v$, and by $FT(\mathcal{P}, A)$ the number of A -forbidden triples in the whole profile \mathcal{P} .

Example 4.2.1. Consider the following profile \mathcal{P} of 2 voters over 4 candidates:

$$>_1: (c_1, c_3, c_4, c_2)$$

$$>_2: (c_2, c_1, c_4, c_3)$$

Let us consider the axis $A = c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4$. The preference $>_1$ is not single-peaked with respect to A (see Figure 4.2.1a) due to the candidate c_2 who is placed between the peak c_1 and the two remaining candidates c_3, c_4 , both preferred to c_2 by the voter v_1 . Indeed, we count here two violations of the single-peakedness definition, i.e. two forbidden triples (c_1, c_2, c_3) and (c_1, c_2, c_4) - see Figure 4.2.1a. Hence, we write $FT(>_1, A) = 2$.

In the same manner, we easily check that $FT(>_2, A) = 1$ - the (unique) forbidden triple here is (c_2, c_3, c_4) (see Figure 4.2.1b). Put together, we have $FT(\mathcal{P}, A) = FT(>_1, A) + FT(>_2, A) = 3$.

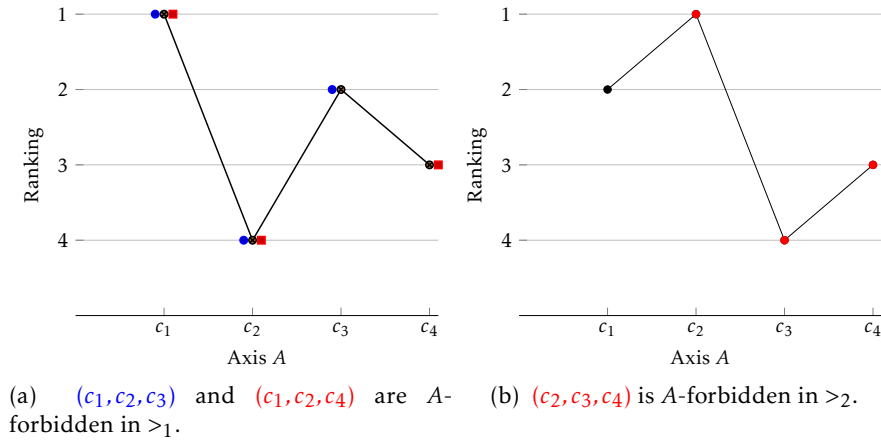


Figure 4.2.1: Illustration of A -forbidden triple definition. Here we have a profile containing three forbidden triples - two forbidden triples in $>_1$ and one forbidden triple in $>_2$.

We can now introduce the notion of k -forbidden triples single-peakedness:

Definition 4.2.4: k -forbidden triples single-peakedness

We say that a profile \mathcal{P} is k -forbidden triples single-peaked consistent if there exists an axis A with respect to which the number of A -forbidden triples in \mathcal{P} is at most k .

We denote by $FT(\mathcal{P})$ the smallest k such that \mathcal{P} is k -forbidden triples single-peaked consistent, and by $A_{FT}(\mathcal{P})$ the axis for which we get $FT(\mathcal{P})$ forbidden triples (actually, there can be several such axes). The distance of \mathcal{P} to single-peakedness is $FT(\mathcal{P})$.

4.3 Positioning with respect to other metrics

The aim of this section is to compare the k -forbidden triples single-peakedness metric with other measures. We start by giving a brief overview of existing metrics of nearly single-peakedness. A more detailed, formal and elaborate version can be found in (Erdélyi et al., 2017):

- **k -Voter Deletion (VD):** A profile \mathcal{P} is k -VD single-peaked with respect to an axis A if there exists a set of k preferences $\{>_{i_1}, >_{i_2}, \dots, >_{i_k}\}$ such that $\mathcal{P} \setminus \{>_{i_1}, >_{i_2}, \dots, >_{i_k}\}$ is single-peaked with respect to A .
- **k -Candidate Deletion (CD):** A profile \mathcal{P} is k -CD single-peaked with respect to an axis A if there exists a subset of k candidates $\mathcal{C}' \subset \mathcal{C}$ such that the profile restricted to $\mathcal{C} \setminus \mathcal{C}'$ is single-peaked with respect to A restricted to $\mathcal{C} \setminus \mathcal{C}'$.
- **k -Local Candidate Deletion (LCD):** A profile \mathcal{P} is k -LCD single-peaked with respect to an axis A if for each voter v there exists a subset of k candidates $\mathcal{C}^v \subset \mathcal{C}$ such that the preference $>_v$ restricted to $\mathcal{C} \setminus \mathcal{C}^v$ is single-peaked with respect to A restricted to $\mathcal{C} \setminus \mathcal{C}^v$.
- **k -Additional Axes (AA):** A profile \mathcal{P} is k -AA single-peaked if there exist k axes A_1, A_2, \dots, A_k such that each preference $>_v$ is single-peaked with respect to at least one of them.
- **k -Global Swaps (GS):** A profile \mathcal{P} is k -GS single-peaked with respect to an axis A if it can be made single-peaked with respect to A by performing at most k swaps of consecutive candidates in the profile.
- **k -Local Swaps (LS):** A profile \mathcal{P} is k -LS single-peaked with respect to an axis A if each preference $>_v$ can be made single-peaked with respect to A by performing at most k swaps of consecutive candidates in it (in other words, we allow k swaps per vote, while we required k swaps for the whole profile in case of GS).
- **k -Candidate Partition (CP):** A profile \mathcal{P} is k -CP single-peaked with respect to an axis A if there exists a partition of \mathcal{C} into k sets $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k$ such that for each $i \in \{1, \dots, k\}$, the profile restricted to \mathcal{C}_i is single-peaked with respect to the axis A restricted to \mathcal{C}_i .
- **k -Clones (CL)** A *clone* is a set of candidates that are ranked consecutively in all preferences (but not necessarily in the same order). By replacing a clone set by a unique candidate, the profile may become single-peaked. With such a substitution (also called *decloning*), we will loose $l - 1$ candidates if the size of the clone set is l .

A profile \mathcal{P} is k -CL single-peaked if there exist k clones such that at most k candidates are lost in the decloning procedure leading to a single-peaked profile.

- **k -Width (WI)** A profile \mathcal{P} is k -WI single-peaked if there exists a set of clone sets leading to a single-peaked profile (when applying the decloning procedure) with each clone set involving at most k candidates.

Erdélyi et al. (2017) then gives the Hasse diagram in which measure X is above measure Y if for each profile \mathcal{P} , we have $Y(\mathcal{P}) \leq X(\mathcal{P})$ - in other words, Y is upper bounded by X , which is denoted by $Y <_l eq X$. Intuitively, when a metric Y is upper bounded by a metric X , the metric X is somehow more expressive, with a larger scale of possible values. For instance, the maximal value taken for VD is obviously $n - 1$, as any profile of only one voter is single-peaked, while for GS , an example where more than $n - 1$ swaps are needed can be found. Indeed, we have $VD \leq GS$, and GS should provide a finer grading of profiles.

The Hasse diagram, which can be found in (Erdélyi et al., 2017), is represented in Figure 4.3.1. It has been proved in the original paper that this diagram is complete in the following sense: if there is no arc (or path) between two metrics X and Y , then one cannot be upper bounded by the other. Proposition 4.3.1 makes it possible to add FT to the diagram in Figure 4.3.1:

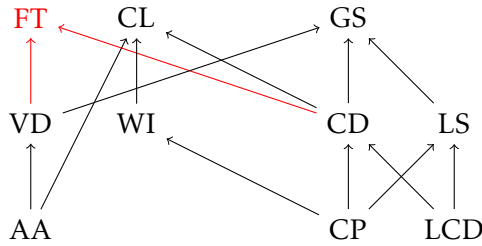


Figure 4.3.1: The Hasse diagram given by Erdélyi et al. (2017): X is above Y (and there is an arc from Y to X) if Y is upper bounded by X . The new metric FT is added in red into the diagram.

Proposition 4.3.1

The metrics upper bounded by FT are VD , CD , AA , CP and LCD .
 For $X \in \{CL, GS, WI, LS\}$, neither X is upper bounded by FT , nor the opposite.

Proof. Let us start by proving the first part of the proposition. By transitivity of Hasse diagram, it is sufficient to prove that VD and CD are upper bounded by FT :

- Let $A_{FT}(\mathcal{P})$ be an optimal axis with respect to which \mathcal{P} is k -FT single-peaked. There are at most k voters that are not single-peaked with respect to $A_{FT}(\mathcal{P})$. Hence, the profile is k -VD single-peaked.
- Let \mathcal{P} be k -CD single-peaked with respect to an axis A . There are k candidates $c_{i_1}, c_{i_2}, \dots, c_{i_k}$ that must be removed to make \mathcal{P} single-peaked w.r.t. A . That

means that each of these candidates violates at least once the definition of single-peakedness: for each $l \in \{1, \dots, k\}$ there is a voter v such that $c_v^* >_v c_j >_v c_{i_k}$ with $(c_v^*, c_{i_k}, c_j) \in R_A$. In other words, there is at least one forbidden triple caused by each of the removed candidates.

For $X \in \{CL, GS, WI, LS\}$, we need now to find two profiles \mathcal{P}_X and \mathcal{P}'_X such that $X(\mathcal{P}_X) < FT(\mathcal{P}_X)$ and $FT(\mathcal{P}'_X) < X(\mathcal{P}'_X)$:

- $X = CL$ or WI : To show that we can have $X(\mathcal{P}) < GS(\mathcal{P})$, Erdélyi et al. (2017) consider the following profile of $3n$ voters over 3 candidates:

$$\begin{aligned} n \text{ voters} &: (c_1, c_2, c_3) \\ n \text{ voters} &: (c_1, c_3, c_2) \\ n \text{ voters} &: (c_2, c_3, c_1) \end{aligned}$$

We now show that this counterexample still works if we replace GS by FT . There are indeed 3 candidates ranked last, so no matter the axis, there will always be n voters not compatible with it (we recall that in any single-peaked profile, at most two candidates can be ranked last). We have therefore $FT(\mathcal{P}) \geq n$. On the other hand, the profile becomes single-peaked if we declone the clone set $\{c_2, c_3\}$ - we have $CL(\mathcal{P}) = WI(\mathcal{P}) = 1$.

Similarly, the profile \mathcal{P}' proposed by Erdélyi et al. (2017) to show that we can have $GS(\mathcal{P}') < X(\mathcal{P}')$ also shows that we can have $FT(\mathcal{P}') < X(\mathcal{P}')$. For convenience, we recall their counterexample here - it is a profile of 4 preferences over m candidates:

$$\begin{aligned} >_1 &: (c_1, c_2, \dots, c_{m-1}, c_m) \\ >_2 &: (c_m, c_{m-1}, \dots, c_2, c_1) \\ >_3 &: (c_1, c_2, \dots, c_{m-2}, c_m, c_{m-1}) \\ >_4 &: (c_{m-1}, c_{m-2}, \dots, c_1, c_m) \end{aligned}$$

Again, there are 3 candidates ranked last, so the profile is not single-peaked. However, it is 1-FT single-peaked w.r.t. $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$, as there is only one forbidden triple (c_1, c_m, c_{m-1}) for the preference $>_3$. Regarding the metrics CL and WI , at least two of the three candidates ranked at least once in last position (i.e., c_1, c_m and c_{m-1}) must be in the same clone set to make the profile single-peaked. It is easy to check that in such a case, $CL(\mathcal{P}') \geq m - 1$ and $WI(\mathcal{P}') \geq m - 1$.

- $X = GS$ or LS : Let us consider the profile \mathcal{P} of $2n + 1$ preferences over 6 voters (with n “sufficiently great”²) - see Figure 4.3.2a for an illustration:

$$\begin{aligned} n \text{ voters} &: (c_1, c_2, c_3, c_4, c_5, c_6) \\ n \text{ voters} &: (c_6, c_5, c_4, c_3, c_2, c_1) \\ >_{2n+1} &: (c_6, c_1, c_2, c_3, c_4, c_5) \end{aligned}$$

²The exact value will be given below.

The profile \mathcal{P} is not single-peaked, as there are 3 candidates ranked last. The second group preference is the reverse order of the first group preference - therefore, the only axis (up to a reversal) w.r.t. which these two groups of preferences are both single-peaked is $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_6$. For any other axis, one of these groups will not be compatible, so we will have at least n forbidden triples or swaps. Hence, if n is greater than the distance of the preference $>_{2n+1}$ from A , A is the unique optimal solution (up to a reversal).

Regarding GS, one can easily check that the optimal way (in terms of number of swaps) to make $>_{2n+1}$ compatible with A is to swap c_6 consecutively with c_5, c_4, c_3, c_2 and c_1 - in other words, the most preferred candidate of v_{2n+1} becomes her worst ranked candidate. We have therefore $GS(\mathcal{P}) = 5$. As $>_{2n+1}$ is the only preference not compatible with A , we also have $LS(\mathcal{P}) = 5$.

On the other hand, (c_6, c_i, c_j) is a forbidden triple for each $i < j \leq 5$. Hence, there are 10 forbidden triples. Thus, $FT(\mathcal{P}) = 10$, and $GS(\mathcal{P}) < FT(\mathcal{P})$ (resp. $LS(\mathcal{P}) < FT(\mathcal{P})$).

We give now a profile \mathcal{P}' such that $FT(\mathcal{P}') < GS(\mathcal{P}')$ (resp. $FT(\mathcal{P}') < LS(\mathcal{P}')$). For this purpose, let us consider the following profile of $2n+1$ preferences (again, with n sufficiently great) over 6 voters (see Figure 4.3.2b for an illustration):

$$\begin{aligned} n \text{ voters} &: (c_1, c_2, c_3, c_4, c_5, c_6) \\ n \text{ voters} &: (c_6, c_5, c_4, c_3, c_2, c_1) \\ >_{2n+1} &: (c_3, c_1, c_4, c_5, c_6, c_2) \end{aligned}$$

As in the previous case, for n sufficiently great, the axis $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_6$ is the only optimal axis (up to a reversal) for both FT and GS, resp. LS, metrics. Indeed, the only preference not compatible with A is $>_{2n+1}$. The only forbidden triple is (c_3, c_2, c_1) , so $FT(\mathcal{P}') = 1$. On the other hand, 4 swaps need to be done in order to make this preference compatible with A - c_2 needs to be consecutively swapped with c_6, c_5, c_4 and c_1 . We have hence $GS(\mathcal{P}') = 4$. This reasoning also holds for LS metrics, as there is only one preference not compatible with A . Therefore, we have $FT(\mathcal{P}') < GS(\mathcal{P}')$ (resp. $FT(\mathcal{P}') < LS(\mathcal{P}')$).

□

A Curious Tortoise Intervention



This proof points out some differences and particularities in the behaviour of both FT and GS metrics: To prove that $GS < FT$, we use the preference $>_{2n+1}: (c_6, c_1, c_2, c_3, c_4, c_5)$ on the axis $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_6$. If the candidate c_6 was deleted, the restricted preference would be perfectly single-peaked w.r.t. A . So the voter agrees with the structure imposed by

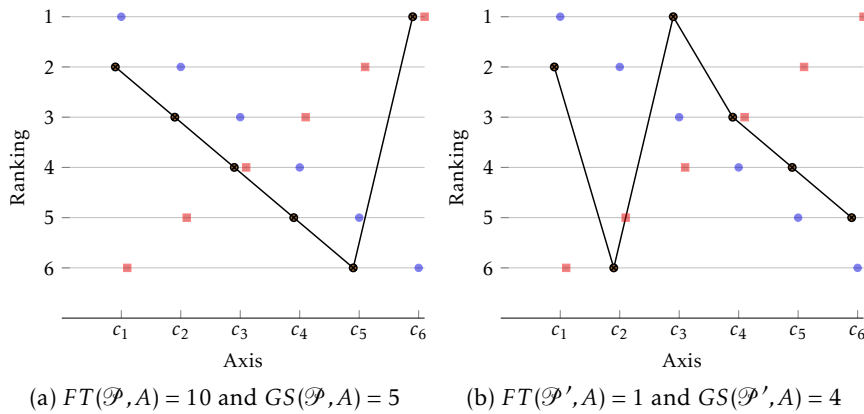


Figure 4.3.2: Two profiles showing that FT cannot be upper bounded by GS (resp. LS) and vice-versa.



A on this subset of candidates. But she must rank the candidate c_6 as the last one to stay compatible with this structure. This seems to be a strong contradiction, as c_6 is actually her most preferred candidate - it seems quite irrational to agree with a structure which ranks your most preferred candidate as your worst one. More generally, a voter is rarely wrong about her most preferred candidate, but GS is here moving it down in the preference to make it compatible with A . From this point of view, this preference seems intuitively quite far from A . The FT metrics seems to better take into account this issue: basically, it takes two consecutively ranked candidates c_i, c_j (on the same side of the peak), and it counts the number of candidates ranked worse than both c_i, c_j and lying between them on the axis - visually, the wider the “depression” between c_i, c_j is, the less the preference fits to the axis. See Figure 4.3.3 for an illustration.

On the other hand, while FT considers the “width” of incompatibilities with the axis, it is blind to the depth of them. This is used above to prove that $FT < GS$. We have here the preference $\succ_{2n+1}: (c_3, c_1, c_4, c_5, c_6, c_2)$ and the axis $A = c_1 \prec c_2 \prec \dots \prec c_6$. We note that the worst candidate of this voter, i.e. c_2 , is ranked between her two most preferred candidates c_1 and c_3 . This is intuitively quite a big deviation from the structure, however, the FT metrics only counts forbidden triples without taking into account the difference between ranking positions of candidates forming a given triple. For instance, the preferences $(c_3, c_1, c_4, c_5, c_6, c_2)$ and $(c_3, c_1, c_2, c_5, c_6, c_4)$ contain both a unique forbidden triple (c_3, c_2, c_1) and (c_3, c_4, c_1) w.r.t. A , but in-



tuitively, the second preferences seems to be more compatible with A than the first one. Note that the metric GS considers this issue. See Figure 4.3.4 for an illustration.

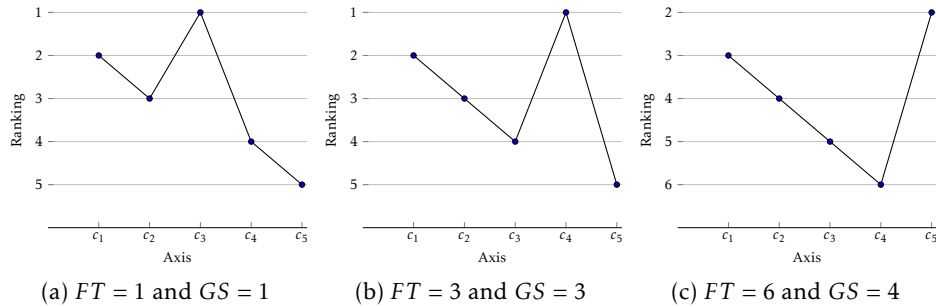


Figure 4.3.3: The FT measure penalizes a lot “wider depressions” caused by forbidden triples. This issue seems to be less important for the GS measure.

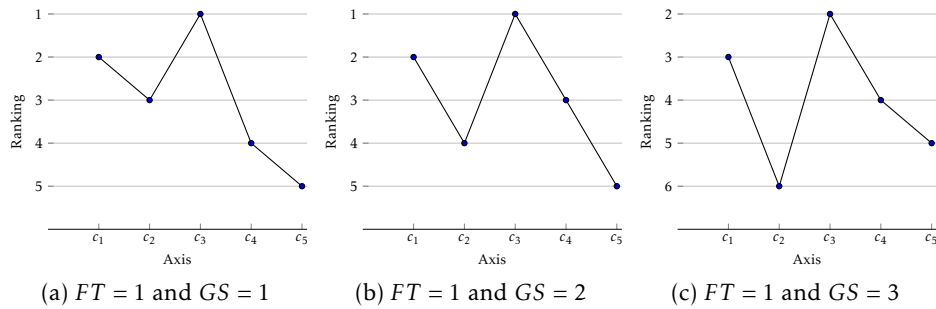


Figure 4.3.4: The FT measure does not take into account the depth of triples, contrary to the GS measure. This is due to the fact that the FT measure treats independently the left side and the right side of the peak.

4.4 Axiomatic Examination

We have seen at the end of the previous section some differences between GS and FT . In this section, our aim is to tackle this question in a more formal way. Concretely, we highlight some common points and differences between k -forbidden triples single-peakedness and other notions of nearly single-peakedness. For comparison purposes, we focus only on the notions resulting in a complete axis - hence, the notions considered here are k -voter deletion single-peakedness (VD), k -global swaps single-peakedness (GS), k -local swaps single-peakedness (LS), and of course k -forbidden triples single-

peakedness (FT).³ We first focus on common points of all considered metrics: we give a list of natural and desirable properties (here called axioms) one could expect to hold for a nearly single-peakedness metric. Actually, we will see that all these properties hold for single-peaked profiles, but some of them do not hold anymore for any of the considered nearly single-peakedness metric. This gives a reader an idea of what is lost by relaxing the structure of single-peakedness.

In a second time, we focus on differences between the considered metrics. In particular, we give a set of three properties such that each of them holds (or does not hold) for each metrics except one. It turns out that these properties allow us to differentiate the metrics, as there are no two metrics verifying the same subset of them. Again, all of these properties hold for single-peaked profiles, which helps us to better understand what is lost using each of the metrics.

We introduce now some notations used in what follows. Given a distance measure X , a profile \mathcal{P} and an axis A , we denote by $X(\mathcal{P}, A)$ the distance of \mathcal{P} to single-peakedness w.r.t. A . Given X and \mathcal{P} , an optimal axis is one that minimizes $X(\mathcal{P}, A)$. By abuse of notation, we write $X(>_v, A)$ for $X(\{>_v\}, A)$.

For the convenience of the reader, we formally define $GS(\mathcal{P}, A)$, $LS(\mathcal{P}, A)$ and $VD(\mathcal{P}, A)$:

$$\begin{aligned} GS(\mathcal{P}, A) &= \sum_{>_v \in \mathcal{P}} GS(>_v, A), \\ LS(\mathcal{P}, A) &= \max_{>_v \in \mathcal{P}} LS(>_v, A), \\ VD(\mathcal{P}, A) &= \sum_{>_v \in \mathcal{P}} VD(>_v, A), \end{aligned}$$

where $GS(>_v, A)$ (resp. $LS(>_v, A)$) is the minimum number of swaps of consecutive candidates required to make $>_v$ single-peaked w.r.t. A , and $VD(>_v, A) = 0$ if $>_v$ is single-peaked w.r.t. A , otherwise $VD(>_v, A) = 1$. We denote by $swap(c, c')$ the swap of two candidates c, c' that are consecutive in $>_v$.

Example 4.4.1. *To illustrate these measures, we go back to the profile \mathcal{P} from Example 4.2.1:*

$$\begin{aligned} >_1: & (c_1, c_3, c_4, c_2) \\ >_2: & (c_2, c_1, c_4, c_3) \end{aligned}$$

We recall that with respect to the axis $A = c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4$, we have $FT(>_1, A) = 2$ and $FT(>_2, A) = 1$ (see Figure 4.2.1). Let us now determine $X(>_1, A)$, $X(>_2, A)$ and $X(\mathcal{P}, A)$ for $X \in \{GS, LS, VD\}$:

³Note that other metrics mentioned above do not result in a complete axis: for instance, the candidate deletion metric (CD) will give an axis on a subset of candidates and this axis can be completed by adding removed candidates arbitrarily - this leads to many different axis (all equivalent for CD); it is not obvious how to compare such a class of axes with the output given by FT (which is a complete axis).

- $X = VD$: Obviously, we have $VD(>_i, A) = 1$ for $i \in \{1, 2\}$, as none of the preferences is compatible with A . $VD(\mathcal{P}, A) = 2$ - at least two preferences need to be removed to make the profile compatible with A .
- $X = GS$: We have $GS(>_1, A) = 2$: making $\text{swap}(c_2, c_4)$, we obtain a preference $>_1'$: (c_1, c_3, c_2, c_4) . Applying $\text{swap}(c_2, c_3)$ on $>_1'$ yields a preference $>_1''$: (c_1, c_2, c_3, c_4) single-peaked with respect to A . It is easy to check that we cannot make $>_1$ compatible with A by performing only one swap, so $GS(>_1, A) = 2$.

We can make $>_2$ compatible with A by swapping c_3 and c_4 . Therefore, $GS(>_2, A) = 1$. We have $GS(\mathcal{P}, A) = 2 + 1 = 3$.

- $X = LS$: We have $LS(>_i, A) = GS(>_i, A)$, as the both measures count the number of swaps of consecutive candidates required to make $>_i$ single-peaked w.r.t. A . However, $LS(\mathcal{P}, A) = \max\{LS(>_1, A), LS(>_2, A)\} = 2$.

4.4.1 Common points of nearly single-peakedness measures and comparison with single-peakedness

As said above, we will first introduce three properties which either hold for FT , GS , LS and VD together, or for none of them:

Property 1 (Reinforcement). Let \mathcal{P}_1 and \mathcal{P}_2 be two preference profiles on the same set of candidates (with $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$). Suppose that there exists an axis A^* optimal for both of them. Then A^* is also optimal for the profile $\mathcal{P}_1 \cup \mathcal{P}_2$.

Property 2 (Hereditiy). Let \mathcal{P} be a profile over a set \mathcal{C} of candidates and $\mathcal{C}' \subset \mathcal{C}$. If an axis is optimal for \mathcal{P} , its restriction to \mathcal{C}' stays optimal for the profile restriction to \mathcal{C}' .

Property 3 (Monotony). Let $>_v, >_{v'}$ be two preference relations over \mathcal{C} with the same peak, let X be a measure of single-peakedness and let $X(>_v, A) \leq X(>_{v'}, A)$ be their respective distances to single-peakedness w.r.t. an axis A . Let $>_{v''}$ be another preference relation such that $>_{v''}$ is between $>_v$ and $>_{v'}$ (which means that $a >_v b$ and $a >_{v'} b \Rightarrow a >_{v''} b$). Then:

$$X(>_v, A) \leq X(>_{v''}, A) \leq X(>_{v'}, A).$$

As all these properties hold for single-peaked preferences, we get an idea of what is lost by relaxing single-peakedness (at least for the existing distance measures in the literature). For instance, from the computational viewpoint, the failure of the hereditiy property prevents the development of a dynamic programming procedure based on examining subsets of candidates for minimizing $X(\mathcal{P}, A)$.

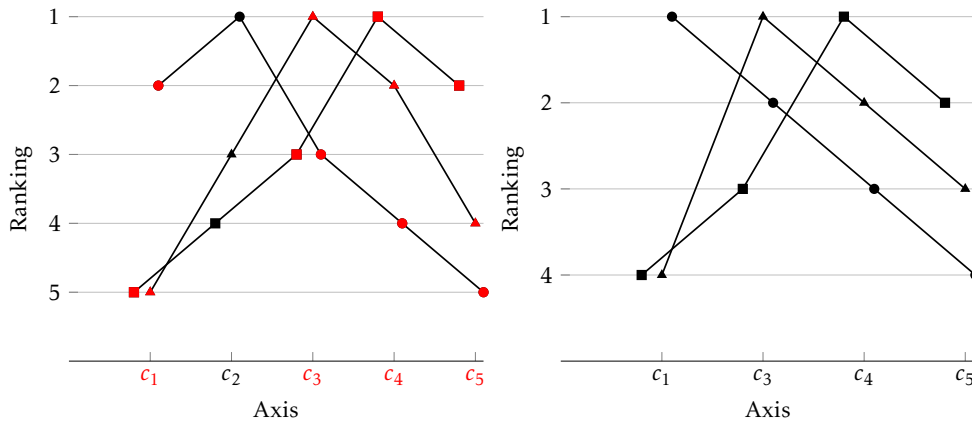


Figure 4.4.1: A subprofile of a single-peaked profile remains single-peaked.

A Curious Tortoise Intervention



Let us comment these properties in a less formal way. The reinforcement property simply states that if two different groups of voters agree on their optimal axis, it remains optimal in the profile that merges these two groups together. This property seems very natural and actually holds for any considered measure.

The heredity property basically states that a subsolution of an optimal solution should remain optimal. While it may seem appealing at first sight and it is verified by single-peaked profiles, we will see that it does not hold for nearly single-peaked measures.

Finally, the monotony property assumes that we have two preferences $>_v$ and $>_{v'}$, and a preference $>_{v''}$ between them. Intuitively, one might see $>_{v''}$ as a “compromise” between $>_v$ and $>_{v'}$. It seems then natural that the distance of $>_{v''}$ from the axis A would also be between $X(>_v, A)$ and $X(>_{v'}, A)$. In other words, one could expect the distance function to be monotonous while “transforming” $>_v$ into $>_{v'}$. However, this is not true for any of the considered notions of nearly single-peakedness.

We first show that these three properties hold for single-peaked profiles, i.e. for profiles \mathcal{P} such that $X(\mathcal{P}, A^* = 0)$ for any single-peakedness measure X and an optimal axis A^* :

- *Reinforcement*: If all the preferences in \mathcal{P}_1 and in \mathcal{P}_2 are single-peaked w.r.t. A , then $\mathcal{P}_1 \cup \mathcal{P}_2$ is obviously single-peaked w.r.t. A .
- *Heredity*: If a preference relation $>$ over \mathcal{C} is single-peaked w.r.t. an axis A , then its restriction to \mathcal{C}' is single-peaked w.r.t. the restriction of A to \mathcal{C}' .

- *Monotony*: As $>$ and $>'$ share the same peak and $>''$ is between $>$ and $>'$, the peak of $>''$ is the same as the one of $>$ and $>'$; for the pairs of candidates that are both on the left or on the right side of the peak on A , $>$ and $>'$ share the same preference consistent with A , and therefore also $>''$; hence $>''$ is single-peaked w.r.t. A .

Let us now see what happens for different notions of nearly single-peakedness:

Proposition 4.4.1

Reinforcement holds for any $X \in \{FT, GS, LS, VD\}$. On the other hand, Heredity and Monotony do not hold for any of these measures of nearly single-peakedness.

Proof.

- **Reinforcement:**

Assume that A^* is an optimal axis for $(\mathcal{C}, \mathcal{P}_1)$ and $(\mathcal{C}, \mathcal{P}_2)$. Let A be an arbitrary chosen axis. For $X \in \{FT, GS, VD\}$, $X(\mathcal{P}, A) = \sum_{> \in \mathcal{P}} X(>, A)$, thus $X(\mathcal{P}_1 \cup \mathcal{P}_2, A) = X(\mathcal{P}_1, A) + X(\mathcal{P}_2, A)$.

By optimality of A^* for \mathcal{P}_1 and \mathcal{P}_2 , we have $X(\mathcal{P}_1, A^*) \leq X(\mathcal{P}_1, A)$ and $X(\mathcal{P}_2, A^*) \leq X(\mathcal{P}_2, A)$. Consequently, $X(\mathcal{P}_1 \cup \mathcal{P}_2, A^*) \leq X(\mathcal{P}_1 \cup \mathcal{P}_2, A)$, hence A^* is optimal for $(\mathcal{C}, \mathcal{P}_1 \cup \mathcal{P}_2)$. For LS , we have that $LS(\mathcal{P}, A) = \max_{> \in \mathcal{P}} LS(>, A)$ and $LS(\mathcal{P}_1 \cup \mathcal{P}_2, A) = \max\{LS(\mathcal{P}_1, A), LS(\mathcal{P}_2, A)\}$. The proof is then similar.

- **Heredity:**

– To prove that the property is false for FT and GS , consider the following profile \mathcal{P} with 5 voters and 4 candidates:

$$\begin{aligned} >_1: (c_4, c_3, c_2, c_1) \\ >_2: (c_4, c_3, c_2, c_1) \\ >_3: (c_1, c_2, c_3, c_4) \\ >_4: (c_3, c_4, c_1, c_2) \\ >_5: (c_4, c_3, c_1, c_2) \end{aligned}$$

By enumerating all 12 axis (up to a reversal), we find out that the unique optimal solution is the axis $A^* = c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4$. For $X \in \{FT, GS\}$, we have $dX(\mathcal{P}, A^*) = 2$ due to $>_4$ and $>_5$ (containing forbidden triples (c_3, c_2, c_1) , resp. (c_4, c_2, c_1)). However, if we delete the candidate c_4 , the restricted axis $A' = c_1 \triangleleft c_2 \triangleleft c_3$ is no more optimal. Denoting by \mathcal{P}' the profile \mathcal{P} where c_4 has been deleted, we still have indeed $X(\mathcal{P}', A') = 2$, while $X(\mathcal{P}', A) = 1$ for the axis A defined by $c_1 \triangleleft_A c_3 \triangleleft_A c_2$ (only $>_3$ is problematic).

– To prove that the property is false for *LS*, consider the profile \mathcal{P} with 2 preferences over 5 candidates:

$$\begin{aligned} >_1: (c_5, c_1, c_2, c_3, c_4) \\ >_2: (c_5, c_4, c_3, c_2, c_1). \end{aligned}$$

Consider the axis $A = c_2 \prec c_1 \prec c_5 \prec c_4 \prec c_3$. As \mathcal{P} is not single-peaked and $LS(\mathcal{P}, A) = 1$ (by swapping c_3 and c_4 in $>_1$, and c_1 and c_2 in $>_2$), the axis A is optimal for *LS*. However, if we delete c_5 from \mathcal{P} , the restricted axis $A' = c_2 \prec c_1 \prec c_4 \prec c_3$ is no more optimal. Denoting by \mathcal{P}' the profile \mathcal{P} where c_5 has been deleted, we still have indeed $LS(\mathcal{P}', A') = 1$, while \mathcal{P}' is single-peaked w.r.t. $c_1 \prec c_2 \prec c_3 \prec c_4$.

– To prove that the property is false for *VD*, consider a profile \mathcal{P} with $n = 4p + 3$ voters (where p is a multiple of 3) and 12 candidates:

$$\begin{aligned} p \text{ voters of type 1 : } & (c_1, c_3, c_2, c_6, c_4, c_5, c_7, c_9, c_8, c_{10}, c_{12}, c_{11}) \\ p \text{ voters of type 2 : } & (c_2, c_3, c_1, c_4, c_6, c_5, c_7, c_9, c_8, c_{10}, c_{12}, c_{11}) \\ 2p/3+1 \text{ voters of type 3 : } & (c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}, c_{12}) \\ 2p/3+1 \text{ voters of type 4 : } & (c_1, c_2, c_3, c_5, c_6, c_4, c_8, c_9, c_7, c_{10}, c_{12}, c_{11}) \\ 2p/3+1 \text{ voters of type 5 : } & (c_1, c_2, c_3, c_5, c_4, c_6, c_8, c_9, c_7, c_{11}, c_{12}, c_{10}). \end{aligned}$$

A profile cannot be single-peaked if there exists a triple of voters v_i, v_j, v_k and a triple of candidates c, c', c'' such that c (resp. c', c'') is the least preferred in $\{c, c', c''\}$ according to $>_i$ (resp. $>_j, >_k$)⁴. It implies that three different types of voters at least have to be deleted to make the election single-peaked since:

- voters of types $\{1, 2, 3\}$, $\{1, 2, 4\}$ and $\{1, 2, 5\}$ are made incompatible by triple c_1, c_2, c_3 ;
- voters of types $\{2, 3, 4\}$, $\{1, 3, 4\}$, $\{2, 4, 5\}$ and $\{1, 4, 5\}$ are made incompatible by triple c_4, c_5, c_6 ;
- voters of types $\{2, 3, 5\}$ and $\{1, 3, 5\}$ are made incompatible by triple c_7, c_8, c_9 ;
- voters of types $\{3, 4, 5\}$ are made incompatible by triple c_{10}, c_{11}, c_{12} .

The optimal solution consists in deleting all voters of types 3, 4 and 5 (the less numerous ones); the voters of types 1 and 2 are for instance single-peaked w.r.t. axis A defined by

$$c_6 \prec c_1 \prec c_3 \prec c_2 \prec c_4 \prec c_5 \prec c_7 \prec c_9 \prec c_8 \prec c_{10} \prec c_{12} \prec c_{11},$$

hence $VD(\mathcal{P}, A) = 2p + 3$. Now, assume that candidate c_{12} is removed from the election, and denote by \mathcal{P}' the obtained profile. Then the voters of types $\{3, 4, 5\}$

⁴We recall that this result is a consequence of the characterization of the single-peaked domain by forbidden minors - see (Ballester and Haeringer, 2011)).

become compatible in \mathcal{P}' ; they are for instance single-peaked w.r.t. axis A' defined by

$$c_{10} \triangleleft c_7 \triangleleft c_6 \triangleleft c_5 \triangleleft c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4 \triangleleft c_8 \triangleleft c_9 \triangleleft c_{11},$$

hence $VD(\mathcal{P}', A') = 2p$. Note that the other triples of types remain incompatible, thus A' is optimal for \mathcal{P}' . The voters of types $\{3, 4, 5\}$ are not single-peaked w.r.t. the axis obtained from A by removing c_{12} , thus the restriction of A to $\mathcal{C} \setminus \{c_{12}\}$ is suboptimal for \mathcal{P}' .

• **Monotony:**

To prove that the property does not hold for any of the studied measures, consider the axis $A = c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4 \triangleleft c_5$ and the following preferences (see also Figure 4.4.2):

$$>_1: (c_3, c_2, c_1, c_4, c_5, c_6)$$

$$>_2: (c_3, c_2, c_1, c_6, c_5, c_4)$$

$$>_3: (c_3, c_1, c_2, c_4, c_5, c_6)$$

It can easily be checked that $>_1$ is between $>_2$ and $>_3$. For $X \in \{FT, GS, LS, VD\}$, we have $X(>_1, A) = 0$, $X(>_2, A) = 3$ and $X(>_3, A) = 1$, which contradicts the property.

A Curious Tortoise Intervention



The counter-example of this proof (see also Figure 4.4.2) points out why the first intuition was wrong: the preference $>_2$ is compatible with the axis on the left of the peak, while the preference $>_3$ is compatible with the axis on the right of the peak. The preference $>_1$ that is between $>_2$ and $>_3$ is therefore single-peaked with respect to the axis. Indeed, the issue is that the left-peak and right-peak side of preferences are somehow independent, so there are two independent ways to be incompatible with the axis.

□

4.4.2 Differences between nearly single-peakedness measures

In this section, we will study another three properties in order to give some insights on the differences between VD , GS , LS and FT . Actually, each of these properties holds for a single distance measure among the VD , GS , LS and FT . As already said above, we note that all these properties hold for (perfectly) single-peaked preferences, which gives us a more precise idea of what is lost using each of the metrics.

Property 4 (Unpopularity). *Let \mathcal{P} be a preference profile and $c \in \mathcal{C}$ a candidate such that:*

- *c is never ranked in first position;*

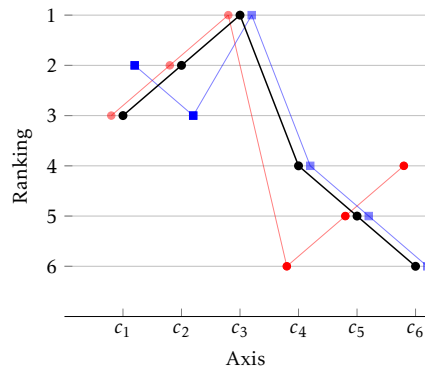


Figure 4.4.2: The monotony property does not hold for nearly single-peakedness: $>_2$ is compatible with the axis on the left of the peak c_3 and $>_3$ is compatible with the axis on the right of the peak c_3 , so $>_1$ (that is between $>_2$ and $>_3$) is perfectly single-peaked.

- c is ranked in last position by more than half of the voters.

Then there exists an optimal axis where c is one of the extremities.

Property 5 (Archimedean). Let \mathcal{P} be a preference profile, X a single-peaked measure and A an axis. If k preference relations in \mathcal{P} are not single-peaked w.r.t. A , then $X(\mathcal{P}, A) \geq k$.

Property 6 (Stability). Let \mathcal{P} be a preference profile and $>$ a preference relation on \mathcal{C} . There exists an optimal axis that stays optimal for the profile $\mathcal{P} \cup \{>\}$. In particular, if the optimal axis for \mathcal{P} is unique, it stays optimal for $\mathcal{P} \cup \{>\}$.

A Curious Tortoise Intervention



The unpopularity property can be seen as a generalization of the well-known result that in a single-peaked profile, a candidate ranked last at least once is necessarily an extremity of any optimal axis. This is in fact a very strong constraint, as in a real-world election with a great number of voters, it is very unlikely only two candidates are ranked last by at least one voter.

This restriction is circumvented by considering nearly single-peaked preferences. It is interesting to note that in the case of the *FT* measure, there is still some property related to the unpopularity of candidates. The unpopularity property states that beyond a certain level of unpopularity (a candidate never ranked in first position and often last) a candidate can hardly be viewed as intermediate between others, and thus there should be an optimal axis where she is an extreme candidate. Note that, by definition, there is at most one unpopular candidate. The *FT* measure is the only one for which the property holds.

	<i>FT</i>	<i>GS</i>	<i>LS</i>	<i>VD</i>
Reinforcement	True	True	True	True
Heredity	False	False	False	False
Monotony	False	False	False	False
Unpopularity	True	False	False	False
Archimedean	True	True	False	True
Stability	False	False	False	True

Table 4.4.1: Properties of distance measures.



The archimedean property states that $X(\mathcal{P}, A)$ is lower bounded by the number of preferences not compatible with A - intuitively, the distance of a profile from an axis strictly increases with the number of non compatible preferences. The *LS* measure is the only one for which the property does not hold because it is not defined from a sum operation over the preferences but from a maximum. This behaviour is specific to local measures and can be an advantage as well as an inconvenient: the good point is that a local measures will most likely tend to find a more consensual axis: an axis with respect to which almost all the preferences are almost single-peaked may be more interesting than an axis which is perfect for a part of the voters and non relevant for another part. On the other hand, the problem of local measures is that they do not take into account the size of these parts - an axis perfectly compatible with the great majority of preferences and not compatible with a very few preferences may be more suitable than an axis from which all preferences are quite near but none of them is perfectly compatible.

The stability property requires some robustness of the solution: the preference relation of a single voter should not change a strictly optimal axis into a suboptimal one. This property may seem quite desirable, especially for large profiles. However, it only holds for the *VD* measure because each preference relation is viewed as either compatible or not with an axis, while the other measures use a degree of compatibility of each preference relation with the considered axis.

An overview of which properties are satisfied by which measures is given in Table 4.4.1. The proofs are detailed below.

Proposition 4.4.2: Unpopularity

The unpopularity property only holds for the FT measure, and does not hold for GS, LS nor VD .

Proof.

- Let us first prove that this property holds for FT . Let c be a candidate never ranked in first position, and ranked in last position by at least half of the voters. Let A be an arbitrary axis such that c is not one of its extremities. Let us denote by n_1 the number of candidates on the left of c in A , and by n_2 the number of candidates on the right of c . We define two axes A_l and A_r obtained from A by putting c respectively on the extreme left position of A_l , and on the extreme right position of A_r . We prove that at least one of the axes A_l, A_r is at least as good as A . To do so, for each voter v , we count the difference of the number of forbidden triples with respect to A and with respect to A_l and A_r . It consists in counting for each of the axes the number of triples involving the candidate c . In fact, as c is never ranked first and the restrictions of A, A_l and A_r on $\mathcal{C} \setminus \{c\}$ lead to the same axis, we observe that the triple (c_i, c_j, c_k) (with c_i, c_j, c_k different from c) is forbidden with respect to A if and only if it is forbidden with respect to A_l (resp. A_r). Let v be an arbitrary voter and c^* be the most preferred candidate in $>_v$. Four configurations are possible:

- (i) v ranks c in last position and c^* is on the left of c in A ;
- (ii) v ranks c in last position and c^* is on the right of c in A ;
- (iii) v does not rank c in last position and c^* is on the left of c in A ;
- (iv) v does not rank c in last position and c^* is on the right of c in A .

Table 4.4.2 expresses $FT(>_v, A_l)$ and $FT(>_v, A_r)$ in function of $FT(>_v, A)$ - if the exact value cannot be given, an upper bound (representing the worst case) is given. For v of type (i), (c^*, c, c') is forbidden (w.r.t. A) if and only if c' is on the right of c in A (there are n_2 such positions). The candidate c is not involved in any forbidden triple with respect to an axis A_l or A_r , as it is placed on the extremity. The same reasoning applies to a voter v of type (ii). For a voter v of type (iii), in the worst case c is not involved in any forbidden triple with respect to A , but moving it on the left (resp. right) extremity will create up to $n_1 - 1$ (resp. n_2) new forbidden triples. We reason the same way for type (iv).

type	$FT(>_v, A_l)$	$FT(>_v, A_r)$
(i)	$FT(>_v, A) - n_2$	$FT(>_v, A) - n_2$
(ii)	$FT(>_v, A) - n_1$	$FT(>_v, A) - n_1$
(iii)	$\leq FT(>_v, A) + n_1 - 1$	$\leq FT(>_v, A) + n_2$
(iv)	$\leq FT(>_v, A) + n_1$	$\leq FT(>_v, A) + n_2 - 1$

Table 4.4.2: Values of $FT(>_v, A_l)$ and $FT(>_v, A_r)$ in function of $FT(>_v, A)$, according to the type of v .

Assume first that $n_1 \leq n_2$. We prove that A_l is always at least as good as A , i.e. $FT(\mathcal{P}, A_l) \leq FT(\mathcal{P}, A)$, which is written:

$$\sum_{>_v \in \mathcal{P}} FT(>_v, A_l) \leq \sum_{>_v \in \mathcal{P}} FT(>_v, A).$$

Thanks to Table 4.4.2 it is sufficient to prove that:

$$m^{(i)}n_2 + m^{(ii)}n_1 \geq (m^{(iii)} + m^{(iv)})n_1 - m^{(iii)}$$

where m^t is the number of voters of type t . By assumption, $m^{(i)} + m^{(ii)} \geq \frac{m}{2}$. As we assume that $n_1 \leq n_2$, the inequality holds.

If $n_1 \geq n_2$, we prove in the same manner that A_r is always at least as good as A .

- Let us now prove that the property is not true for *GS*. Let us consider a preference profile \mathcal{P} over 5 candidates such that there are :
 - 2 voters with preference $(c_1, c_2, c_3, c_4, c_5)$,
 - 3 voters with preference $(c_1, c_4, c_3, c_5, c_2)$,
 - 2 voters with preference $(c_3, c_4, c_1, c_5, c_2)$,
 - 2 voters with preference $(c_5, c_2, c_1, c_4, c_3)$.

By enumerating all 24 axes (up to a reversal) having the unpopular candidate 2 on one extremity, we can prove that the axis $c_5 \triangleleft c_2 \triangleleft c_1 \triangleleft c_4 \triangleleft c_3$ is (strictly) better than any of these. We note that this counter-example works also to prove that the property is false for *LS*.

- Finally, to prove that the property is false for *VD*, let us consider the following

counter-example :

- $>_1: (c_1, c_2, c_3, c_4, c_5, c_6, c_d)$
- $>_2: (c_1, c_3, c_2, c_6, c_5, c_4, c_d)$
- $>_3: (c_2, c_3, c_1, c_4, c_6, c_5, c_d)$
- $>_4: (c_2, c_1, c_3, c_5, c_4, c_6, c_d)$
- $>_5: (c_3, c_1, c_2, c_6, c_4, c_5, c_d)$
- $>_6: (c_3, c_2, c_1, c_5, c_6, c_4, c_d)$
- $>_7: (c_1, c_2, c_3, c_d, c_4, c_5, c_6)$
- $>_8: (c_1, c_2, c_3, c_d, c_4, c_5, c_6)$
- $>_9: (c_6, c_5, c_4, c_d, c_3, c_2, c_1)$
- $>_{10}: (c_6, c_5, c_4, c_d, c_3, c_2, c_1)$

Voters 7, 8, 9 and 10 are compatible with the axis $c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_d \triangleleft c_4 \triangleleft c_5 \triangleleft c_6$. We prove that there is no axis with the unpopular candidate c_d at an extremity that allows us to delete at most 6 voters. The voters 1 to 6 have been created in such a way that at most 3 of them are compatible with the same axis. Hence, if there is an axis with c_d at an extremity such that 4 voters (or more) are single-peaked with respect to it, one of the preferences 7 to 10 must be compatible with it. Assume that $>_7$ - and then $>_8$ - are compatible with the axis. The extremities are then c_d, c_6 . Then, $>_9$ and $>_{10}$ are not compatible with the axis which is of the form $c_d \triangleleft \{c_1, c_2, c_3\} \triangleleft c_4 \triangleleft c_5 \triangleleft c_6$. At least two preferences among $\{>_1, \dots, >_6\}$ must be compatible with the axis, however, only $>_1$ can be single-peaked with respect to the axis in construction. Let us assume that $>_9$ - and then $>_{10}$ - are compatible with the axis. Analogously, the axis is of the form $c_d \triangleleft \{c_4, c_5, c_6\} \triangleleft c_3 \triangleleft c_2 \triangleleft c_1$. As there are no two preferences among $\{>_1, \dots, >_6\}$ ranking $\{c_4, c_5, c_6\}$ in the same order, only one of these can be compatible with the axis.

□

Proposition 4.4.3: Archimedean

The archimedean property holds for the *FT*, *GS* and *VD* measures. It does not hold for the *LS* measure.

Proof. The property holds for *FT* and *GS* because $GS(>_v, A) \geq 1$ (resp. $FT(>_v, A) \geq 1$) if $>_v$ is not single-peaked w.r.t. to A . It obviously holds for *VD* by definition of this measure.

To prove that the property does not hold for *LS*, consider the following profile \mathcal{P} :

- $>_i$ for $i \in \{1, \dots, p\} : (c_1, c_2, c_3)$
- $>_i$ for $i \in \{p+1, \dots, 2p\} : (c_1, c_3, c_2)$
- $>_i$ for $i \in \{2p+1, \dots, 3p\} : (c_2, c_3, c_1)$

For any axis A , there are p voters not compatible with it, while

$$LS(\mathcal{P}, A) = \max_{i \in \{1, \dots, 3p\}} LS(>_i, A) = 1$$

because $LS(>_i, A) \leq 1$ for all i and there exists $k \in \{0, p, 2p\}$ such that $LS(>_i, A) = 1$ for $i \in \{k+1, \dots, k+p\}$. \square

Proposition 4.4.4: Stability

The stability property only holds for the VD measure, and it does not hold for FT , GS and LS .

Proof. We first prove that the property holds for the VD measure. Let \mathcal{P} be a preference profile over the set of candidates \mathcal{C} and A^* be an optimal axis on \mathcal{C} such that $VD(\mathcal{P}, A^*) = k$. Given an additional preference relation $>$ on \mathcal{C} , there are two possibilities:

- if there exists A such that $VD(\mathcal{P} \cup \{>\}, A) = k$, then A is also optimal for \mathcal{P} because $VD(\mathcal{P}, A) \leq VD(\mathcal{P} \cup \{>\}, A)$;
- otherwise $VD(\mathcal{P} \cup \{>\}, A^*) = k + 1$ (only one vote is added), thus A^* stays optimal for $\mathcal{P} \cup \{>\}$.

To prove that the property does not hold for the other measures, consider the following profile \mathcal{P} :

$$\begin{aligned} >_1: (c_1, c_2, c_3, c_4) \\ >_2: (c_4, c_3, c_2, c_1) \end{aligned}$$

This profile is single-peaked with respect to the unique axis $A^* = c_1 \prec c_2 \prec c_3 \prec c_4$, which is therefore optimal for the three measures. Consider now the profile obtained by adding preference $>_3 = (c_1, c_4, c_3, c_2)$ to \mathcal{P} . For $X \in \{FT, GS\}$, it can easily be checked that $X(\mathcal{P} \cup \{>\}, A^*) = 3$ while $X(\mathcal{P} \cup \{>\}, A) = 2$ for the axis $A = c_2 \prec c_1 \prec c_4 \prec c_3$. Thus A^* is not optimal for $\mathcal{P} \cup \{>\}$. The same counter-example works for LS ; the only difference is that $LS(\mathcal{P} \cup \{>\}, A) = 1$. \square

4.5 A Maximum likelihood estimator

In the same manner as Condorcet interpreted the ranking of candidates returned by Kemeny's voting rule as a maximum likelihood estimation of a "true" ranking (Young, 1988), where the votes (individual preferences of the voters) are viewed as noisy observations of this true ranking, the axis A minimizing the number of A -forbidden triples can be interpreted as a maximum likelihood estimation of an "hidden" axis, where each vote is viewed as a noisy estimate of a preference single-peaked w.r.t. A . To this purpose, the following assumptions are required:

	c_1	c_2	c_3	c_4
$c_1 > c_2$	–	–	6	4
$c_1 > c_3$	–	10	–	6
$c_1 > c_4$	–	10	15	–
$c_2 > c_3$	18	–	–	4
$c_2 > c_4$	19	–	12	–
$c_3 > c_4$	22	15	–	–

Table 4.5.1: A voting matrix of a profile of 100 preferences over 4 candidates. We assume that for each c_i , there are 25 preferences with c_i as peak. The value in each cell is the number of pairwise preferences corresponding to the header of the row, among the 25 voters whose peak corresponds to the header of the column. For instance, the cell at the intersection of the row “ $c_2 > c_4$ ” and the column “ c_3 ” states that there are 12 preferences with peak c_3 that rank c_2 better than c_4 . The notation “–” stands for 25 or 0 depending on the peak.

1. Each voter has a peak (a candidate that she prefers to all the other candidates).
2. Let A be the hidden axis. For each voter v with peak c^* :
 - for any triple c_i, c_j, c_k such that $c_i \prec_A c_j \prec_A c_k$, if $c^* = c_i$ (resp. $c^* = c_k$), then $c_j >_v c_k$ (resp. $c_j >_v c_i$) with probability $p > 1/2$ (the same for all voters); this requirement simply means that the triple is more likely to be compatible with the “real” axis than forbidden.
 - for any pair c_i, c_j of candidates such that $c_i \prec_A c^* \prec_A c_j$ or $c_j \prec_A c^* \prec_A c_i$, the probability that $c_i >_v c_j$ is $1/2$. In other words, the candidates on the left side of the peak and on the right side of the peak are treated independently.
3. Every voter’s judgment on every pair of candidates that does not involve her peak is independent of her judgment on every other pair.
4. Each voter’s judgment is independent of the other voters’ judgments.

Each voter judges every pair of candidates and the results can be summarized in a table where one counts the number of times each pairwise preference is observed.

Example 4.5.1. Consider for instance Table 4.5.1, that provides a synthesized view of pairwise preferences for an election with 100 voters and 4 candidates. The peaks of the voters are assumed to be uniformly distributed among the candidates (each candidate c_i is the peak of 25 voters). The framed numbers are those that matter for counting the number of A -forbidden triples in the profile, if A is defined by $c_1 \prec_A c_2 \prec_A c_3 \prec_A c_4$. The number of A -forbidden triples is here:

$$(25-18)+(25-19)+(25-22)+(25-15)+6+4+6+4=46$$

(for instance, as 18 voters of peak c_1 judge that $c_2 > c_3$, the other $25 - 18 = 7$ voters of peak c_1 judge that $c_3 > c_2$, which is not compatible with A).

If assumptions 1–4 hold, the likelihood $L(A)$ to observe the voting matrix in Table 4.5.1 if the profile is nearly single-peaked with respect to the axis $c_1 \prec_A c_2 \prec_A c_3 \prec_A c_4$ is:

$$\begin{aligned} L(A) &= \binom{25}{18} p^{18} (1-p)^7 \binom{25}{19} p^{19} (1-p)^6 \binom{25}{22} p^{22} (1-p)^3 \\ &\quad \binom{25}{10} \left(\frac{1}{2}\right)^{10} \binom{1}{2}^{15} \binom{25}{10} \binom{1}{2}^{10} \binom{1}{2}^{15} \binom{25}{15} p^{15} (1-p)^{10} \\ &\quad \binom{25}{19} p^{19} (1-p)^6 \binom{25}{15} \binom{1}{2}^{15} \binom{1}{2}^{10} \binom{25}{12} \binom{1}{2}^{12} \binom{1}{2}^{13} \\ &\quad \binom{25}{21} p^{21} (1-p)^4 \binom{25}{19} p^{19} (1-p)^6 \binom{25}{21} p^{21} (1-p)^4 \\ &= K \left(\frac{1}{2}\right)^{100} p^{154} (1-p)^{46} \end{aligned}$$

where K is the product of binomial coefficients. The likelihood $L(A')$ that the profile is nearly single-peaked w.r.t. the axis A' defined by $c_2 \prec_{A'} c_1 \prec_{A'} c_4 \prec_{A'} c_3$ is:

$$\begin{aligned} L(A') &= \binom{25}{18} \left(\frac{1}{2}\right)^{18} \binom{1}{2}^7 \binom{25}{19} \binom{1}{2}^{19} \binom{1}{2}^6 \binom{25}{3} p^3 (1-p)^{22} \\ &\quad \binom{25}{10} p^{10} (1-p)^{15} \binom{25}{10} p^{10} (1-p)^{15} \binom{25}{10} p^{10} (1-p)^{15} \dots \\ &= K \left(\frac{1}{2}\right)^{100} p^{66} (1-p)^{134} \end{aligned}$$

where K is here again the product of binomial coefficients. Note indeed that, for all $4!/2 = 12$ axes (the division by 2 is because an ordering and the reverse ordering correspond to the same axis), the binomial coefficients are the same because $\binom{n}{k} = \binom{n}{n-k}$. Besides, the product of fractions $1/2$ is $(1/2)^{100}$ for all axes. Hence, axis A is more likely than A' because $p > 1/2$. More generally, the following formula is obtained for the likelihood of an axis:

Proposition 4.5.1

If assumptions 1–4 hold and the peaks are equally distributed in the considered sample of voters (i.e., each candidate is the peak of n/m voters), the likelihood of an axis A on $m \geq 3$ candidates is:

$$K \left(\frac{1}{2} \right)^{n(m-1)(m-2)/6} p^{n(m-1)(m-2)/3-t(A)} (1-p)^{t(A)}$$

where the value of K only depends on the voting matrix and is the same for all axes, and $t(A)$ is the number of A -forbidden triples.

Proof. Let us consider the component, say α , on row $c_i > c_j$ and column (peak) c_k of the voting matrix. In the likelihood expression of an axis A , this entry yields the term:

- $\binom{n/m}{\alpha} 1/2^{n/m}$ if c_k is between c_i and c_j in A ;
- $\binom{n/m}{\alpha} p^\alpha (1-p)^{n/m-\alpha}$ if c_i is between c_k and c_j in A ;
- $\binom{n/m}{\alpha} p^{n/m-\alpha} (1-p)^\alpha$ if c_j is between c_k and c_i in A .

The term $1/2^{n/m}$ (first case) occurs exactly once for each triple $\{c_i, c_j, c_k\}$ of candidates (when in A the peak is between the two others). As there are $m(m-1)(m-2)/6$ triples of candidates, the product of these terms is therefore $1/2^{n(m-1)(m-2)/6}$.

In the second case (resp. third case), the exponent $n/m-\alpha$ (resp. α) of $(1-p)$ is the number of A -forbidden triples involving c_i, c_j, c_k when c_k is the peak. The total number of A -forbidden triples is $t(A)$, hence, when making the product of the terms, the sum of the exponents of $(1-p)$ is $t(A)$.

The sum of the exponents of p can then be inferred from the total number of triples involving c_i, c_j, c_k when c_k is the peak and c_k is not between c_i and c_j in A . This number of triples is equal to n/m multiplied by the number of components of the voting matrix such that the peak c_k is not between c_i and c_j in A (the number of “framed numbers” in the voting matrix). The number of such components is equal to $\sum_{i=1}^m \left(\binom{i-1}{2} + \binom{m-i}{2} \right)$, where i is the position of c_k on the axis, and $\binom{i-1}{2}$ (resp. $\binom{m-i}{2}$) is the number of pairs $\{c_i, c_j\}$ on the left (resp. right) of c_k . This is equivalent to $2 \sum_{i=0}^{m-1} \binom{i}{2} = m(m-1)(m-2)/3$. Multiplying by n/m , one deduces that there are $n(m-1)(m-2)/3$ triples where the peak c_k is not between c_i and c_j . Consequently, there are $n(m-1)(m-2)/3 - t(A)$ triples that are not A -forbidden, which is the exponent of p . \square

The Proposition 4.5.1 yields then the following corollary:

Corollary 4.5.1

Assuming that $p > 1/2$, we have $(1-p) < p$ and hence, if the peaks are equally distributed in the sample of voters, the most likely axis is the one with the lowest exponent of $(1-p)$, which is the one that minimizes the number $t(A)$ of A -forbidden triples.

For instance, if one computes $t(A)$ for all axes A based on the data of Table 4.5.1, one finds that the answer is $c_1 \prec_A c_2 \prec_A c_3 \prec_A c_4$.

4.6 Computational aspects

So far, we have studied theoretical properties of the FT measure and its positioning with respect to existing measures. We will now see how do we deal with this measure in practice. We first show that, given a preference profile \mathcal{P} , it is actually NP-hard to find an axis A minimizing the number of A -forbidden triples in \mathcal{P} . Nevertheless, we give an ILP formulation of the problem which will later reveal quite efficient in practice.

4.6.1 FT Single-peaked Evaluation and Consistency

Given a preference profile \mathcal{P} , we study here the computational complexity of determining $FT(\mathcal{P}, A)$ for a given axis A (evaluation problem), and that of determining if $FT(\mathcal{P}) \leq k$ for a given integer k (consistency problem). We show that the evaluation problem can be handled in polynomial time while the consistency problem is NP-complete.

The evaluation problem is formulated as follows:

FT SINGLE-PEAKED EVALUATION

Input: A profile \mathcal{P} and an axis A .

Output: $FT(\mathcal{P}, A)$.

Clearly, this problem can be solved in $\mathcal{O}(nm^2)$, where $n = |\mathcal{P}|$ and $m = |\mathcal{C}|$, by brute force enumeration of all triples. This complexity can be improved to $\mathcal{O}(nm\sqrt{\log m})$, as stated in the following proposition:

Proposition 4.6.1

The FT single-peaked evaluation problem can be solved in $\mathcal{O}(nm\sqrt{\log m})$.

Proof. Let $\succ_v \in \mathcal{P}$ and c^* be the candidate at the top of \succ_v . Let \mathcal{C}_l (resp. \mathcal{C}_r) denote the set of candidates on the left (resp. right) of c^* on A . Formally, $c \prec_A c^*$ (resp. $c^* \prec_A c$) for any $c \in \mathcal{C}_l$ (resp. $c \in \mathcal{C}_r$). We denote by \prec_l and \succ_l (resp. \prec_r and \succ_r) the restrictions of

\prec_A and \succ to \mathcal{C}_l (resp. \mathcal{C}_r). We have:

$$FT(\succ, A) = \delta_{KT}(\prec_l, \prec_l) + \delta_{KT}(\succ_r, \prec_r)$$

where $\delta_{KT}(\cdot, \cdot)$ is the classical Kendall-tau distance function (number of pairwise disagreements between two orders) and \prec_l is the reverse of \succ_l . Now, noticing that

$$FT(\mathcal{P}, A) = \sum_{\succ \in \mathcal{P}} FT(\succ, A),$$

the $\mathcal{O}(nm\sqrt{\log m})$ complexity follows because the Kendall-tau distance between two permutations of size m can be computed in $\mathcal{O}(m\sqrt{\log m})$ (Chan and Pătraşcu, 2010). \square

A Curious Tortoise Intervention



The proof of Proposition 4.6.1 sheds light on the difference between *FT* and *GS* measures. While the *FT* measure treats the sets \mathcal{C}_l and \mathcal{C}_r independently, these sets interact in the case of the *GS* measure. Note that this issue has been already illustrated in Figure 4.3.4, without providing a more detailed explanation.

Consequently, given a preference profile \mathcal{P} , it is easy to evaluate its distance from a given axis A in terms of the *FT* measure - as we have just seen, it equals the sum of the *FT*-distances of the restrictions of \mathcal{P} on sets \mathcal{C}_l (resp. \mathcal{C}_r) to the restricted axis \prec_l (resp. \succ_r).

Regarding the *GS* measure, the evaluation problem is much less straightforward - given a preference \succ and an axis A , one needs to find a preference \succ' single-peaked w.r.t. A which minimizes the Kendall tau distance $d_{KT}(\succ, \succ')$ to evaluate the *GS*-distance of the preference \succ from the axis A . Erdélyi et al. (2017) proposed a dynamic programming algorithm to solve this problem. Its complexity is in $O(m^3)$. Therefore, given a profile \mathcal{P} of n preferences, and an axis A , the evaluation of the *GS* measure is in $O(n \cdot m^3)$.

The consistency problem is formulated as follows:

FT SINGLE-PEAKED CONSISTENCY

Input: A profile \mathcal{P} and an integer k .

Output: Yes if $FT(\mathcal{P}) \leq k$, otherwise no.

The following result holds for this decision problem (yes/no question):

Theorem 4.6.1

The *FT* single-peaked consistency problem is NP-complete.

Proof. By Proposition 4.6.1, the problem is clearly in NP. We show that it is NP-complete by a reduction from the max cut problem:

MAX CUT

Input: An undirected graph $G = (V, E)$ and an integer k .

Output: Yes if there exists a partition of V into two sets S, T such that at least k edges have one endpoint in S and one in T , otherwise no.

Given such an instance of the max cut problem, with $V = \{v_1, \dots, v_n\}$, we build the following instance of the FT single-peaked consistency problem:

- There are $2n + 1$ candidates: a candidate c_0 , and for each vertex v_i we have two candidates c_i and \bar{c}_i .
- We have M voters (for a sufficiently large M , to be specified later), called α -voters, with the following preference: $c_0 > c_1 > \bar{c}_1 > c_2 > \bar{c}_2 > \dots > c_n > \bar{c}_n$.
- Similarly, we have M voters, called β -voters, with the following preference: $c_0 > \bar{c}_1 > c_1 > \bar{c}_2 > c_2 > \dots > \bar{c}_n > c_n$.
- For each edge (v_i, v_j) with $i < j$, we have an edge-voter e_{ij} whose preference is built from an α -voter by moving up \bar{c}_j just below c_i , and moving down \bar{c}_i just above c_j : $c_0 > c_1 > \dots > \bar{c}_{i-1} > c_i > \bar{c}_j > c_{i+1} > \dots > \bar{c}_{j-1} > \bar{c}_i > c_j > \dots$

Let $K = 2 \sum_{(v_i, v_j) \in E} |j - i - 1|$, and fix $M = K + m + 1$. We show that there is a cut (S, T) with at least k edges if and only if there is an axis with at most $K + |E| - k$ forbidden triples.

Let us call an axis *regular* if (1) c_0 is in the middle position $n + 1$ and (2) for each i , both c_i and \bar{c}_i are at distance i from c_0 (i.e., one in position $n + 1 - i$ and one in position $n + 1 + i$).

There is a bijection h between partitions (S, T) of V and regular axes, where in the axis $A = h(S, T)$ c_i is on the left of c_0 and \bar{c}_i on the right iff $v_i \in S$. For instance, if there are 5 vertices and $S = \{v_2, v_5\}$ ($T = \{v_1, v_3, v_4\}$), then $h(S, T)$ is the axis

$$c_5 \triangleleft \bar{c}_4 \triangleleft \bar{c}_3 \triangleleft c_2 \triangleleft \bar{c}_1 \triangleleft c_0 \triangleleft c_1 \triangleleft \bar{c}_2 \triangleleft c_3 \triangleleft c_4 \triangleleft \bar{c}_5$$

Let (S, T) be a partition of V inducing a cut with at least k edges. Let us count the number of forbidden triples in the axis $h(S, T)$. Note first that α -voters and β -voters are compatible with any regular axis, so they induce no forbidden triples. Let us consider an edge-voter e_{ij} (with $i < j$):

- For each ℓ such that $i < \ell < j$, there is exactly one forbidden triple $\{c_0, c_\ell, \bar{c}_i\}$ or $\{c_0, \bar{c}_\ell, \bar{c}_i\}$, and exactly one forbidden triple $\{c_0, c_\ell, \bar{c}_j\}$ or $\{c_0, \bar{c}_\ell, \bar{c}_j\}$. This gives in total $2|j - i - 1|$ forbidden triples.
- If c_i and c_j are on the same side of c_0 in the axis, then there is a forbidden triple $(c_0, \bar{c}_i, \bar{c}_j)$. Otherwise, if c_i and c_j are not on the same side of c_0 , there is no new forbidden triple.

Considering all the edges, the first item correspond in total to K forbidden triples. For the second item, each edge *not* in the cut gives one more forbidden triple. If there are (at least) k edges in the cut, then there are (at most) $K + m - k$ forbidden triples.

Conversely, let us assume that there is an axis A with at most $M + m - k$ forbidden triples. We first notice that, as $M > K + m$, A is necessarily regular. Indeed, if for some i , candidates c_i and \bar{c}_i were on the same side of c_0 , then either α -voters or β -voters induce forbidden triples, so at least M of them. Similarly, if for some $i < j$ a candidate c_j or \bar{c}_j were closer to c_0 than c_i or \bar{c}_i , then again α -voters or β -voters induce at least M forbidden triples.

Since A is regular, we consider the cut (S, T) such that $A = h(S, T)$. As before, for each voter e_{ij} there are exactly $2|j - i - 1|$ forbidden triples involving candidates c_ℓ or \bar{c}_ℓ for $i < \ell < j$, so K in total. Again, if c_i and c_j are on the same side of c_0 there is one more forbidden triple.

Since there are (at most) $K + m - k$ forbidden triples, there must be at least k edges in the cut (S, T) . \square

4.6.2 An Integer Program to Determine $FT(\mathcal{P})$

The FT single-peaked consistency problem can be formulated as an Integer Program (IP). For each pair $\{c_i, c_j\}$ of candidates (with $i, j \in \{1, \dots, m\}$), we introduce a binary variable x_{ij} describing their relative position on the sought axis A . More precisely, the constraints of type 1 and type 2 detailed below will ensure that:

$$x_{ij} = \begin{cases} 1 & \text{if } c_i \triangleleft_A c_j, \\ 0 & \text{otherwise.} \end{cases}$$

Additionally, for each voter $v \in \{1, \dots, n\}$ and each pairwise preference $c_i \succ_v c_j$ with $\pi(v) \notin \{i, j\}$, where $c_{\pi(v)}$ is the peak of v , we define a binary variable z_{vij} related to the triple $(c_{\pi(v)}, c_i, c_j)$. More precisely, the constraints of type 3 and type 4 detailed below will ensure that:

$$z_{vij} = \begin{cases} 1 & \text{if } (c_{\pi(v)}, c_i, c_j) \text{ is } A\text{-forbidden in } v, \\ 0 & \text{otherwise.} \end{cases}$$

The sum of variables z_{vij} is the number of forbidden triples in the profile \mathcal{P} . The IP objective function is therefore:

$$\min \sum_{v=1}^n \sum_{i=1}^m \sum_{j=1}^m z_{vij}.$$

We now detail the four types of constraints in the program:

1. For each pair $\{c_i, c_j\}$ of candidates, one and only one of the variables $\{x_{ij}, x_{ji}\}$ equals 1 (the relation \triangleleft_A is antisymmetric - we cannot have $c_i \triangleleft_A c_j$ and $c_j \triangleleft_A c_i$).
2. For each tuple (c_i, c_j, c_k) , if $x_{ik} = 1$ and $x_{kj} = 1$ then $x_{ij} = 1$ (the relation \triangleleft is transitive - $c_i \triangleleft_A c_k$ and $c_k \triangleleft_A c_j$ imply $c_i \triangleleft_A c_j$).
3. For each tuple (v, i, j) such that $\pi(v) \notin \{i, j\}$ and $c_i >_v c_j$, if $c_{\pi(v)} \triangleleft_A c_j$ and $c_j \triangleleft_A c_i$ then $z_{vij} = 1$ ($(c_{\pi(v)}, c_i, c_j)$ is A -forbidden in v , on the right of the peak).
4. For each tuple (v, i, j) such that $\pi(v) \notin \{i, j\}$ and $c_i >_v c_j$, if $c_i \triangleleft_A c_j$ and $c_j \triangleleft_A c_{\pi(v)}$ then $z_{vij} = 1$ ($(c_{\pi(v)}, c_i, c_j)$ is A -forbidden in v , on the left of the peak).

Putting together the objective function and the constraints, we obtain the following IP:

$$\min \sum_{v=1}^n \sum_{i=1}^m \sum_{j=1}^m z_{vij}$$

subject to:

$$\begin{cases} x_{ij} + x_{ji} = 1 & \forall \{c_i, c_j\} & (1) \end{cases}$$

$$\begin{cases} x_{ij} \geq x_{ik} + x_{kj} - 1 & \forall (c_i, c_j, c_k) \text{ with } i \neq j \neq k & (2) \end{cases}$$

$$\begin{cases} z_{vij} \geq x_{\pi(v)j} + x_{ji} - 1 & \forall v, i, j \text{ with } c_i >_v c_j & (3) \end{cases}$$

$$\begin{cases} z_{vij} \geq x_{j\pi(v)} + x_{ij} - 1 & \forall v, i, j \text{ with } c_i >_v c_j & (4) \end{cases}$$

$$x_{ij} \in \{0, 1\} \forall i, j, \quad z_{vij} \in \{0, 1\} \forall v, i, j$$

A Curious Tortoise Intervention



Note that in practice, it is actually sufficient to define the variables z_{vij} as continuous variables (in other words, we replace $z_{vij} \in \{0, 1\}$ by $z_{vij} \geq 0$). This reduces the number of binary variables.

Example 4.6.1. For a better visualisation, let us consider the following profile of two preferences over 4 candidates:

$$>_1: (c_1, c_2, c_3, c_4)$$

$$>_2: (c_4, c_1, c_2, c_3)$$

We do not give the antisymmetric and transitivity constraints, and we will only give the constraints of type 3 and 4 for the first voter v_1 :

- A -forbidden triples constraints on the right of the peak: We have $\pi(v_1) = 1$. If c_3 is on the right side of the peak (i.e., $x_{13} = 1$), we cannot have c_3 between c_2 and c_1 (i.e., c_2 is on the right of c_3 , so we have $x_{32} = 1$). This is written:

$$z_{v_1 23} \geq x_{13} + x_{32} - 1.$$

By the same reasoning, we find

$$z_{v24} \geq x_{14} + x_{42} - 1,$$

and

$$z_{v34} \geq x_{14} + x_{43} - 1.$$

- *A-forbidden triples constraints on the left of the peak: Analogously, if c_3 is on the left of the peak (i.e., $x_{31} = 1$), we cannot have c_2 on the left of c_3 (i.e., $x_{23} = 1$), in which case c_3 would be between c_1 and c_2 . This is written:*

$$z_{v23} \geq x_{31} + x_{23} - 1.$$

By the same reasoning, we find

$$z_{v24} \geq x_{41} + x_{24} - 1.$$

and

$$z_{v34} \geq x_{41} + x_{34} - 1.$$

The formulation is straightforward, however, it can become less efficient when the number of voters increases due to the important number of variables that need to be created for each voter. It is possible that a formulation where the number of variables does not depend on the number of voters could be found, for example by adapting the ILP proposed by Kratica et al. (2010).

Note that this formulation is also valid for determining an axis when the preference profile only consists of top orders (i.e., when each voter does not provide a complete ranking but only the upper part of his/her preferences). In this case, one can just assume that, for any candidate c that is inside the top order of a voter v , and any candidate c' that is outside the top order, the preference is $c >_v c'$; besides, we can not decide whether $c >_v c'$ or $c' >_v c$ if both c and c' are outside the top order - the both relation between c and c' are possible.

Finally, note that the optimal value of the ILP continuous relaxation always equals 0, by setting $x_{ij} = \frac{1}{2}$ for each ordered pair (i, j) .

4.7 Experimental study

We carried out numerical tests on real and randomly generated preference profiles in order to compare experimentally the various distance measures discussed above, except the *LS* distance measure, whose logic differs from that of the others. For measuring the *VD* distance, we used the C++ code developed by Sui et al. (2013), made available

on the web⁵. For measuring the *FT* distance, we used the Gurobi software to solve the IP formulation given in Section 4.6.2. Finally, for measuring the *GS* distance, we used a brute force algorithm - to the best of our knowledge, no more efficient algorithm is known for this problem so far.

We study the quality of optimal axes on real data, compared to reference axes whose design is detailed below. To evaluate the quality of an axis, we measure the proportion of the betweenness relation (see Definition 4.2.2) that is common to the optimal axis and the reference axis. This proportion is expressed in percentage in the sequel, and we call it *recognition rate*. More formally, let us recall that given an axis A , we introduce in Definition 4.2.2 the relation R_A as a set of triples (c_i, c_j, c_k) such that $i < k$ and c_j is between c_i and c_k on the axis A (we recall that c_j and c_k do not need to be consecutive on A). Given two axes A and A' defined on the same set of candidates, the proportion of the betweenness relation, denoted by $\delta(A, A')$, is defined as follows:

$$\delta(A, A') = |R_A \cap R_{A'}| / |R_A|.$$

Note that $|R_A| = |R_{A'}|$. More precisely, for any axis A on the set of m candidates, we have

$$|R_A| = 2 \sum_{i=1}^m (i-1) \cdot (m-i)$$

Indeed, let us consider the i -th leftmost candidate c of A . There are $i-1$ candidates on the left of her, and $m-i$ candidates on the right of her. This yields $2 \cdot (i-1) \cdot (m-i-1)$ triples in which c is in the middle position (we recall that R_A is defined for ordered tuples - in other words, $(c_i, c_j, c_k) \in R_A$ if and only if (c_k, c_j, c_i)).

In practice, it is easy to see that without loss of generality, the tuples (c_i, c_j, c_k) and (c_k, c_j, c_i) can be considered as a unique tuple in the evaluation of δ . Therefore, we considered in this section the following version of R_A :

$$R_A = \{(c_i, c_j, c_k) \in \mathcal{C}^3 : i < k \text{ and } c_i \prec_A c_j \prec_A c_k \text{ or } c_k \prec_A c_j \prec_A c_i\}.$$

This yields $|R_A| = \sum_{i=1}^m (i-1) \cdot (m-i)$.

Example 4.7.1. To illustrate the notion of recognition rate, let us consider three axis defined on the set of candidates $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$:

$$\begin{aligned} A &= c_1 \prec_A c_2 \prec_A c_3 \prec_A c_4 \\ A' &= c_1 \prec_{A'} c_3 \prec_{A'} c_2 \prec_{A'} c_4 \\ A'' &= c_3 \prec_{A''} c_1 \prec_{A''} c_2 \prec_{A''} c_4 \end{aligned}$$

⁵<http://www.cs.toronto.edu/~lex/code/asprgen.html>

We have:

$$\begin{aligned} R_A &= \{(1, 2, 3), (1, 2, 4), (1, 3, 4), (2, 3, 4)\}, \\ R_{A'} &= \{(1, 2, 4), (3, 2, 4), (1, 3, 2), (1, 3, 4)\}, \\ R_{A''} &= \{(2, 1, 3), (3, 1, 4), (1, 2, 4), (3, 2, 4)\}. \end{aligned}$$

We see that

$$|R_A| = |R_{A'}| = |R_{A''}| = (1 - 1) \cdot (4 - 1) + (2 - 1) \cdot (4 - 2) + (3 - 1) \cdot (4 - 3) + (4 - 1) \cdot (4 - 4) = 4.$$

Moreover, we have $|R_A \cap R_{A'}| = 2$, $|R_A \cap R_{A''}| = 1$ and $|R_{A'} \cap R_{A''}| = 2$. This yields $\delta(A, A') = 0.5$, $\delta(A, A'') = 0.25$ and $\delta(A', A'') = 0.5$. In other words, A is nearer (in terms of the distance δ) to A' than to A'' , while A' is at the same distance from both A and A'' .

To go further and better understand the impact of the characteristics of the profiles on the numerical results obtained, we also study the quality of optimal axes on profiles randomly generated according to diverse probability distributions for structured preferences.

4.7.1 Numerical tests on real data

The real data sets were taken from the 2007 Glasgow city council election, the 2002 Irish general election as well as a 2017 voting experiment during the French presidential election. The two first data sets are available on the PrefLib website⁶. The French presidential election data set is available on the website of the experiment called *Voter autrement*⁷ (Bouveret et al., 2018).

Both the Glasgow election and the Irish election were separated by voting districts (with one list of candidates per district): 21 wards for the Glasgow election, one of which has been discarded in our tests because there were too many candidates - 13 - for measuring the GS distance; 42 constituencies for the Irish election, among which we investigate here only 2 constituencies where electronic voting machines were used (Dublin North and Dublin West). Each ward (resp. constituency) involved different candidates and voters, and elected 3 or 4 councillors (resp. between 3 and 5 deputies) using the Single Transferable Vote (STV) system. This implies that some political parties had several candidates for the same voting district. A ballot consists in a ranking of the k most preferred candidates of a voter, for varying values of k according to the voter. In order to fit the data with our setting, we restricted ourselves to the ballots for which $k = m$ (complete rankings of the candidates). The number of candidates in the Glasgow (resp. Irish) data set ranges from 8 (resp. 9) to 11 (resp. 12), and the number of complete votes from 320 (resp. 3662) to 1003 (resp. 3800).

⁶<https://www.preflib.org/data/index.php>

⁷<https://zenodo.org/record/1199545>

In the *Voter autrement* data set, one file was usable for our purpose (file `stv111.csv`, here also reporting the results of an experiment about STV), with 11 candidates from as many distinct political parties and 4068 complete votes.

For each election (at the level of a ward, a constituency or a country), we built a reference left-right axis on the candidates. To do so, we used Wikipedia as external source. The free encyclopedia provides indeed a political position (of course debatable) for each political party (e.g., left wing, right wing, centre, centre right, etc.). We assumed that the political position of an affiliated candidate corresponds to that of the belonging party, and we built an axis over the affiliated candidates based on these positions. We excluded the non-affiliated candidates from the data sets because we were not able to define a political position for them. Actually, the “Wikipedia axis” is not unique since several parties can be labeled by the same political position on Wikipedia, or some parties can have several candidates in an election. For instance, a Wikipedia axis reads $((c_1, c_3), c_2, (c_4, c_5))$, where candidates $\{c_1, c_3\}$ as well as $\{c_4, c_5\}$ have indistinguishable political positions. This corresponds to a set of $2 \cdot 2 = 4$ compatible axes:

$$\begin{aligned} c_1 \triangleleft c_3 \triangleleft c_2 \triangleleft c_4 \triangleleft c_5, & \quad c_3 \triangleleft c_1 \triangleleft c_2 \triangleleft c_4 \triangleleft c_5, \\ c_1 \triangleleft c_3 \triangleleft c_2 \triangleleft c_5 \triangleleft c_4, & \quad c_3 \triangleleft c_1 \triangleleft c_2 \triangleleft c_5 \triangleleft c_4. \end{aligned}$$

Note that indistinguishable political positions do *not* mean here that the candidates share the same position on the political spectrum, but that we have a partial knowledge of the exact axis. The sets of candidates with indistinguishable political positions (as $\{c_1, c_3\}$ and $\{c_4, c_5\}$ above) are called *blocks* in the following. Given a distance measure X (in $\{VD, FT, GS\}$) and a profile \mathcal{P} , the recognition rate is formulated in the following manner to take into account blocks:

$$\min\{\delta(A_X(\mathcal{P}), A') : A' \text{ compatible with the Wikipedia axis}\}$$

where $A_X(\mathcal{P})$ is an optimal axis according to X . We recall that, given two axes A, A' , we denote by $\delta(A, A')$, and call *recognition rate*, the proportion of the betweenness relation that is common to A and A' .

Apart from the recognition rate, we also distinguish three classes of results for the axis obtained by minimizing a given distance measure:

- C (Correct): The optimal axis is compatible with the Wikipedia axis, e.g. $c_3 \triangleleft c_1 \triangleleft c_2 \triangleleft c_4 \triangleleft c_5$ for $((c_1, c_3), c_2, (c_4, c_5))$.
- EE (Exchanged Extremities): The optimal axis can be made compatible with the Wikipedia axis by exchanging the block on the left extremity with the one on the right, e.g. $c_5 \triangleleft c_2 \triangleleft c_6 \triangleleft c_4 \triangleleft c_3 \triangleleft c_1$ for $((c_1, c_3), c_2, c_6, c_4, c_5)$.
- F (False): The optimal axis is called false otherwise.

d	C	EE	F	Recognition Rate
2007 Glasgow city council				
VD	2	1	17	57.25%
FT	5	5	10	67%
2002 Irish general election				
VD	0	0	2	75.35%
FT	0	0	2	71.5%
2017 French presidential election				
VD	0	0	1	58.8%
FT	0	0	1	74.6%

Table 4.7.1: Results obtained on real election data.

We have chosen to distinguish class EE because the experiments revealed a difficulty in recognizing the two extreme positions. To get an intuition of what is going on, consider for instance a profile where the two “extreme” candidates are ranked in the two last positions by a large number of voters, in an arbitrary order, and the voters who rank one of them in first position do not want to rank anyone else. In such a situation, the voting data do not provide much information that might help to distinguish who is left wing and who is right wing.

The results obtained are summarized in Table 4.7.1. Note that only the results for the VD and FT measures are given in the table, because the brute force algorithm used for the GS measure was not able to compute an axis in a reasonable amount of time for more than 8 candidates.

Regarding the two profiles with 8 candidates in the Glasgow data set, the results obtained with the GS measure are in class EE, while with FT one result is in class C and the other in class EE (the result is in class F in both cases with VD).

Table 4.7.1 also indicates how many times each class occurs for the VD and FT measures (over 20 preference profiles for the Glasgow city council election, 2 for the Irish general election, and 1 for the French presidential election), as well as the average recognition rate. More detailed tables are given in Appendix 4.A.

Let us detail now in a more down-to-earth manner the results obtained on the voting data from the French election.

The Wikipedia axis is $((c_1, c_{10}, c_{11}), c_6, c_9, (c_7, c_5), (c_2, c_4), c_8)$,⁸ with one non-affiliated can-

⁸For the following candidates: c_1 - Arthaud, c_2 - Asselineau, c_3 - Cheminade, c_4 - Dupont-Aignan, c_5 - Fillon, c_6 - Hamon, c_7 - Lassalle, c_8 - Le Pen, c_9 - Macron, c_{10} - Mélenchon, c_{11} - Poutou.

didate (candidate 3) excluded from the voting data. The axis A_{VD} minimizing VD is $c_5 \triangleleft c_2 \triangleleft c_9 \triangleleft c_6 \triangleleft c_{10} \triangleleft c_{11} \triangleleft c_1 \triangleleft c_7 \triangleleft c_4 \triangleleft c_8$, which is not compatible with the Wikipedia axis. The axis A_{FT} minimizing the FT measure is $c_2 \triangleleft c_1 \triangleleft c_{11} \triangleleft c_{10} \triangleleft c_6 \triangleleft c_9 \triangleleft c_7 \triangleleft c_4 \triangleleft c_5 \triangleleft c_8$ - if we follow the classification introduced above, this axis should be also classified as not recognized. Nevertheless, we observe here that A_{FT} is much better than A_{VD} , in the sense that, by swapping candidates 4 and 5 and moving candidate 2 in A_{FT} , an axis compatible with the Wikipedia axis can be obtained, while many more fixes are needed in A_{VD} . Note that this observation is supported by recognition rate values - the recognition rate for the axis A_{VD} is 58.8%, while the recognition rate for the axis A_{FT} is 74.6%

The results tend to show that the recognition ability of the FT measure is better than that of VD . When the FT measure is used, an axis perfectly compatible with Wikipedia is recognized in 21.7% of cases, and one reaches 50% if one adds the cases when the extremities are swapped. Yet, one should not conclude from this latter observation that the proposed measures are unsuitable for recognizing an axis from voting data, as it may come from the data themselves, if the preferences are too poorly structured (as for instance in some local elections).

To assess if there is a positive correlation between the “level of structure” of a profile \mathcal{P} and the recognition rate obtained with the FT measure, we evaluated the level of structure of \mathcal{P} as the ratio $FT(\mathcal{P})/E[FT(\mathcal{P}, A)]$, where $E[FT(\mathcal{P}, A)]$ denotes the expected value of $FT(\mathcal{P}, A)$ if A is an axis chosen at random according to the uniform distribution. The closer this ratio is to 0 (resp. to 1), the more (resp. the less) structured are the input preferences (note that $FT(\mathcal{P}) = 0$ if the preferences are perfectly single-peaked). This ratio is easy to compute from $FT(\mathcal{P})$ because $E[FT(\mathcal{P}, A)]$ is simply $n \binom{m-1}{2} / 3$, where n (resp. m) is the number of voters (resp. candidates). For a preference relation $\succ \in \mathcal{P}$ of peak c^* , there are indeed $\binom{m-1}{2}$ triples $c^* \succ c_i \succ c_j$ to consider, and the probability that a triple is forbidden is $1/3$ because it happens if c_j is between c^* and c_i on A . Thus $E[FT(\succ, A)] = \binom{m-1}{2} / 3$. As $E[FT(\mathcal{P}, A)] = \sum_{\succ \in \mathcal{P}} E[FT(\succ, A)]$ by linearity of expectation, the result follows: $E[FT(\mathcal{P}, A)] = n \binom{m-1}{2} / 3$.

The average value of this ratio is 0.60 (resp. 0.58) for the 2007 Glasgow city council election (resp. 2002 Irish general election), and 0.33 for the 2017 French presidential election. We recall that lower the ratio if, more the preferences are structured (in sense of single-peakedness). In other words, French presidential election data seem much more structured than Glasgow and Irish elections. This result is compliant with intuition, as it seems easily conceivable that preferences in local elections (as for the 20 wards of Glasgow) are less sensitive to the left-right spectrum than in national elections (as indicated above, the Irish general election yields a national result, but the election is divided into local voting districts with one list of candidates per district).

4.7.2 Numerical tests on synthetic data

In order to deepen the analysis of the recognition abilities of the VD , FT and GS measures, we also generated synthetic election data. The aim is to model situations where the preferences are noisy but there is a strong structure behind them. Given an axis A on the candidates, each preference relation in a synthetic profile is generated in two steps:

1. A candidate c^* is drawn uniformly at random in \mathcal{C} and an auxiliary preference relation $>_0$ of peak c^* single-peaked w.r.t. A is generated uniformly at random.
2. A preference relation $>$ is drawn from the Mallows model centered around $>_0$.

We recall that the Mallows model defines a probability distribution on rankings. A central ranking $>_0$ has the highest probability, and the probability of other rankings decreases in a Gaussian manner with the Kendall-tau distance from $>_0$. Formally, given a dispersion parameter $\theta \geq 0$, the probability $P(>)$ of a ranking $>$ is proportional to $e^{-\theta d_{KT}(>, >_0)}$, where d_{KT} is the Kendall-tau distance. If $\theta = 0$, the uniform distribution is obtained. The greater the value of θ , the higher the probabilities of the rankings around $>_0$.

It is known that using the Mallows model with parameters $>_0$ and θ is equivalent to generating a binary relation R where, for each pair c_i, c_j of candidates, if $c_i >_0 c_j$, then $c_i R c_j$ with probability $p = e^\theta / (1 + e^\theta)$, otherwise $c_j R c_i$; if the obtained binary relation R is transitive then stop and return the corresponding ranking, otherwise repeat the process until R is transitive. For the sake of interpretability, in the tables, we give the value of p instead of θ .

We used the PerMallows R package⁹ for generating rankings according to the Mallows model. For a fast generation of the profiles, the number of voters is set to 100 and the number m of candidates varies from 5 to 9. The above probability p takes its values in $\{0.7, 0.75, 0.8, 0.85, 0.9\}$ (obtained by setting $\theta = \ln(p/(1-p))$ in the Mallows model). For each couple (m, p) of parameter values, 100 instances were generated and we counted the number of instances for which the axis is perfectly recognized. As the brute force algorithm for the GS measure is not usable in practice for more than 7 candidates (due to excessive computation times for 100 voters), only the results for VD and FT are given from 8 candidates.

The results are reported in Table 4.7.2. It appears that the FT and GS measures behaves quite similarly, while the VD measure is the one for which axis A is the most often recognized. This result was quite unexpected because it is well-known, as mentioned in the introduction, that an optimal axis for VD explains only a few percentage of voter preferences in real election data. Nevertheless, the good behaviour of VD can

⁹<https://cran.r-project.org/package=PerMallows>

be simply explained by the manner in which the preferences are generated here: the probability that a voter preference is perfectly compatible with A is low but is the highest among all the preferences, thus the law of large numbers plays in favor of the VD measure, and this with all the more intensity as probability p is high.

To refine the analysis, we also studied the obtained recognition rates for the various measures, since it is a smoother criterion than the previous one. The results are reported in Table 4.7.3. The differences between the three considered measures are then much narrower, which means that, for the instances where the axis A is not perfectly recognized by using the FT or GS measure, the optimal axis is very similar to it.

$d \backslash p$	0.7	0.75	0.8	0.85	0.9
5 candidates					
VD	69 %	91 %	100 %	100 %	100 %
FT	42 %	79 %	96 %	99 %	98 %
GS	40 %	78 %	98 %	100 %	100 %
6 candidates					
VD	56 %	92 %	100 %	100 %	100 %
FT	23 %	61 %	94 %	100 %	98 %
GS	29 %	65 %	94 %	100 %	100 %
7 candidates					
VD	39 %	85 %	100 %	100 %	100 %
FT	9 %	46 %	93 %	98 %	100 %
GS	19 %	56 %	94 %	99 %	100 %
8 candidates					
VD	26 %	74 %	98 %	100 %	100 %
FT	5 %	29 %	81 %	98 %	100 %
9 candidates					
VD	12 %	58 %	91 %	100 %	100 %
FT	3 %	16 %	67 %	95 %	100 %

Table 4.7.2: Percentage of profiles where the axis is perfectly recognized, w.r.t. distance measure d and probability p . The preferences in the profiles are generated with steps 1-2.

We have then studied what happens when one sticks with the assumptions made in the interpretation of an optimal FT axis as a maximum likelihood estimator. In this purpose, we used the following alternative to step 2 in the generation of preferences:

2'. Let $>'_0$ denote the ranking $>_0$ where the peak c^* has been removed. A preference relation $>'$ is drawn from the Mallows model centered around $>'_0$. Return the preference

$d \backslash p$	0.7	0.75	0.8	0.85	0.9
5 candidates					
<i>VD</i>	83,5 %	95,5 %	100 %	100 %	100 %
<i>FT</i>	72,4 %	90,2 %	97,9 %	99,4 %	98,8 %
<i>GS</i>	74,1 %	91,1 %	99,1 %	100 %	100 %
6 candidates					
<i>VD</i>	88,5 %	97,8 %	100 %	100 %	100 %
<i>FT</i>	82,2 %	93,1 %	98,4 %	100 %	99,5 %
<i>GS</i>	82,9 %	93,5 %	98,3 %	100 %	100 %
7 candidates					
<i>VD</i>	83,5 %	96 %	100 %	100 %	100 %
<i>FT</i>	75 %	89,1 %	99,1 %	99,8 %	100 %
<i>GS</i>	77,8 %	90,6 %	99,2 %	99,9 %	100 %
8 candidates					
<i>VD</i>	77,5 %	93,7 %	99,5 %	100 %	100 %
<i>FT</i>	72 %	86,7 %	97,3 %	99,9 %	100 %
9 candidates					
<i>VD</i>	72 %	58 %	91 %	100 %	100 %
<i>FT</i>	69,1 %	83,2 %	96,2 %	99,7 %	100 %

Table 4.7.3: Recognition rates w.r.t. measure d and probability p . The preferences are generated with steps 1-2.

relation $>'$ with c^* added in first position.

It amounts to using the Mallows model to make noisy the preferences on the left side and on the right side of the peak on the axis A , without changing the peak. The results are reported in Tables 4.7.4 and 4.7.5. It appears clearly that optimizing the FT or GS measure is then the most effective manner to recognize the axis because, contrary to the VD measure, they are not sensitive to the number of candidates nor to the probability $p \geq 0.7$: they almost systematically recognize perfectly the axis A from the generated preference profile.

$d \backslash p$	0.7	0.75	0.8	0.85	0.9
5 candidates					
VD	99 %	99 %	100 %	100 %	100 %
FT	100 %	98 %	97 %	98 %	98 %
GS	100 %	100 %	100 %	100 %	100 %
6 candidates					
VD	83 %	95 %	100 %	100 %	100 %
FT	100 %	99 %	99 %	100 %	100 %
GS	100 %	100 %	100 %	100 %	100 %
7 candidates					
VD	72 %	95 %	100 %	100 %	100 %
FT	100 %	100 %	100 %	100 %	100 %
8 candidates					
VD	48 %	83 %	100 %	100 %	100 %
FT	100 %	100 %	100 %	100 %	100 %
9 candidates					
VD	30 %	65 %	97 %	100 %	100 %
FT	100 %	100 %	100 %	100 %	100 %

Table 4.7.4: Percentage of profiles where the axis is perfectly recognized, w.r.t. distance measure d and probability p . The preferences in the profiles are generated with steps 1-2'.

Another type of situations that can make single-peaked preferences noisy is the presence of clones among the candidates. Clones are relatively similar candidates (e.g., candidates belonging to the same political party), that share the same political positioning in the eyes of voters (see e.g., Elkind et al., 2012). In our numerical tests, we studied the robustness to the presence of clones of each of the considered measures. Let us call *weak axis* an axis where several candidates are clones, and describe such an axis with the same notation used for Wikipedia axes. We considered weak axes where the blocks contained approximately the same number of candidates - that means there were no

$d \backslash p$	0.7	0.75	0.8	0.85	0.9
5 candidates					
<i>VD</i>	99 %	99 %	100 %	100 %	100 %
<i>FT</i>	100 %	99 %	98 %	99 %	99 %
<i>GS</i>	100 %	100 %	100 %	100 %	100 %
6 candidates					
<i>VD</i>	95.2 %	98.7 %	100 %	100 %	100 %
<i>FT</i>	100 %	100 %	99 %	100 %	100 %
<i>GS</i>	100 %	100 %	100 %	100 %	100 %
7 candidates					
<i>VD</i>	94 %	98.7 %	100 %	100 %	100 %
<i>FT</i>	100 %	100 %	100 %	100 %	100 %
8 candidates					
<i>VD</i>	86 %	95 %	100 %	100 %	100 %
<i>FT</i>	100 %	100 %	100 %	100 %	100 %
9 candidates					
<i>VD</i>	78.5 %	89.1 %	99 %	100 %	100 %
<i>FT</i>	100 %	100 %	100 %	100 %	100 %

Table 4.7.5: Recognition rates w.r.t. measure d and probability p . The preferences are generated with steps 1-2'.

political position shared by (considerably) more candidates than the others. More precisely, we worked with the axes to simulate elections with the presence of clones among the candidates:

- for 6 candidates: $((c_1, c_2), (c_3, c_4), (c_5, c_6))$,
- for 7 candidates: $((c_1, c_2), (c_3, c_4, c_5), (c_6, c_7))$,
- for 8 candidates: $((c_1, c_2), (c_3, c_4), (c_5, c_6), (c_7, c_8))$,
- for 9 candidates: $((c_1, c_2), (c_3, c_4, c_5), (c_6, c_7, c_8), (c_9, c_{10}))$

(without loss of generality, as the labeling of the candidates does not matter). For each weak axis, number of voters and number of candidates, we generated 1000 profiles compatible with the corresponding weak axis. To do so, for each new voter, we first picked a (complete) axis A compatible with the corresponding weak axis at random uniform. Then the preference $>_v$ was chosen, still at random uniform, from the set of all preferences single-peaked with respect to A .

We computed an optimal axis according to each of the three measures for up to 7 candidates, and only the FT and VD measures beyond 7 candidates. For each measure, we counted the number of times the returned axis was compatible with the weak axis. Regarding the FT and GS measures, the optimal axis was compatible with the weak axis in all tests performed, independently of the number of voters or candidates. In contrast, the VD measure is much less robust to the presence of clones: the percentages of profiles for which the optimal axis for VD was compatible with the weak axis are given in Table 4.7.6, where “#cand.” stands for “number of candidates”.

As in the case of real-world tests, we computed the average level of structure as the ra-

#cand. \ #voters	100	200	500	1000
6	47 %	25 %	23 %	10 %
7	39 %	23 %	19 %	9 %
8	17 %	8 %	0 %	0 %
10	8 %	1 %	0 %	0 %

Table 4.7.6: Percentages of profiles for which the optimal axis for VD was compatible with the weak axis.

tion $FT(\mathcal{P})/E[FT(\mathcal{P}, A)]^{10}$ on 100 randomly generated profiles (using steps 1-2 for generating preferences), for a number $n \in \{100, 200, 300\}$ of voters, a number $m \in \{5, \dots, 10\}$ of candidates, and a probability p varying from 0.7 to 0.9. The results are reported in Table 4.7.7. There is a clear anticorrelation between the value of p (the closer to 1, the more single-peaked the generated preferences) and the computed level of structure.

¹⁰See p. 134 for more details

$m \backslash p$	0.7	0.75	0.8	0.85	0.9
100 voters					
5	0.735	0.66	0.57	0.42	0.279
6	0.75	0.66	0.558	0.441	0.297
7	0.762	0.696	0.588	0.459	0.309
8	0.762	0.702	0.606	0.462	0.318
9	0.768	0.714	0.603	0.477	0.318
10	0.771	0.711	0.621	0.486	0.336
200 voters					
5	0.759	0.651	0.546	0.411	0.267
6	0.762	0.681	0.564	0.438	0.288
7	0.786	0.705	0.588	0.453	0.306
8	0.789	0.717	0.597	0.465	0.315
9	0.792	0.726	0.603	0.48	0.321
10	0.792	0.726	0.618	0.492	0.333
300 voters					
5	0.753	0.675	0.549	0.411	0.288
6	0.765	0.696	0.576	0.429	0.3
7	0.786	0.711	0.588	0.453	0.306
8	0.786	0.72	0.612	0.465	0.318
9	0.795	0.729	0.597	0.477	0.324
10	0.807	0.732	0.618	0.48	0.336

Table 4.7.7: The average number of triples normalized by $\frac{n}{3} \binom{m-1}{2}$

4.8 Conclusion

In this chapter, we have proposed a new distance measure to single-peakedness, based on counting the number of violations of Black’s definition of single-peaked preferences.

In Section 4.4, we have undertaken an axiomatic comparison with other popular single-peakedness measures. More precisely, we have identified some theoretical properties that are (or not) guaranteed by each of the considered measures. In particular, we have seen that all these properties were guaranteed by perfectly single-peaked preferences, which provides some insights on what is lost by relaxing the notion of single-peakedness.

In Section 4.5, we have suggested a probabilistic model for nearly single-peaked preferences where the probability that a voter prefers a candidate to another depends on the position of his/her most preferred candidate on the left-right political spectrum. It has allowed us to interpret the axis minimizing the number of violations as a maximum likelihood estimate of the “correct” axis.

From the computational viewpoint, we have shown in Section 4.6 that determining an optimal axis for this measure is NP-hard. We have then presented an IP formulation of the problem. Besides, we have carried out numerical experiments on real and synthetic data, that show that the proposed measure compares favourably to other popular measures. In particular, it has an operational IP formulation while no procedure is known for the minimization of the number of swaps in the preferences to make them single-peaked; and it is more robust to noise in preferences than the minimization of the number of votes to delete to make the preference profile single-peaked.

Finally, we have held various numerical experiments in Section 4.7 on both real-world and synthetic data to evaluate the efficiency of the *FT* measure in practice, and also to provide a comparison with the *VD* measure and *GS* measure. Very briefly, the *GS* measure provides very good results, but, to our best knowledge, there does not exist an algorithm for the *GS* single-peaked consistency that would be efficient in practice. This makes this method (nearly) unusable, especially for the instance involving more than 8 candidates. On the other hand, *VD* single-peaked efficiency can be computed very efficiently, however, the results quality seems to be a step behind the results provided by the *GS* measure. The *FT* measure seems to be a good compromise between the *GS* measure and the *VD* measure, combining the good computational efficiency with promising results quality.

For future work, it would be interesting to extend the maximum likelihood approach to other distance measures and preference structures, in order to get a better intuition of which measure and which structure to consider in a given voting context. The same analytical objective may also be pursued by pushing further the axiomatic study.

Chapter appendix

4.A Detailed results on real data

File	m	n	Wikipedia axis	Optimal axis for VD	Type	Rec. rate
01	9	548	((5,9),(8,3),(7,6,1),2)	2 < 7 < 6 < 1 < 3 < 8 < 9 < 5	C	100 %
02	11	479	((9,5),(6,4,2),(10,7,8,11),(1,3))	1 < 5 < 10 < 8 < 7 < 2 < 4 < 6 < 11 < 9 < 3	F	48 %
03	10	320	((6,7),(8,9),(10,2,1,4),5)	1 < 2 < 6 < 7 < 10 < 8 < 9 < 5 < 4	F	37 %
04	11	368	((9,10),(4,6,7),(3,8,11),(1,5))	1 < 3 < 4 < 6 < 7 < 8 < 9 < 10 < 11 < 5	F	45 %
05	10	681	((1,8),(10,6,5),(4,2,7),(3,9))	3 < 2 < 10 < 6 < 5 < 4 < 1 < 8 < 7 < 9	F	45 %
06	10	520	((3,9),(10,1,4),(2,6,7),(5,8))	8 < 7 < 6 < 1 < 4 < 10 < 2 < 3 < 5 < 9	F	58 %
08	10	520	((1,6),(9,7,2),(3,5,8),(4,10))	1 < 3 < 8 < 5 < 2 < 7 < 9 < 10 < 6 < 4	F	46 %
09	11	369	((11,8),(3,4,1),(7,6,2),(5,9))	8 < 5 < 2 < 1 < 3 < 4 < 6 < 7 < 9 < 11	F	45 %
10	9	771	((9,1),(7,5,4),(8,3,2),6)	1 < 7 < 5 < 4 < 3 < 2 < 6 < 8 < 9	F	61 %
11	10	567	((4,6),(7,8),(3,5,9,10),1)	1 < 8 < 7 < 5 < 3 < 9 < 10 < 6 < 4	EE	54 %
12	8	983	((8,3),(4,5),(6,1,2),7)	7 < 6 < 5 < 4 < 1 < 2 < 3 < 8	F	64 %
13	11	433	((11,10),(8,6,3),(9,7,2),1,5)	1 < 7 < 3 < 6 < 8 < 2 < 5 < 9 < 10 < 11	F	47 %
14	8	1003	((7,2),(5,8),(6,1,3),4)	4 < 6 < 3 < 8 < 5 < 1 < 2 < 7	F	80 %
15	9	576	((3,8),(7,4),(9,5,1),6,2)	2 < 3 < 1 < 4 < 7 < 9 < 5 < 6 < 8	F	44 %
16	10	625	((8,7),(6,3,1),(4,10),(5,2))	8 < 7 < 5 < 4 < 6 < 3 < 1 < 2 < 10	F	49 %
17	9	904	((8,5),(3,4),(7,2,6),(9,1))	1 < 9 < 7 < 2 < 6 < 3 < 4 < 5 < 8	C	100 %
18	9	709	((9,1),6,(7,5,3),4)	1 < 3 < 4 < 5 < 6 < 7 < 9	F	40 %
19	11	370	((1,3),(10,9,7),(8,4,6),(5,11),2)	1 < 2 < 3 < 4 < 5 < 6 < 7 < 8 < 9 < 10 < 11	F	44 %
20	9	687	((3,6),(7,8,9),(1,2,5),4)	4 < 5 < 1 < 2 < 3 < 6 < 7 < 8 < 9	F	61 %
21	10	342	((2,1),(8,10),(9,4,5),(3,6),7)	2 < 1 < 5 < 4 < 8 < 10 < 9 < 3 < 6 < 7	F	77 %

Table 4.A.1: Results with VD for 20 wards of the Glasgow city council election. File N is ED-00008-000000NN in the PrefLib library, m the number of candidates, n the number of complete votes, *Type* the class of result, *Rec. rate* the recognition rate. For conciseness, we only give the indices of candidates in axis descriptions (c_i is denoted by i).

File	m	n	Wikipedia axis	Optimal axis for FT	Class	Rec. rate
01	9	548	((5, 9), (8, 3), (7, 6, 1), 2)	2 < 8 < 3 < 6 < 1 < 7 < 5 < 9	EE	52 %
02	11	479	((9, 5), (6, 4, 2), (10, 7, 8, 11), (1, 3))	10 < 5 < 9 < 8 < 7 < 4 < 2 < 6 < 11 < 1 < 3	F	55 %
03	10	320	((6, 7), (8, 9), (10, 2, 1, 4), 5)	6 < 4 < 2 < 5 < 10 < 8 < 9 < 7 < 1	F	42 %
04	11	368	((9, 10), (4, 6, 7), (3, 8, 11), (1, 5))	1 < 10 < 9 < 11 < 8 < 4 < 6 < 7 < 3 < 5	F	47 %
05	10	681	((1, 8), (10, 6, 5), (4, 2, 7), (3, 9))	3 < 9 < 10 < 6 < 5 < 4 < 2 < 1 < 7 < 8	F	51 %
06	10	520	((3, 9), (10, 1, 4), (2, 6, 7), (5, 8))	9 < 3 < 4 < 1 < 10 < 6 < 2 < 7 < 5 < 8	C	100 %
08	10	520	((1, 6), (9, 7, 2), (3, 5, 8), (4, 10))	6 < 1 < 3 < 5 < 2 < 7 < 9 < 8 < 4 < 10	F	67,5 %
09	11	369	((11, 8), (3, 4, 1), (7, 6, 2), (5, 9))	11 < 8 < 7 < 6 < 4 < 3 < 1 < 2 < 5 < 9	F	67,5 %
10	9	771	((9, 1), (7, 5, 4), (8, 3, 2), 6)	6 < 8 < 2 < 3 < 4 < 5 < 7 < 1 < 9	C	100 %
11	10	567	((4, 6), (7, 8), (3, 5, 9, 10), 1)	4 < 6 < 9 < 10 < 5 < 3 < 7 < 8 < 1	EE	54 %
12	8	983	((8, 3), (4, 5), (6, 1, 2), 7)	8 < 3 < 6 < 2 < 1 < 4 < 5 < 7	EE	52 %
13	11	433	((11, 10), (8, 6, 3), (9, 7, 2), 1, 5)	11 < 10 < 9 < 7 < 8 < 6 < 3 < 2 < 1 < 5	F	67,5 %
14	8	1003	((7, 2), (5, 8), (6, 1, 3), 4)	7 < 2 < 8 < 5 < 1 < 3 < 6 < 4	C	100 %
15	9	576	((3, 8), (7, 4), (9, 5, 1), 6, 2)	3 < 8 < 7 < 4 < 5 < 1 < 9 < 6 < 2	C	100 %
16	10	625	((8, 7), (6, 3, 1), (4, 10), (5, 2))	8 < 7 < 6 < 3 < 1 < 4 < 10 < 5 < 2	C	100 %
17	9	904	((8, 5), (3, 4), (7, 2, 6), (9, 1))	1 < 9 < 4 < 3 < 6 < 2 < 7 < 5 < 8	EE	61 %
18	9	709	((9, 1), 6, (7, 5, 3), 4)	4 < 5 < 7 < 6 < 3 < 1 < 9	F	86 %
19	11	370	((1, 3), (10, 9, 7), (8, 4, 6), (5, 11), 2)	11 < 6 < 10 < 9 < 7 < 8 < 4 < 3 < 1 < 5 < 2	F	44 %
20	9	687	((3, 6), (7, 8, 9), (1, 2, 5), 4)	4 < 9 < 8 < 7 < 5 < 2 < 1 < 3 < 6	EE	46 %
21	10	342	((2, 1), (8, 10), (9, 4, 5), (3, 6), 7)	3 < 5 < 9 < 10 < 8 < 4 < 2 < 1 < 6 < 7	F	47,5 %

Table 4.A.2: Results with FT for 20 wards of the Glasgow city council election. File NN is ED-00008-000000NN in the PrefLib library, m the number of candidates, n the number of complete votes, *Class* the class of result, *Rec. rate* the recognition rate.



Multidimensional Euclidean preferences

Chapter 5

An introduction to multidimensional preferences

In the previous part, we have presented several relaxations of the single-peaked domain, in order to make the domain “richer” in terms of expressivity - indeed, we recall that the single-peaked domain restriction relies on strong assumptions, and hence is very difficult to observe in real-world situations.

In this part, we present another way to relax classical domain restrictions. In Chapter 1, we have already emphasized the idea that in case of single-peaked preferences, the axis can be interpreted as a decision criterion. However, in real-world decision problems, there are usually more than one criterion to take into account. It seems then really natural to simply add more dimensions to the structure¹. But is it really so simple? Well, let us see...

5.1 Some words about multidimensional single-peaked preferences

Even though we do not deal with multidimensional single-peaked preferences in this thesis, we present here briefly this domain restriction. Multidimensional single-peaked preferences were introduced by Barberà et al. (1993) for the first time. They consider that the set of candidates form a multidimensional grid (in other words, *every* point in the grid is a candidate). More formally, let $d \geq 1$ be a dimension. For each $i \in \{1, \dots, d\}$, we define an interval of m_i integers $I_i = \llbracket a_i, b_i \rrbracket$. The set of candidates \mathcal{C} is then written as the cartesian product of these intervals :

$$C = I_1 \times I_2 \times \dots \times I_d$$

We can now define multidimensional single-peakedness, after recalling the definition of the so-called city block (or Manhattan) norm, also known as ℓ_1 -norm: given

¹More precisely, we focus in this part on a multidimensional “extension” of Euclidean preferences.

a d -dimensional real space \mathbb{R}^d and a vector $x = \{x_1, \dots, x_d\} \in \mathbb{R}^d$, we define $\|x\|_{\ell_1} = |x_1| + |x_2| + \dots + |x_d|$.

Definition 5.1.1: Multidimensional Single-Peakedness (Barberà et al., 1993)

Let $d \geq 1$ be an integer, and for each $i \in \{1, \dots, d\}$, let us consider an interval of m_i integers $I_i = \llbracket a_i, b_i \rrbracket$. Given a voter v , we denote by $>_v$ her preference and by c_v^* her most preferred candidate. The preference $>_v$ over the set of candidates $C = I_1 \times I_2 \times \dots \times I_d$ is d -single-peaked with respect to the box $B = I_1 \times I_2 \times \dots \times I_d$ if for each couple of candidates c, c' such that

$$\|c_v^* - c'\|_{\ell_1} = \|c_v^* - c\|_{\ell_1} + \|c - c'\|_{\ell_1},$$

we have $c >_v c'$.

A preference profile \mathcal{P} is d -single-peaked with respect to the box B if each preference of \mathcal{P} is d -single-peaked with respect to B .

In other words, this definition generalizes the classical idea of single-peakedness: if c lies *between* c_v^* and c' , then c is preferred to c' by v . Here, the notion of *betweenness* is defined using the ℓ_1 norm - c is *between* c_v^* and c' if it lies in the “bounding box” of c_v^* and c' . See Figure 5.1.1 for illustration.

A Curious Tortoise Intervention



In particular, if $d = 1$, we have $\mathcal{C} = I_1$, with $T_1 = \llbracket a_1, b_1 \rrbracket$. In other words, there are m_1 candidates $a_1, (a_1 + 1), (a_1 + 2), \dots, b_1$. This interval can be actually seen as an axis $a_1 \triangleleft a_1 + 1 \triangleleft \dots \triangleleft b_1$. We get hence a generalization of the classical one-dimensional single-peakedness.

Later, this definition has been generalized by Sui et al. (2013) to profiles over a set of candidates C that can be *embedded* in a grid (i.e., non every point in the grid is a candidate). Thus, contrary to the previous case, we have a set of m candidates $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ and d axis A_1, A_2, \dots, A_d on \mathcal{C} (i.e., the axes are not intervals of integers, but (total) orderings of \mathcal{C}). Denoting by $p_i(c_j)$ the position of c_j in the axis A_i , the candidate c_i is mapped to the grid point $(p_1(c_i), p_2(c_i), \dots, p_m(c_i))$. The idea remains then similar - c is preferred to c' whenever it lies in the “bounding box” of c_v^* and c' - see Figure 5.1.2 for illustration.

Definition 5.1.2: Multidimensional Single-Peakedness (Sui et al., 2013)

Let $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ be a set of candidates, $d \geq 1$ be an integer and A_1, A_2, \dots, A_d be d different axis (i.e., linear orders \triangleleft_{A_i} for $i \in \{1, 2, \dots, d\}$) on \mathcal{C} . A preference $>_v$ is d -single-peaked with respect to $A = (A_1, A_2, \dots, A_d)$ if $c >_v c'$ whenever we have, for each axis A_i (where $i \leq d$), $c' \triangleleft_{A_i} c \triangleleft_{A_i} c_v^*$ or $c_v^* \triangleleft_{A_i} c \triangleleft_{A_i} c'$.

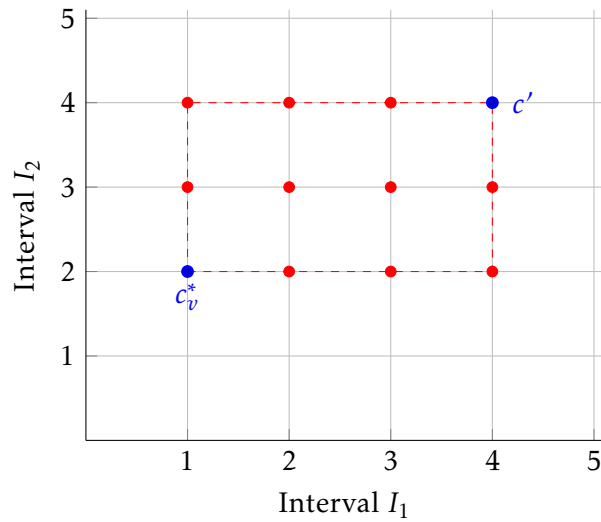


Figure 5.1.1: Multidimensional single-peakedness by Barberà et al. (1993): for any candidate c in the bounding box of c_v^* and c' , we have $c >_v c'$.

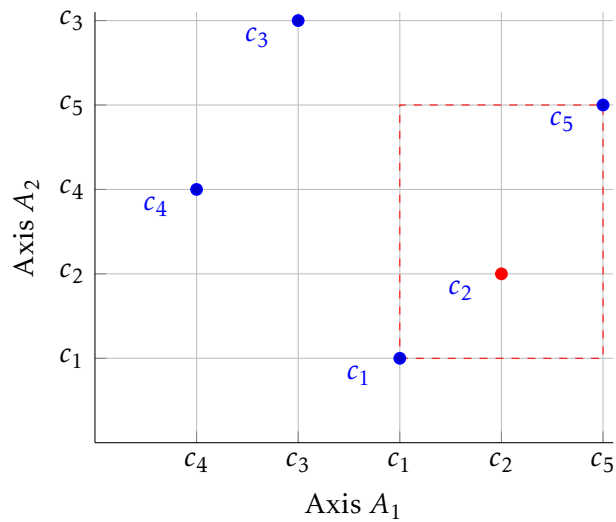


Figure 5.1.2: Multidimensional single-peakedness by Sui et al. (2013) (the example is taken from there): c_2 is in the bounding box of c_1 and c_5 . Therefore, for any preference $>_v$ with $c_v^* = c_1$ (resp. c_5), we have $c_2 >_v c_5$ (resp. $c_2 >_v c_1$).

Finally, since good things come in threes, Elkind et al. (2022) recently noticed that this second definition was not closed under alternative deletion - indeed, if we remove the most preferred candidate c_v^* of the voter v , the definition may not be satisfied for the new most preferred candidate of v - see for instance the Example 20 of Elkind et al. (2022). That is why a new definition is proposed, in which all triples of voters are considered, and not only those involving c_v^* :

Definition 5.1.3: Multidimensional (Hereditary) Single-Peakedness (Elkind et al., 2022)

Let $\mathcal{C} = \{c_1, c_2, \dots, c_m\}$ be a set of candidates, $d \geq 1$ be an integer and A_1, A_2, \dots, A_d be d different axis (i.e., linear orders \prec_{A_i} for $i \in \{1, 2, \dots, d\}$) on \mathcal{C} . A preference $>_v$ is d -single-peaked with respect to $A = (A_1, A_2, \dots, A_d)$ if for any triple of candidates c, c', c'' , such that $c \prec_{A_i} c' \prec_{A_i} c''$ for every $i \leq d$, we do not have both $c >_v c'$ and $c'' >_v c'$.

It goes without saying that the domain of multidimensional single-peaked preferences is much less restrictive than the classical one-dimensional version. Indeed, the restriction is maybe even weaker than one would like it to be. Concretely, Elkind et al. (2022) showed that any profile with at most $2^{2^{d-1}}$ candidates is d -single-peaked. From the practical point of view, the empirical tests carried out on real-world elections by Sui et al. (2013) showed that up to 65% of votes could be explained by using 2-single-peaked preferences, and a nearly perfect fit was obtained by combining with a low degrees of local candidate deletion measure.

It is worth mentioning that many nice properties of classical single-peakedness are lost by adding more dimensions. For instance, the transitivity of the majority relation cannot be guaranteed for 2-single-peaked elections.

5.2 Multidimensional Euclidean preferences

The next two chapters focus on multidimensional Euclidean preferences (also known as spatial preferences or multidimensional unfolding models). Although the adjective *Euclidean* usually involves the use of the ℓ_2 norm (which is also called *Euclidean* norm), several different norms are considered in our works. By abuse of language, no matter what norm $\|\cdot\|$ is used, we call the resulting domain restriction *Euclidean preferences (with respect to the norm $\|\cdot\|$)*, in order to refer to the original idea of spatial preferences.

5.2.1 The State of the Art

The study of spatial models is an important stream of research in social choice, pioneered by the works of Hotelling (1929) and Downs (1957). The most widely studied Euclidean preferences are those that are derived by measuring the distances with the ℓ_2 norm. In seminal works, Bennett and Hays (1960), as well as Hays and Bennett (1961)

discussed some structural questions - among others, they established that the maximal size (number of distinct votes) of a Euclidean profile in function of the number m of candidates is equal to $\sum_{k=m-d}^m |s(m, k)|$ where $s(m, k)$ are the (unsigned) Stirling numbers of the first kind (the same result has been found by Good and Tideman (1977)). Also, given a preference profile, the authors asked themselves which is the minimal dimension d such that the profile is d -Euclidean. They gave some techniques to obtain bounds on d . Some decades later, Bogomolnaia and Laslier (2004) showed that $d \geq \min\{m, n-1\}$ is necessary to ensure that any profile of n votes on m candidates is d -Euclidean (i.e., Euclidean in dimension d). They also showed that any profile of at most 2 voters or 3 candidates is 2-Euclidean. These results were deepened by Bulteau and Chen (2022): they proved that any profile of 3 voters on at most 7 candidates is 2-Euclidean. Finally, let us mention that Kamiya et al. (2011) studied the question of counting and enumerating d -Euclidean profiles of maximal size² in function of the number m of candidates. They provided a formula for the number of profiles of maximal size if $d = m - 2$, and they were able to enumerate them for $m = 4$ (and $d = 2$). Regarding the computational aspects, Peters (2017) studied the recognition problem (i.e., deciding whether or not a preference profile is d -Euclidean) for $d > 1$. He proved that the recognition problem is equivalent to the existential theory of reals, and thus $\exists\mathbb{R}$ -complete. Actually, some Euclidean preference profiles require exponentially many bits in order to specify any Euclidean embedding, so it is even not clear if the decision version of the problem is in NP.

We note that there are also some works focusing on Euclidean preferences using ℓ_1 and ℓ_∞ norms: Recently, Chen et al. (2022) showed that each preference profile with m alternatives and n voters is d -Euclidean with respect to the norm ℓ_1 whenever $d \geq \min\{n, m-1\}$. Also, they studied the smallest non-Euclidean profiles (still with respect to the norm ℓ_1) in case of $d = 2$. Peters and Lackner (2017) proved that the recognition problem is in NP for d -Euclidean preferences with respect to the ℓ_1 norm and the ℓ_∞ norm; the precise complexity of these problems (in particular, if it is in P or not) remains an open question.

From a more operational point of view, spatial representations are used in particular in voting advice applications (e.g., Wahl-O-Mat in Germany, Smartvote in Switzerland, Vote Compass in the United States, and many others in multiple countries), i.e., online tools that helps the voter to choose the candidate closest to her political stances, and actually often provides her a full ranking of candidates according to her answers to a survey on a range of policy statements. The answers are indeed converted into positions on different dimensions, each position reporting on the level of agreement on a particular policy statement. The ℓ_1 norm is typically used when there are many dimensions, while the ℓ_2 norm is used when the number of dimensions is lower (Moreno et al., 2022; Isotalo, 2020). For an overview of the topic of voting advice applications, the reader may refer to the survey by Garzia and Marschall (2019).

²in terms of the number of pairwise distinct preferences

5.2.2 Definitions

We denote by $d \geq 2$ the space dimension and by $\|\cdot\| : \mathbb{R}^d \rightarrow \mathbb{R}^+$ a norm application. With these notations, we can define *d-Euclidean preferences with respect to the norm $\|\cdot\|$* :

Definition 5.2.1: *d-Euclidean preferences with respect to the norm $\|\cdot\|$*

Let $d \geq 1$ be an integer. A preference profile P of a set V of n voters over a set C of m candidates is *d-Euclidean with respect to the norm $\|\cdot\|$* if there exists a mapping $f : V \cup C \rightarrow \mathbb{R}^d$ such that for each $v \in V$ and each couple of candidates $c_i, c_j \in C$:

$$c_i >_v c_j \Leftrightarrow \|f(v) - f(c_i)\| < \|f(v) - f(c_j)\|.$$

Note that in the sequel we will assume, w.l.o.g., that no couple of voters have the same preference (if so, we can simply remove one of them). The norms considered in our works are norms ℓ_1 , ℓ_2 and ℓ_∞ . For convenience, let us give their definitions - we denote by $x = (x_1, x_2, \dots, x_d)$ a vector of \mathbb{R}^d :

$$\|x\|_{\ell_1} = |x_1| + |x_2| + \dots + |x_d|$$

$$\|x\|_{\ell_2} = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}$$

$$\|x\|_{\ell_\infty} = \max\{x_1, x_2, \dots, x_d\}$$

Note that in which follows, the used norm will always be specified. Nevertheless, if no confusion is possible, we will only write *d-Euclidean preferences* instead of *d-Euclidean preferences with respect to the norm $\|\cdot\|$* .

Example 5.2.1. *To illustrate the notion of multidimensional Euclidean preferences, let us consider a profile of 3 preferences over 3 candidates:*

$$>_1: (c_1, c_3, c_2)$$

$$>_2: (c_3, c_2, c_1)$$

$$>_3: (c_1, c_2, c_3)$$

This profile is 2-Euclidean with respect to the norm ℓ_2 : we can for instance set $f(v_1) = (2, 5)$, $f(v_2) = (11, 2)$, $f(v_3) = (2, 9)$, $f(c_1) = (3, 3)$, $f(c_2) = (8, 6)$ and $f(c_3) = (6, 2)$. We check easily that the definition of a 2-Euclidean profile is satisfied, as illustrated on Figure 5.2.1 for the first two voters. Intuitively, the idea is still the same as in the case of 1-Euclidean preferences: the voters and candidates are embedded in the same space, and the closer the voter v is to the candidate c , the better c is ranked in $>_v$.

The mapping f is called a *d-Euclidean representation of P* . Obviously, such a mapping is not necessarily unique. Let us interpret the definition in a slightly different way: imagine that the position $f(c)$ is fixed for each candidate $c \in C$. Our goal is now to extend f on $C \cup V$ (i.e., define the positions of the voters $v \in V$) in such a way that f

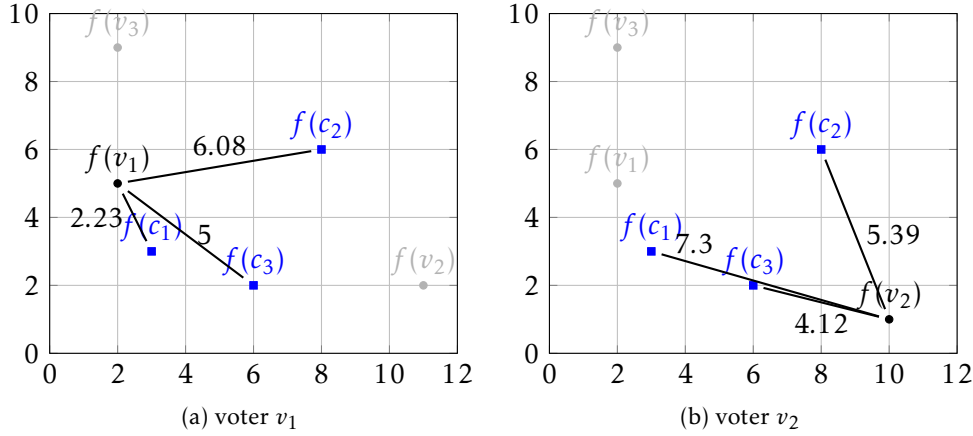


Figure 5.2.1: The mapping f satisfies the definition of a 2-Euclidean profile with respect to the norm ℓ_2 (here it is only illustrated for voters v_1 and v_2).

fulfils the condition of Definition 5.2.1. For each voter v , we define the set $D_f^\ell(v)$ of possible positions $f(v)$:

$$D_f^\ell(v) = \{p \in \mathbb{R}^d : \forall c_i, c_j \in C, c_i >_v c_j \Rightarrow \|p - f(c_i)\|_\ell < \|p - f(c_j)\|_\ell\}$$

Clearly, mapping f is a d -Euclidean representation of P (with respect to the norm ℓ) if and only if $D_f^\ell(v) \neq \emptyset$ for each $v \in V$. If no confusion is possible, we only write $D_f(v)$ instead of $D_f^\ell(v)$. We note that, following this idea, to build a Euclidean representation of P , we somehow only need to define the positions of the candidates in such a way that each $D_f^\ell(v)$ is non-empty. For this purpose, we introduce the following notations:

Definition 5.2.2

Let $d \geq 1$ be an integer, $\|\cdot\|_\ell$ be a norm of \mathbb{R}^d and $f : C \rightarrow \mathbb{R}^d$ be a mapping. For a pair $\{c_i, c_j\} \subseteq C$ of candidates mapped in positions $f(c_i)$ and $f(c_j)$, the set of points $p \in \mathbb{R}^d$ such that $\|f(c_i) - p\|_\ell = \|f(c_j) - p\|_\ell$ is called *the boundary hypersurface of c_i and c_j* (or just hypersurface in what follows), and is denoted by $H_f^\ell(c_i, c_j)$. We denote then by $D_f^\ell(c_i, c_j)$ the set of points $p \in \mathbb{R}^d$ such that $\|f(c_i) - p\|_\ell < \|f(c_j) - p\|_\ell$, and by $D_f^\ell(c_j, c_i)$ the set of points $p \in \mathbb{R}^d$ such that $\|f(c_i) - p\|_\ell > \|f(c_j) - p\|_\ell$.

In no confusion is possible, we only write $H_f(c_i, c_j)$ instead of $H_f^\ell(c_i, c_j)$. With the notations of Definition 5.2.2, we have $D_f^\ell(v) = \bigcap_{c_i >_v c_j} D_f^\ell(c_i, c_j)$. The notations are illustrated in Figure 5.2.2.

Definition 5.2.1 can be now reformulated as follows :

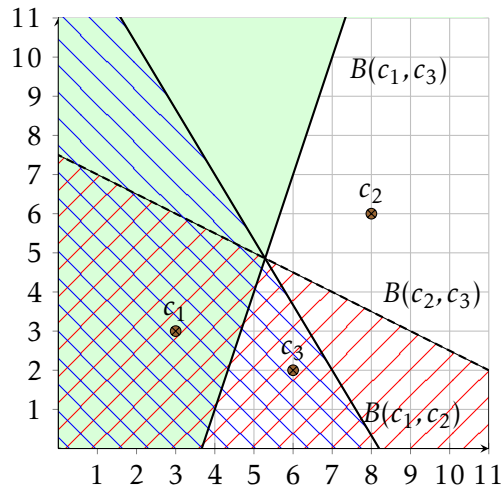


Figure 5.2.2: A Euclidean representation of the profile consisting of the 6 possible preferences on 3 candidates c_1, c_2, c_3 . The area with a solid background corresponds to $D_f(c_1, c_3)$, and the area with a downward (resp. upward) diagonal hatching to $D_f(c_1, c_2)$ (resp. $D_f(c_3, c_2)$). The intersection of these areas corresponds to $D_f(v)$ for $c_1 >_v c_2 >_v c_3$.

Definition 5.2.3

A preference profile P is d -Euclidean with respect to the norm ℓ if there exists a mapping $f : C \rightarrow \mathbb{R}^d$ such that the set $D_f^\ell(v)$ induced by f is non-empty for each $v \in V$.

5.2.3 Weighted majority tournaments and Kemeny rankings with 2-Euclidean preferences

As for the other domain restrictions, we are interested in determining whether d -Euclidean preferences (with respect to a given norm) allow to circumvent some NP-hardness results in social choice problems. Unfortunately, no such result has been identified so far to the best of our knowledge, and several results were proved to remain NP-hard for d -Euclidean preferences. We prove in this section that the Kemeny ranking problem, NP-hard in general but polynomial-time solvable if the preferences are 1-Euclidean³, is still NP-hard for 2-Euclidean preferences under norms ℓ_1, ℓ_2 and ℓ_∞ ; we then mention at the end of the section other NP-hardness results for d -Euclidean preferences.

Organization of this section. Several notions first need to be introduced:

1. We first define the Kemeny ranking problem.

³actually, single-peaked is sufficient

2. Then we introduce the *Feedback arc set problem* (FAS): A classical way to prove that the Kemeny ranking problem is NP-hard in the general case is to provide a polynomial time reduction from FAS to the Kemeny ranking problem. FAS is a problem formulated on (general) directed graphs. It turns out that the decision version of this problem remains NP-hard on bipartite weighted tournaments, which is a special class of directed graphs the definition of which we will recall below.

Once these notions are introduced, we prove the NP-hardness of the Kemeny ranking problem for 2-Euclidean preferences. Actually, we prove that any weighted tournament (resp. bipartite weighted tournament) is *inducible* (see Definition 5.2.5 below) by a 2-Euclidean profile with respect to the norm ℓ_1 (resp. ℓ_2, ℓ_∞). As FAS is known to remain NP-hard on bipartite weighted tournaments, it follows that the Kemeny ranking remains NP-hard for 2-Euclidean preferences.

Kemeny ranking Given two preferences $>_1$ and $>_2$, let $d_{KT}(>_1, >_2)$ denote the Kendall-tau distance between $>_1$ and $>_2$, i.e., the number of pairs of candidates $\{c_i, c_j\}$ such that $c_i >_1 c_j$ and $c_j >_2 c_i$, or vice versa. The Kendall-tau distance $KT(>, \mathcal{P})$ between a ranking $>$ and a profile \mathcal{P} is then defined as

$$KT(>, \mathcal{P}) = \sum_{>_v \in \mathcal{P}} d_{KT}(>, >_v)$$

Now, we are able to define the Kemeny ranking problem:

Definition 5.2.4: Kemeny ranking problem

In the Kemeny ranking problem, given a preference profile \mathcal{P} , we want to determine a ranking $>$ on the candidates that minimizes $KT(>, \mathcal{P})$. Such a ranking is called *Kemeny ranking*.

In the decision version of the Kemeny ranking problem, given some integer k , we want to determine whether there exists a ranking $>$ such that $KT(>, \mathcal{P}) \leq k$, or not. As stated before, we recall that this problem is known to be NP-complete (Bartholdi et al., 1989).

Example 5.2.2. Let us consider the following profile \mathcal{P} of 3 preferences on 4 candidates:

$$\begin{aligned} >_1: (c_1, c_2, c_3, c_4) \\ >_2: (c_3, c_2, c_4, c_1) \\ >_3: (c_2, c_3, c_4, c_1) \end{aligned}$$

Let us consider the ranking $c_2 > c_3 > c_4 > c_1$. We have:

$$KT(>, \mathcal{P}) = d_{KT}(>, >_1) + d_{KT}(>, >_2) + d_{KT}(>, >_3) = 3 + 1 + 0 = 4$$

Let us now consider another ranking $c_1 >' c_2 >' c_3 >' c_4$. We have:

$$KT(>', \mathcal{P}) = d_{KT}(>', >_1) + d_{KT}(>', >_2) + d_{KT}(>', >_3) = 0 + 4 + 3 = 7$$

It appears therefore that $>$ is nearer to \mathcal{P} than $>'$ (in terms of d_{KT}). Actually, $>$ is a Kemeny ranking for \mathcal{P} . We can check that \mathcal{P} is single-peaked w.r.t. the axis $A = c_1 \triangleleft c_2 \triangleleft c_3 \triangleleft c_4$. It is well-known (see Black (1958)) that the majority relation of a single-peaked profile with an odd number of voters is transitive. But when the majority relation is transitive, the Kemeny ranking is unique and corresponds to the majority relation.

It is easy to check that $>$ is the majority relation of \mathcal{P} . In particular, we see that the Kemeny ranking problem is polynomial time solvable for single-peaked preferences.

Feedback arc set problem In the decision version of the feedback arc set problem (FAS), we are given a directed graph G and an integer k , and we want to determine whether we can delete (at most) k arcs in G in such a way that the resulting graph is acyclic. This problem is NP-complete (Ausiello et al., 1980). It is well known that it remains NP-hard in bipartite graphs, as one can replace each arc $e_i = (u, v)$ by two arcs $(u, w_i), (w_i, v)$ where w_i is a new vertex associated to e_i , and obtain an equivalent instance in a bipartite graph.

Weighted tournaments A tournament is a directed graph $G = (N, A)$ where for each pair $\{u_i, u_j\}$ of vertices, there is exactly one arc - either (u_i, u_j) or (u_j, u_i) . The name *tournament* refers to a situation where a game (without tie) was organized between each pair of nodes $\{u_i, u_j\}$ and the arc represents who won the game (i.e., the arc (u_i, u_j) represents the situation in which u_i wins over u_j).

Weighted tournaments are then a generalization where each game is won by a certain margin which defines the integer weight $w(u_i, u_j) \geq 0$ of arc (u_i, u_j) . Note that this generalization allows ties - $w(u_i, u_j) = 0$ if u_i and u_j are ex æquo - in this case, there is actually no arc between u_i and u_j . A weighted tournament is bipartite if the partial digraph with edges with non-zero weights is bipartite.

Definition 5.2.5: Inducible weighted tournament

A weighted tournament on a set $N = \{u_1, \dots, u_m\}$ of nodes is *inducible* if there exists a preference profile \mathcal{P} of n preferences on a set $C = \{c_1, \dots, c_m\}$ of candidates such that for any pair $\{u_i, u_j\}$ of nodes, $w(u_i, u_j) = |\{v \in V : c_i >_v c_j\}| - |\{v \in V : c_j >_v c_i\}|$.

It is known (Debord, 1987), (McGarvey, 1953) that a weighted tournament is inducible if and only if all the weights are of the same parity. In the following, we refer to odd (resp. even) weighted tournaments if all the weights are odd (resp. even). Note that when $|N| \geq 3$, a bipartite weighted tournament is necessarily even (as there is at least one pair $\{u_i, u_j\}$ with $w(u_i, u_j) = 0$).



Figure 5.2.3: The tournament induced by the profile from Example 5.2.2.

Debord’s theorem apply to 2-Euclidean preferences We actually prove a version of Debord’s theorem (Debord, 1987) (refining McGarvey’s theorem (McGarvey, 1953)) for 2-Euclidean profile (the proof is on page 161):

Theorem 5.2.1

Every weighted tournament with weights of the same parity is inducible by a 2-dimensional ℓ_2 -Euclidean profile. Every weighted bipartite tournament with even weights is inducible by a 2-Euclidean profile w.r.t. norm ℓ_1 , and by a 2-Euclidean profile w.r.t. norm ℓ_2 .

Basically, this theorem states that every weighted (bipartite) tournament can be seen as a 2-Euclidean profile (under one of considered norms). Thus, essentially, hardness results for computational social choice problems that can be formulated on the (weighted) majority tournament are still true if preferences are 2-dimensional Euclidean because this assumption is not restrictive with regard to the weighted majority tournament.⁴ In particular, this is the case of the Kemeny ranking problem - as said above, FAS is NP-hard on (bipartite) weighted tournaments, so the classical reduction from FAS to the Kemeny ranking problem can be used in case of 2-Euclidean preferences:

Corollary 5.2.1

Under norms ℓ_1 , ℓ_2 and ℓ_∞ , the Kemeny ranking problem on 2-Euclidean preferences is NP-hard. This is true even if a 2-dimensional representation of preferences is given in the input.

Another example of a problem that remains NP-hard for 2-Euclidean preferences is the *Slater rule problem*. The Slater rule asks for a consensus ranking which minimizes the number of disagreements with pairwise majority comparisons (Slater, 1961). While

⁴In other words, the hypothesis of 2-Euclidean preferences does not provide us a restriction on a “good” subset of weighted majority tournaments on which these problems might become polynomial time solvable.

this rule is often considered as a tournament solution concept (the Slater set consists of the winning candidates), it also defines a consensus ranking given a preference profile (Conitzer, 2006). Using Theorem 5.2.1, we get the following result:

Corollary 5.2.2

Under norms ℓ_2 , ℓ_1 and ℓ_∞ , the Slater ranking problem on 2-Euclidean preferences is NP-hard. This is true even if a 2-dimensional representation of preferences is given in the input.

A classical way to build a profile that induces a given weighted tournament G is to convert the nodes of G into candidates and the arcs into voters. More precisely, consider an even weighted tournament⁵, and suppose that we build a profile such that:

- There is one candidate c_i for each vertex u_i ;
- For each arc (u_i, u_j) , there are $w(u_i, u_j)/2$ (identical) copies of 2 voter f_{ij} and g_{ij} .
- All copies of f_{ij} and g_{ij} prefer c_i to c_j ;
- For any other pair $\{c, c'\}$ of candidates, exactly one voter among f_{ij} and g_{ij} prefers c to c' (and one prefers c' to c).

Then such a profile clearly induces the desired weighted tournament. If preferences are unrestricted, such properties for the preferences of f_{ij} and g_{ij} can be obtained for instance by following the approach proposed by McGarvey (1953): it consists in defining the preferences

$$f_{ij} = (c_i, c_j, \underbrace{c_1, c_2, \dots, c_n}_{\text{except } c_i \text{ and } c_j})$$

and

$$g_{ij} = (\underbrace{c_n, c_{n-1}, \dots, c_1}_{\text{except } c_i \text{ and } c_j}, c_i, c_j)$$

In other words, g_{ij} is the reverse ranking of f_{ij} , except the relative order of c_i and c_j which remains the same in both rankings.

In the sequel, we show that we can still obtain the previous properties with an Euclidean profile under norms ℓ_1 , ℓ_2 or ℓ_∞ .

⁵If the tournament is odd, then we can add one voter to the construction, and use pairs f_{ij} and g_{ij} to adjust the weights as needed.

Proof of Theorem 5.2.1 under ℓ_1 We start with a weighted bipartite tournament (thus necessarily even, as mentioned earlier) G with vertex set $L \cup R$ and arc set A (each arc having one extremity in L and one in R). We denote by n the number of vertices, and by m the number of arcs. We build an instance where candidates and voters lie on a square, whose sides are parallel to the axes, (see Figure 5.2.4). More precisely:

- Each vertex u_i corresponds to a candidate c_i . If $u_i \in L$ (resp. $u_i \in R$), c_i will be on the vertical left side (resp. right side) of the square. We will say that $c_i \in L$ (resp. $c_i \in R$) if $u_i \in L$ (resp. $u_i \in R$)
- Each arc (u_i, u_j) correspond to two voters f_{ij} and g_{ij} . Point f_{ij} will be on the horizontal upper side of the square, while g_{ij} will be on the horizontal lower side of the square.

Let us consider an arc (u_i, u_j) , with $u_i \in L$ and $u_j \in R$. We call A_{ij} the point on the upper horizontal side such that $\|c_i - A_{ij}\|_{\ell_1} = \|c_j - A_{ij}\|_{\ell_1}$ (note that such point indeed exists on the horizontal side of the square). Similarly, we call B_{ij} the point on the lower horizontal side such that $\|c_i - B_{ij}\|_{\ell_1} = \|c_j - B_{ij}\|_{\ell_1}$.

We put two voters f_{ij} and g_{ij} which are positioned on the edges at ε (to be specified) to the left of A_{ij} and B_{ij} , respectively. If the arc had been (u_j, u_i) , then the voters f_{ji} and g_{ji} would have been at ε to the right of A_{ij} and B_{ij} .

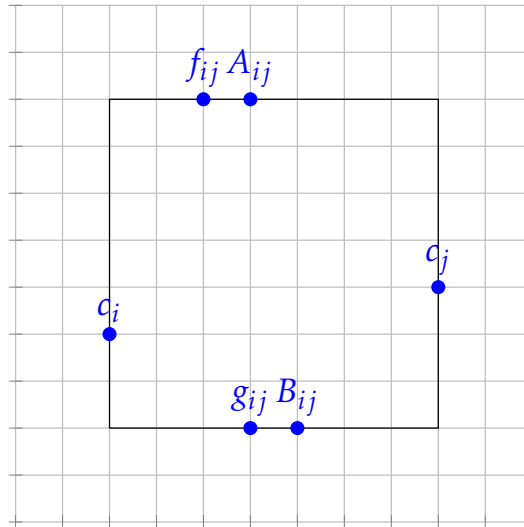


Figure 5.2.4: The construction with two vertices $u_i \in L$, $u_j \in R$, and an arc (u_i, u_j) .

Assume that we choose the vertical positions of candidates in such a way that all A_{ij} are distinct (and equivalently, all B_{ij} are distinct), see below for an explicit construction. Note that as B_{ij} is the symmetric of A_{ij} with respect to the center of the square, the order of A -points on the upper side is the inverse order of B -points on the lower side.

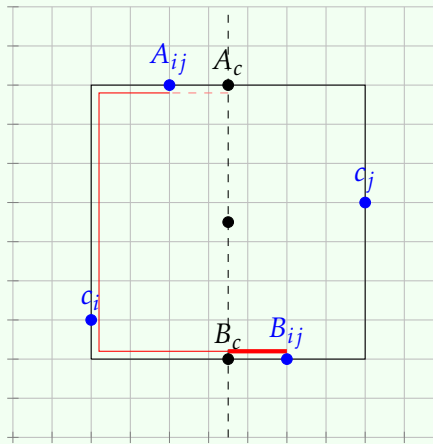
A Curious Tortoise Intervention



Wait... why is B_{ij} the symmetric of A_{ij} with respect to the center of the square? Well, it is actually very easy: let us denote by d_A (resp. d_B) the ℓ_1 -distance between A_{ij} and c_i, c_j (resp. B_{ij} and c_i, c_j), and let Δ be the side length. We have

$$2d_A + 2d_B = 4\Delta.$$

In other words, $d_A + d_B = 2\Delta$, or also the half-perimeter of the square. Let A_c (resp. B_c) be the midpoint of the horizontal upper side (resp. horizontal lower side) of the square. Suppose without loss of generality that A_{ij} is on the left of A_c . Then B_{ij} must be “shifted” by the same distance to the right of B_c so that we still have $d_A + d_B = 2\Delta$:



Then we choose ε sufficiently small so that between f_{ij} and A_{ij} there is no other A -point. Similarly, between g_{ij} and B_{ij} there is no other B -point.

Let us consider an arc (u_i, u_j) with $u_i \in L$ and $u_j \in R$ (the other case being completely symmetric). Then:

- Both voters f_{ij} and g_{ij} prefer c_i to c_j (as A_{ij} is equidistant from c_i and c_j , the same for B_{ij}).
- For any other pair $\{c, c'\}$ of candidates, exactly one voter among f_{ij} and g_{ij} prefers c to c' (and one prefers c' to c). This is easy to see if both c and c' belong to L , or if both belong to R . If $c \in L$ and $c' \in R$, we use the fact that A -points and B -points are symmetric with respect to the center of the square. Therefore, if A_{ij} is on the left of $A_{cc'}$, B_{ij} is on the right of $B_{cc'}$. But as there is no other A -point (resp. B -point) between A_{ij} and f_{ij} (resp. B_{ij} and g_{ij}), we also have f_{ij} on the left of $A_{cc'}$ and g_{ij} on the right of $B_{cc'}$. In other words, f_{ij} prefers c to c' and g_{ij} prefers c' to c (see

Figure 5.2.5). The case of A_{ij} on the right of $A_{cc'}$ (and so B_{ij} on the left of $B_{cc'}$) is treated symmetrically.

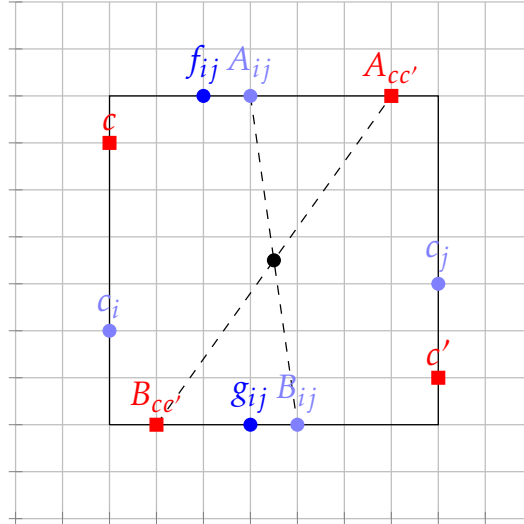


Figure 5.2.5: $c \in L$ and $c' \in R$: f_{ij} is on the left of $A_{cc'}$, so she prefers c to c' . Voter g_{ij} is on the right of $B_{cc'}$, so she prefers c' to c .

Thus, this construction fulfills the conditions 1-4 given above, and yields a profile inducing the desired (bipartite) tournament. We now give an explicit (polynomial time) construction which ensures that A -points are distinct. Consequently, all B -points are distinct as they are symmetric to A -points with respect to the center of the square.

Explicit construction:

We consider a square with side lengths $\Delta = 2^{n+1}$, where n is the number of vertices of the graph. Let us consider that the bottom left corner of the square has coordinates $(0,0)$.

We set the y -coordinate of candidate c_i to $y_i = 2^i$. Then the x -coordinate x_{ij} of A_{ij} is such that $x_{ij} + \Delta - y_i = \Delta - x_{ij} + \Delta - y_j$, meaning that:

$$x_{ij} = \frac{\Delta + y_i - y_j}{2}.$$

Then we can verify that these y -values are such that all A -points are distinct. Indeed, for any distinct pairs $\{i, j\}$ and $\{k, \ell\}$ of indices, $y_i + y_j \neq y_k + y_\ell$. To see this, if say ℓ is the largest among the indices, then:

- If $j = \ell$, then $i \neq k$ (as the pairs are distinct), and $y_i + y_j \neq y_k + y_\ell$.

- If $i, j < \ell$, then $y_i + y_j \leq 2(2^{\ell-1}) = 2^\ell = y_\ell < y_\ell + y_k$.

Then all the values $x_{ij} = \frac{y_i - y_j}{2}$ are distinct. Note that as y -values and Δ are even integers, x_{ij} is an integer, and we can choose $\varepsilon = \frac{1}{2}$ (and multiply everything by 2 if we want integers).

As the coordinates can be encoded with a polynomial number of bits, the reduction is polynomial. \square

Proof of Theorem 5.2.1 under ℓ_∞ We use a construction which is similar to the case of ℓ_1 , but positioning candidates and voters on a square which is oriented as in Figure 5.2.6. The diagonal of the square has length 2Δ with $\Delta = 2^{n+1}$.

We position a candidate $c_i \in L$ on the lower left side, at position $(-2^i, 2^i - \Delta)$. A candidate $c_j \in R$ is on the upper right side, at position $(2^j, \Delta - 2^j)$.

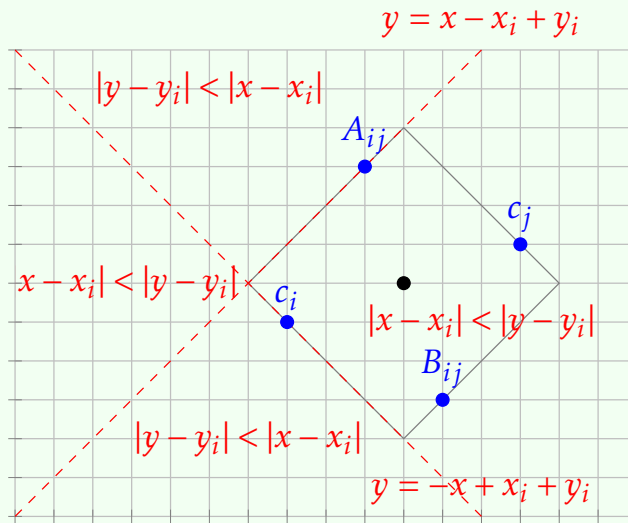
Then we define two points A_{ij} and B_{ij} , respectively on the upper left side and on the lower right side, both being equidistant (under ℓ_∞) from c_i and c_j . Namely, the coordinates of A_{ij} are $(\frac{2^i+2^j}{2} - \Delta, \frac{2^i+2^j}{2})$. Point B_{ij} is the symmetric of A_{ij} with respect to the center O of the square. See Figure 5.2.6 for an illustration.

A Curious Tortoise Intervention



How did we compute the coordinates of A_{ij} , and how can we be sure that such a point always exists? Well, let us do some calculations...

Suppose that we have an arc (v_i, v_j) , with $v_i \in L$ and $v_j \in R$. We have $c_i = (-2^i, 2^i - \Delta)$ and $c_j = (2^j, \Delta - 2^j)$. We want to find a point $A_{ij} = (x_A, y_A)$ equidistant from c_i and c_j under norm ℓ_∞ . If we denote by (x_i, y_i) , resp. (x_j, y_j) , the coordinates of c_i , resp. c_j , it is easy to see that $|x_A - x_i| < |y_A - y_i|$, resp. $|x_A - x_j| > |y_A - y_j|$ - see the scheme below for a better visualization:



We have thus

$$y_A - (2^i - \Delta) = 2^j - x_A.$$

As the point A_{ij} is on the upper left side of the square, we have

$$y_A = x_A + \Delta.$$

Put together, we get

$$x_A + \Delta - 2^i + \Delta = 2^j - x_A.$$

Finally, we obtain

$$x_A = \frac{2^i + 2^j}{2} - \Delta,$$

and

$$y_A = \frac{2^i + 2^j}{2}.$$

Moreover, $-\Delta \leq x_A \leq 0$, so A_{ij} is on the upper left side of the square. Similar computations can be done for B_{ij} .

Nevertheless, there is a more elegant way to obtain the positions of A_{ij} and B_{ij} . If you compare Figure 5.2.4 and Figure 5.2.6, maybe you see how to do it. Otherwise, we will come back to this point in Chapter 7.

As previously, if there is an arc (v_i, v_j) with $v_i \in L$ and $v_j \in R$, we create two voters f_{ij} and g_{ij} , point f_{ij} being positioned on the edge of A_{ij} at ε to the bottom/left of A_{ij} , and g_{ij} being positioned on the edge of B_{ij} at ε to the bottom/left of B_{ij} . If there is an arc (v_j, v_i) with $v_i \in L$ and $v_j \in R$, then f_{ji} and g_{ji} are positioned on the edges at ε to the right/up of A_{ij} and B_{ij} .

The choice of the coordinates of candidates ensure that all the A -points and B -points are distinct (for the same reason as in the proof for norm ℓ_1), and integral, so we can choose $\varepsilon = 1/2$ to ensure that there is no A -points between f_{ij} and A_{ij} (neither B -points between g_{ij} and B_{ij}). \square

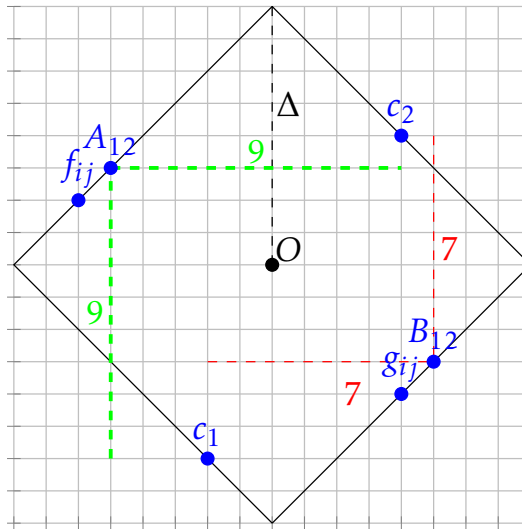


Figure 5.2.6: The construction with only two vertices $v_1 \in L$ and $v_2 \in R$, and the arc (v_i, v_j) .

Proof of Theorem 5.2.1 under ℓ_2 We now start with an even weighted tournament, and will position candidates and voters on a circle, centered at point O of coordinates $(0, 0)$. More precisely (see Figure 5.2.7):

- Each vertex v_i corresponds to a candidate c_i positioned on the circle.
- Let us call D_{ij} the line of equidistant points (under ℓ_2) between c_i and c_j , and A_{ij} and B_{ij} the two points of D_{ij} on the circle. Each arc (v_i, v_j) correspond to two voters f_{ij} and g_{ij} , both positioned on the circle. Point f_{ij} is on the same side of D_{ij} as c_i , with an angle ε between A_{ij} and f_{ij} . Similarly, g_{ij} is on the same side of D_{ij} as c_i , with an angle ε between B_{ij} and g_{ij} .

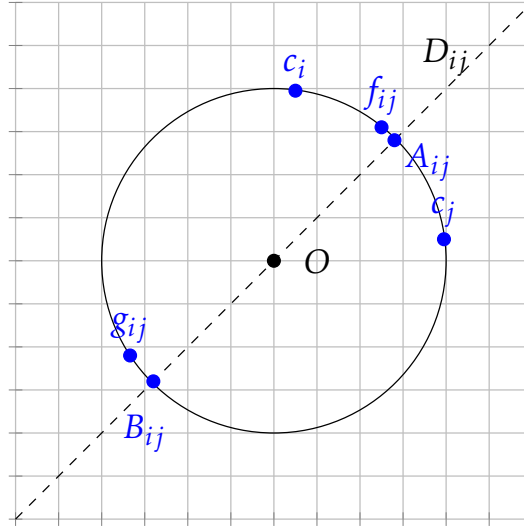


Figure 5.2.7: The construction with two vertices v_i, v_j and an arc (v_i, v_j) .

As previously, suppose that we choose the positions of candidates in such a way that all the points A_{ij} and B_{ij} are distinct.

Then we choose ε sufficiently small such that each A -point or B -point lies neither between f_{ij} and A_{ij} , nor between g_{ij} and B_{ij} .

Let us consider an arc (v_i, v_j) . Then:

- Both voters f_{ij} and g_{ij} prefer c_i to c_j (as A_{ij} is equidistant from c_i and c_j , the same for B_{ij}).
- For any other pair $\{c, c'\}$ of candidates, exactly one voter among f_{ij} and g_{ij} prefers c to c' (and one prefers c' to c). This follows from the fact that all D -lines intersect in O , and that we have fixed ε in a way that there is no other A -point (resp. B -point) between f_{ij} and A_{ij} (resp. g_{ij} and B_{ij}). Therefore, f_{ij} and g_{ij} cannot be on the same side of the D -line corresponding to $\{c, c'\}$.

Explicit construction Let us call Θ_i the angle (polar coordinate, in radian) of c_i (i.e., the angle between the horizontal axis and $\overrightarrow{Oc_i}$). Then we shall choose Θ_i in such a way that all the points A_{ij} and B_{ij} are distinct. This appears as soon as $(\Theta_i + \Theta_j)$ are distinct, as the angle of the line D_{ij} is $\frac{\Theta_i + \Theta_j}{2}$.

Let us fix $\Theta_i = \frac{2^i}{2^n} = 2^{i-n}$. By the same reasoning as in the proof for norm ℓ_1 , all $(\Theta_i + \Theta_j)$ are distinct (note that $0 \leq \Theta_i \leq \pi/2$ so $(\Theta_i + \Theta_j)$ are indeed distinct modulo 2π). We can fix $\varepsilon = 1/2^{n+1}$, to fulfill the property for f_{ij} and g_{ij} .

We note that the actual preference profile can be easily built from this embedding of points in the 2-dimensional space. Indeed, if $i > j$, voter f_{ij} has angle $\frac{2^i+2^j+1/2}{2^n}$, and she prefers c_k to c_l iff $|2^k - a_{ij}| < |2^l - a_{ij}|$, where $a_{ij} = 2^i + 2^j + 1/2$ (if $i < j$ it is the same with $a_{ij} = 2^i + 2^j - 1/2$). Voter g_{ij} has the reverse preference on all pairs but $\{c_i, c_j\}$. Thus, the construction is polynomial time. \square

Some concluding words We have shown that the result of McGarvey (1953) and Debord (1987) about inducible weighted tournaments is still true for 2-Euclidean preferences under ℓ_2 norm, and that every even weighted bipartite tournament is inducible by 2-Euclidean preferences under ℓ_1 norm and ℓ_∞ norm. These results allowed us to answer an open problem when the input preferences are 2-Euclidean (under ℓ_1 , ℓ_2 or ℓ_∞).

It would be interesting to investigate the impact of 2-Euclidean preferences on the complexity of other NP-hard social choice problems that cannot be formulated on an induced weighted tournament; for instance, Godziszewski et al. (2021) showed that computing the result of a number of multiwinner voting rules remains NP-hard with 2-dimensional Euclidean preferences, without resorting to weighted majority tournaments. Recently, Chen and Roy (2022) showed that the 2-Euclidean stable roommates problem is NP-hard for $d \geq 3$. As we can see, so far, there is no known NP-hard problem which would become polynomial time solvable for 2-Euclidean (resp. d -Euclidean preferences). This question is worth investigating.

The expressivity of d -Euclidean domain is studied in the following chapters. More precisely, we study in Chapter 6 the recognition problem of 2-Euclidean preferences with respect to norm ℓ_2 , while Chapter 7 is dedicated to the study of 2-Euclidean preferences with respect to norms ℓ_1 and ℓ_2 . As already discussed in the introduction, even though we cannot expect some appealing theoretical properties or computational guarantees for 2-Euclidean preferences, it is always interesting to identify an underlying structure of a given preference profile. For instance, such a knowledge can be useful for data analysis, voting advice applications or recommendation systems.

Chapter 6

A Heuristic algorithm for recognizing 2-Euclidean preferences with respect to the norm ℓ_2

6.1 Introduction

This chapter deals with the recognition of 2-Euclidean preferences with respect to the norm ℓ_2 . Given a specific domain restriction and a set of preferences (also called *preference profile* hereafter), a recognition algorithm aims at deciding whether the preferences belong or not to the domain restriction, and if possible also provides a concise certificate of membership or non-membership. Recognition algorithms have been proposed for various domain restrictions in social choice, among which single-peaked preferences on an axis (Bartholdi III and Trick, 1986) or on a tree (Trick, 1989) or on a circle (Peters and Lackner, 2020), single-crossing preferences (Doignon and Falmagne, 1994), intermediate preferences on median graphs (Clearwater et al., 2015), etc.

In contrast with all these results, the domain of d -Euclidean preferences, especially when $d \geq 2$, remains much less understood from the theoretical point of view. Despite the fact that Euclidean preferences were widely studied (we refer the reader to the Section 5.2.1), the recognition problem appears to be very challenging and remains widely open from an algorithmic perspective:

- From a theoretical viewpoint, Chen et al. (2017) conjectured that the domain cannot be characterised by finitely many forbidden minors, contrary to, for example, the single-peaked domain (Ballester and Haeringer, 2011). Peters (2017) showed the validity of this conjecture. More than that, he proved that for any $d \geq 2$, the set of d -Euclidean preference profiles does not admit a good characterisation by forbidden minors. He also showed that the recognition problem (under ℓ_2) is in $\exists\mathbb{R}$, and that some Euclidean preference profiles require exponentially many bits in order to specify any Euclidean embedding.

- From a practical viewpoint, dealing with exact resolution, Peters (2017) pointed out that using an ETR-solver to recognize Euclidean profiles for $d = 2$ or $d = 3$ reveals unfeasible in practice, and to the best of our knowledge no efficient algorithm is known even for small size instances. Several attempts, starting with Kruskal (1964), were made to find an efficient heuristic algorithm to give an *approximate* multidimensional Euclidean representation of a given profile (which may or may not be Euclidean). However, these algorithms encountered all the same problems : they tend to give a degenerate solution, typically all candidates placed on a circle and all voters forming a cluster around the circle center. As this kind of solution is generally not far (in terms of the sum of distance errors) from an optimal solution, it does not provide any information about the structure behind preferences, and typically does not allow to determine if preferences are d -Euclidean or not. Some recent approaches (see for example Borg and Groenen (2005) or Busing et al. (2005)) tried to overcome these difficulties.

In this chapter, we propose a new approach to recognize 2-Euclidean preferences. The underlying idea is simple: we fix the candidates' positions randomly in the plane, we determine the set of votes that are compatible with these positions, and then we check if our input profile is included in this set of votes. If so, the input is 2-Euclidean. We repeat this test a certain number of times (since the random positions of candidates might not be the good ones) to detect 2-Euclidean profiles. We complement this with a test aiming at detecting when an input profile is *not* 2-Euclidean. Although the naive implementation of this idea is not very efficient, we propose several theoretical and algorithmic improvements to make it more operational. We made some experiments both on real datasets and synthetic ones. In addition, this algorithmic tool allowed us to provide some new insights on questions about 2-Euclidean preferences, such as the number of inclusion-wise maximal profiles for a given number of candidates.

The chapter is organized as follows: we introduce some preliminaries in Section 6.2 (we recall that more detailed introduction on multidimensional Euclidean preferences can be found in Chapter 5, Section 5.2). Then we present our algorithm in Section 6.3. Experimental results are provided and discussed in Section 6.4.

6.2 Preliminaries

We will use in this chapter the notions introduced in Section 5.2.2. We recall briefly that given a preference profile \mathcal{P} over the set of candidates C , we need to find a mapping $f : C \rightarrow \mathbb{R}^2$ such that for each preference $\succ_v \in \mathcal{P}$, the area D_v^f corresponding to this preference is non-empty. If such a mapping exists, the profile \mathcal{P} is 2-Euclidean.

Actually, each mapping f defines a preference profile associated to it, and in practice, we need just to ensure that the input profile \mathcal{P} is a subprofile, up to a renaming of candidates. Let us now formalize this idea - to begin, we give a definition of profile \mathcal{P}_f

associated to the mapping f :

Definition 6.2.1

Let C be a set of candidates, and $f : C \rightarrow \mathbb{R}^2$ be a mapping. We define the *profile* \mathcal{P}_f associated to f as the set of all rankings v such that $D_f(v)$ is non-empty.

We have then the following property, used in our algorithm.

Property 7. A profile \mathcal{P} is 2-Euclidean if and only if there exists a mapping $f : C \rightarrow \mathbb{R}^2$ such that $\mathcal{P} \subseteq \mathcal{P}_f$.

Furthermore, we will use the notation $[\mathcal{P}]$ to refer to any profile obtained by a permutation of the names of the candidates in \mathcal{P} . By extension, we will write $\mathcal{P} \subseteq [\mathcal{P}_f]$ (resp. $\mathcal{P} = [\mathcal{P}_f]$) to state that \mathcal{P} is included in (resp. is equal to) a profile obtained by renaming the candidates of \mathcal{P}_f . For instance,

$$\left\{ \begin{array}{l} c_2 > c_3 > c_1, \\ c_1 > c_3 > c_2 \end{array} \right\} = \left[\left\{ \begin{array}{l} c_1 > c_2 > c_3, \\ c_3 > c_2 > c_1 \end{array} \right\} \right]$$

because the profile on the left hand side is obtained from the profile on the right hand side by renaming c_1 (resp. c_2, c_3) in c_2 (resp. c_3, c_1). Property 7 remains clearly true if we replace $\mathcal{P} \subseteq \mathcal{P}_f$ by $\mathcal{P} \subseteq [\mathcal{P}_f]$.

Finally, let us define the notion of (inclusion-wise) *maximal* 2-Euclidean profile (we recall that the preferences are pairwise distinct):

Definition 6.2.2: Maximal profile

A 2-Euclidean profile $\mathcal{P} = \{>_1, >_2, \dots, >_n\}$ is *maximal* if for any vote $>_v \notin \mathcal{P}$, the profile $\mathcal{P}' = \mathcal{P} \cup \{>_v\}$ is not 2-Euclidean.

Let us call a representation function f *degenerate* if either three points $f(c_i)$, $f(c_j)$ and $f(c_k)$ are aligned, or at least 4 bisectors intersect in the same point. By slightly moving the positions of candidates, it is easy to see that for any 2-Euclidean profile there exists a non-degenerate representation of it. Note that if the representation function f is non-degenerate, then the profile \mathcal{P}_f is maximal.¹

6.3 2-Euclidean recognition heuristic

Assume that we have a profile \mathcal{P} in input, and we want to decide whether it is 2-Euclidean or not. We propose here an algorithm based on the following principles.

¹Indeed, it can be shown that then $|\mathcal{P}_f| = \text{ub}(m)$, where $\text{ub}(m) = m(3m-10)(m-1)(m+1)/24 + m(m-1) + 1$ is the maximal number of pairwise distinct votes in a 2-Euclidean preference profile on m candidates (this can be easily derived (see for instance Escoffier et al. (2022a)) from a result by Bennett and Hays (1960)).

First, as some necessary conditions for a profile to be 2-Euclidean have been identified in the literature, the algorithm checks if these conditions are fulfilled. If it is not the case, the profile is not 2-Euclidean and the algorithm returns NO. Otherwise, the algorithm tries to “guess” the set of positions $f(c_1), f(c_2), \dots, f(c_m)$ of candidates. It then builds the (maximal) 2-Euclidean profile \mathcal{P}_f associated to the representation f , and finally checks if $\mathcal{P} \subseteq [\mathcal{P}_f]$. If this is the case, the profile is 2-Euclidean and the algorithm returns YES. Otherwise, we reiterate this process of guessing positions of candidates. Finally, if none of the tests succeeded, then the status is undefined and the algorithm returns UNKNOWN. The pseudocode of this heuristic is given in Algorithm 2.

Algorithm 2 `is_euclid(\mathcal{P})`

Input : a preference profile \mathcal{P}
Output: NO if \mathcal{P} is not 2-Euclidean, YES if \mathcal{P} is 2-Euclidean, UNKNOWN if not decided
if there is a NO-certificate for \mathcal{P} **then**
 return NO
end if
while timeout not reached **do**
 Generate a random 2-Euclidean representation f of candidates
 Build the profile \mathcal{P}_f associated to f
 if $\mathcal{P} \subseteq [\mathcal{P}_f]$ **then**
 return YES
 end if
end while
return UNKNOWN

We now detail how each step of this heuristic is performed, explaining the main ideas used to make it as efficient as possible.

6.3.1 NO-Certificates

As explained above, the first step of the algorithm is to detect, as much as possible, profiles that are *not* 2-Euclidean. To do so, we use in our algorithm two necessary conditions (also called NO-certificates in which follows, as they guarantee that the profile is not 2-Euclidean) known in the literature, based on the size of the profile and on some forbidden substructures. We will discuss possible improvements of this step of the algorithm in the global conclusion of the thesis.

- **Cardinality condition:** Using the maximum number $ub(m)$ of pairwise distinct votes in a 2-Euclidean profile on m candidates (see footnote 1 p. 171), the algorithm simply outputs NO if $n > ub(m)$. Note that this is actually tested for any restriction of the profile to a subset of candidates. Given a profile \mathcal{P} over a set C of m candidates (which is supposed to contain only pairwise distinct prefer-

ences), and given $S \subseteq C$, we denote by $\mathcal{P}|_S$ the restriction of \mathcal{P} to S .² Any subset S of candidates such that $|\mathcal{P}|_S| > \text{ub}(|S|)$ is a NO-certificate for \mathcal{P} .

- **Condition on profile restrictions on 4 candidates:** Kamiya et al. (2011) showed that for 4 candidates, there are only 3 maximal 2-Euclidean profiles $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ (up to a permutation of candidates). Consequently, a profile \mathcal{P} on 4 candidates is 2-Euclidean if and only if there exists $i \in \{1, 2, 3\}$ for which $\mathcal{P} \subseteq [\mathcal{P}_i]$. We use this characterization to derive a NO-certificate, as follows. We say that \mathcal{P}' is a k -restriction of \mathcal{P} if there exists $S \subseteq C$ of cardinal k such that $\mathcal{P}' = \mathcal{P}|_S$. We will note by \mathcal{P}_k the set of all k -restrictions of a given profile \mathcal{P} : $\mathcal{P}_k = \{\mathcal{P}|_S : S \subseteq C, |S| = k\}$.

Obviously, if \mathcal{P} is 2-Euclidean, then for each k , any k -restriction of \mathcal{P} is also 2-Euclidean. We use the characterization of 2-Euclidean profiles on 4 candidates to identify non-Euclidean profiles \mathcal{P} on m candidates: we generate all $\binom{m}{4}$ elements of \mathcal{P}_4 and check if they are subprofiles of $[\mathcal{P}_1]$, $[\mathcal{P}_2]$ or $[\mathcal{P}_3]$, one of the 3 maximal 2-Euclidean profiles on 4 candidates. Any 4-restriction of \mathcal{P} that is not 2-Euclidean is a NO-certificate for \mathcal{P} .

6.3.2 The random generation of the representation f

The most straightforward idea consists in picking up the positions $f(c_1), f(c_2), \dots, f(c_m)$ according to a uniform distribution in the square $[0, M] \times [0, M]$ (for some constant M). However, this turns out to be inefficient in practice because it does not take into account the input profile P , which yields a low chance that the positions are correctly guessed. This led us to adapt the random generation process by observing that, if a candidate c_i is ranked last by at least one voter v , then for any 2-Euclidean representation f , point $f(c_i)$ must be a vertex of the convex hull of the set $\{f(c_1), f(c_2), \dots, f(c_m)\}$. This is shown in Lemma 6.3.1. We therefore generate positions of candidates by imposing that the number of vertices of the convex hull of the set of positions of candidates is at least the number of candidates ranked last by at least one voter in P .

Lemma 6.3.1

Let $C = \{c_1, \dots, c_m\}$ be a set of candidates and $f : C \rightarrow \mathbb{R}^2$ an injective mapping of the candidates in the plane. Let $c \in C$ and $\{i_0, i_1, \dots, i_k\} \subseteq \{1, \dots, m\}$ such that $f(c_{i_0})$ is a convex combination of $f(c_{i_1}), f(c_{i_2}), \dots, f(c_{i_k})$. For each possible position $f(v)$ of a voter v in the plane (i.e., inducing a strict order $>_v$), there exists $i_v \in \{i_1, i_2, \dots, i_k\}$ such that $c_{i_0} >_v c_{i_v}$. In particular, c_{i_0} is never ranked last in the profile \mathcal{P}_f associated to f .

Proof. By definition, $f(c_{i_0}) = \lambda_1 f(c_{i_1}) + \dots + \lambda_k f(c_{i_k})$ with $\lambda_1 + \dots + \lambda_k = 1$ and $\lambda_i \geq 0$ for each $i \in \{1, \dots, k\}$. Let $f(v)$ be any possible position of a voter v in the plane. By the

²We recall that $\mathcal{P}|_S$ is a copy of \mathcal{P} in which we have kept only the candidates of S - see Definition 1.1.2.

Cauchy-Schwarz inequality, we have:

$$\begin{aligned}
 \|f(v) - f(c_{i_0})\| &= \|f(v) - (\lambda_1 f(c_{i_1}) + \dots + \lambda_k f(c_{i_k}))\| \\
 &\leq \|\lambda_1(f(v) - f(c_{i_1})) + \dots + \lambda_k(f(v) - f(c_{i_k}))\| \\
 &\leq \lambda_1 \|f(v) - f(c_{i_1})\| + \dots + \lambda_k \|f(v) - f(c_{i_k})\| \\
 &\leq (\lambda_1 + \dots + \lambda_k) \max_{i \in \{i_1, \dots, i_k\}} \|f(v) - f(c_i)\| \\
 &= \max_{i \in \{i_1, \dots, i_k\}} \|f(v) - f(c_i)\|
 \end{aligned}$$

By setting $i_v = \arg \max_{i \in \{i_1, \dots, i_k\}} \|f(v) - f(c_i)\|$, we have thus $\|f(v) - f(c_{i_0})\| \leq \|f(v) - f(c_{i_v})\|$. As $>_v$ is a strict order, this inequality is actually strict, and, by Definition 5.2.1, $c_{i_0} >_v c_{i_v}$. In particular, c_{i_0} is not ranked last by v . \square

Hence, to generate a random 2-Euclidean representation, we take into account the number of vertices that the convex hull of the set of points $\{f(c_1), f(c_2), \dots, f(c_m)\}$ should have. We call this number the *size* of the convex hull in the following. More precisely, we proceed as follows:

1. We go through the input profile \mathcal{P} and we determine the set $C_L \subseteq C$ of the candidates ranked last at least once in \mathcal{P} .
2. We pick randomly, according to a given probability distribution π , an integer $k \in \{L, \dots, m\}$, where $L = |C_L|$. In fact, as there are L candidates ranked last at least once in \mathcal{P} , the convex hull of any representation of \mathcal{P} must contain at least L vertices (and possibly more if \mathcal{P} is not maximal).
3. Using Valtr's algorithm (see Valtr (1995)), we generate uniformly at random a polygon with k vertices in a square $[0, M] \times [0, M]$ (for some constant M). We assume that the set of positions of vertices corresponds to $\{f(c_1), f(c_2), \dots, f(c_k)\}$. We choose then (uniformly at random) $m - k$ points inside the polygon in order to fix the remaining positions $f(c_{k+1}), \dots, f(c_m)$. We recall that, according to Lemma 6.3.1, only the candidates c_1, \dots, c_k can be ranked last in the associated profile \mathcal{P}_f .

A Curious Tortoise Intervention



Still by using Lemma 6.3.1, the above procedure could be generalized: for the moment, we only care about the last ranked candidates, but we continue to pick the remaining points at random (inside the polygon). However, if a candidate is ranked second to last by a voter $>_v$, it cannot lie inside the polygon determined by the set of candidates ranked better than her by $>_v$. And of course, this generalizes for any position.

It might be interesting to see if this observation (not implemented in our code) could improve the algorithm performance in practice.

Let us now specify the probability distribution π used in step 2 above. We propose here three different distributions (the efficiency of which will be compared in the experimental study):

- **Uniform distribution:** We pick the size of the convex hull between L and m using the uniform distribution π_U . Formally, for each $k \in \{L, \dots, m\}$, we have $\pi_U(k) = 1/(m - |C_L| + 1)$.
- **Vertex-based distribution:** The idea is to fit the probability distribution of the size of the convex hull that we would get by picking m points uniformly at random points in the plane (this distribution of the sizes has no reason to be uniformly distributed). Denoting by $\pi_m(k)$ the probability of having a convex hull of size exactly k for m randomly drawn points, we define the conditional probability distribution π_m^L as follows, as we want to consider only instances with a convex

$$\text{hull of size at least } L: \pi_m^L(k) = \begin{cases} 0 & \text{if } k < L, \\ \frac{\pi_m(k)}{\sum_{i=L}^m \pi_m(i)} & \text{if } k \geq L. \end{cases}$$

To evaluate the values $\pi_m(k)$, we used a Monte Carlo simulation: we generated N random instances (with $N = 10^6$). Denoting by n_k the number of instances for which the size of the convex hull was equal to k , we set $\pi_m(k) = \frac{n_k}{N}$.

- **Profile-based distribution:** The idea is to fit the probability distribution of the number of candidates ranked last in the set of maximal 2-Euclidean profiles on m candidates, distribution that we estimate here also using a Monte Carlo simulation. More formally, let $E_m = \{\mathcal{P}_1, \dots, \mathcal{P}_q\}$ (with $\mathcal{P}_j \not\subseteq [\mathcal{P}_i]$ for $j \neq i$) denote the set of maximal 2-Euclidean profiles on m candidates, i.e., a profile \mathcal{P} on m candidates is 2-Euclidean if and only if there exists $i \in \{1, \dots, q\}$ for which $\mathcal{P} \subseteq [\mathcal{P}_i]$, and let n_k denote now the number of profiles of E_m with exactly k candidates ranked last. Ideally, we would like to randomly draw the size of the convex hull by considering the probability $n_k/|E_m|$ of having exactly k candidates ranked last in a maximal 2-Euclidean profile. This probability has no reason to coincide with $\pi_m(k)$ because some profiles $[\mathcal{P}_f]$ may occur more frequently than others when a 2-Euclidean representation f of candidates is randomly drawn. As there is no known characterization of maximal 2-Euclidean profiles for $m \geq 5$ (and no known way to generate all of them), the cardinality of E_m is unknown for $m \geq 5$. Thus, we approximate each probability $n_k/|E_m|$ by using again a Monte Carlo simulation method. More precisely, we start by generating a set $\tilde{E}_m \subseteq E_m$ (as large as possible) of maximal 2-Euclidean profiles (see the next subsection for a description of how to generate \mathcal{P}_f from a randomly drawn f). Then we compute \tilde{n}_k , the number of profiles of \tilde{E}_m with exactly k candidates ranked last, and the probability is approximated by $\tilde{\pi}_m(k) = \frac{\tilde{n}_k}{|\tilde{E}_m|}$. Finally, as we want to consider only instances with a convex hull of size at least L , we use the following distribution:

$$\tilde{\pi}_m^L(k) = \begin{cases} 0 & \text{if } k < L, \\ \frac{\tilde{\pi}_m(k)}{\sum_{i=L}^m \tilde{\pi}_m(i)} & \text{if } k \geq L. \end{cases}$$

6.3.3 The profile \mathcal{P}_f associated to a mapping f

The previous method generates a random representation function $f: C \rightarrow \mathbb{R}^2$. We now describe how, given the points $f(c_1), f(c_2), \dots, f(c_m)$, we determine the (unique) maximal 2-Euclidean profile \mathcal{P}_f associated to f . As any preference area D_v^f borders at least one intersection point of bisectors, we examine the different intersection points to determine the set of all preference areas induced by f . If we assume w.l.o.g. that f is non-degenerate³, it amounts to consider all triples (item (1) below) and pairs (item (2) below) of candidates, because four or more bisectors do not intersect in a non-degenerate representation.

1. **“Triple-intersection preferences”**: For each triple of candidates c_i, c_j, c_k , we compute the point I_{ijk} which is the circumcenter of the triangle $\{f(c_i), f(c_j), f(c_k)\}$. As f is non-degenerate (i.e., the points $f(c_i), f(c_j)$ and $f(c_k)$ are not aligned), this point exists and is unique. We define a fictitious voter v_{ijk} such that $f(v_{ijk}) = I_{ijk}$. By circumcenter definition, v_{ijk} is indifferent between candidates c_i, c_j and c_k because $\|f(v_{ijk}) - f(c_i)\| = \|f(v_{ijk}) - f(c_j)\| = \|f(v_{ijk}) - f(c_k)\|$. Let us denote by d (resp. d_c) this common distance (resp. $\|f(v_{ijk}) - f(c)\|$). The preference of v_{ijk} is of the form

$$R_1 >_{v_{ijk}} \{c_i, c_j, c_k\} >_{v_{ijk}} R_2$$

where R_1 (resp. R_2) is the ranking on candidates $c \in C$ such that $d_c < d$ (resp. $d_c > d$), induced by the distances d_c .

As said above, I_{ijk} is the only point at equal distance from c_i, c_j and c_k . It is easy to see that there exists $\varepsilon > 0$ such that moving v_{ijk} within the open ball $\text{Ball}(I_{ijk}, \varepsilon)$ will only impact the order of c_i, c_j and c_k in $>_{v_{ijk}}$, without changing the order of the other candidates (see Figure 6.3.1). More formally, for any strict order R_{ijk} on the set $\{i, j, k\}$, there is a point $p \in \text{Ball}(I_{ijk}, \varepsilon)$ such that the preference order $>_{v_p}$ of a voter v_p positioned in p is of the form $R_1 >_{v_p} R_{ijk} >_{v_p} R_2$. Each point I_{ijk} thus allows us to determine 6 preferences of the profile \mathcal{P}_f , one per each of the 6 possible strict orders R_{ijk} .

2. **“Double-intersection preferences”**: We compute the point $I_{ij,kl} = B(c_i, c_j) \cap B(c_k, c_l)$ for each quadruple of candidates c_i, c_j, c_k, c_l (see Figure 6.3.2 for an illustration). We define a fictitious voter $v_{ij,kl}$ such that $f(v_{ij,kl}) = I_{ij,kl}$. Let us define $d_{ij} = \|f(v_{ij,kl}) - f(c_i)\| = \|f(v_{ij,kl}) - f(c_j)\|$, and $d_{kl} = \|f(v_{ij,kl}) - f(c_k)\| = \|f(v_{ij,kl}) - f(c_l)\|$. Assuming w.l.o.g. that $d_{ij} < d_{kl}$, the preference of $v_{ij,kl}$ is of the form

$$R_1 >_{v_{ij,kl}} \{c_i, c_j\} >_{v_{ij,kl}} R_2 >_{v_{ij,kl}} \{c_k, c_l\} >_{v_{ij,kl}} R_3$$

where R_1 (resp. R_2, R_3) is the ranking on candidates c such that $d_c < d_{ij}$ (resp. $d_{ij} < d_c < d_{kl}$, $d_c > d_{kl}$). Using the same arguments as in the case of 3-intersections,

³In practice, we can get a degenerate representation while using random generation. However, it is easy to detect, so we can simply reject it. As this phenomenon almost never occurs, it has no impact on the performance of the algorithm in practice.

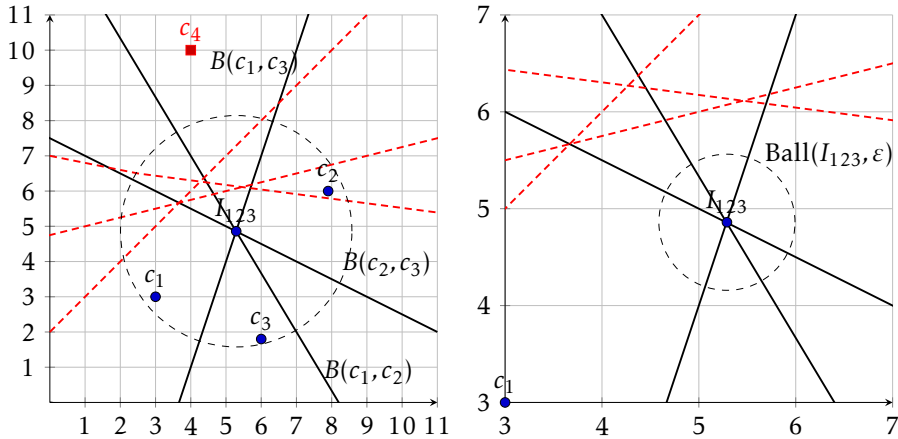


Figure 6.3.1: Left part: there are 6 preference areas around the intersection point I_{ijk} . Right part: there exists a neighbourhood of I_{ijk} not crossed by any other perpendicular bisector.

this allows us to determine 4 preferences of the profile \mathcal{P}_f , namely strict orders of the form $R_1 > R_{ij} > R_2 > \pi_{kl} > R_3$, where R_{ij} (resp. R_{kl}) is $c_i > c_j$ (resp. $c_k > c_l$) or the opposite.

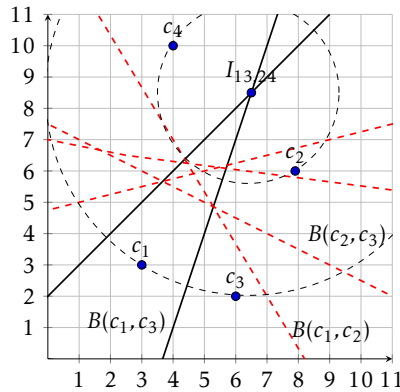


Figure 6.3.2: 2-intersection $I_{13,24}$ with 4 preference areas around it.

As each preference area D_v can be adjacent to more than one intersection point, the two steps (1) and (2) above will give us some duplicates. The procedure is therefore completed by deleting duplicate preferences in the resulting profile.

6.3.4 Testing if the profile \mathcal{P} is a subprofile of $[\mathcal{P}_f]$

Finally, we need to check if $\mathcal{P} \subseteq [\mathcal{P}_f]$. This operation reveals to be very time-consuming (thus making the heuristic non-operational) if it is not optimized. The following procedure allows us to greatly alleviate the computational burden. Let us denote by $>_{v_0}$

a preference arbitrarily chosen in \mathcal{P} . Assume that \mathcal{P} is a subprofile of $[\mathcal{P}_f]$. Then $>_{v_0}$ necessarily corresponds to some preference $>_v$ of \mathcal{P}_f , up to a permutation σ of the candidates in \mathcal{P}_f . If $c_{i_1} >_{v_0} c_{i_2} >_{v_0} \dots >_{v_0} c_{i_m}$ and $c_{j_1} >_v c_{j_2} >_v \dots >_v c_{j_m}$, then the permutation is defined by $\sigma(c_{j_k}) = c_{i_k}$ for $k \in \{1, \dots, m\}$. By applying the permutation σ to all the preferences of \mathcal{P}_f , we have then $\mathcal{P} \subseteq \mathcal{P}_f^\sigma$, where \mathcal{P}_f^σ denotes the profile obtained from \mathcal{P}_f by permuting the candidates according to σ . Denoting by σ_v the permutation obtained for a preference v in \mathcal{P}_f (for the same choice of v_0 in \mathcal{P}), testing whether $\mathcal{P} \subseteq [\mathcal{P}_f]$ thus amounts to testing if there exists $v \in \mathcal{P}_f$ such that $\mathcal{P} \subseteq \mathcal{P}_f^{\sigma_v}$.

In practice, we actually perform symmetrically, i.e., we test if there exists $v \in \mathcal{P}$ such that $\mathcal{P}^{\sigma_v} \subseteq \mathcal{P}_f$ for an arbitrarily chosen preference $v_0 \in \mathcal{P}_f$ (testing whether $[\mathcal{P}] \subseteq \mathcal{P}_f$ is equivalent to testing whether $\mathcal{P} \subseteq [\mathcal{P}_f]$). This indeed allows us to take advantage of precomputing all profiles of $[\mathcal{P}]$, which avoids repeated computation of profile permutations. We store these profiles in a lookup table where they can be found quickly. Instead of performing $O(m^4)$ profile permutations for *each* profile \mathcal{P}_f (because there are $O(m^4)$ votes v in \mathcal{P}_f , by the result of Bennett and Hays (1960)), only $m!$ permutations of \mathcal{P} are computed *once* in the precomputation phase. For the small values of m we are working with (profiles involving up to 9 candidates), it represents a significant saving in computation time. Indeed, by storing the permuted profiles of \mathcal{P} in a lookup table, we can consider about a *thousand times more* profiles \mathcal{P}_f per a fixed period of time than without this table.

6.4 Experimental study

The heuristic algorithm has been implemented in C++. To analyse how it performs in practice, several numerical tests were carried out on an Intel Xeon X5677 (3.46 GHz base, 3.73 GHz turbo). Besides the computation time, we paid a special attention to the *recognition rate* of the algorithm, defined as the fraction of instances for which it was able to conclude (yes or no) within a given timeout, fixed at one hour in the results we present. The section is organized as follows. We first present our results on real-world data from the PrefLib library (Mattei and Walsh (2013)) in Section 6.4.1. In Section 6.4.2, we give the results on synthetic data: we start by explaining how the data were generated, and then we study the recognition rate and execution time in function of the number of voters and candidates. While the recognition rate turned out to be high in most cases both on the real-world and synthetic datasets, we also noted some weak points for specific values of m and n , where many instances remained undecided. To detect whether this was mainly due to the no-test or the yes-test, we made also some experiments on 2-Euclidean profiles, to study the recognition rate on these instances. Moreover, we make a brief comparison with the Gurobi Quadratic Constraint Optimizer, which is outperformed by our heuristic in both running time and recognition rate. Finally, in Section 6.4.3, we discuss several observations made from our experiments about the number of maximal profiles over 5 and 6 candidates.

6.4.1 Experimental study on real-world data

As mentioned earlier, PrefLib is a reference library that contains several types of preference data. For our experiments, we focused only on the complete strict order datasets (SOC data files). There are 7741 such files.

About 4% of the instances were detected as 2-Euclidean and 91.5% as non-Euclidean, while 4.5% remained undecided after the timeout of one hour. More detailed information about the recognition rates on the different datasets as well as some characteristics of instances are summarized in Table 6.4.1.

As a first comment, these tests show that a large majority of profiles in Preflib are not 2-Euclidean, even though it contains many datasets of small/medium size. Moreover, *all* the profiles that were recognized as 2-Euclidean contain at most 3 candidates, or at most 2 voters, or at most 3 voters and 7 candidates. These profiles are “trivially” 2-Euclidean, in the sense that Bultheau and Chen (2022) proved that any such profile is 2-Euclidean (and in our tests we actually directly answered yes on such profiles).

Second, the results show that our algorithm performs well on these datasets, as it was able to conclude for 95.5% of the instances⁴. Looking towards possible improvements, we focused on the undecided instances. Two types emerged:

1. First, the algorithm performs badly on instances with 3 voters, when the number of candidates is large. This might be surprising at first sight but there is actually a clear explanation: as any profile with 4 candidates and 3 voters is 2-Euclidean, our NO-certificate based on the subprofiles on 4 candidates cannot conclude negatively when there are 3 voters. Hence the NO-test is not efficient, and in datasets 00049 and 00056 many instances with 3 voters and a large number of candidates (like 100) remained undecided. We note that this problem basically disappears for more voters, even for instances with 4 voters (most of the instances with 4 voters were detected as non-Euclidean).
2. Then, we note that our algorithm seems to struggle on profiles with about 10 voters and 10-20 candidates (see the results for datasets 00006 and 00042). Strengthening the NO-certificate by adding more forbidden subprofiles would probably help to solve these instances.

As a final remark on these instances, we saw that all the instances that violate the cardinality condition (first NO-certificate) also contain a forbidden subprofile on 4 candidates (second NO-certificate). Thus, the most important NO-certificate is the second one, the cardinality condition being interesting only to conclude negatively in an instant for profiles with a large number of voters.

⁴and several undecided instances were actually solved within slightly more than the timeout of one hour.

d	m	n	ALL	YES	NO	?	RR
04	3 (4)	6 (24)	200	100	100	0	1
06	14-24	7-9	20	0	4	16	0.2
09	9 (7)	123 (70)	2	0	2	0	1
11	103-242	5	3	0	3	0	1
12	11	30	1	0	1	0	1
14	10	4926	1	0	1	0	1
15	10 - 242	3-5	79	0	38	41	0.48
24	4	24	4	0	4	0	1
25	4	24	4	0	4	0	1
32	6	15	1	0	0	1	0
35	15	41-42	6	0	6	0	1
41	885	130	1	0	1	0	1
42	1-16	1-19	96	6	1	89	0.07
43	5-177	6-28	123	0	121	2	0.98
44	266-1080	8-12	36	0	34	2	0.94
45	55-71	36-45	29	0	29	0	1
46	38-208	18-19	4	0	4	0	1
47	13-28	33-54	362	0	362	0	1
48	7-160	1-31	642	1	637	4	0.99
49	1-102	1-146	610	14	546	50	0.92
50	216	12	1	0	1	0	1
51	22-117	12-16	12	0	12	0	1
52	5-32	5-21	67	0	59	8	0.88
53	6-24	14-76	454	0	454	0	1
54	15-353	5-131	951	0	947	4	0.99
55	47-353	50-992	53	0	53	0	1
56	6-353	1-21	3979	191	3662	126	0.97
all			7741	312	7086	343	0.955

Table 6.4.1: The recognition rates (RR) on different Preflib datasets. Column m (resp. n) gives the range of the numbers of candidates (resp. voters) in the dataset 000d. Column “ALL” gives the number of instances in the dataset, whereas column “YES” (resp. “NO”, “?”) gives the number of times the algorithm returns YES (resp. NO, UNKNOWN). Only two types of combinations m, n appear in datasets 00004 and 00009, hence the separate notation convention.

6.4.2 Experimental study on synthetic data

To generate a random profile of n preferences over m candidates, we draw n rankings uniformly at random from the $m!$ possible rankings (impartial culture assumption). In all the tests, the timeout was again set at one hour (note that it was also the time bound set in Peters (2017) for the experiments with the ETR-solver `nlsat` by Jovanović and de Moura (2012)). For each couple of values (n, m) , the results are averaged over 10000 instances (preference profiles).

Probability distribution used to position the candidates

As discussed in the Section 6.3.2, in the heuristic, several probability distributions have been considered to pick up randomly m two-dimensional points representing the positions of candidates, namely the *uniform distribution* (on the size of the convex hull, imposed to be greater than or equal to the number L of candidates ranked last by at least one voter), the *vertices-based* distribution and the *profile-based* distribution.

The profile-based distribution outperformed the two others for 5 candidates. For a larger number of candidates, as it was computationally hard to estimate this distribution with a sufficient precision, we compared the performances of the heuristic with the uniform distribution on the one side and the vertices-based distribution of the other side. It turns out that, in general, the time needed to recognize a profile (i.e., detect whether it is 2-Euclidean or not) is shorter using the uniform distribution. Nevertheless, this has hardly any impact on the recognition rate for profiles with up to 6 candidates because the recognition time remains below the timeout set to one hour. The difference in the recognition rates obtained for the two distributions grows for profiles with more candidates, as the recognition time then approaches the timeout.

Phase transition

We now give the recognition rates (proportion of positive, negative and undecided instances) for profiles with 5, 6, 7 and 8 candidates, depending on the number of voters. The results are given in Figure 6.4.1. We observe a phase transition: below some threshold (that depends on m) almost all the inputs are 2-Euclidean, whereas above some other threshold almost all the inputs are not, and there is a phase transition in between. We note that the phase transition is done very quickly: it occurs between 7 and 9 voters for profiles with 5 candidates, between 5 and 7 voters with 6 candidates, between 5 and 6-7 voters with 7 candidates and between 4 and 6 voters for profiles with 8 candidates.

Concerning the recognition rates, quite unsurprisingly, we observe that the largest proportions of undecided instances occur in the phase transition. Figure 6.4.1 gives the curves of the recognition rate in function of the number of voters, for profiles involving 5 to 8 candidates. Clearly, the proportion of undecided instances in the phase transition increases with the number of candidates (see the peaks of the curves of the proportions of undecided instances in the four plots).

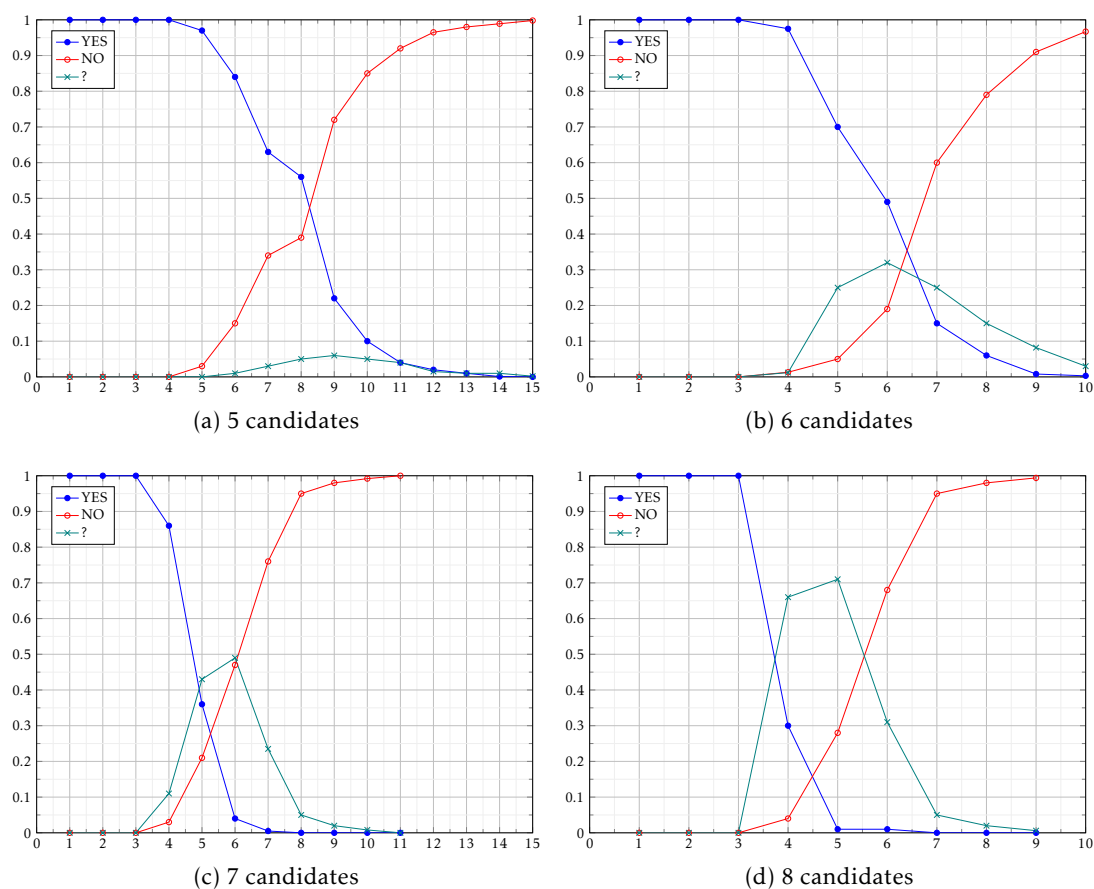


Figure 6.4.1: Recognition rates for random profiles on 5, 6, 7 and 8 candidates, w.r.t. the number of voters.

Running times

First note that testing the NO-certificate is performed very quickly (typically less than 1 second for instances with up to 8 candidates and 10 voters). Thus, for the couples (m, n) that are beyond the phase transition in Figure 6.4.1, as almost all the instances are recognized as non-Euclidean, the median running time of the algorithm is very small (typically less than 1s). We recall that the median of a set of ordered values is the value such that 50% of the values in the set are greater, and 50% are smaller.

In any case, the NO-certificate part of the algorithm does not significantly impact the global running time. Thus, we found more interesting to focus on the (median) running time of the YES-part of the heuristic. To do this, we measure the running time of the algorithm on the yes-instances (i.e., recognized as 2-Euclidean). The results are given in Table 6.4.2. We note that the running time is very far from the timeout of

3600 seconds. The dash mark means that we could not get meaningful information: indeed, they correspond to cases where the proportion of 2-Euclidean profiles is very small (after the phase transition), and we could not produce enough 2-Euclidean instances to get a relevant value of the median (note that, as said before, in these cases the NO-certificates allow anyway to reach a very small median running time for the heuristic).

$m \backslash n$	4	5	6	7	8	9	10
5	0	0	0.003	0.015	0.082	0.68	0.81
6	0	0.2	0.3	24	85	392	443
7	3.543	8.6	19.9	193.3	-	-	-
8	6.2	31.275	-	-	-	-	-

Table 6.4.2: The median running time [s] for random instances where the answer is YES.

Performance of the heuristic on 2-Euclidean input profiles

The problem of the random generation of input profiles with the impartial culture assumption is that we do not know if the generated profile is 2-Euclidean or not. Hence, we cannot distinguish the non-Euclidean profiles not detected by the NO-certificate from the 2-Euclidean profiles not recognized by the heuristic. In consequence, we do not know if we should focus more on improving the YES-part of the heuristic (finding a representation) or on finding out a more efficient NO-certificate to reduce the proportion of undecided instances. For this reason, we did another set of tests, with only 2-Euclidean profiles as input, to have some ideas on the performance of the YES-part of the heuristic. These 2-Euclidean inputs are generated as follows: We pick m random points in the plane that represent the candidate positions and we calculate the associated profile \mathcal{P}_f . We then randomly draw n (distinct) preferences of \mathcal{P}_f , to obtain a 2-Euclidean profile of n preferences over m candidates.

Regarding the recognition rate, any 2-Euclidean profile of at most 25 voters over 5 candidates has been recognized within the fixed timeout (one hour). For 6 candidates, all the instances with at most 8 voters were also recognized, and then the recognition rate decreased very slowly with the number of voters for instances involving between 9 and 25 voters, while remaining greater than 97%. The recognition rates for profiles with 7, 8 and 9 candidates are given in Table 6.4.3 (upper table). For 9 candidates, the recognition rate decreases from 7 voters, due to the fact that we often reach the timeout. Globally, we note that the recognition rates are pretty good (including the “grey areas” in the phase transitions), which makes us believe that the heuristic performs quite well for recognizing 2-Euclidean profiles with up to 9 candidates, and that one should focus more on improving the NO-certificates to reduce the proportion of undecided profiles.

The medians of running times are summarized in Table 6.4.3 (lower table). The table stops at 10 voters, but we could go further. Indeed, for 5 candidates, whatever

the number of voters, the median time did not exceed 10 seconds. For 6 candidates, the profiles with up to 25 voters could be recognized in the median time of 60 seconds, and for 7 candidates, the profiles with up to 16 candidates were recognized in the median time of 300 seconds. Comparing the recognition times of Table 6.4.2 to those of Table 6.4.3 (lower table), we note that the second ones are better. This is not surprising as the profiles are not generated in the same way. The way they are generated here, with random points in the plane, is closer to the way the heuristic recognizes them.

$m \backslash n$	4	5	6	7	8	9	10
7	1	1	1	1	0.99	0.99	0.97
8	1	1	0.98	0.95	0.89	0.81	
9	1	0.96	0.81	0.63	0.21		

$m \backslash n$	4	5	6	7	8	9	10
5	0	0	0	0	0.0006	0.0006	0.001
6	0	0.004	0.01	0.03	0.05	0.09	0.13
7	0.04	0.26	0.94	2.76	7	16.75	29.6
8	3.5	7.5	38.1	108	415	1417	
9	43.3	254	592	691	1824		

Table 6.4.3: The recognition rate (upper table) and median running time [s] (lower table) for 2-Euclidean input profiles.

Comparison with the Gurobi Quadratic Constraint Optimizer

We have implemented a Gurobi QCP model to recognize 2-Euclidean profile⁵. The model is based on the following idea: given a preference profile \mathcal{P} , we add a quadratic constraint for each inequality implied by Definition 5.2.1. Let $A \leq B$ be a symbolic description of such a constraint γ . We add a positive variable d_γ on the right side of the constraint γ , and we minimize the sum of variables d_γ over the set of all constraints. If the profile is 2-Euclidean, all the constraints can be satisfied and the optimum will then be equal to 0. In the opposite case, there is at least one constraint γ_1 that cannot be satisfied - necessarily, $d_{\gamma_1} > 0$, so the optimum value will be strictly positive.

At first, we have performed a set of experiments with random profiles as input. However, *no* profile was detected as non-Euclidean within the timeout of one hour by the solver (the solver was only able to identify some 2-Euclidean instances). So the solver *is not able to identify non-Euclidean instances*, in contrast with our heuristic.

Then, we evaluated the efficiency of the solver on 2-Euclidean instances. The results

⁵To the best of our knowledge, there is no algorithm available in the literature that would guarantee to return an *exact* solution if it exists.

are summarised in Table 6.4.4. We give here only the results on profiles involving 5 to 7 voters and 5 or 6 candidates: in fact, the solver recognized almost no profile within the timeout of 1 hour with more voters or candidates. We observe that our heuristic significantly outperforms the QCP-model. Although we do not pretend that this is the most efficient model, it is a positive and encouraging news for future development and improvement of our heuristic.

	Heuristic		Gurobi QCPsolver	
	5 candidates			
voters	RR	median [s]	RR	median [s]
5	1	0	0.53	14
6	1	0	0.49	66
7	1	0	0.34	118
	6 candidates			
voters	RR	median [s]	RR	median [s]
5	1	0.004	0.22	186
6	1	0.01	0.2	234
7	1	0.03	0.06	926

Table 6.4.4: Recognition rates (RR) and median running times of the heuristic and the Gurobi QCP model, for 2-Euclidean input profiles.

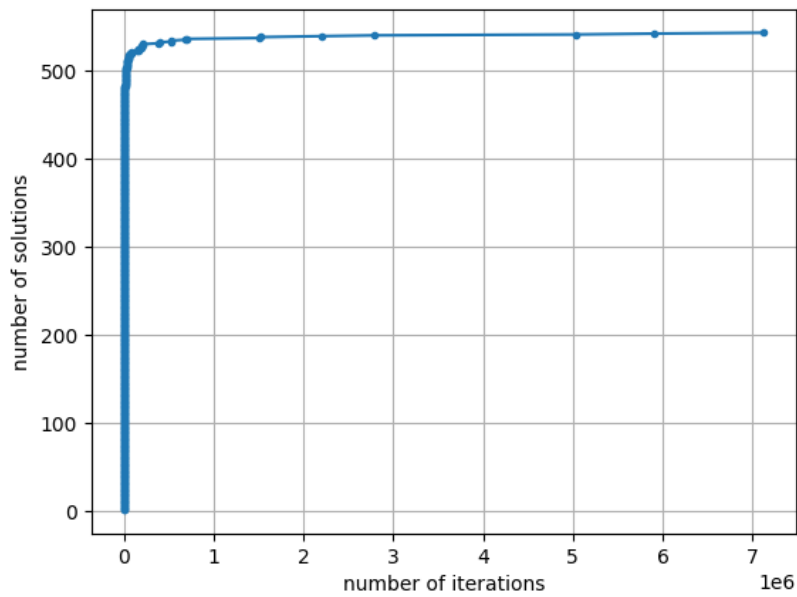
6.4.3 Maximal 2-Euclidean profiles for 5 and 6 candidates

As explained before, with 4 candidates, it is known that there are 3 maximal 2-Euclidean profiles (up to a permutation of candidates). This gives a simple characterization of 2-Euclidean profiles on 4 candidates. As far as we know, nothing is known on the number of maximal 2-Euclidean profiles with more than 4 candidates. For $m = 5$ or 6 , we use the idea of our heuristic to build a set of maximal 2-Euclidean profiles, thus providing a lower bound on this number. More precisely, we repeat (until a timeout is reached) the following procedure. We first generate m random points in the plane (using the uniform distribution on the convex hull size). If the positions are non-degenerated, we built the (maximal) 2-Euclidean profile associated to these positions. We add this profile to our set, after having tested that it is not already contained in it (up to a permutation of candidates).

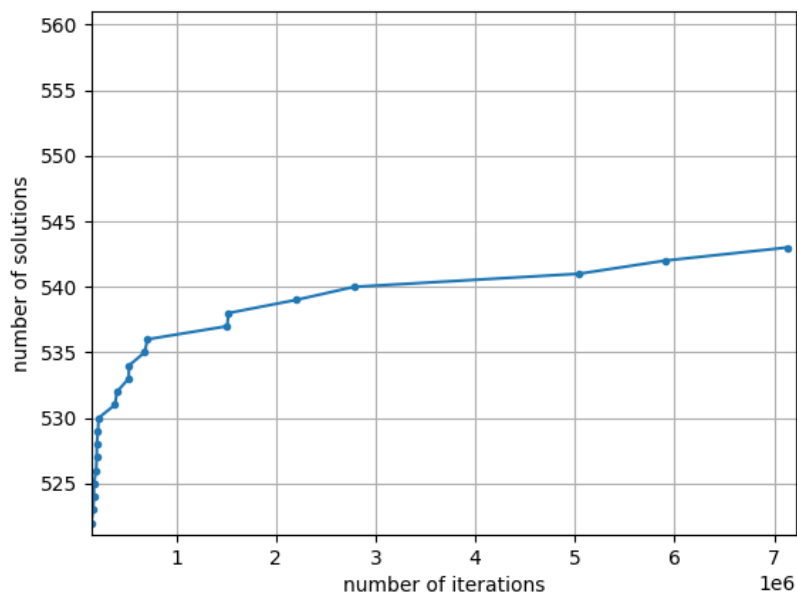
With this heuristic, we could find 543 maximal profiles on 5 candidates, and about 230 000 maximal profiles on 6 candidates! These lower bounds already indicate that the number of maximal profiles seem to grow incredibly fast with the number of candidates.

For 5 candidates, we conjecture that 543 is the exact number, as all these 543 profiles were found within 3 hours and the timeout was set up to one week (see Figure 6.4.2

giving the convergence speed). Note that if the conjecture is true, then we would get a characterization of the 2-Euclidean profiles on 5 candidates. One could this would in turn give a potentially powerful NO-certificate using subprofiles on 5 candidates: we will discuss this point in the global conclusion.



(a) total overview



(b) detail

Figure 6.4.2: The number of discovered maximal profiles over 5 candidates in function of the number of iterations

As another point, our experiments indicate that, not surprisingly, some profiles appear much more frequently than others: while 519 solutions were discovered in 93 792 iterations, 710 192 iterations were needed to get 536 solutions, and we made 7 129 863 iterations to get all 543 solutions. We recall that if the distribution over k solutions were uniform, the expected number of iterations would be approximately $k \log k$.

For 6 candidates, we think that the lower bound of 230 000 is actually very far from the correct number. Indeed, we found these profiles in about 250 000 iterations only, clearly without reaching convergence (we stopped here as with such a high number of profiles the test of uniqueness becomes time-consuming).

6.5 Conclusion

In this chapter, we provided a heuristic for recognizing 2-Euclidean profiles, and evaluated its performance on both real-world and synthetic datasets. A first question that arises from our work concerns the phase transition we observed in the experiments. It would be very interesting to find some mathematical arguments proving bounds on this phenomenon. This would allow us to see how this transition evolves with the number of candidates.

A second question concerns the characterization of 2-Euclidean profiles on 5 candidates. We were able to generate a set of 543 maximal 2-Euclidean profiles, but we do not know if this list is exhaustive or not. Finding such a characterization would possibly improve the NO-test of the heuristic (hence in particular the recognition rate), which seems to be, from our experiments, the key point to address. More generally, we can think of adding forbidden structures in the NO-certificates.

Finally, it would be of course a natural extension of our work to provide a heuristic able to recognize d -Euclidean preferences for $d = 3$ or bigger.

Chapter 7

Euclidean preferences in the plane under ℓ_1 , ℓ_2 and ℓ_∞ norms

7.1 Introduction

We introduced in Chapter 5 the concept of multidimensional Euclidean preferences as well as the state of art of this domain. We have then proposed in Chapter 6 a recognition heuristic for 2-Euclidean preferences. In this chapter, we focus on 2-Euclidean preferences with respect to the norms ℓ_1 and ℓ_∞ , with some results generalized for any dimension $d \geq 2$. Indeed, as already said in Chapter 5, most of the existing works focus on Euclidean preferences with the ℓ_2 norm, while the domain of d -Euclidean preferences using the norm ℓ_1 (resp. ℓ_∞) remains quite unexplored. However, d -Euclidean preferences w.r.t. the norm ℓ_1 are often used in voting advice applications (e.g., Wahl-O-Mat in Germany, Smartvote in Switzerland, Vote Compass in the United States, and many others in multiple countries), i.e., online tools that help the voter to choose the candidate closest to her political stances, and actually often provides her a full ranking of candidates according to her answers to a survey on a range of policy statements. The answers are indeed converted into positions on different dimensions, each position reporting on the level of agreement on a particular policy statement. The ℓ_1 norm is typically used when there are many dimensions (Moreno et al., 2022; Isotalo, 2020).

In this work, we consider the norms ℓ_2 , ℓ_1 and ℓ_∞ , and we try to identify the (structural) differences between them. For instance, we are interested in the following questions:

- Are there forbidden structures that make a profile not Euclidean, under some of the three norms?
- Given a set of m candidates, what is the maximal size (in terms of the number of pairwise distinct preferences) of profiles that are Euclidean?
- Are there some differences or similarities between the norms in the expressivity of Euclidean preferences? Put another way, are there profiles that are Euclidean

with one norm and not with another one?

As the majority of this chapter deals with the dimension 2, we will generally omit the reference to the dimension, and say *Euclidean preferences* for *2-Euclidean preferences*. Each time we use $d \geq 2$, we mention it explicitly to avoid any confusion. On the other hand, and contrary to the previous chapter, we work with three different norms. That is why we shorten the notation introduced in Chapter 5: we say *ℓ -Euclidean preferences* instead of *Euclidean preferences with respect to the norm ℓ* .

The chapter is organized as follows: In Section 7.2, we define the analogue of ℓ_2 bisector (in \mathbb{R}^2) for the norm ℓ_1 , called *boundary hypersurface*. We then observe that a preference profile is ℓ_1 -Euclidean in \mathbb{R}^2 if and only if it is ℓ_∞ -Euclidean.

In Section 7.3, we prove that this equivalence between ℓ_1 -Euclidean preferences and ℓ_∞ -Euclidean preferences no more holds in \mathbb{R}^d for $d \geq 3$. Actually, we prove that in an ℓ_∞ -Euclidean profile there are at most $2d$ candidates ranked last by at least one voter, while there are at most 2^d such candidates for an ℓ_1 -Euclidean profile, and that these upper bounds are tight. We note that this provides a strong difference with ℓ_2 -Euclidean preferences where each candidate can be ranked last at least once for any $d \geq 2$. Also, it is an interesting generalization of the case of 1-dimensional Euclidean preferences, where it is well known that at most 2 candidates can be ranked last. We also give some first geometric insights on the differences between Euclidean profiles under ℓ_1 (resp. ℓ_∞) and ℓ_2 . We then focus on the case $d = 2$. As said just above, in this case a profile is ℓ_1 -Euclidean if and only if it is ℓ_∞ -Euclidean. That is why from Section 7.4, we will only work with norms ℓ_1 and ℓ_2 . We first present in Section 7.4 some geometric properties of ℓ_1 -Euclidean representations that will be useful for the results of subsequent sections. As it can easily be seen that every profile with 2 or 3 candidates is ℓ_1 -Euclidean (thus ℓ_∞ -Euclidean) and ℓ_2 -Euclidean (Bogomolnaia and Laslier, 2004), we focus in Section 7.5 on the case of $m = 4$ candidates. We first give an explicit example of a profile which is ℓ_1 -Euclidean but not ℓ_2 -Euclidean. We then focus on the maximal size (in terms of the number of pairwise distinct preferences) of profiles on 4 candidates that are Euclidean. It is known since the work of Bennett and Hays (1960) that the maximal size is 18 for ℓ_2 . We show that this maximal size is exactly 19 for ℓ_1 . Then, we give a new proof that a profile on 4 candidates is ℓ_2 -Euclidean if and only if it is a subprofile of one of three voter-maximal two-dimensional Euclidean profiles (involving 18 voters). Kamiya et al. (2011) proved the same result, but they rely on a link they establish with the problem of enumerating chambers of hyperplane arrangements (for an introduction to the topic, see, e.g., Stanley et al. (2004)), while we use simpler and purely geometrical arguments.

Finally, in Section 7.6, we focus on the case $m \geq 5$, namely on the the maximal size of profiles which are Euclidean. We show that, despite the strong restriction on the number of candidates ranked last by some voter (at most $2d = 4$), the maximal size of

an ℓ_1 -Euclidean profile is $\Theta(m^4)$, i.e., of the same order of magnitude as for ℓ_2 , as shown by Bennett and Hays (1960).

7.2 Hypersurfaces under ℓ_1 (resp. ℓ_∞) in the plane

We introduce in this section the notion of boundary hypersurfaces under ℓ_1 that will be studied in more details in the next sections. After defining formally the hypersurfaces, we present their different types and give their classification. We also prove a result that ensures that without loss of generality, only one of the three possible types can be considered, which will simplify significantly our further studies in the remaining of this chapter.

7.2.1 Reminder on notations

For convinience, let us briefly recall the main notions that will be used in this chapter - the complete and formal introduction to the topic can be found in Chapter 5, Section 5.2.

To prove that a given preference profile \mathcal{P} is ℓ -Euclidean (with $\ell \in \{\ell_1, \ell_2, \ell_\infty\}$, and $d = 2^1$), we need to find a mapping $f : C \rightarrow \mathbb{R}^d$ such that for each voter v , the area $D_f(v)$ corresponding to the preference $>_v$ is non-empty. We recall that

$$D_f^\ell(v) = \{f(v) \in \mathbb{R}^d : \forall \{c_1, c_2\} \subseteq C, c_1 >_v c_2 \Rightarrow \|f(v) - f(c_1)\|_\ell < \|f(v) - f(c_2)\|_\ell\}.$$

In other words, given preference $>_v$ and two candidates c_i, c_j such that $c_i >_v c_j$, any point of the (non-empty) area $D_f^\ell(v)$ is nearer to c_i than c_j .

So in practice, given the positions of candidates $f(c_1), f(c_2), \dots, f(c_m)$, we need to identify and enumerate the non-empty areas. We have seen in Chapter 6 how to do this with $d = 2$ and $\ell = \ell_2$ - it is sufficient to draft, for each couple of candidates c_i, c_j , the bissector of the segment of extremities $f(c_i)$ and $f(c_j)$. It divides the plane into two half-planes, the first one where c_i is preferred to c_j , and the second one where c_j is preferred to c_i . Each (non-empty) preference area is then a (non-empty) intersection of such half-planes. In particular, we have seen that the areas are convex.

Now we would like to extend this idea to ℓ_1 -Euclidean preferences. But how does “an ℓ_1 -bissector” look like? Before we answer this question, let us introduce some additional notations. We will call *ℓ -boundary hypersurfaces* (or only *boundary hypersurfaces* if no confusion is possible) the “ ℓ -bissectors” by which the preference areas are delimited:

¹However, the definition can be of course announced for any $d \geq 2$.

Definition 7.2.1

For a pair $\{c_1, c_2\} \subseteq C$ of candidates mapped in positions $f(c_1)$ and $f(c_2)$, the set of points $p \in \mathbb{R}^d$ such that $\|f(c_1) - p\|_\ell = \|f(c_2) - p\|_\ell$ is called *the ℓ -boundary hypersurface of c_1 and c_2* (or just hypersurface in what follows), and is denoted by $H_f^\ell(c_1, c_2)$. We denote then by $D_f^\ell(c_1, c_2)$ the set of points $p \in \mathbb{R}^d$ such that $\|f(c_1) - p\|_\ell < \|f(c_2) - p\|_\ell$, and by $D_f^\ell(c_2, c_1)$ the set of points $p \in \mathbb{R}^d$ such that $\|f(c_1) - p\|_\ell > \|f(c_2) - p\|_\ell$.

For convinience, we recall that

$$D_f^\ell(v) = \bigcap_{c_i >_v c_j} D_f^\ell(c_i, c_j)$$

For conciseness, and only if no confusion is possible, we will omit the representation function f and/or the norm ℓ in the notions introduced in Definition 7.2.1. Thus, we will write $H(c_i, c_j)$, resp. $D(c_i, c_j)$ and $D(v)$, instead of $H_f^\ell(c_i, c_j)$, resp. $D_f^\ell(c_i, c_j)$ and $D_f^\ell(v)$. Also, by abuse of notation, the terms area and (its corresponding) preference ranking will be used interchangeably. Finally, for ease of notation, when the position of candidates are fixed, c_i will denote both the candidate and her position in \mathbb{R}^2 (i.e., $f(c_i)$ in the above notation).

7.2.2 Types of hypersurfaces

We first focus on the description of the hypersurfaces in sense of the norm ℓ_1 (resp. ℓ_∞) separating two points c_1 and c_2 , with $d = 2$. We recall that under the ℓ_2 norm, the hypersurface (called bisector in this case) is always a straight line. Actually, it can be defined as the intersection of ℓ_2 -balls centred respectively in c_1 and c_2 , with the radius varying over \mathbb{R}^+ . This point of view can be applied to ℓ_1 (resp. ℓ_∞) hypersurfaces, by simply replacing ℓ_2 -balls by ℓ_1 -balls (resp. ℓ_∞ -balls).

Formally, if we denote by $\mathcal{S}^\ell(x, \delta)$ the ℓ -ball centred in $x \in \mathbb{R}^2$ of rayon δ , the ℓ_2 -hypersurface formalizes as follows:

$$H^\ell(c_1, c_2) = \{\mathcal{S}^\ell(c_1, \delta) \cap \mathcal{S}^\ell(c_2, \delta) \mid \delta > 0\}. \quad (7.1)$$

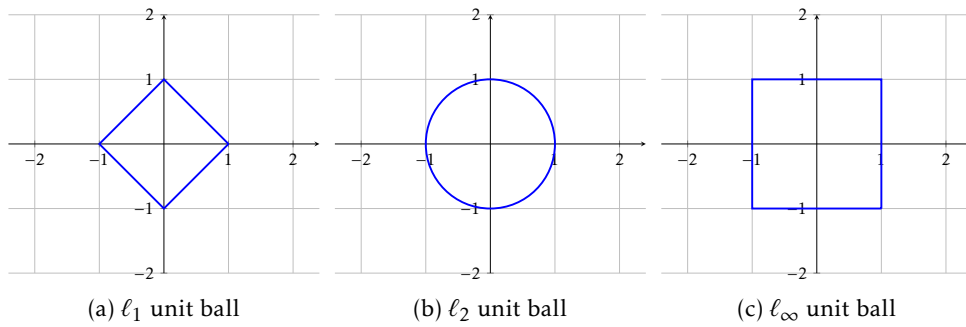


Figure 7.2.1: Unit balls for ℓ_1 , ℓ_2 and ℓ_∞ norms.

It is well-known that a ℓ_2 -ball is a circle, and that two circles (centred in two different points) do not intersect, or intersect at exactly one or two points. A ℓ_1 -ball is a square. We see that, contrary to the previous case, the intersection of two squares (of the same size) centred in two different points may be a segment - indeed, they can “share” (a part of) their sides. In other words, this intersection depends on the relative positions of the centres of the squares. Therefore, the shape of hypersurface depends on the relative positions of c_1 and c_2 when using the ℓ_1 metrics, as we will now show. We denote by (x_1, y_1) (resp. (x_2, y_2)) the coordinates of c_1 (resp. c_2), and we use the notations $\Delta x = |x_1 - x_2|$ and $\Delta y = |y_1 - y_2|$.

1. Let us first consider the case $\Delta x \neq \Delta y$, with $\Delta x > 0$ and $\Delta y > 0$. This case is illustrated in Figure 7.2.2.

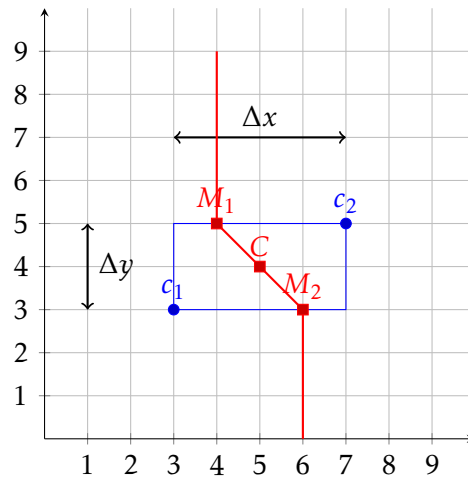


Figure 7.2.2: A (boundary) hypersurface separating c_1 and c_2 : $\Delta x \neq \Delta y, \Delta x > 0, \Delta y > 0$.

Without loss of generality, assume that $\Delta x > \Delta y$ (the case $\Delta x < \Delta y$ can be treated

analogously). The positions c_1 and c_2 can be seen as two opposite vertices of a rectangle (see Fig. 7.2.2).

- By definition of the ℓ_1 metrics, there are two points M_1, M_2 on the rectangle boundary that belong to $H(c_1, c_2)$: these are the points that are at distance $\frac{\Delta x + \Delta y}{2}$ from both points c_1 and c_2 . Points M_1 and M_2 are symmetric with respect to the rectangle centre, and we observe that the segment $[M_1, M_2]$ belongs to $H(c_1, c_2)$ - in fact, we note that this segment is the intersection of the balls of rayon $\frac{\Delta x + \Delta y}{2}$ centred respectively in c_1 and c_2 :

$$[M_1, M_2] = \mathcal{S}^{\ell_1}(c_1, \frac{\Delta x + \Delta y}{2}) \cap \mathcal{S}^{\ell_1}(c_2, \frac{\Delta x + \Delta y}{2})$$

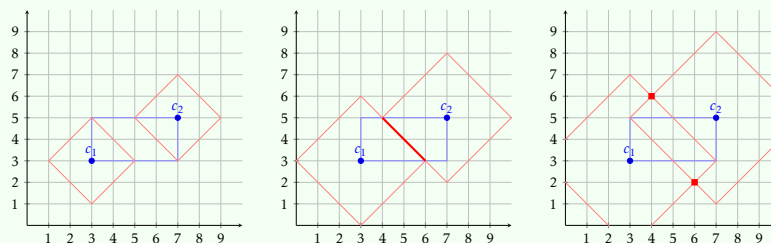
- The half-line $\{(x_{M_1}, y) : y \geq y_{M_1}\}$ also belongs to $H(c_1, c_2)$, where x_{M_1} and y_{M_1} denote the coordinates of M_1 , as for $y \geq y_{M_1}$, each point (x_{M_1}, y) is at distance $\frac{\Delta x + \Delta y}{2} + y - y_{M_1}$ both from c_1 and c_2 .
- Similarly, the half-line $\{(x_{M_2}, y) : y \leq y_{M_2}\}$ belongs to $H(c_1, c_2)$.

To sum it up, we have identified three parts of $H(c_1, c_2)$: two vertical half-lines connected by a diagonal segment. We can easily prove that for each $z \in \mathbb{R}^2$ that does not belong to one of these parts, we have $\|z - c_1\|_{\ell_1} \neq \|z - c_2\|_{\ell_1}$. More precisely, the points to the left-hand side of the hypersurface are closer to c_1 , while the ones on the right-hand side are closer to c_2 .

A Curious Tortoise Intervention



Without giving the formal proof of the fact that these three parts determine completely the hypersurface, we can convince ourselves about it by returning to the formal definition. Indeed, the hypersurface is the intersection of ℓ_1 -balls of the same diameter δ (which we vary over \mathbb{R}^+) centred in c_1 and c_2 . If $\delta < \frac{\Delta x + \Delta y}{2}$, the balls are too small to intersect. With $\delta = \frac{\Delta x + \Delta y}{2}$, the intersection of balls correspond to the segment $[M_1, M_2]$. Finally, with δ greater than this value, the balls will only intersect in two different points:



2. Let us now consider the case $\Delta x = \Delta y > 0$. This case is illustrated in Figure 7.2.3.

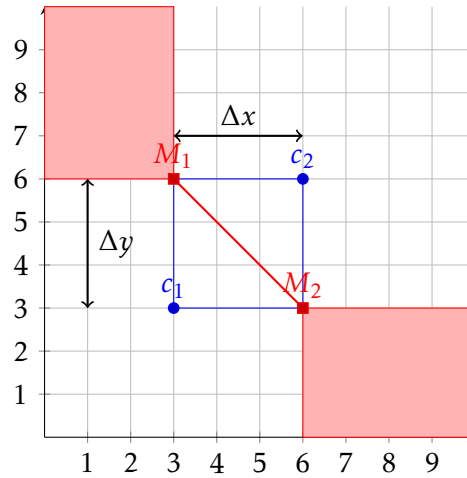



Figure 7.2.3: The hypersurface separating c_1 and c_2 if $\Delta x = \Delta y > 0$.

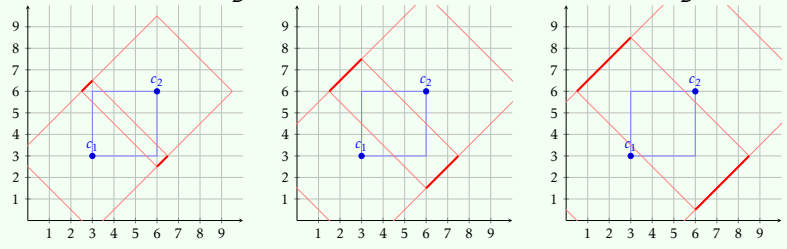
In this special case, the rectangle is a square where c_1 and c_2 are opposite vertices, and M_1 and M_2 are the two other opposite vertices. The hypersurface $H(c_1, c_2)$ is then composed of the three following parts:

- (a) the quadrant $\{(x, y) \in \mathbb{R}^2 : x \leq x_{M_1}, y \geq y_{M_1}\}$,
- (b) the segment $[M_1, M_2]$,
- (c) the quadrant $\{(x, y) \in \mathbb{R}^2 : x \geq x_{M_1}, y \leq y_{M_1}\}$.

A Curious Tortoise Intervention



As in the previous case, if $\delta < \frac{\Delta x + \Delta y}{2}$, the ℓ_1 -balls centered respectively in c_1 and c_2 do not intersect, and the intersection is a segment $[M_1, M_2]$ if $\delta = \frac{\Delta x + \Delta y}{2}$. Let see what happens if $\delta > \frac{\Delta x + \Delta y}{2}$:



3. Consider now the case $\Delta x = 0$ or $\Delta y = 0$. Then the segment $[M_1, M_2]$ is reduced to a unique point, and the hypersurface is then the same as for the ℓ_2 metrics (i.e., a straight line at equal ℓ_2 distance from c_1 and c_2), as shown at Figure 7.2.4.

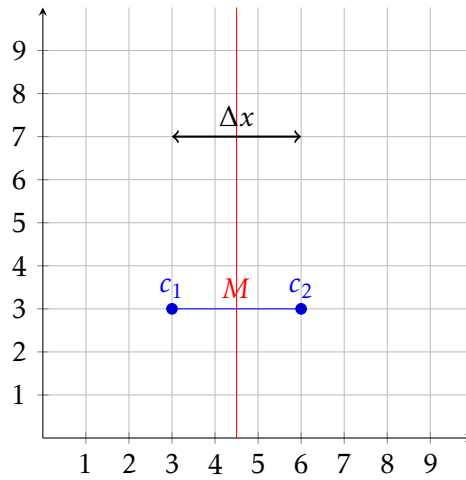


Figure 7.2.4: The hypersurface separating c_1 and c_2 if $\Delta y = 0$. If $\Delta = 0$, the hypersurface is a horizontal line.

This is rather bad news if we want to study ℓ_1 -Euclidean preferences: it becomes tricky to identify non-empty areas from given positions $f(c_1), f(c_2), \dots, f(c_m)$ if the hypersurfaces form depends on these positions. Fortunately, we can prove that any ℓ_1 -Euclidean profile admits a representation f such that all hypersurfaces are of type 1 (i.e., $\Delta x \neq \Delta y$ and $\Delta x, \Delta y > 0$). Therefore, all hypersurfaces can be assumed to be of type 1 in the rest of this chapter (when dealing with $d = 2$):

Lemma 7.2.1

Let \mathcal{P} be an ℓ_1 -Euclidean profile. There exists a representation of \mathcal{P} in which all hypersurfaces are of type 1, i.e., $\Delta x \neq \Delta y$ and $\Delta x, \Delta y > 0$.

Proof. In an ℓ_1 -Euclidean representation of a preference profile, as we consider only strict preferences, we have for any candidates c_i, c_j and voter v :

$$\left| \|f(v) - f(c_i)\|_{\ell_1} - \|f(v) - f(c_j)\|_{\ell_1} \right| > 0 \quad (7.2)$$

Then, let us denote by ε_d the minimum difference in absolute value of distances as in (7.2), over all pairs $\{c_i, c_j\}$ of candidates and voters v . Moreover, let (x_i, y_i) be the position of candidate c_i in the representation, and S_x (resp. S_y, S_{xy}) the set of pairs of candidates $\{c_i, c_j\}$ with $|x_i - x_j| > 0$ (resp. $|y_i - y_j| > 0$, $\left| |x_i - x_j| - |y_i - y_j| \right| > 0$). We define also:

$$\begin{aligned}\varepsilon_x &= \min_{\{c_i, c_j\} \in S_x} |x_i - x_j|, \\ \varepsilon_y &= \min_{\{c_i, c_j\} \in S_y} |y_i - y_j|, \\ \text{and } \varepsilon_{xy} &= \min_{\{c_i, c_j\} \in S_{xy}} \left| |x_i - x_j| - |y_i - y_j| \right|.\end{aligned}$$

If there is a pair $\{c_i, c_j\}$ such that $x_i = x_j$ (in other words, $\{c_i, c_j\} \notin S_x$), we can move one of these candidates, say c_i , by adding ε to x_i with $\varepsilon = \frac{1}{2} \min\{\varepsilon_d, \varepsilon_x, \varepsilon_y, \varepsilon_{xy}\}$. We then get $|x_i - x_j| > 0$. We note that after this operation, we have $S_x \leftarrow S_x \cup \{\{c_i, c_j\}\}$ and S_y and S_{xy} are not modified. An analogous reasoning can be done for every pair $\{c_i, c_j\}$ of candidates such that $\{c_i, c_j\} \notin S_y$ (by moving one candidate on the y -axis), resp. $\{c_i, c_j\} \notin S_{xy}$ (by moving one candidate on one axis). This way, by iterating these modifications, we finally get a representation without the degenerated cases $\Delta x = \Delta y$, $\Delta x = 0$, or $\Delta y = 0$. \square

7.2.3 Classification of the hypersurfaces under ℓ_1 in the plane

We can now go further into the classification of the different hypersurfaces of type 1. There are 4 case to distinguish, depending on whether $y_1 < y_2$ or $y_1 > y_2$, and whether $\Delta x < \Delta y$ or $\Delta y > \Delta x$. This is illustrated in Figure 7.2.5; the four cases are called H^- , V^- , H^+ and V^+ .

First, notice that if $\Delta x < \Delta y$, both half-line parts of the hypersurface are horizontal. In the opposite case, when $\Delta x > \Delta y$, these half-lines are vertical.

Now, let us look at the segment $[M_1, M_2]$ of the hypersurface. In the following, the numbering of the quadrants of the Cartesian coordinate system goes counter-clockwise starting from the upper right quadrant. Without loss of generality, assume that $x_1 < x_2$, where $c_1 = (x_1, y_1)$ and $c_2 = (x_2, y_2)$. If $y_1 < y_2$, the segment $[M_1, M_2]$ is parallel to the II-IV quadrant diagonal, also called the “minus diagonal” (see the upper part of Figure 7.2.5). If $y_1 > y_2$, the segment $[M_1, M_2]$ is parallel to the I-III quadrant diagonal, also called the “plus diagonal” (see the lower part of Figure 7.2.5).

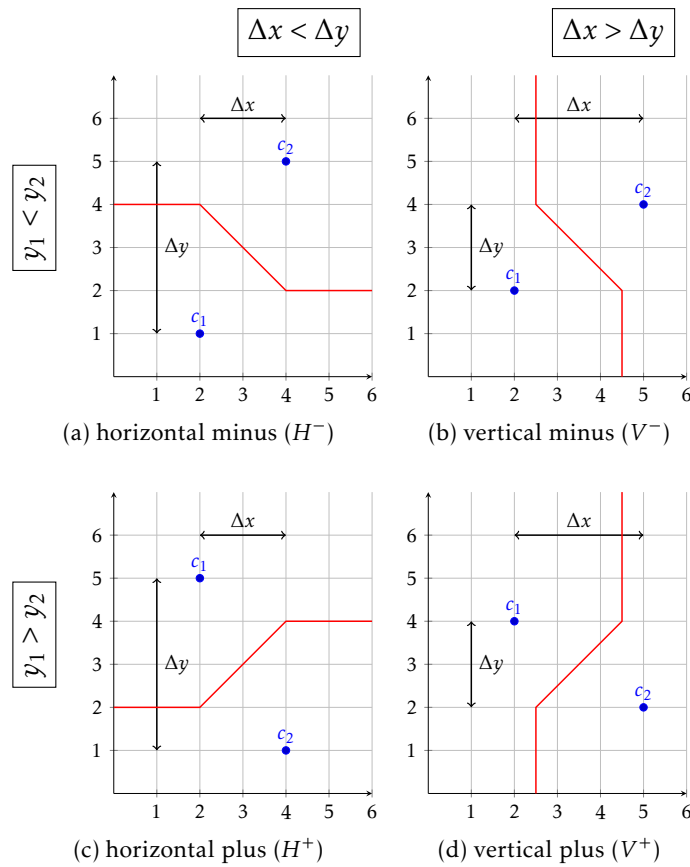


Figure 7.2.5: The different ℓ_1 -hypersurfaces for $x_1 < x_2$.

7.2.4 Hypersurfaces under ℓ_∞ in the plane

Let us do the same classification for the ℓ_∞ -norm. Actually, it can be done very easily - it is sufficient to notice that a ℓ_∞ -ball is also a square, just differently rotated. More formally, for $d = 2$, for all $\delta \geq 0$ and $p \in \mathbb{R}^2$, the spheres $\mathcal{S}_\delta^{\ell_1}(p)$ and $\mathcal{S}_{\delta/\sqrt{2}}^{\ell_\infty}(p)$ are homothetic via the rotation of 45° (see Figure 7.2.6). Using the characterisation of $H(c_1, c_2)$ in Equation (7.1), this yields the following observation, already noted by Lee and Wong (1980).

Observation 7.2.1: (Lee and Wong, 1980)
 A preference profile is ℓ_1 -Euclidean in \mathbb{R}^2 if and only if it is ℓ_∞ -Euclidean in \mathbb{R}^2 .

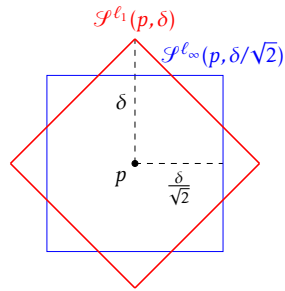
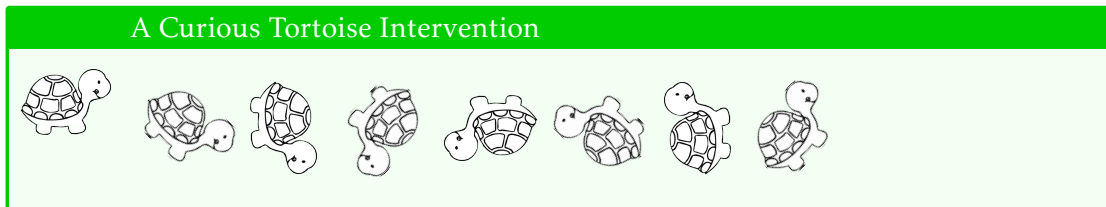


Figure 7.2.6: For $d = 2$, the spheres $\mathcal{S}^{\ell_1}(p, \delta)$ and $\mathcal{S}^{\ell_\infty}(p, \frac{\sqrt{2}}{2}\delta)$ are homothetic via the rotation of 45° .



Therefore, all results for $d = 2$ will be announced using the norm ℓ_1 in the remaining of this chapter, keeping in mind that they also hold for ℓ_∞ .

A very natural question is to ask if this equivalence generalizes for $d \geq 3$. We will prove in the following section that it does not.

7.3 Relations between different norms

We start this section by comparing an ℓ_1 and an ℓ_2 representation of the profile on 3 candidates that includes all $3! = 6$ possible strict preferences (see Figure 7.3.1).

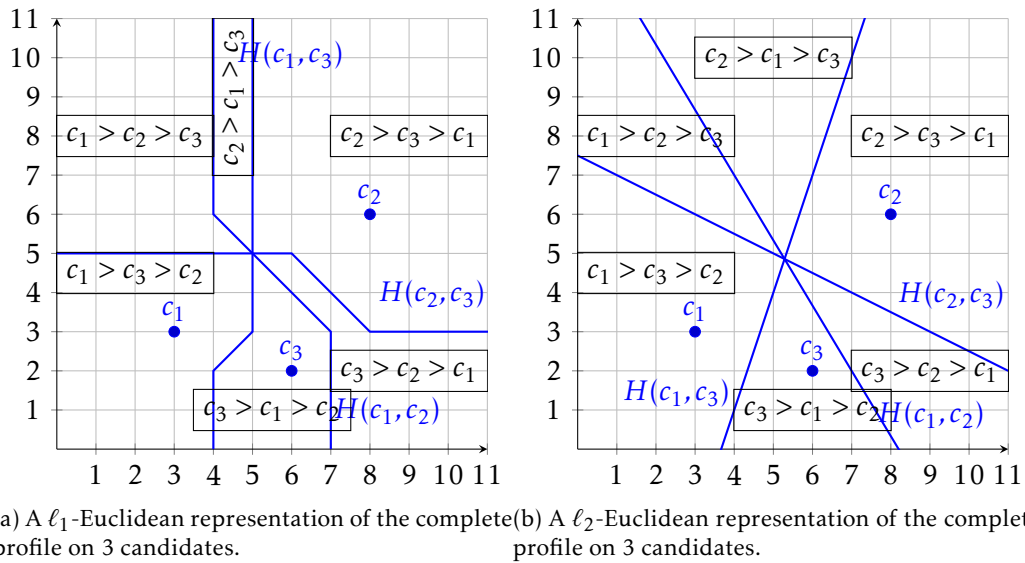


Figure 7.3.1

Although in this case the profile is both ℓ_1 and ℓ_2 -Euclidean, a natural question is if there is a profile that would be ℓ_1 -Euclidean but not ℓ_2 -Euclidean, and vice-versa.

Put together with what we have seen in the previous section, there are two questions to be answered:

- Do we have the equivalence between ℓ_1 -norm and ℓ_∞ -norm for $d \geq 3$?
- Is there a profile that is ℓ_1 -Euclidean but not ℓ_2 -Euclidean, and vice-versa?

The first of them is studied in Section 7.3.1, while the second one in Section 7.3.2.

7.3.1 Relation between norms ℓ_1 and ℓ_∞

Let us show that the domain of ℓ_1 and ℓ_∞ -Euclidean profiles are *not* equivalent for $d \geq 3$. This is actually a corollary of the following proposition, which provides a structural property of ℓ_∞ -Euclidean and ℓ_1 -Euclidean profiles in \mathbb{R}^d .

Proposition 7.3.1

In an ℓ_∞ -Euclidean profile in \mathbb{R}^d , at most $2d$ candidates are ranked last by at least one voter. In an ℓ_1 -Euclidean profile in \mathbb{R}^d , at most 2^d candidates are ranked last by at least one voter. These bounds are tight for all d .

Proof. Let us first consider an ℓ_∞ -Euclidean profile \mathcal{P} , and a corresponding mapping f . We denote by $f_i(x)$ the position of candidate/voter x on the i^{th} coordinate. For

$i = 1, \dots, d$, let us denote by $c_{j_i}^m$ and $c_{j_i}^M$ the candidates that have minimal and maximal i^{th} coordinate. There are at most $2d$ of them (some candidates may be extremal on several coordinates). Take a candidate c which is *not* among these extremal candidates, and take a voter v . We show that c cannot be ranked last by v . Let i be such that $\|f(c) - f(v)\|_{\ell_\infty} = |f_i(c) - f_i(v)|$.

If $f_i(c) \geq f_i(v)$, then

$$\|f(c) - f(v)\|_{\ell_\infty} = f_i(c) - f_i(v) \leq f_i(c_{j_i}^M) - f_i(v) \leq \|f(c_{j_i}^M) - f(v)\|_{\ell_\infty}.$$

As the two distances cannot be the same (no tie in the preferences), c is ranked before $c_{j_i}^M$ by v .

If $f_i(c) < f_i(v)$, then

$$\|f(c) - f(v)\|_{\ell_\infty} = f_i(v) - f_i(c) \leq f_i(v) - f_i(c_{j_i}^m) \leq \|f(c_{j_i}^m) - f(v)\|_{\ell_\infty}.$$

Again, c is ranked before $c_{j_i}^m$ by v . In both cases, c is not ranked last.

To show the tightness of the bound, we consider a profile on $2d$ candidates where $f(c_{2i-1})$ is -1 on coordinate i and 0 on all other coordinates, and $f(c_{2i}) = -f(c_{2i-1})$. There are also $2d$ voters, with $f(v_i) = f(c_i)$ for $i = 1, \dots, 2d$. Then it is easy to see that c_{2i-1} is ranked last by v_{2i} , and c_{2i} is ranked last by v_{2i-1} .²

Let us now focus on ℓ_1 -Euclidean profiles. For each vector u in $\{-1, 1\}^d$, let c_u be a candidate which maximizes $u \cdot f(c) = \sum_{i=1}^d u_i \cdot f_i(c)$. As previously, consider a candidate c which is not among these (at most) 2^d extreme candidates, and take a voter v . We show that c cannot be ranked last by v . By definition, $\|f(c) - f(v)\|_{\ell_1} = \sum_{i=1}^d |f_i(c) - f_i(v)|$. Define the vector u as $u_i = 1$ if $f_i(c) \geq f_i(v)$ and $u_i = -1$ otherwise. Then:

$$\begin{aligned} \|f(c) - f(v)\|_{\ell_1} &= \sum_{i=1}^d u_i \cdot (f_i(c) - f_i(v)) = u \cdot f(c) - u \cdot f(v) \\ &\leq u \cdot f(c_u) - u \cdot f(v) \leq \sum_{i=1}^d |f_i(c_u) - f_i(v)| = \|c_u - v\|_{\ell_1}. \end{aligned}$$

As the two distances must be different (no tie in the preferences), c is not ranked last by v .

To show the tightness, let us consider the following profile on 2^d candidates and 2^d voters. For each vector $u \in \{-1, 1\}^d$, we define a candidate c_u with $f(c_u) = u$, and a voter v_u with $f(v_u) = -u$. Then we have $\|f(c_u) - f(v_u)\|_{\ell_1} = 2d$, while if $u \neq u'$ we have $\|f(c_{u'}) - f(v_u)\|_{\ell_1} \leq 2(d-1)$ (as $f_i(c_{u'}) = f_i(v_u)$ on at least one coordinate i). So c_u is ranked last by v_u .³

□

²Note that, defined like this, there are some ties in the distances among the candidates that are not ranked last, but these can be easily broken by slightly moving the positions, for instance moving c_{2i-1} (resp. c_{2i}) by $-\epsilon_i$ (resp. $+\epsilon_i$) on the i^{th} coordinate, with $\epsilon_i \neq \epsilon_j$ for $i \neq j$.

³As previously, the ties between distances among candidates that are not ranked last can be removed by slightly moving the positions.

A Curious Tortoise Intervention



Note that for $d = 1$, we get two known result: there are at most two candidates ranked last in any ℓ_∞ -Euclidean profile ($2d = 2 \cdot 1 = 2$) and any ℓ_1 -Euclidean profile ($2^d = 2^1 = 2$). Indeed, all three norms $\ell_1, \ell_2, \ell_\infty$ coincide for $d = 1$. But a ℓ_2 -Euclidean profile is in particular single-peaked, so there are at most two candidates ranked last.

For $d = 2$, we have $2 \cdot 2 = 2^2$, which is consistent with Observation 7.2.1.

A Curious Tortoise Intervention



It's me again! Notice that this proposition proves that there are some ℓ_1 -Euclidean profiles that are not ℓ_∞ -Euclidean, as more candidates can be ranked last in an ℓ_∞ -Euclidean profile.

Would there be any ℓ_∞ -Euclidean profile that would not be ℓ_1 -Euclidean? Even though I am convinced such a profile exists, I was not able (so far!) to find it...

7.3.2 Relation between norms ℓ_1 and ℓ_2

In this section, we show that for $d = 2$, there are ℓ_2 -Euclidean profiles that are not ℓ_1 -Euclidean, and vice-versa.

To build an ℓ_2 -Euclidean profile that is not ℓ_1 -Euclidean, we can use Proposition 7.3.1 given in the previous section - there are at most 4 candidates ranked last at least once in any ℓ_1 -Euclidean profiles. However, it is easy to build an ℓ_2 -Euclidean profile in which each candidate is ranked last at least once. For instance, we can fix the values $f(c_1), f(c_2), \dots, f(c_m)$ in a way that they form a regular polygon - hence, these points will all lie on a circle. We add then m voters on this circle with the circular angle between $f(c_i)$ and $f(v_i)$ equals to $\pi - \varepsilon$, where $\varepsilon < \frac{\pi}{m}$. Therefore, c_i is ranked last by v_i . In other words, each candidate is ranked last at least once. Note that the role of ε is to obtain a preference profile without ties.

We now give a profile that is ℓ_1 -Euclidean but not ℓ_2 -Euclidean. Let us consider the following profile \mathcal{P} of 9 preferences over 4 candidates:

$$\begin{aligned}
 >_{v_1}: (c_4, c_3, c_1, c_2) & >_{v_2}: (c_3, c_4, c_1, c_2) \\
 >_{v_3}: (c_4, c_3, c_2, c_1) & >_{v_4}: (c_3, c_4, c_2, c_1) \\
 >_{v_5}: (c_2, c_1, c_4, c_3) & >_{v_6}: (c_2, c_1, c_3, c_4) \\
 >_{v_7}: (c_1, c_2, c_4, c_3) & >_{v_8}: (c_1, c_2, c_3, c_4) \\
 >_{v_9}: (c_2, c_3, c_1, c_4) &
 \end{aligned}$$

Let us show that this profile is not ℓ_2 -Euclidean in \mathbb{R}^2 . By contradiction, assume that a ℓ_2 -Euclidean representation in \mathbb{R}^2 exists. The points c_1, c_2, c_3 form necessarily a (non-degenerate) triangle, as 5 different rankings over $\{c_1, c_2, c_3\}$ are present in the profile, and at most 4 can be represented if c_1, c_2, c_3 are aligned in \mathbb{R}^2 . Figure 7.3.2 illustrates the different preference areas obtained from candidates c_1, c_2, c_3 forming a triangle.

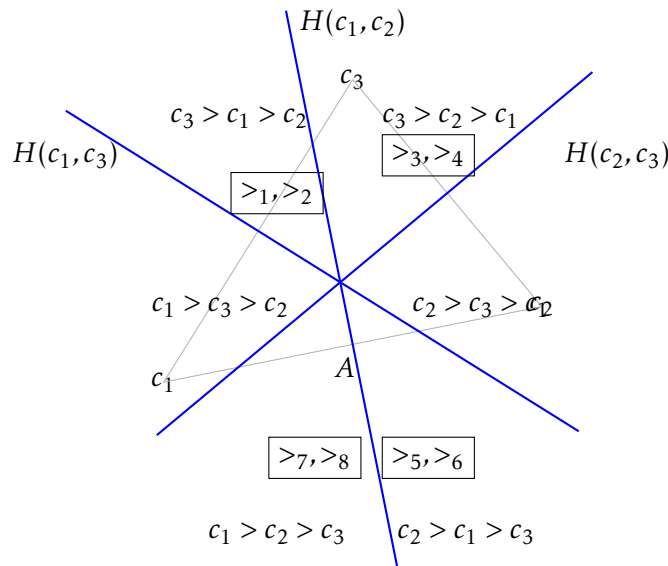


Figure 7.3.2: The different preference areas obtained from candidates c_1, c_2 and c_3 forming a triangle. Each preference R_v is a subset of an area, the precise contours of which depends on the position of c_4 in \mathbb{R}^2 .

Note that for each $i \in \{1, 2, 3, 4\}$, we obtain $>_{2i-1}$ from $>_{2i}^4$ by swapping c_3 and c_4 . Thus, $H(c_3, c_4)$ has to go through the area $c_3 > c_1 > c_2$ to separate $>_1$ and $>_2$, through the area $c_3 > c_2 > c_1$ to separate $>_3$ and $>_4$ and finally through the areas $c_2 > c_1 > c_3$ and $c_1 > c_2 > c_3$ to separate $>_5$ and $>_6$, and $>_7$ and $>_8$ (see Figure 7.3.2 for more clarity). This is not possible, as any straight line can cross at most 3 of these 4 areas. Indeed, if a straight line crosses both the area containing $\{>_7, >_8\}$ and the one containing $\{>_5, >_6\}$,

⁴These pairs correspond to rows in the display of the profile given above

then it must intersect $H(c_1, c_2)$ below point A . Similarly, if it crosses both the area containing $\{>_1, >_2\}$ and the one containing $\{>_3, >_4\}$, then it must intersect $H(c_1, c_2)$ above point A . Thus, to cross the 4 areas, it must intersect $H(c_1, c_2)$ twice, a contradiction.

Hence, no ℓ_2 -Euclidean representation of \mathcal{P} exists in \mathbb{R}^2 .

However, this profile is ℓ_1 -Euclidean: we give a possible representation of it in Figure 7.3.3.

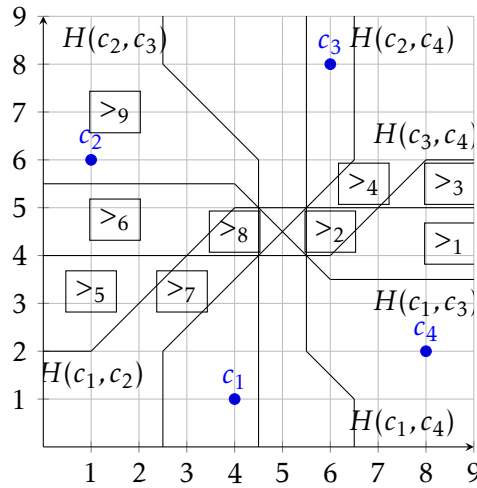
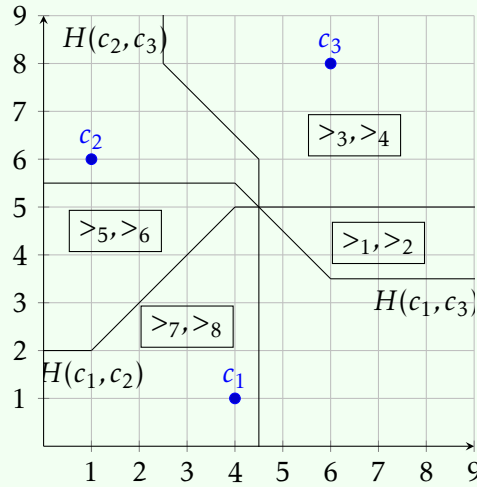


Figure 7.3.3: A ℓ_1 -Euclidean representation of the profile \mathcal{P}

A Curious Tortoise Intervention



Let us see how is it possible to find an ℓ_1 -representation of \mathcal{P} . We proceed as in the case of ℓ_2 -representation, and we start with the candidates c_1, c_2 and c_3 forming a triangle:



As in the case of ℓ_2 -representation, we need $H(c_3, c_4)$ to go through area containing $>_1$ and $>_2$, $>_3$ and $>_4$, $>_5$ and $>_6$, and finally $>_7$ and $>_8$. In particular, it must intersect $H(c_1, c_2)$ in two distinct points. We note that this time, it is possible (see carefully Figure 7.3.3 again).

7.4 Properties of hypersurfaces under ℓ_1 in the plane

It will come as no surprise that many geometrical properties holding for ℓ_2 do not hold for ℓ_1 -hypersurfaces. Let us mention some of them that will be useful for the rest of this paper.

7.4.1 Intersection of two hypersurfaces

It is well-known that given two distinct straight lines (i.e., ℓ_2 -hypersurfaces), the intersections of these lines is either empty (if the lines are parallel) or contains a unique point. In the case of ℓ_1 -hypersurfaces, more situations may arise, as stated in the following proposition (several examples of possible intersections are given in Figure 7.4.1 for illustration. The proof, as well as complete figures illustrating the different situations, are given in Appendix 7.A.1).

Proposition 7.4.1

The intersection of two distinct ℓ_1 -hypersurfaces is either empty, or contains a unique point, or two distinct points, or an infinite number of points.

The following result states that, to recognise a ℓ_1 -Euclidean profile, we can assume without loss of generality that the last case of Proposition 7.4.1 (corresponding to a

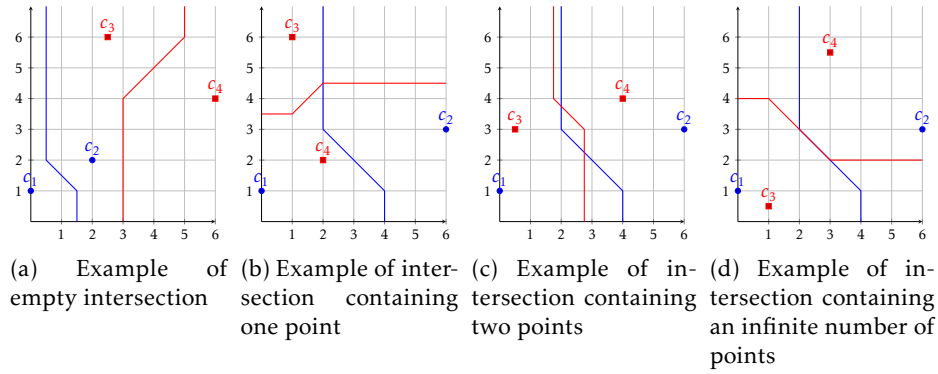


Figure 7.4.1: The intersection of two distinct ℓ_1 -hypersurfaces : several examples.

degenerate case) never occurs. Thus, in the remainder of the chapter, we assume w.l.o.g. that hypersurfaces intersect in at most 2 points.

Lemma 7.4.1

Let \mathcal{P} be an ℓ_1 -Euclidean profile. There exists a representation of \mathcal{P} in which any pair of hypersurfaces intersect in at most 2 points.

Proof. The proof is similar to that of Lemma 7.2.1. Let us suppose that for a given ℓ_1 -Euclidean representation f , there are two hypersurfaces $H(c_i, c_j)$ and $H(c_k, c_l)$ that intersect in infinitely many points. We will show that it is always possible to slightly change the position of one of the points $f(c_i), f(c_j), f(c_k)$ and $f(c_l)$ so that the hypersurfaces intersect in at most two points, without modifying the types of hypersurfaces and the set of representation areas.

Suppose that one of the hypersurfaces, say $H(c_i, c_j)$, is vertical (the other case being symmetrical). We move the point $f(c_i)$. To do so, we denote

$$\varepsilon_t = \min_{c_k \in C \setminus \{c_i\}} \left| |x_i - x_k| - |y_i - y_k| \right|$$

and, as in the Lemma 7.2.1,

$$\varepsilon_d = \min_{v \in V} \min_{c_i, c_j \in C} \left| \|f(v) - f(c_i)\|_{\ell_1} - \|f(v) - f(c_j)\|_{\ell_1} \right|.$$

As we consider only strict preferences, $\varepsilon_d > 0$. Also, thanks to Lemma 7.2.1 that excluded a degeneration $\Delta x = \Delta y$, we have $\varepsilon_t > 0$. Let $\varepsilon = \frac{1}{2} \min\{\varepsilon_d, \varepsilon_t\}$.⁵ We can now move the point $f(c_i)$ by adding ε to x_i . As $\varepsilon < \varepsilon_d$, we do not change the set of preferences corresponding to representation areas. As $\varepsilon < \varepsilon_t$, we do not change the type of any

⁵For completeness, we should also choose ε smaller than ε_{x_i} , ε_{y_i} and ε_{xy} introduced in Lemma 7.2.1, to ensure that we do not create any degeneration excluded by this Lemma while moving the point $f(c_i)$.

hypersurface involving c_i . Finally, as x_i increased by $\varepsilon > 0$, the value of $|x_i - x_j| + |y_i - y_j|$ changes, and the hypersurface (both the vertical extremities and the middle segment) slightly moves to the right on the x -axis. Therefore, $H(c_i, c_j)$ and $H(c_k, c_l)$ no more intersect in an infinity of points. \square

7.4.2 Intersection of three hypersurfaces

Under the norm ℓ_2 , the bisectors of the three sides of a non-degenerate triangle intersect in a unique point, called circumcenter. Otherwise (i.e., in the case where the three points are aligned), they do not intersect. In terms of hypersurfaces, given three points c_1, c_2 and c_3 , the hypersurfaces $H(c_1, c_2)$, $H(c_1, c_3)$ and $H(c_2, c_3)$ intersect in at most one point under ℓ_2 . It is possible to find an analogous result under ℓ_1 -hypersurfaces : the three hypersurfaces also intersect in at most one point. However, the “degenerate” case is different under this norm. To introduce it, the following notion of *diagonal rectangle associated with the positions of 2 candidates* (see Figure 7.4.2 for an illustration) is needed.

Definition 7.4.1

Let c_i and c_j be two candidates, and $(x_i, y_i), (x_j, y_j)$ their positions in the two-dimensional plane. Let us denote by:

- $d_i^+ = \{(x, y) | y = x - x_i + y_i\}$ the “+” diagonal going through the point c_i ,
- $d_i^- = \{(x, y) | y = -x + x_i + y_i\}$ the “-” diagonal going through the point c_i ,
- $d_j^+ = \{(x, y) | y = x - x_j + y_j\}$ the “+” diagonal going through the point c_j ,
- $d_j^- = \{(x, y) | y = -x + x_j + y_j\}$ the “-” diagonal going through the point c_j .

Let us call A the intersection point of d_i^+ and d_j^- and B the intersection point of d_i^- and d_j^+ (see Figure 7.4.2). We call *diagonal rectangle determined by c_i and c_j* the rectangle whose sides are parallel to the diagonals and whose vertices are c_i, A, c_j and B , and we denote by $paral(c_i, c_j)$ the interior of the diagonal rectangle.

Note that as we consider non degenerated profile following Lemma 7.2.1, no point (besides c_i and c_j) lies on one of the 4 diagonals - and in particular on the boarder of the diagonal rectangle. Moreover, c_j does not lie on d_i^+ and d_i^- , and c_i does not lie on d_j^+ and d_j^- .

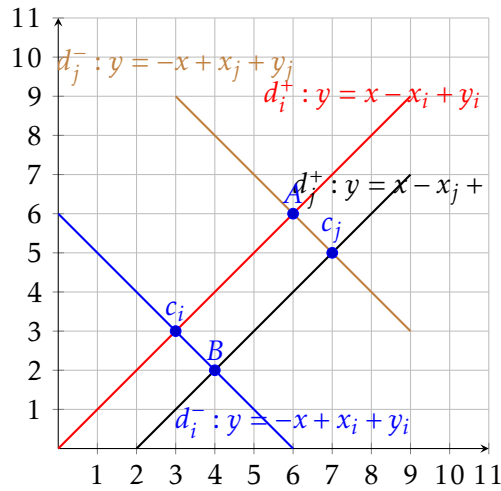


Figure 7.4.2: The diagonal rectangle determined by c_i and c_j .

Then we have the following result:

Proposition 7.4.2

Given three points $c_1 = (x_1, y_1)$, $c_2 = (x_2, y_2)$ and $c_3 = (x_3, y_3)$:

- If c_1 , c_2 or c_3 is inside the diagonal rectangle determined by the two other points, then $H(c_1, c_2)$, $H(c_1, c_3)$ and $H(c_2, c_3)$ do not (pairwise) intersect.
- Otherwise, the intersection of the three hypersurfaces is a unique point.

To prove Proposition 7.4.2, we will proceed as follows:

Step 1: We prove that given three points, the intersection of the three corresponding hypersurfaces is empty if and only if they are all vertical (resp. horizontal). Otherwise, they intersect in a unique point.

Step 2: We finally prove that if one point lies inside the diagonal rectangle determined by two remaining points, then all three corresponding hypersurfaces are vertical (resp. horizontal).

Before we prove the Step 1, let us prove the following easy lemma, that we will use in the proof of Step 1, as well as in some other proofs:

Lemma 7.4.2

Given three points c_1, c_2 and c_3 , we have:

$$H(c_1, c_2) \cap H(c_1, c_3) \cap H(c_2, c_3) = H(c_i, c_j) \cap H(c_j, c_k)$$

for all i, j, k such that $\{i, j, k\} = \{1, 2, 3\}$.

Proof. The left-right inclusion is obvious. For the right-left inclusion, without loss of generality, assume that $i = 1, j = 2$ and $k = 3$, and consider $x \in H(c_i, c_j) \cap H(c_j, c_k)$. Then,

$$\|x - c_1\|_{\ell_1} = \|x - c_2\|_{\ell_1} = \|x - c_3\|_{\ell_1}$$

because

$$\begin{aligned} x \in H(c_i, c_j) &\Rightarrow \|x - c_i\|_{\ell_1} = \|x - c_j\|_{\ell_1}, \\ x \in H(c_i, c_k) &\Rightarrow \|x - c_i\|_{\ell_1} = \|x - c_k\|_{\ell_1}. \end{aligned}$$

Hence, $x \in H(c_1, c_2) \cap H(c_1, c_3) \cap H(c_2, c_3)$. \square

We can now prove Step 1 - as it is an interesting result in itself that will also be used several times in the remaining of this chapter, we announce it as a proposition:

Proposition 7.4.3

Given three points c_1, c_2 and c_3 :

- If $H(c_1, c_2)$, $H(c_1, c_3)$ and $H(c_2, c_3)$ are all vertical (or all horizontal), then the intersection of each pair of hypersurfaces is empty. In particular, the intersection of the 3 hypersurfaces is empty.
- If two of them are vertical and one is horizontal (or vice-versa), then the intersection of the 3 hypersurfaces is a unique point.

Proof. Assume first that the three hypersurfaces are vertical. Let $(x_1, y_1), (x_2, y_2)$ and (x_3, y_3) denote the positions in the plane of c_1, c_2 and c_3 . Without loss of generality, we assume that $x_1 < x_2 < x_3$ and that $H(c_1, c_3)$ is of type V^- .

Given a vertical hypersurface $H(c_i, c_j)$, for each point $(x, y) \in H(c_i, c_j)$ we have $x_i < x < x_j$, because $x_i < x_{M_1} < x_j$ and $x_i < x_{M_2} < x_j$ for the extremities M_1, M_2 of the middle segment of the hypersurface (see Figure 7.2.2, page 193). Thus, $H(c_1, c_2)$ and $H(c_2, c_3)$ do not intersect, as we have $x_1 < x_2 < x_3$. Using Lemma 7.4.2 (page 209), we conclude that $H(c_1, c_2) \cap H(c_1, c_3) \cap H(c_2, c_3) = \emptyset$.

Let us now assume that two hypersurfaces are vertical (resp. horizontal) and the third one is horizontal (resp. vertical). Without loss of generality, we can assume that $H(c_1, c_2)$ is horizontal and both remaining hypersurfaces $H(c_1, c_3)$ and $H(c_2, c_3)$ are vertical. Any vertical hypersurface intersects any horizontal hypersurface in a unique

point (by assuming w.l.o.g. that the representation is non-degenerate, see Lemma 7.4.1). In particular, $|H(c_1, c_2) \cap H(c_1, c_3)| = 1$. Lemma 7.4.2 states that $H(c_1, c_2) \cap H(c_1, c_3) \cap H(c_2, c_3) = H(c_i, c_j) \cap H(c_i, c_k)$ for $\{i, j, k\} = \{1, 2, 3\}$. We have therefore the three hypersurfaces intersecting in a unique point. \square

We can finally prove the Step 2 and conclude the proof of Proposition 7.4.2. As this is very long and quite technical case analysis type proof that does not provide any interesting insight into the topics, we only give a sketch of the proof here. The detailed version can be found in Appendix 7.A

Property 7.4.2 - sketch of proof. To prove the first point, we assume without loss of generality that c_2 lies inside the diagonal rectangle determined by c_1 and c_3 , and we prove that in this case, the three hypersurfaces are all vertical or horizontal. More precisely, we do case distinction on the type of $H(c_1, c_3)$, and we prove that every time it is vertical (resp. horizontal), the remaining two hypersurfaces $H(c_1, c_2)$ and $H(c_2, c_3)$ are also vertical (resp. horizontal). Proposition 7.4.3 then implies that they do not (pairwise) intersect.

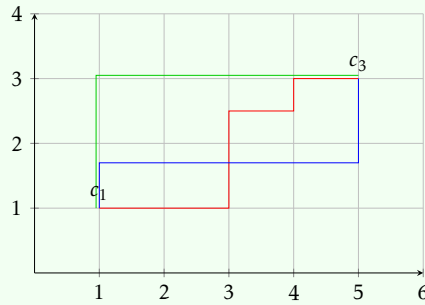
To prove the second point, we suppose that any point does not lie in the diagonal rectangle determined by the remaining two points, and we show that in such a case, there is necessarily at least one horizontal and one vertical hypersurface. As a vertical hypersurface and a horizontal hypersurface intersect in a unique point, Lemma 7.4.2 allows us to conclude that the three hypersurfaces intersect in a unique point. \square

A Curious Tortoise Intervention

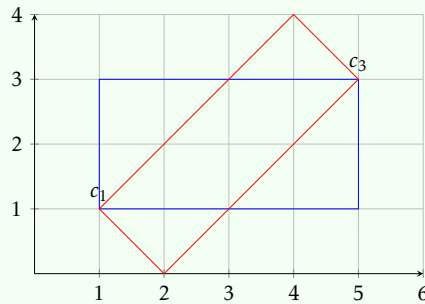


Let us discuss a little bit the difference with ℓ_2 -norm. We have already seen that under ℓ_2 -norm, the hypersurfaces $H(c_1, c_2)$, $H(c_1, c_3)$ and $H(c_2, c_3)$ do not intersect if and only if c_1, c_2 and c_3 are aligned. In other words, if one of these points, let say c_2 , lies on the segment c_1, c_3 . But actually, what does the notion of segment look like in ℓ_1 -geometry?

In ℓ_2 -geometry the segment $[AB]$ is the (unique) shortest path connecting the points A and B . But it is well-known that in ℓ_1 -geometry, there are infinitely many shortest paths from A to B :



So intuitively, one could expect that the hypersurfaces do not intersect under the norm ℓ_1 if and only if one of the points, let say c_2 , lies on (some) shortest path connecting c_1 and c_3 . However, we have just proven that this is not true - actually, it would mean that c_2 lies inside the **rectangle** with sides parallel to the axis determined by c_1 and c_3 , while we have seen that it lies inside *the diagonal rectangle* determined by c_1 and c_3 :



7.4.3 Some corollaries and further results

Propositions 7.4.2 and 7.4.3 have some interesting corollaries that will be useful in the remaining of the chapter. The next result is a direct corollary of Proposition 7.4.3:

Corollary 7.4.1

Given three points c_i, c_j, c_k , the hypersurfaces $H(c_i, c_j)$ and $H(c_i, c_k)$ intersect in at most one point. In other words, given four points c_i, c_j, c_k and c_l , if two hypersurfaces $H(c_i, c_j)$ and $H(c_k, c_l)$ intersect in two different points, then c_i, c_j, c_k and c_l are all distinct.

Proof. By contradiction, assume that $H(c_i, c_j)$ and $H(c_i, c_k)$ intersect in two distinct points. According to Lemma 7.4.2 (page 209), $|H(c_i, c_j) \cap H(c_i, c_k) \cap H(c_j, c_k)| \geq 2$. We get a contradiction with Proposition 7.4.3 which states that if the three hypersurfaces intersect, then the point of intersection is unique. \square

Proposition 7.4.2 has a direct consequence on the preferences within an ℓ_1 -Euclidean profile. It is given in the following corollary, which will be used in Section 7.6 to show an upper bound on the number of candidates ranked last by at least one voter in a ℓ_1 -Euclidean profile.

Corollary 7.4.2

Let $\mathcal{P} = (V, C)$ be an ℓ_1 -Euclidean profile, and consider three candidates $c_1 = (x_1, y_1)$, $c_2 = (x_2, y_2)$ and $c_3 = (x_3, y_3)$ in a given ℓ_1 -Euclidean representation of \mathcal{P} . If $c_2 \in \text{paral}(c_1, c_3)$, then there is no voter $v \in V$ for who both $c_1 >_v c_2$ and $c_3 >_v c_2$. In other words, c_2 is never ranked last among c_1, c_2, c_3 .

Proof. Assume that c_2 is inside the diagonal rectangle determined by c_1 and c_3 . Proposition 7.4.2 implies that $H(c_1, c_2)$, $H(c_1, c_3)$ and $H(c_2, c_3)$ do not (pairwise) intersect. Hence, they are all horizontal, or all vertical (as a vertical hypersurface always intersects a horizontal one). Without loss of generality, assume that all three hypersurfaces are vertical, and that $x_1 < x_2 < x_3$. As each point (x, y) of $H(c_i, c_j)$ satisfies $x_i < x < x_j$, we have $H(c_1, c_2)$ on the left of $H(c_2, c_3)$.

We now show by contradiction that $H(c_1, c_3)$ lies between these two hypersurfaces. Assume the left-to-right order of hypersurfaces is $H(c_1, c_3), H(c_1, c_2), H(c_2, c_3)$. As c_1 lies in the leftmost area, it is necessarily the top-ranked candidate there. The second-ranked candidate in this area must be c_3 , the leftmost hypersurface being $H(c_1, c_3)$. Thus, the ranking of the leftmost area is $c_1 > c_3 > c_2$. By moving from the leftmost to the rightmost area, we obtain consecutively (by crossing the hypersurfaces one by one) the four following rankings: (c_1, c_3, c_2) (the leftmost one), (c_3, c_1, c_2) (after crossing $H(c_1, c_3)$), (c_3, c_2, c_1) (after crossing $H(c_1, c_2)$) and finally (c_2, c_3, c_1) (the rightmost one, after crossing $H(c_2, c_3)$). We get a contradiction: as c_3 lies in the rightmost area (because we have $x_1 < x_2 < x_3$), it must be a top-ranked candidate there.

The case where $H(c_1, c_3)$ is the rightmost hypersurface can be treated similarly. Hence, the only possible order of hypersurfaces is $H(c_1, c_2), H(c_1, c_3), H(c_2, c_3)$, and we see, with similar arguments as previously, that c_2 is never ranked last. \square

A Curious Tortoise Intervention



One can see here a similarity with the definition of multidimensional single-peakedness briefly introduced in Chapter 5: if c_2 lies inside the so called *bounding box* determined by c_1 and c_3 , then c_2 is never ranked last among c_1, c_2 and c_3 .

Before ending this section, we give here one more result that will be useful in what follows: actually, note that Proposition 7.4.1 only gives the possible number of intersection points between two hypersurfaces, however, it does not specify the conditions

in which each of the cases appears. The following result (see Appendix 7.B.2 for the proof) gives a more precise statement, which will be needed in the next sections in order to compute, based on geometrical arguments, the maximal size of ℓ_1 -Euclidean profiles.

Proposition 7.4.4

Given four points c_1, c_2, c_3 and c_4 , there is at most one pair of hypersurfaces $H(c_i, c_j), H(c_k, c_l)$ (with $\{i, j, k, l\} = \{1, 2, 3, 4\}$) intersecting in two distinct points.

This is a technical (but very useful!) result, with a long, not really interesting, proof. Basically, we suppose that there are two pairs of hypersurfaces $H(c_i, c_j), H(c_k, c_l)$ and $H(c_{i'}, c_{j'}), H(c_{k'}, c_{l'})$ (with $\{i, j, k, l\} = \{i', j', k', l'\} = \{1, 2, 3, 4\}$) intersecting in two distinct points. We have hence two pairs of vertical (resp. horizontal) hypersurfaces, and one pair of vertical and one pair of horizontal hypersurfaces. We prove (by a case analysis) that none of these configurations is possible. The detailed proof can be found in Appendix 7.A.

A Curious Tortoise Intervention



What is the number of double-intersections with 5 candidates? With 6 candidates? And with m candidates?

Given the complexity of the proof with only 4 candidates, this questions seems pretty challenging - at least if we keep the same approach to answer it...

7.5 Euclidean profiles on 4 candidates in the plane

As we have seen, all the profiles with 3 candidates are ℓ_2 -Euclidean and ℓ_1 -Euclidean (see Section 7.3). We focus here on the case with 4 candidates. In Section 7.5.1, we study the maximum size of ℓ -Euclidean profiles (for $\ell = \ell_1$ and $\ell = \ell_2$). In Section 7.5.2, we provide a concise characterization of ℓ_2 -Euclidean profiles.

7.5.1 Maximum size of a Euclidean profile on 4 candidates

Bennett and Hays (1960) gave a recursive formula to compute the maximum cardinality of ℓ_2 -Euclidean profiles in \mathbb{R}^d . For $d = 2$ and 4 candidates, their formula gives the following result:

Proposition 7.5.1: (Bennett and Hays, 1960)

The maximum cardinality of a ℓ_2 -Euclidean profile on 4 candidates is 18.

We examine this question for the norm ℓ_1 , and show that the maximum cardinality is 19 (Theorem 7.5.1). The core of the proof is to show that it is at most 19 (Lemma 7.5.1): this is done by counting the (maximal) number of areas delimited by hypersurfaces. For this, we use several results of Section 7.4, as well as Euler’s formula for planar graphs. An explicit construction of a ℓ_1 -Euclidean profile with 19 preferences is then given in Lemma 7.5.2, which shows that the upper bound of Lemma 7.5.1 is tight.

Lemma 7.5.1

Any ℓ_1 -Euclidean profile on 4 candidates has at most 19 (pairwise distinct) preferences.

Sketch of proof: The proof of this lemma can be divided into several steps. A complete version of the proof is given in Appendix 7.B.

Step 1: Transforming a Euclidean geometric representation into a planar graph.

Given a geometric representation, i.e., the positions of candidates, we build a planar graph based on the hypersurfaces between pairs of candidates. For this purposes, we identify the points of intersections of hypersurfaces - this will give us the set of graph vertices. We connect then two vertices of the graph if and only if the corresponding intersections are connected by a portion of a hypersurface in the geometric representation that does not cross any other intersection. This gives us the set of edges. By construction, this graph is planar. See Figure 7.5.1 for an example of construction.

Step 2: Adaptation of the Euler’s formula: We recall the Euler’s formula for planar graphs which establishes the relation between the number of edges, vertices and areas the plane is divided into by a (planar) graph. More precisely, Euler’s formula is:

$$n_f = n_e - n_v + 2$$

with n_e the number of edges, n_v the number of vertices and n_f the number of faces of a planar graph. We then slightly modify Euler’s formula so it applies to the Euclidean representation. Actually, the (unique) “outer” area of a planar graph is divided into 12 areas by unbounded parts of hypersurfaces in the geometric representation (we recall that there are 6 hypersurfaces, each of them having two unbounded parts in the geometric representation). Therefore, the unique outer area of the planar graph must be counted for 12 geometric areas of the representation, which yields the following formula:

$$n_z = n_e - n_v + 13$$

with n_z the number of areas obtained by the representation (i.e., the number of different preferences whose corresponding area is non-empty). For better visualisation, see Figure 7.5.1. We have on the left side an ℓ_1 -Euclidean geometric representation and on the right side its associated planar graph. This planar graph has 7 vertices and 12 edges. The Euler formula gives hence $n_f = 12 - 7 + 2 = 7$. The modified Euler formula gives then $n_z = 12 - 7 + 13 = 18$ areas the plane is divided into by the ℓ_1 -Euclidean representation.

Step 3: Types of intersections we argue that for each Euclidean profile, there exists a Euclidean representation of it in which at most 3 different hypersurfaces intersect in the same point.

Step 4: Number of intersections We enumerate the maximal possible number of intersections (of two hypersurfaces, or of three hypersurfaces).

Step 5: Finally, we derive from the maximal number of intersections (i.e., vertices of the graph) the maximal number of non-empty areas, using the modified Euler’s formula. This gives us an upper bound on the maximal number of pairwise distinct preferences in a given ℓ_1 -Euclidean profile. \square

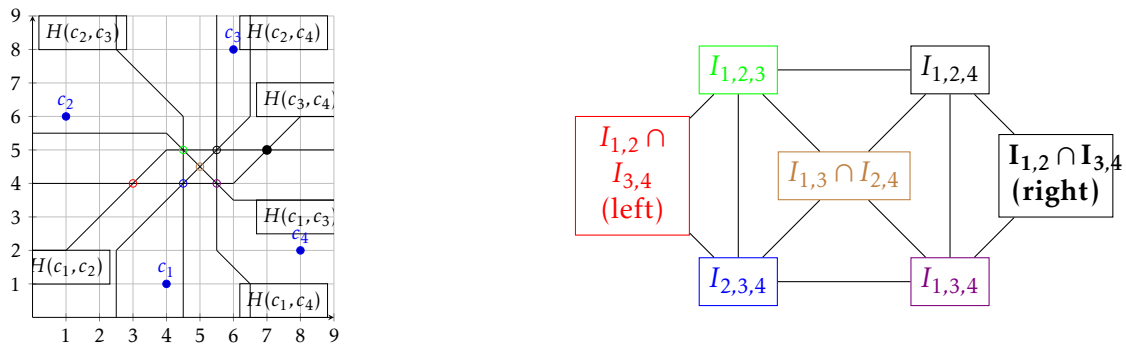


Figure 7.5.1: A ℓ_1 -Euclidean representation of a profile \mathcal{P} , and its corresponding graph. The intersection of $H(c_i, c_j)$, $H(c_i, c_k)$, $H(c_j, c_k)$ (resp. $H(c_i, c_j)$ and $H(c_k, c_l)$) yields a vertex $I_{i,j,k}$ (resp. $I_{i,j} \cap I_{k,l}$).

We will now show that this upper bound is tight - for this purpose, let us consider the following profile \mathcal{P}_0^* with 19 voters and 4 candidates (for more conciseness and readability, preferences are in columns, so for instance the first preference is (c_1, c_2, c_3, c_4)).

$$\mathcal{P}_0^* = \begin{pmatrix} c_1 & c_1 & c_1 & c_1 & c_1 & c_1 & c_2 & c_2 & c_2 & c_2 & c_3 & c_3 & c_3 & c_4 & c_4 & c_4 & c_4 & c_4 & c_4 \\ c_2 & c_2 & c_3 & c_3 & c_4 & c_4 & c_1 & c_1 & c_4 & c_4 & c_1 & c_4 & c_4 & c_1 & c_1 & c_2 & c_2 & c_3 & c_3 \\ c_3 & c_4 & c_2 & c_4 & c_2 & c_3 & c_3 & c_4 & c_1 & c_3 & c_4 & c_1 & c_2 & c_2 & c_3 & c_1 & c_3 & c_1 & c_2 \\ c_4 & c_3 & c_4 & c_2 & c_3 & c_2 & c_4 & c_3 & c_3 & c_1 & c_2 & c_2 & c_1 & c_3 & c_2 & c_3 & c_1 & c_2 & c_1 \end{pmatrix}$$

Lemma 7.5.2

\mathcal{P}_0^* is ℓ_1 -Euclidean.

Proof. Figure 7.5.2 provides a ℓ_1 -Euclidean representation of \mathcal{P}_0^* . Preference p_1 corresponds to (c_1, c_2, c_3, c_4) (the first column in \mathcal{P}_0^*), preference p_7 to (c_2, c_1, c_3, c_4) (the 7th column in \mathcal{P}_0^*) as we cross $H(c_1, c_2)$ to go from p_1 to p_7 , etc. The representation function $f : C \rightarrow \mathbb{R}^2$ leading to Figure 7.5.2 corresponds to the following positions: $f(c_1) = (0, 8), f(c_2) = (10, 10), f(c_3) = (4, 1)$ and $f(c_4) = (8, 3)$. These positions are sufficient to plot the hypersurfaces and to convince ourselves that there are 19 non-empty preference areas. For example, let us place a voter v in the area corresponding to preference p_2 , concretely on the coordinates $(5.5, 8)$. We will check that her preference is indeed p_2 . Denoting by P_v the position of voter v (i.e., the point $(5.5, 8)$), we have

$$\begin{aligned} \|P_v - f(c_1)\|_{\ell_1} &= |5.5 - 0| + |8 - 8| = 5.5, \\ \|P_v - f(c_2)\|_{\ell_1} &= |5.5 - 10| + |8 - 10| = 6.5, \\ \|P_v - f(c_3)\|_{\ell_1} &= |5.5 - 4| + |8 - 1| = 8.5, \\ \|P_v - f(c_4)\|_{\ell_1} &= |5.5 - 8| + |8 - 3| = 7.5. \end{aligned}$$

We see that, indeed, the preference of voter v corresponds to $p_2 = (c_1, c_2, c_4, c_3)$. □

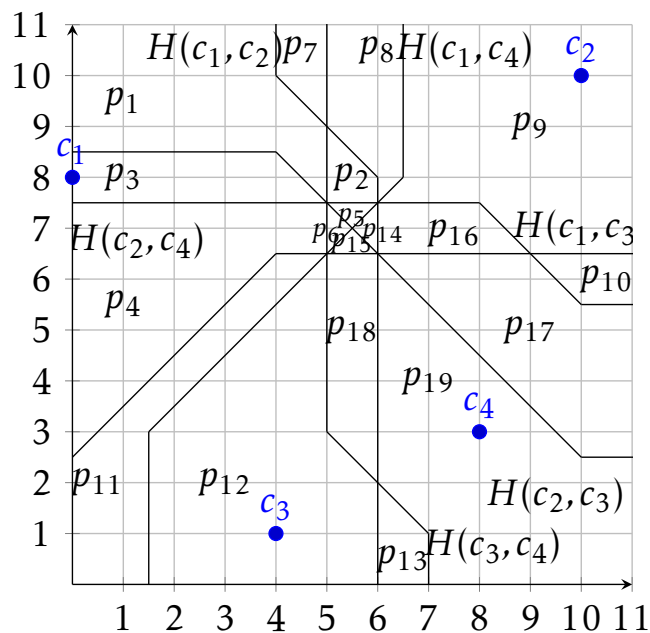


Figure 7.5.2: An ℓ_1 -Euclidean representation of a profile with 4 candidates and 19 (pairwise) distinct preferences.

As a direct consequence of Lemmata 7.5.1 and 7.5.2, we have the following result, which concludes the section.

Theorem 7.5.1

The maximum cardinality of an ℓ_1 -Euclidean profile on 4 candidates is 19.

We note that \mathcal{P}_0^* is another example of a preference profile on 4 candidates which is ℓ_1 -Euclidean but not ℓ_2 -Euclidean (because there are more than 18 preferences). Moreover, we conjecture that \mathcal{P}_0^* is the unique maximal ℓ_1 -Euclidean profile on 4 candidates in terms of inclusion:

Conjecture 7.5.1

Let \mathcal{P} be a ℓ_1 -Euclidean profile of pairwise distinct candidates on 4 candidates. Then $\mathcal{P} \subset \mathcal{P}_0^*$, up to renaming of candidates.

Even though we were not able to prove this conjecture formally, we performed some numerical tests to support (or disprove) the statement. A random ℓ_1 -Euclidean profile \mathcal{P} was generated as follows: first, we picked the positions of candidates c_1, c_2, c_3 and c_4 at random uniform. Then we used a Monte-Carlo based method to generate the ℓ_1 -Euclidean profile corresponding to these positions of candidates. More precisely, we picked 100 000 points at random, and for each point v , we evaluated its distance to the positions of c_1, c_2, c_3 and c_4 , which yielded the preference \succ_v .⁶ \succ_v were then added in \mathcal{P} (if not already in it). Finally, we checked if \mathcal{P} was contained in \mathcal{P}_0^* up to renaming of candidates.

We generated several millions of instances without finding a profile that would disprove the conjecture. A discussion about how to prove formally the conjecture is provided in the global conclusion.

7.5.2 Characterization of ℓ_2 -Euclidean profiles

A central question in structured preferences is to determine whether a given profile is structured or not. As we have seen before, with 4 candidates, any profile with more than 18 (resp. 19) preferences is not ℓ_2 -Euclidean (resp. not ℓ_1 -Euclidean). However, there are smaller profiles which are not ℓ_1 - or ℓ_2 -Euclidean (we have seen two examples of such profiles in Section 7.3).

⁶We can make a comparison with the heuristic presented in Chapter 6: to generate a random ℓ_2 -Euclidean profile, it was sufficient to generate the positions of candidates, from where we were able to enumerate all 2-intersections and 3-intersection and therefore list all preferences compatible with the representation. However, in case of ℓ_1 -Euclidean, the problem of enumerating all intersections of *hypersurfaces* is much more tricky, as the number of intersections as well as types of hypersurfaces depend on the relative position of candidates.

In the sequel, we give a concise description of all ℓ_2 -Euclidean profiles on 4 candidates, that moreover enables to easily determine whether a given profile on 4 candidates is ℓ_2 -Euclidean or not. As noted in the introduction, this result has also been proved by Kamiya et al. (2011). The interest of the proof presented below is that it is based on simpler and purely geometrical arguments. Furthermore, it might be possible to consider the same type of approach to prove a result of the same nature for $m \geq 5$ candidates *in the plane*, while the result by Kamiya et al. only applies for $d = m - 2$ (thus for 4 candidates in the plane). Moreover, this approach could also be used to prove similar results under the norm ℓ_1 .

Back to ℓ_2 -Euclidean profiles on 4 candidates, we show that the number of maximal Euclidean profiles is very small. More precisely, we prove that there are only 3 maximal ℓ_2 -Euclidean profiles \mathcal{P}_1^* , \mathcal{P}_2^* and \mathcal{P}_3^* (up to a permutation of the candidates), each of them of size 18. Thus, a profile is ℓ_2 -Euclidean if and only if it is a subprofile of \mathcal{P}_1^* , \mathcal{P}_2^* or \mathcal{P}_3^* (up to a permutation of the candidates).

We say that two profiles are *isomorphic* if they contain the same set of preferences up to a permutation of the candidates. Let us consider the three following profiles \mathcal{P}_1^* , \mathcal{P}_2^* and \mathcal{P}_3^* :

$$\mathcal{P}_1^* = \begin{pmatrix} c_1 & c_1 & c_1 & c_2 & c_2 & c_2 & c_2 & c_2 & c_2 & c_2 & c_3 & c_3 & c_3 & c_4 & c_4 & c_4 & c_4 & c_4 \\ c_2 & c_2 & c_4 & c_1 & c_1 & c_3 & c_3 & c_4 & c_4 & c_2 & c_2 & c_4 & c_1 & c_1 & c_2 & c_2 & c_3 & c_3 \\ c_3 & c_4 & c_2 & c_3 & c_4 & c_1 & c_4 & c_1 & c_3 & c_1 & c_4 & c_2 & c_2 & c_3 & c_1 & c_3 & c_1 & c_2 \\ c_4 & c_3 & c_3 & c_4 & c_3 & c_4 & c_1 & c_3 & c_1 & c_4 & c_1 & c_1 & c_3 & c_2 & c_3 & c_1 & c_2 & c_1 \end{pmatrix},$$

$$\mathcal{P}_2^* = \begin{pmatrix} c_1 & c_1 & c_1 & c_1 & c_1 & c_1 & c_2 & c_2 & c_3 & c_3 & c_3 & c_3 & c_3 & c_3 & c_4 & c_4 & c_4 & c_4 \\ c_2 & c_2 & c_3 & c_3 & c_4 & c_4 & c_1 & c_3 & c_1 & c_1 & c_2 & c_2 & c_4 & c_4 & c_1 & c_1 & c_3 & c_3 \\ c_3 & c_4 & c_2 & c_4 & c_2 & c_3 & c_3 & c_1 & c_2 & c_4 & c_1 & c_4 & c_1 & c_2 & c_2 & c_3 & c_1 & c_2 \\ c_4 & c_3 & c_4 & c_2 & c_3 & c_2 & c_4 & c_4 & c_4 & c_2 & c_4 & c_1 & c_2 & c_1 & c_3 & c_2 & c_2 & c_1 \end{pmatrix},$$

$$\mathcal{P}_3^* = \begin{pmatrix} c_1 & c_1 & c_1 & c_1 & c_2 & c_2 & c_2 & c_2 & c_2 & c_2 & c_3 & c_3 & c_3 & c_3 & c_4 & c_4 & c_4 & c_4 \\ c_2 & c_2 & c_3 & c_4 & c_1 & c_1 & c_3 & c_3 & c_4 & c_4 & c_1 & c_2 & c_2 & c_4 & c_1 & c_2 & c_2 & c_3 \\ c_3 & c_4 & c_2 & c_2 & c_3 & c_4 & c_1 & c_4 & c_1 & c_3 & c_2 & c_1 & c_4 & c_2 & c_2 & c_1 & c_3 & c_2 \\ c_4 & c_3 & c_4 & c_3 & c_4 & c_3 & c_4 & c_1 & c_3 & c_1 & c_4 & c_4 & c_1 & c_1 & c_3 & c_3 & c_1 & c_1 \end{pmatrix}.$$

Theorem 7.5.2

A profile on 4 candidates is ℓ_2 -Euclidean if and only if it is isomorphic to a subprofile of \mathcal{P}_1^* , \mathcal{P}_2^* or \mathcal{P}_3^* .

The proof of this theorem is very long, however, I decided not to move it to appendix, as it is one of my most favourite proofs of this thesis. To help the reader to follow it without getting lost, we will start by summarising the proof into several steps:

Step 1: We justify that any maximal ℓ_2 -Euclidean profile admits a representation in which there are no parallel hypersurfaces (that are simple lines while using ℓ_2 norm). In particular, each pair of hypersurfaces intersects in a unique point. We have then three different 2-intersections and four different 3-intersections.

Step 2: We focus on the relative positions of the 3-intersections. We note that there are two possible cases - the four 3-intersections form a convex quadrilateral (case 1), or there is one of the 3-intersections inside the triangle of the three others (case 2). See Figure 7.5.3 for an illustration.

We study then the case 1. This part can also be divided into several steps:

Step Case 1-1: We draw the hypersurfaces and justify that whatever the positions of candidates are, the partitionning of the plane will always look like in Figure 7.5.4.

Step Case1-2: We show that, up to renaming candidates, there are two possible labelings of hypersurfaces - see Figure 7.5.5. We justify that these two labelings are symmetric - one can be obtained from the other by renaming candidates. Therefore, one of them can be chosen without loss of generality.

Step Case1-3: Once the hypersurfaces are labeled, we list the preferences associated with the different areas, which gives us the profiles \mathcal{P}_1^* and \mathcal{P}_2^* .

Finally, we deal with the case 2 using the same techniques. We obtain the profile \mathcal{P}_3^* .

The reader can continue directly to Section 7.6, or to dive into the detailed version of proof.

Proof. **Step 1:**

For any ℓ_2 -Euclidean profile there is a representation of it such that no pair of hypersurfaces (which are simple lines of the plane in the present case) are parallel. In fact, there will be two parallel lines if there are (at least) three aligned candidates, or if two pairs of candidates are the extremities of two parallel segments. In each of these cases, we can always slightly move one of the candidates (using the same technique as in Lemma 7.2.1) so that the two concerned lines are no more parallel and such a modified mapping is still a representation of the given profile.

Assuming that, we have:

- One 2-intersection $H(c_i, c_j) \cap H(c_k, c_l)$ for each pair of hypersurfaces with i, j, k, l pairwise distinct. For 4 candidates, it yields three 2-intersections (because there

are three such pairs).

- One 3-intersection $I_{i,j,k} = H(c_i, c_j) \cap H(c_i, c_k) \cap H(c_j, c_k)$ for each triple of hypersurfaces with i, j, k pairwise distinct. For a profile on 4 candidates, it yields four 3-intersections (because there are four such triples).

A Curious Tortoise Intervention



Even though this step can seem quite trivial, it is actually crucial in the proof. We prove that, in particular, for any number of candidates m (here $m = 4$), all maximal profiles are of the same size - this size equals the maximal number of areas the plane can be partitioned into by the set of hypersurfaces (for $m = 4$, this number equals 18). Also, the number of 3-intersections and 2-intersections can be easily counted for any m .

Note that this step does not work for ℓ_1 -norm. Indeed, two ℓ_1 -hypersurfaces are “parallel” if and only if one point lies inside the diagonal rectangle determined by two other points (see Section 7.4.2, or more precisely Proposition 7.4.2, for more details). Given a representation of a profile, it is not sufficient to slightly move one of the points - we must move the point enough to get it outside of the diagonal rectangle. The distance to its new position may be too big, so the set of non-empty areas can change!

For this reason, it is not clear if all maximal ℓ_1 -Euclidean profiles are of the same size...

Step 2:

Let us study the relative positions of the 3-intersections in the plane. There are two possible scenarios (see Figure 7.5.3):

1. The 3-intersections are the vertices of a convex quadrilateral (left part of Figure 7.5.3). By the assumption made in Step 1, no pair of opposite sides of this quadrilateral are parallel.
2. Three of the 3-intersections are the vertices of a triangle, and the fourth one is inside this triangle (right part of Figure 7.5.3).

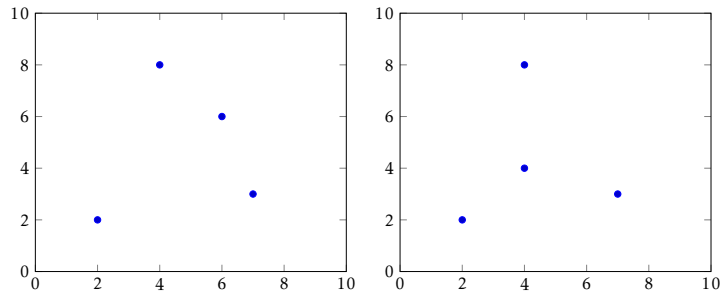


Figure 7.5.3: The possible relative positions of the 3-intersections in the plane.

We will now take a closer look to each of these cases, and we will construct all maximal profiles corresponding to each of them.

Case 1:

Step Case 1-1:

Assume that the 3-intersections are the vertices of a convex quadrilateral (see Figure 7.5.4), as described above. There are $\binom{4}{2} = 6$ hypersurfaces, each of them goes through exactly two 3-intersections (because, for an hypersurface $H(c_i, c_j)$, there are two ways to choose c_k with $k \notin \{i, j\}$). We recall that, without loss of generality, we can assume that there is no pair of parallel hypersurfaces. There are then four hypersurfaces that form the sides of the convex quadrilateral. Each of the two pairs of hypersurfaces corresponding to opposite sides of the quadrilateral results in an intersection outside the quadrilateral, which yields two distinct 2-intersections. The remaining two hypersurfaces represent the diagonals of the quadrilateral, and will hence intersect inside it - it results in the third (and last) 2-intersection. Whatever the positions of the four candidates, if the 3-intersections form a convex quadrilateral, the partitioning of the plane will always look like in Figure 7.5.4a (where $c_1 = (1, 5)$, $c_2 = (4, 2)$, $c_3 = (6, 8)$ and $c_4 = (9, 3)$): one 2-intersection lies inside the convex quadrilateral, and the two remaining 2-intersections (of hypersurfaces forming opposite sides of the quadrilateral) outside of it.

Note that in Figure 7.5.4a some areas are small. For readability reasons, in what follows, we use instead Figure 7.5.4b (with a similar arrangement of areas) where the areas are larger but without the explicit positions of candidates.

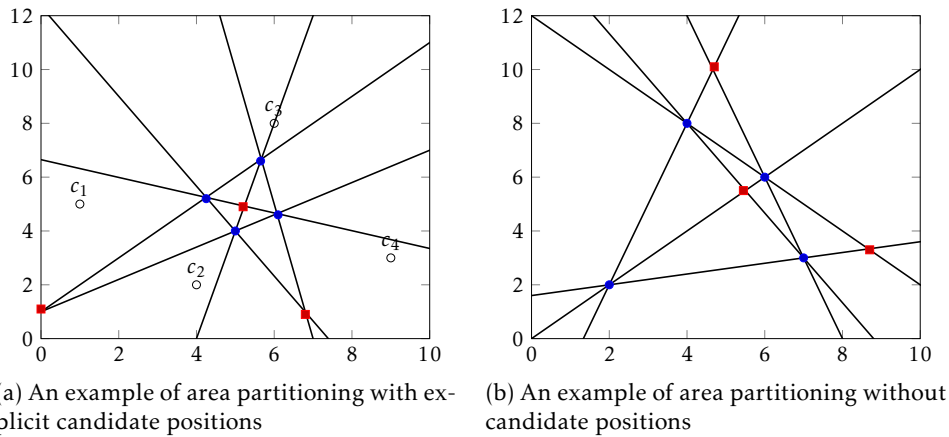


Figure 7.5.4: Case 1: The plane is divided into 18 areas, with the 3-intersections forming a convex quadrilateral. The candidates are plotted with empty circles, 3-intersections with blue circles and 2-intersections with red squares.

Step Case 1-2:

To enumerate all possible maximal profiles corresponding to this configuration of the 3-intersections, the hypersurfaces (and hence the intersections) need to be labeled so we can list the preferences corresponding to the different areas (see Figure 7.5.5). Without loss of generality, we label one of the 3-intersections as $I_{1,2,3}$ (we recall that it corresponds to the intersection $H(c_1, c_2) \cap H(c_1, c_3) \cap H(c_2, c_3)$), and one of the hypersurfaces going through it as $H(c_1, c_2)$. The second 3-intersection involving $H(c_1, c_2)$ is then necessarily $I_{1,2,4}$. The two remaining hypersurfaces going through $I_{1,2,3}$ are $H(c_1, c_3)$ and $H(c_2, c_3)$, that we can arbitrarily label (because it will turn out to be symmetrical). From these labels $I_{1,2,3}$, $I_{1,2,4}$, $H(c_1, c_2)$, $H(c_1, c_3)$ and $H(c_2, c_3)$, we can infer the labels of the two remaining 3-intersections, and so the labels of the remaining hypersurfaces. As mentioned earlier, both ways of labeling $H(c_1, c_3)$ and $H(c_2, c_3)$ are symmetric: it is sufficient to rename c_1 as c_2 and c_2 as c_1 to switch from one complete labeling to the other one (see Figure 7.5.5). Hence, the labels of $H(c_1, c_3)$ and $H(c_2, c_3)$ can be fixed without loss of generality, and there is only one possible complete labeling, up to a renaming of the candidates.

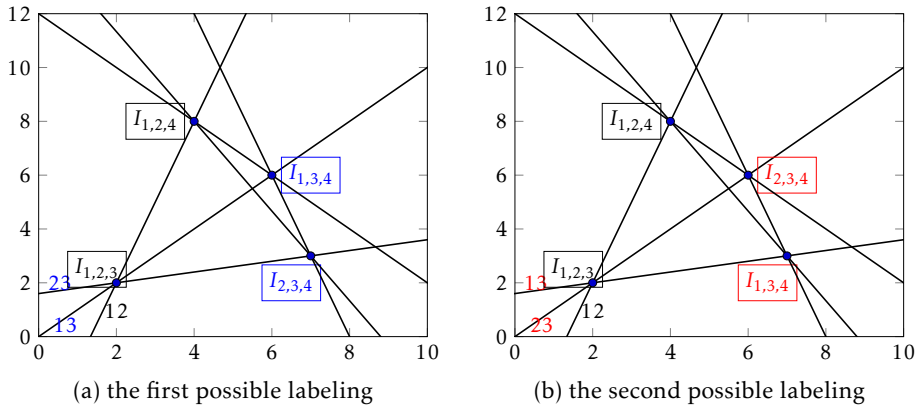


Figure 7.5.5: Labeled representation: $H(c_i, c_j)$ is noted as ij due to lack of space.

Step Case 1-3:

Once the hypersurfaces are labeled, we can list the preferences associated with the different areas. Let us focus on the areas A_1, A_2, A_3 and A_{13} , as well as on the corresponding preferences p_1, p_2, p_3 and p_{13} in Figure 7.5.6. To switch from p_1 to p_2 , candidate c_3 is swapped with c_4 (because $H(c_3, c_4)$ is crossed between areas A_1 and A_2), while c_4 is swapped with c_2 to switch from p_2 to p_3 (as $H(c_2, c_4)$ is crossed between areas A_2 and A_3), and finally c_4 is swapped with c_1 to obtain p_{13} . Necessarily, c_4 is ranked either in the first or in the last position in p_1 (resp. p_4), as it is successively swapped with all the remaining candidates. Hence, the area A_1 corresponds to one of the following preferences:

- $p_1 = (c_1, c_2, c_3, c_4)$,
- $p'_1 = (c_4, c_3, c_2, c_1)$.

Once at least one preference is known, we can list all the preferences of the profile. Both profiles $\mathcal{P} = \{p_1, \dots, p_{18}\}$ and $\mathcal{P}' = \{p'_1, \dots, p'_{18}\}$ are listed in Table 7.5.1. Profile \mathcal{P} corresponds to \mathcal{P}_1^* in the statement of the theorem, while \mathcal{P}' corresponds to \mathcal{P}_2^* .

Note that, for each $1 \leq i \leq 18$, p_i is the “opposite” of p'_i . Nevertheless, \mathcal{P}' can not be obtained from \mathcal{P} by renaming the candidates: indeed, while in \mathcal{P} candidates c_1 and c_3 are each ranked first 3 times, and c_2 and c_4 ranked first 6 times, in \mathcal{P}' in contrast, we have c_1 and c_3 that are ranked first 6 times, c_2 ranked first 2 times and c_4 ranked first 4 times.

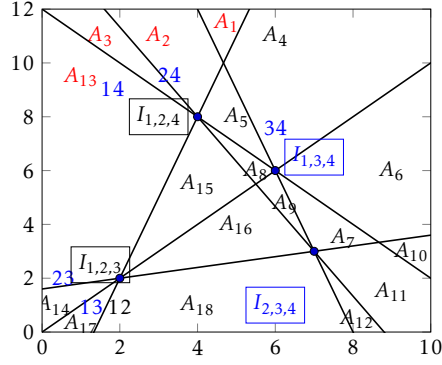


Figure 7.5.6: Listing the different areas A_1, \dots, A_{18} into which the plane is divided in case 1.

i	p_i	p'_i	i	p_i	p'_i
1	(c_1, c_2, c_3, c_4)	(c_4, c_3, c_2, c_1)	10	(c_3, c_2, c_1, c_4)	(c_4, c_1, c_2, c_3)
2	(c_1, c_2, c_4, c_3)	(c_3, c_4, c_2, c_1)	11	(c_3, c_2, c_4, c_1)	(c_1, c_4, c_2, c_3)
3	(c_1, c_4, c_2, c_3)	(c_3, c_2, c_4, c_1)	12	(c_3, c_4, c_2, c_1)	(c_1, c_2, c_4, c_3)
4	(c_2, c_1, c_3, c_4)	(c_4, c_3, c_1, c_2)	13	(c_4, c_1, c_2, c_3)	(c_3, c_2, c_1, c_4)
5	(c_2, c_1, c_4, c_3)	(c_3, c_4, c_1, c_2)	14	(c_4, c_1, c_3, c_2)	(c_2, c_3, c_1, c_4)
6	(c_2, c_3, c_1, c_4)	(c_4, c_1, c_3, c_2)	15	(c_4, c_2, c_1, c_3)	(c_3, c_1, c_2, c_4)
7	(c_2, c_3, c_4, c_1)	(c_1, c_4, c_3, c_2)	16	(c_4, c_2, c_3, c_1)	(c_1, c_3, c_2, c_4)
8	(c_2, c_4, c_1, c_3)	(c_3, c_1, c_4, c_2)	17	(c_4, c_3, c_1, c_2)	(c_2, c_1, c_3, c_4)
9	(c_2, c_4, c_3, c_1)	(c_1, c_3, c_4, c_2)	18	(c_4, c_3, c_2, c_1)	(c_1, c_2, c_3, c_4)

Table 7.5.1: The two maximal profiles $\mathcal{P} = \{p_1, \dots, p_{18}\}$ and $\mathcal{P}' = \{p'_1, \dots, p'_{18}\}$ obtained in case 1.

Case 2:

To begin, we need to label the intersections and hypersurfaces as we did it in the previous case. This yields the representation given in Figure 7.5.7.

Let us denote by T the triangle consisting of areas A_5, A_6, A_7, A_8, A_9 and A_{10} in Figure 7.5.7. Using the same argument as in the previous case, we note that there are two possible rankings for area A_1 (see Figure 7.5.7, and the succession of areas A_1, A_2, A_4 and A_{15}):

- $p_1 = (c_1, c_2, c_3, c_4)$,
- $p'_1 = (c_4, c_3, c_2, c_1)$.

However, if $p'_1 = (c_4, c_3, c_2, c_1)$, candidate c_2 is ranked in last position inside the triangle T : in fact, none of the hypersurfaces crossing the triangle involves c_2 . Let us now discuss the position of c_2 to show that p'_1 is not feasible:

- Denoting by $D(c_i, c_j)$ the set of points that are closer to c_i than to c_j , we have:

$$D(c_2, c_1) = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_{11} \cup A_{15}.$$

In fact, c_1 is preferred to c_2 in triangle T . Therefore, c_1 must lie on the same side of $H(c_1, c_2)$ as this triangle, and c_2 must then lie on the opposite side of $H(c_1, c_2)$, i.e. on the same side as the area A_1 .

- Analogously, candidate c_2 is necessarily on the same side of $H(c_2, c_3)$ as the area A_3 :

$$D(c_2, c_3) = A_3 \cup A_{11} \cup A_{12} \cup A_{13} \cup A_{14} \cup A_{18}.$$

- Finally, candidate c_2 is necessarily on the same side of $H(c_2, c_4)$ as the area A_4 :

$$D(c_2, c_4) = A_4 \cup A_{14} \cup A_{15} \cup A_{16} \cup A_{17} \cup A_{18}.$$

As $c_2 \in D(c_2, c_i)$ for each $i \in \{1, 3, 4\}$, and as $D(c_2, c_1) \cap D(c_2, c_3) \cap D(c_2, c_4) = \emptyset$, we cannot have $p'_1 = (4, 3, 2, 1)$.

The case $p_1 = (1, 2, 3, 4)$ is feasible, leading to the profile described in Table 7.5.2, which corresponds to profile \mathcal{P}_3^* in the statement of the theorem. \square

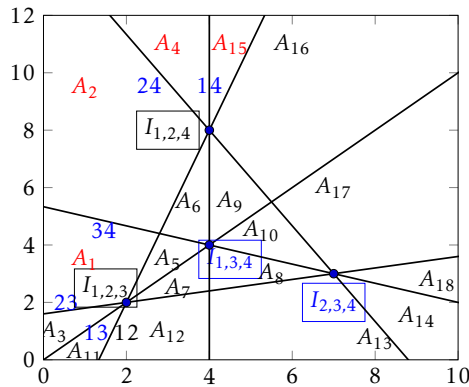


Figure 7.5.7: Listing the different areas A_1, \dots, A_{18} into which the plane is divided in case 2.

i	p_i	i	p_i
1	(c_1, c_2, c_3, c_4)	10	(c_2, c_4, c_3, c_1)
2	(c_1, c_2, c_4, c_3)	11	(c_3, c_1, c_2, c_4)
3	(c_1, c_3, c_2, c_4)	12	(c_3, c_2, c_1, c_4)
4	(c_1, c_4, c_2, c_3)	13	(c_3, c_2, c_4, c_1)
5	(c_2, c_1, c_3, c_4)	14	(c_3, c_4, c_2, c_1)
6	(c_2, c_1, c_4, c_3)	15	(c_4, c_1, c_2, c_3)
7	(c_2, c_3, c_1, c_4)	16	(c_4, c_2, c_1, c_3)
8	(c_2, c_3, c_4, c_1)	17	(c_4, c_2, c_3, c_1)
9	(c_2, c_4, c_1, c_3)	18	(c_4, c_3, c_2, c_1)

Table 7.5.2: The maximal profile obtained in case 2.

7.6 Euclidean profiles on $m \geq 5$ candidates in the plane

Let us now focus on the general case, by giving some results on the relative expressive power of ℓ_2 -Euclidean and ℓ_1 -Euclidean preference profiles. We first note that, as shown in Proposition 7.3.1, at most 4 candidates are ranked in last position (by at least one voter), regardless of the number of candidates in ℓ_1 -Euclidean profiles. This is in sharp contrast to the ℓ_2 -Euclidean case, in which profiles where each candidate is ranked last at least once can easily be built, as mentioned in the introduction.

This property might indicate that being Euclidean is much more restrictive for ℓ_1 than for ℓ_2 . We show however that if we are interested in the maximum size of a Euclidean profile, then there is no such difference. We show indeed that the maximum size of a ℓ_1 -Euclidean profile on m candidates is $\Theta(m^4)$ (Theorem 7.6.1), which is the same asymptotic bound as the one found by Bennett and Hays (1960) for ℓ_2 .

Actually, a precise formula can be easily derived from their result: this maximal size is precisely $\frac{m(3m-10)(m-1)(m+1)}{24} + m(m-1) + 1$. While such a precise formula seems to be tricky to establish for ℓ_1 and is left as an open question, we show that the asymptotical bound is the same:

Theorem 7.6.1

The maximum size of an ℓ_1 -Euclidean profile in \mathbb{R}^2 over m candidates is in $\Theta(m^4)$.

Sketch of proof. We first show that the size of such a profile is in $O(m^4)$. There are $\frac{m(m-1)}{2}$ hypersurfaces. With a non-degenerated profile, each pair of hypersurfaces intersects at most twice, hence, there are at most $2\left(\frac{m(m-1)}{2}\right)^2$ points of intersections. As in the case of norm ℓ_2 , we have (in a non-degenerated profile) at most 3 hypersurfaces intersecting in one point. If a point is at the intersection of 2 (resp 3) hypersurfaces, it is incident to 4 areas (resp. 6 areas). Then, as each area has at least one intersection point in its border,

the number of areas is upper bounded by 6 times the number of intersection points, i.e., in $O(m^4)$.

Let us now show a profile for which this bound is reached. The idea is quite straightforward: as there are only vertical or horizontal hypersurfaces, and as each vertical and each horizontal hypersurface intersect, the positions of candidates c_1 to c_m will be iteratively fixed in such a way that approximately half of hypersurfaces are vertical and half are horizontal. The number of intersections will then be in $\Theta(m^4)$, and the construction ensures that the number of areas is in $\Theta(m^4)$.

The explicit construction of this profile is deferred to Appendix 7.C.1. □

7.7 Conclusion

In this chapter, we have studied 2-Euclidean preferences w.r.t. the norm ℓ_1 and the norm ℓ_∞ . We have started the chapter by defining, in Section 7.2, boundary hypersurfaces in \mathbb{R}^2 and under the norm ℓ_1 . We have then recalled that there is an equivalence between the 2-Euclidean preferences w.r.t. the norm ℓ_1 and the 2-Euclidean preferences w.r.t. the norm ℓ_∞ . We have then shown in Section 7.3 that this equivalence does not hold anymore for d -Euclidean preferences with $d \geq 3$. More precisely, we have shown that there are at most $2d$ candidates ranked last by at least one voter in a d -Euclidean profile w.r.t. the norm ℓ_∞ , and at most 2^d candidates ranked last by at least one voter in a d -Euclidean profile w.r.t. the norm ℓ_1 .

From Section 7.4, we focused only on 2-Euclidean preferences under the norm ℓ_1 (or ℓ_∞ , as there is the equivalence between ℓ_1 -Euclidean preferences and ℓ_∞ -Euclidean preferences in dimension 2). In Section 7.4, we have studied several geometric properties of ℓ_1 -Euclidean representations. In Section 7.5, we focus on ℓ_1 -Euclidean profiles on 4 candidates. We prove that a (inclusion-wise) maximal ℓ_1 -Euclidean profile contains at most 19 (pairwise distinct) preferences, and we conjecture that there is actually a unique maximal ℓ_1 -Euclidean profile on 4 candidates. This is in contrast with the case of the norm ℓ_2 : it is well-known that there are three maximal ℓ_2 -Euclidean profiles on 4 candidates, each of them containing 18 (pairwise distinct) preferences (see Kamiya et al. (2011)). We provide an alternative proof of this results at the end of Section 7.5.

Finally, we focus on ℓ_1 -Euclidean profiles on at least 5 candidates. In particular, we show that the maximal size of an ℓ_1 -Euclidean profile is $\Theta(m^4)$, i.e., of the same order of magnitude as for ℓ_2 , as shown by Bennett and Hays (1960).

Because of their novelty, multiple avenues of research regarding ℓ_1 -Euclidean preference profiles can be considered. For instance, the conjecture that there is a unique maximal ℓ_1 -Euclidean preference profile for four candidates remains to be proved. A broader research question is to investigate the existence of a general formula giving

the maximal size of a ℓ_1 -Euclidean preference profile (as there is for ℓ_2). Regarding the computational aspects, Peters (2017) proved that the problem of recognising ℓ_1 -Euclidean preference profiles in \mathbb{R}^d is in NP, but a more specific complexity class remains to be determined, and efficient recognition procedures are still to be proposed. Also, it would be natural to extend our works on d -Euclidean preferences with respect to the norms ℓ_1 and ℓ_∞ for $d \geq 3$. These and other possible research directions are discussed in the global conclusion.

Chapter appendix

7.A Missing proofs of Section 7.4

7.A.1 Proof of Proposition 7.4.1

Proposition 7.4.1. *The intersection of two distinct ℓ_1 -hypersurfaces is either empty or contains a unique point, two distinct points or an infinite number of points.*

Proof. We can assume, without loss of generality, that the hypersurfaces are given by two distinct pairs of points. Let us denote by $c_i = (x_i, y_i), i \in \{1, 2, 3, 4\}$ these points and their coordinates. Still without loss of generality, let $H(c_1, c_2)$ be of type V^- . There are four basic cases to distinguish (see Figs 7.A.1–7.A.4 for illustrations):

1. $H(c_1, c_2)$ is of type V^- and $H(c_3, c_4)$ of type H^+ (see Figure 7.A.1):
In this case, the hypersurfaces intersect in a unique point as the half-lines (resp. the middle segments) of $H(c_1, c_2)$ and $H(c_3, c_4)$ are of opposite orientations.
2. $H(c_1, c_2)$ is of type V^- and $H(c_3, c_4)$ of type H^- (see Figure 7.A.2):
As in the previous case, there will be at least one intersection as a horizontal hypersurface and a vertical hypersurface always intersect. Contrary to the above, the middle segments of both hypersurfaces have the same orientation, so they can overlap: in such a case, the intersection contains this overlapping segment, thus an infinite number of points.
3. $H(c_1, c_2)$ is of type V^- and $H(c_3, c_4)$ of type V^+ (see Figure 7.A.3):
In this case, the hypersurfaces may not intersect: let us assume that $\max\{x_1, x_2\} < \min\{x_3, x_4\}$. By definition, we have $x \in [\min\{x_i, x_j\}, \max\{x_i, x_j\}]$ for each $(x, y) \in H(c_i, c_j)$. The above inequality then implies that the intersection of $H(c_1, c_2)$ and $H(c_3, c_4)$ is empty (graphically, $H(c_1, c_2)$ will be “on the left” of $H(c_3, c_4)$ - see the first case of Figure 7.A.3).
The hypersurfaces may also intersect in a unique point: a middle segment of one of the hypersurfaces can intersect one of the half-lines of the second hypersurface, or its middle segment, as the middle segments are not of the same type (see the second case of Figure 7.A.3).
Finally, as the half-lines of both hypersurfaces are of the same type, one of the half-lines of $H(c_1, c_2)$ may (partially) overlap one of the half-lines of $H(c_3, c_4)$ (see

the third case of Figure 7.A.3). In this case, the intersection will contain an infinity of points.

4. Both hypersurfaces $H(c_1, c_2)$ and $H(c_3, c_4)$ are of type V^- (see Figure 7.A.4): This is the most complex case. For the same reason as above, the hypersurfaces may not intersect. They may also intersect in a unique point if the middle segment of one hypersurface intersects one of the half-lines of the second one (see the first case of the Figure 7.A.4). As the types of half-lines and middle segments are both the same for $H(c_1, c_2)$ and $H(c_3, c_4)$, they can also intersect in two distinct points if the middle segment of $H(c_1, c_2)$ intersects one of the half-lines of $H(c_3, c_4)$ and the middle segment of $H(c_3, c_4)$ intersects one of the half-lines of $H(c_1, c_2)$ (see the second case of Figure 7.A.4). Finally, the intersection can contain an infinity of points: as the half-lines are of the same type, a half-line of $H(c_1, c_2)$ may (partially) overlap a half-line of $H(c_3, c_4)$. In addition, the middle segments being also of the same type, they can (partially) overlap. See cases 3 and 4 of Figure 7.A.4 (the case when both the half-lines and the middle segments overlap is not presented in Figure 7.A.4, but it is obviously possible).

□

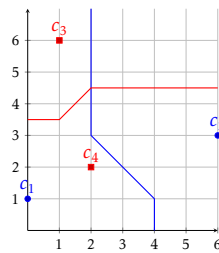


Figure 7.A.1: Intersection of two hypersurfaces: $H(c_1, c_2)$ is of type V^- and $H(c_3, c_4)$ of type H^+ .

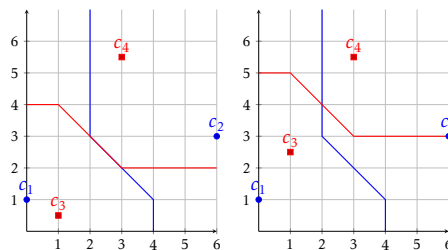


Figure 7.A.2: Intersection of two hypersurfaces: $H(c_1, c_2)$ is of type V^- and $H(c_3, c_4)$ of type H^- .

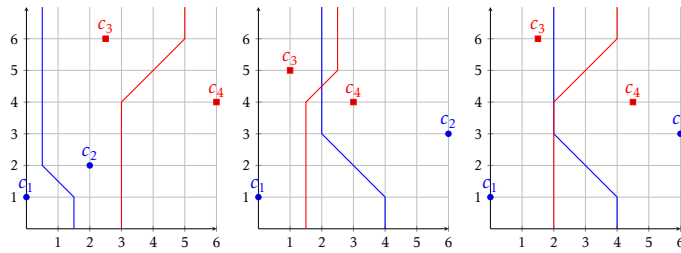


Figure 7.A.3: Intersection of two hypersurfaces: $H(c_1, c_2)$ is of type V^- and $H(c_3, c_4)$ of type V^+ .

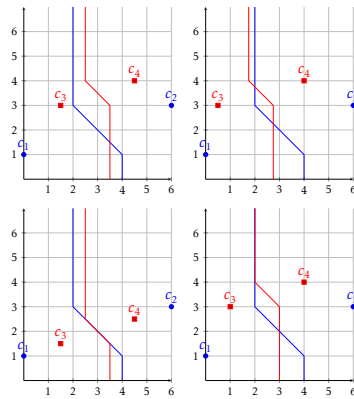


Figure 7.A.4: Intersection of two hypersurfaces: both hypersurfaces $H(c_1, c_2)$ and $H(c_3, c_4)$ are of type V^- .

7.A.2 Missing proof of Property 7.4.2

Proposition 7.4.2. Given three points $c_1 = (x_1, y_1)$, $c_2 = (x_2, y_2)$ and $c_3 = (x_3, y_3)$:

- If c_1 , c_2 or c_3 is inside the diagonal rectangle determined by the two other points, then $H(c_1, c_2)$, $H(c_1, c_3)$ and $H(c_2, c_3)$ do not (pairwise) intersect.
- Otherwise, the intersection of the three hypersurfaces is a unique point.

Proof of Proposition 7.4.2. To prove the first point, we assume without loss of generality that c_2 lies inside the diagonal rectangle determined by c_1 and c_3 , and we prove that in this case, the three hypersurfaces are all vertical or horizontal - Proposition 7.4.3 then implies that they do not (pairwise) intersect.

Up to exchanging the roles of c_1 and c_3 , we can assume, still without loss of generality, that $x_1 < x_3$. There are then 4 cases to distinguish (see Figure 7.A.5):

- $H(c_1, c_3)$ is of type V^- (see Figure 7.A.5a).
In this case, we have $y_1 < y_3$ (see the classification of hypersurfaces given in Figure 7.2.5). Moreover, we have $x_1 < x_2 < x_3$. In the diagonal rectangle given in

Figure 7.A.5a, c_2 lies then above the diagonal d_1^- (i.e, $y_2 > -x_2 + x_1 + y_1$) and below the diagonal d_1^+ (i.e, $y_2 < x_2 - x_1 + y_1$). Put together, we get

$$-x_2 + x_1 < y_2 - y_1 < x_2 - x_1.$$

In other words, $|y_1 - y_2| < x_2 - x_1$, hence $H(c_1, c_2)$ is vertical. We show similarly that $H(c_2, c_3)$ is vertical, as c_2 lies above the diagonal d_3^+ and below the diagonal d_3^- . All three hypersurfaces being vertical, they do not (pairwise) intersect.

- (b) Let us now suppose that $H(c_1, c_3)$ is of type H^- (see Figure 7.A.5b). We have $y_1 < y_2 < y_3$. As c_2 lies above diagonals d_1^- and d_1^+ , we have $y_2 > -x_2 + x_1 + y_1$ and $y_2 > x_2 - x_1 + y_1$. Put together, we have $y_2 - y_1 > x_1 - x_2$ and $y_2 - y_1 > -(x_1 - x_2)$ - in other words, $y_2 - y_1 > |x_1 - x_2|$. Therefore, the hypersurface $H(c_1, c_2)$ is horizontal. We show similarly that $H(c_2, c_3)$ is horizontal, as c_2 lies below diagonals d_3^+ and d_3^- , so we obtain $y_3 - y_2 > |x_2 - x_3|$.
- (c) We suppose here that $H(c_1, c_3)$ is of type V^+ . We have $y_1 > y_3$ and $x_1 < x_2 < x_3$. Analogously to the previous case, we show that $x_2 - x_1 > |y_2 - y_1|$, so $H(c_1, c_2)$ is vertical, and that $x_3 - x_2 > |y_2 - y_3|$, which implies that $H(c_2, c_3)$ is also vertical.
- (d) Finally, we consider $H(c_1, c_3)$ of type H^+ . We have $y_1 > y_2 > y_3$. As in previous cases, we show that $H(c_1, c_2)$ is horizontal as c_2 lies below diagonals d_1^+ and d_1^- , and $H(c_2, c_3)$ is also horizontal as c_2 lies above diagonals d_3^+ and d_3^- .

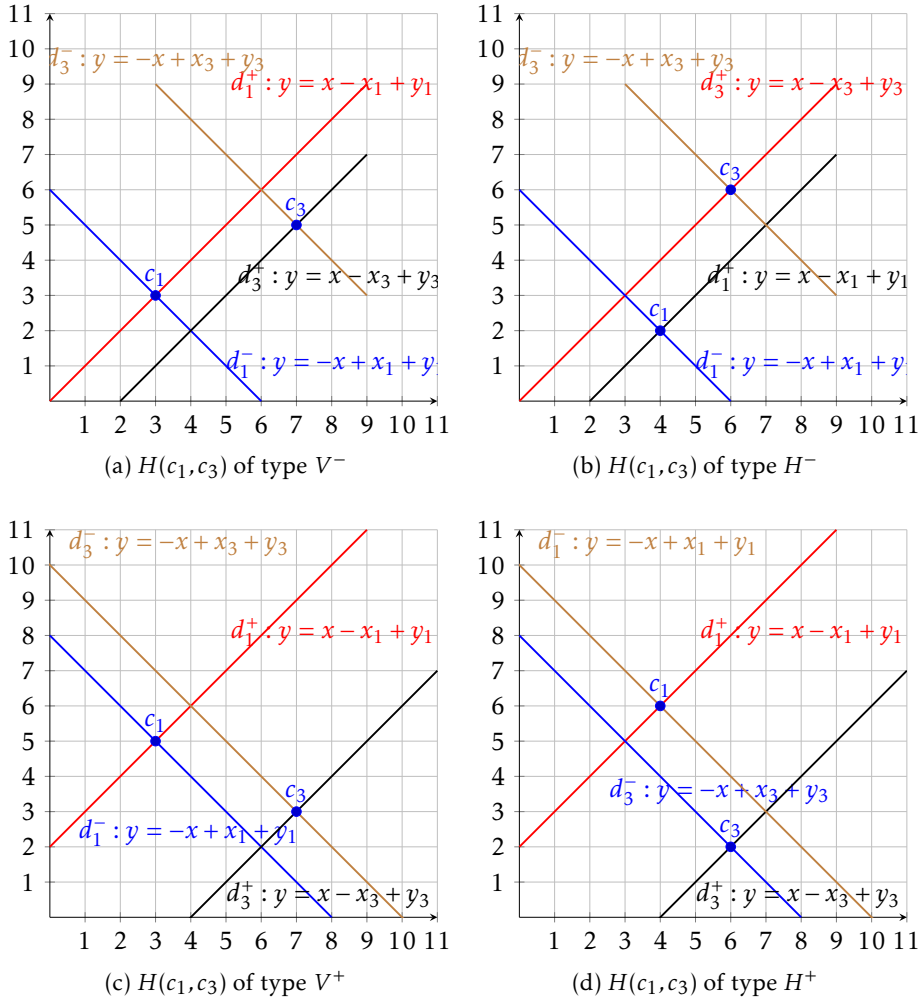


Figure 7.A.5: The diagonal rectangle determined by c_1 and c_3 with $x_1 < x_3$: 4 cases to distinguish.

To prove the second point of the proposition, we suppose that any point does not lie in the diagonal rectangle determined by the remaining two points, and we will show that in such a case, there is at least one horizontal and one vertical hypersurface. As a vertical hypersurface and a horizontal hypersurface intersect in a unique point, Lemma 7.4.2 allows us to conclude that the three hypersurfaces intersect in a unique point.

Suppose first that $H(c_1, c_3)$ is vertical (see Figure 7.A.6a). The diagonals d_1^+, d_1^-, d_3^+ and d_3^- divide the plane into 9 areas A_1, A_2, \dots, A_9 . The point c_2 does not lie in A_1 (resp. A_5, A_9) because c_1 (resp. c_2, c_3) does not lie in the diagonal rectangle determined by the remaining two points. If c_2 lies in area A_2, A_3, A_4 or A_7 , the hypersurface $H(c_1, c_2)$ is horizontal, so it intersects the vertical hypersurface $H(c_1, c_3)$. If c_2 lies in

one of the remaining areas A_6 or A_8 , the hypersurface $H(c_2, c_3)$ is horizontal, so it intersects $H(c_1, c_3)$. We note that whether the oblique middle-segment of the hypersurface $H(c_1, c_3)$ is ascending (case (c) of Figure 7.A.5) or descending (case (a)) has no impact on this reasoning and it can therefore be used without change for both cases (a) and (c) of Figure 7.A.5.

Analogously, we treat the case in which $H(c_1, c_3)$ is horizontal: c_2 cannot lie in areas A_3 , A_5 and A_7 as any point does not lie within the diagonal rectangle determined by the remaining two points. If c_2 lies in A_1 , A_2 , A_6 or A_9 , the hypersurface $H(c_1, c_2)$ is vertical. If it lies in one of the two remaining areas A_4 or A_8 , the hypersurface $H(c_2, c_3)$ is vertical. To sum up, there is always at least one horizontal and one vertical hypersurface.

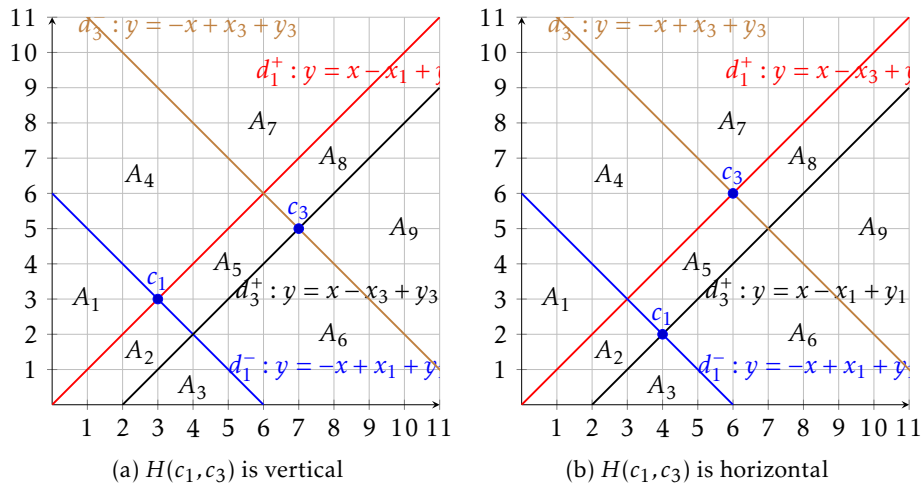


Figure 7.A.6: The diagonal rectangle determined by c_1 and c_3 with $x_1 < x_3$, and the possible placements of c_2 .

□

7.B Missing proofs of Section 7.5

7.B.1 Detailed proof of Lemma 7.5.1

Proof. Step 1: Transforming an ℓ_1 -Euclidean geometric representation into a planar graph:

To prove this proposition, given an ℓ_1 -Euclidean representation of a profile \mathcal{P} , we define a graph whose vertices are all hypersurface intersections, and where there is an edge between two intersections (denoted by I_1 and I_2) if and only if both I_1 and I_2 lie on the same hypersurface, and there is any other intersection on the segment of extremities I_1 and I_2 (see Figure 7.B.1 for an illustration).

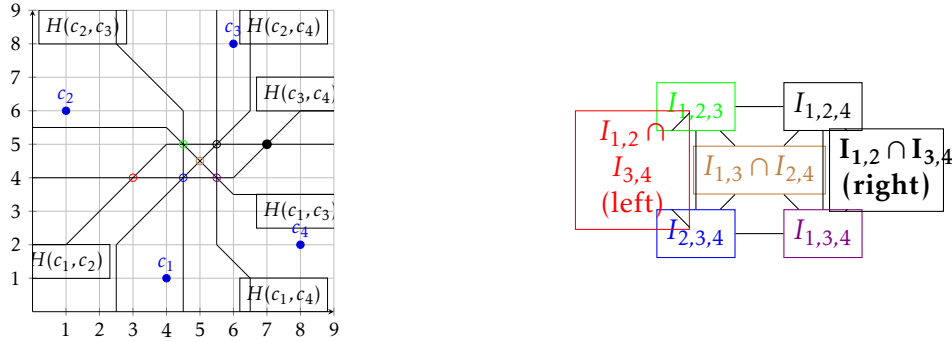


Figure 7.B.1: A ℓ_1 -Euclidean representation of a profile \mathcal{P} , and its corresponding graph. The intersection of $H(c_i, c_j)$, $H(c_i, c_k)$, $H(c_j, c_k)$ (resp. $H(c_i, c_j)$ and $H(c_k, c_l)$) yields a vertex $I_{i,j,k}$ (resp. $I_{i,j} \cap I_{k,l}$).

The corresponding graph is by construction planar. We note that each inner face of the graph corresponds to a bounded area in the representation of the profile, while unbounded areas in the representation of the preference profile are all merged into the outer face of the planar graph.

Step 2: Adaptation of the Euler’s formula:

We can then use Euler’s formula in the corresponding graph. It states that the number of faces of a planar graph is $n_f = n_e - n_v + 2$, where n_e is the number of edges and n_v the number of vertices.

Let us denote by n_z the number of areas in the ℓ_1 -Euclidean representation of the profile. Note that each area corresponds to a single preference, so $n \leq n_z$. For 4 candidates, there are 6 hypersurfaces, leading to at most 12 unbounded areas. As mentioned above, these 12 unbounded areas are merged into the outer face of the planar graph. As the bounded areas yield $n_f - 1$ inner faces, we have $n_z \leq n_f - 1 + 12 = n_f + 11$, and therefore (by Euler’s formula):

$$n_z \leq n_e - n_v + 13 \tag{7.3}$$

Types of intersections:

If k different hypersurfaces intersect in a common point, we call this point a k -intersection. We can assume, without loss of generality, that there are only 2-intersections and 3-intersections:

Let f be a representation of a given ℓ_1 -Euclidean profile containing a 4-intersection I . As 3 points give only 3 different hypersurfaces, the 4 hypersurfaces intersecting in I involve the four points $f(c_1), f(c_2), f(c_3), f(c_4)$ corresponding to the positions of the four candidates c_1, c_2, c_3, c_4 . By definition, I is equidistant from all candidates - more formally, we have $\|f(c) - I\|_{\ell_1} = \delta > 0$ for each $c \in \{c_1, c_2, c_3, c_4\}$. As in the Lemma 7.2.1,

we define:

$$\varepsilon_d = \min_{v \in V} \min_{c_i, c_j \in C} \left| \|f(v) - f(c_i)\|_{\ell_1} - \|f(v) - f(c_j)\|_{\ell_1} \right|.$$

As we consider only strict preferences, $\varepsilon_d > 0$. We can then add $\varepsilon = \frac{\varepsilon_d}{2}$ to x_1 .⁷ Doing that, I will be no more equidistant from all four points and therefore, there will no more be a 4-intersection in such a modified representation. By iterating the process, all k -intersections can be excluded for any $k \geq 4$.

Step 4: Number of intersections

As there are 4 candidates, there are at most four 3-intersections:

- $I_{123} = H(c_1, c_2) \cap H(c_1, c_3) \cap H(c_2, c_3)$,
- $I_{124} = H(c_1, c_2) \cap H(c_1, c_4) \cap H(c_2, c_4)$,
- $I_{134} = H(c_1, c_3) \cap H(c_1, c_4) \cap H(c_3, c_4)$,
- $I_{234} = H(c_2, c_3) \cap H(c_2, c_4) \cap H(c_3, c_4)$.

By Lemma 7.4.2, we have covered all intersections of type $H(c_i, c_j) \cap H(c_i, c_k)$. That means, all 2-intersections will be of type $H(c_i, c_j) \cap H(c_k, c_l)$ with i, j, k, l pairwise distinct. There are 3 pairs of hypersurfaces of this type:

- $H(c_1, c_2) \cap H(c_3, c_4)$,
- $H(c_1, c_3) \cap H(c_2, c_4)$,
- $H(c_1, c_4) \cap H(c_2, c_3)$.

Each of these three pairs can give us one 2-intersection. In addition, Proposition 7.4.4 implies that at most one of these pairs of hypersurfaces can intersect twice. To sum up, we have at most four 2-intersections. Therefore $n_v \leq 8$ (at most four 3-intersections and four 2-intersections).

Step 5:

If $n_v = 8$, there are four 2-intersections and four 3-intersections. Each 2-intersection generates four outgoing half-lines, and each 3-intersection generates six outgoing half-lines. We then get $4 \cdot 4 + 4 \cdot 6 = 40$ outgoing half-lines. However, 12 of them are delimiting outer non-bounded areas, so they are not responsible for any graph edge. Therefore, $40 - 12 = 28$ half-lines are left for forming edges. We observe that each of these half-lines is used in the creation of exactly one edge, and that each edge is a segment corresponding to the common part of exactly two half-lines (as each edge has two extremities

⁷More precisely, ε should be smaller than the minimum of ε_d and $\min\{\varepsilon_x, \varepsilon_y, \varepsilon_{xy}\}$ as defined in Lemma 7.2.1, to ensure that we do not create one of the degenerations excluded by this Lemma.

which are two different intersections). Thus, we have $n_e = 28/2 = 14$. Finally, using Equation 7.3:

$$n_z \leq 14 - 8 + 13 = 19.$$

It is easy to check that if $n_v < 8$, then $n_z < 19$: in fact, each 2-intersection (resp. 3-intersection) generate four (resp. six) outgoing half-lines. In both cases at most a half of them are delimiting outer non-bounded areas - which means that at least half of them has another 2-intersection or 3-intersection lying on it. Therefore, each vertex allows to create at least two edges, so in the Euler formula the benefit of deleting a vertex is outweighed by the drawback of deleting two edges. Thus, in any case, $n_z \leq 19$. The size n of the profile therefore satisfies $n \leq n_z \leq 19$. \square

7.B.2 Missing proof of Proposition 7.4.4

Proof. For $i \in \{1, 2, 3, 4\}$, we denote by (x_i, y_i) the position of candidate c_i in the plane. Assume there are two pairs of hypersurfaces intersecting in two distinct points. Thanks to Corollary 7.4.1, we can assume, w.l.o.g., that the first pair involves the hypersurfaces $H(c_1, c_2)$ and $H(c_3, c_4)$. Moreover, still w.l.o.g, we can assume that they are both of type V^- , and that $H(c_1, c_2)$ is “on the left” of $H(c_3, c_4)$ (as in Figure 7.B.2), and that $x_1 < x_2$ and $x_3 < x_4$. According to the classification of hypersurfaces (see Figure 7.2.5, page 198), as $H(c_1, c_2)$ and $H(c_3, c_4)$ are of type V^- , we have $y_1 < y_2$ and $y_3 < y_4$. Note that we necessarily have:

$$\{x_1, x_3\} < \{x_2, x_4\} \tag{7.4}$$

and

$$y_2 > \{y_1, y_4\} > y_3. \tag{7.5}$$

Equation (7.4) directly follows from the fact that for each point (x, y) of a hypersurface $H(c_i, c_j)$, we have $x \in [x_i, x_j]$: indeed, if $x_3 > x_2$, we would have $x_1 < x_2 < x_3 < x_4$ and the x-coordinate of each point of $H(c_1, c_2)$ would be smaller than the x-coordinate of each point of $H(c_3, c_4)$. In other words, the hypersurfaces would not intersect. An analogous reasoning can be done to show that $x_1 < x_4$. Equation (7.5) follows from the fact that the hypersurfaces do not even intersect if these inequalities are not satisfied.

Furthermore, Equation (7.5) means that when two vertical hypersurfaces $H(c_1, c_2)$ and $H(c_3, c_4)$ intersect twice, if the highest point in $\{c_1, \dots, c_4\}$ belongs to $\{c_1, c_2\}$ (resp. $\{c_3, c_4\}$) then the lowest point belongs to $\{c_3, c_4\}$ (resp. $\{c_1, c_2\}$).

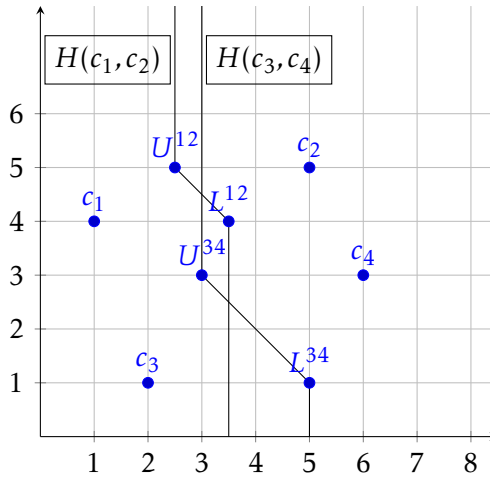


Figure 7.B.2: Two intersecting hypersurfaces.

Assume first that the second pair of hypersurfaces that intersect twice are also vertical. As $y_3 = \min\{y_1, y_2, y_3, y_4\}$ and $y_2 = \max\{y_1, y_2, y_3, y_4\}$, by the discussion above $H(c_2, c_3)$ and $H(c_1, c_4)$ cannot intersect twice vertically. So the unique possibility is that $H(c_1, c_3)$ and $H(c_2, c_4)$ intersect twice. This is, however, not possible: any point in $H(c_1, c_3)$ has x -coordinate in $[x_1, x_3]$, any point in $H(c_2, c_4)$ has x -coordinate in $[x_2, x_4]$, but $[x_1, x_3] \cap [x_2, x_4] = \emptyset$ by Equation (7.4).

Suppose now that the second pair of hypersurfaces that intersect twice are horizontal. This pair can be either $H(c_1, c_4)$ and $H(c_2, c_3)$, or $H(c_1, c_3)$ and $H(c_2, c_4)$.

- Let us first consider the case where it is $H(c_1, c_4)$ and $H(c_2, c_3)$, which is illustrated in Figure 7.B.3. Let us look at the preference p_1 in the upper left part. We have $c_1 > c_2$ and $c_3 > c_4$ (by the positions of $H(c_1, c_2)$ and $H(c_3, c_4)$). As $H(c_2, c_3)$ is horizontal, we have $c_2 > c_3$ (because $y_2 > y_3$ from Equation (7.5), see also Figure 7.B.2). Therefore $p_1 = (c_1, c_2, c_3, c_4)$. As $H(c_1, c_4)$ and $H(c_2, c_3)$ are horizontal, $H(c_2, c_3)$ is necessarily *above* $H(c_1, c_4)$ on the (infinite) left part of the figure, since starting from $p_1 = (c_1, c_2, c_3, c_4)$ and going down we need to cross $H(c_2, c_3)$ before $H(c_1, c_4)$: in fact, going down from the area corresponding to p_1 , we will not cross nor $H(c_1, c_2)$ neither $H(c_3, c_4)$ as they are vertical. We can only cross the remaining hypersurfaces $H(c_1, c_3)$, $H(c_1, c_4)$, $H(c_2, c_3)$ and $H(c_2, c_4)$. However, the hypersurface $H(c_1, c_3)$ (resp. $H(c_1, c_4)$, $H(c_2, c_4)$) cannot be the first hypersurface to be crossed, as c_2 is ranked between c_1 and c_3 (resp. c_2 and c_3 are ranked between c_1 and c_4 , c_3 between c_2 and c_4). Therefore, the first hypersurface to be crossed is necessarily $H(c_2, c_3)$ (c_2 and c_3 are ranked one next to other in p_1 , so they can be swapped).

Similarly, we get $p_2 = (c_4, c_3, c_2, c_1)$ in the lower right part, thus $H(c_2, c_3)$ is *below* $H(c_1, c_4)$ on the (infinite) right part of the figure, since starting from $p_2 =$

(c_4, c_3, c_2, c_1) and going up we need to cross $H(c_2, c_3)$ before $H(c_1, c_4)$, using the same reasoning as in the case of p_1 .

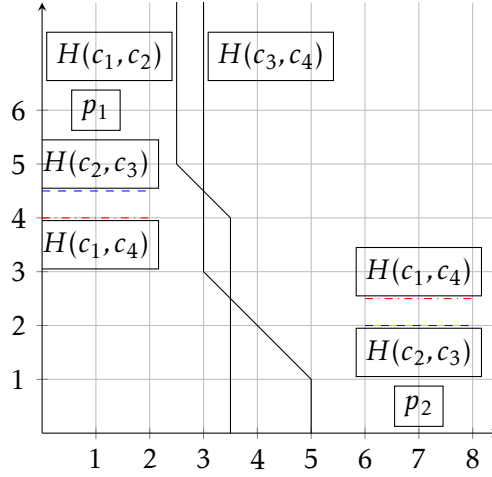


Figure 7.B.3: Relative positions of $H(c_1, c_4)$ and $H(c_2, c_3)$.

Hence, $H(c_2, c_3)$ and $H(c_1, c_4)$ cannot intersect twice (otherwise the same hypersurface would be above the other one both on the left part and the right part of the figure).

- Let us finally focus on the case where $H(c_1, c_3)$ and $H(c_2, c_4)$ are horizontal and intersect twice. They are necessarily both of type H^+ or both of type H^- (otherwise, they cannot intersect in two different points). We will show that none of these two cases is possible - in other words, that we cannot have $H(c_1, c_3)$ and $H(c_2, c_4)$ intersecting twice.

1. Firstly, let us assume that $H(c_1, c_3)$ and $H(c_2, c_4)$ are of type H^- :

We recall that $y_2 > \{y_1, y_4\} > y_3$ and $\{x_1, x_3\} < \{x_2, x_4\}$ (see Equations 7.5 and 7.4). However, the information on the type of $H(c_1, c_3)$ and $H(c_2, c_4)$ allows us to complete these partial orders on the coordinates of the candidates: using the classification of hypersurfaces (see Figure 7.2.5), we must have $x_3 < x_1$ and $x_4 < x_2$ if the hypersurfaces are of type H^- . Put together, we have $x_3 < x_1 < x_4 < x_2$. Moreover, the necessary order on y -coordinates is $y_2 > y_1 > y_4 > y_3$ - if the order would be $y_2 > y_4 > y_1 > y_3$, $H(c_1, c_3)$ and $H(c_2, c_4)$ would not intersect as any point of $H(c_i, c_j)$ has its y -coordinate in $[x_i, x_j]$. Let us denote by $U^{ij} = (U_x^{ij}, U_y^{ij})$ (resp. $L^{ij} = (L_x^{ij}, L_y^{ij})$) the upper extreme point (resp. the lower extreme point) of the middle-segment of $H(c_i, c_j)$. As $H(c_1, c_2)$ and $H(c_3, c_4)$ are of type V^- and intersect twice, the given orders on both x -coordinates and y -coordinates of candidates implies that $U_x^{12} < U_x^{34}$ and $L_x^{12} < L_x^{34}$ (see Figure 7.B.2). When we express the

segment extremities positions using the candidates coordinates, these two inequalities rewrite, after simplifying, as follows:

$$\begin{aligned} x_2 + x_1 + y_1 - y_2 &< x_4 + x_3 + y_3 - y_4 \\ x_1 + x_2 + y_2 - y_1 &< x_3 + x_4 + y_4 - y_3 \end{aligned}$$

If we sum the both inequalities, we get:

$$2(x_4 + x_3) > 2(x_1 + x_2),$$

in other words,

$$x_4 - x_1 > x_2 - x_3.$$

But this is in contradiction with the order on x -coordinates which states that $x_3 < x_1 < x_4 < x_2$. Therefore, the hypersurfaces $H(c_1, c_3)$ and $H(c_2, c_4)$ cannot be of type H^- .

2. Let us now assume that $H(c_1, c_3)$ and $H(c_2, c_4)$ are of type H^+ :

This case is illustrated in Figure 7.B.4. Note that the “upper” horizontal part of $H(c_1, c_3)$ (starting at the x -position x_3) is below the line (c_1, L^{12}) , as $y_3 < y_1$. Similarly, the lower horizontal part of $H(c_2, c_4)$ (ending at the x -position x_2) is above the line (U^{34}, c_4) as $y_2 > y_4$. If $H(c_1, c_3)$ and $H(c_2, c_4)$ intersect, then $H(c_1, c_3)$ is above $H(c_2, c_4)$ in the central part, see Figure 7.B.4.

Then in the (non empty) rectangle delimited by the 4 hypersurfaces (in the center of Figure 7.B.4), we have: $c_1 > c_2$, $c_4 > c_3$, $c_2 > c_4$ and $c_3 > c_1$, which yields $c_1 > c_2 > c_4 > c_3 > c_1$, a contradiction.

To conclude, we have proved by contradiction that $H(c_1, c_3)$ and $H(c_2, c_4)$ can neither be both of type H^+ nor both of type H^- . Therefore, they cannot intersect twice.

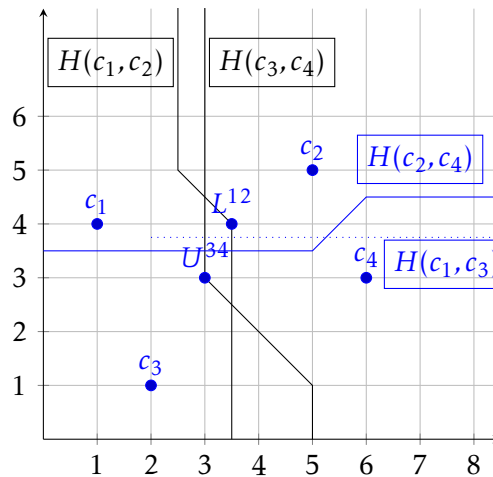


Figure 7.B.4: The case where $H(c_1, c_3)$ and $H(c_2, c_4)$ are horizontal and intersect twice.

□

7.C Missing proofs of Section 7.6

7.C.1 Missing part in the proof of Theorem 7.6.1

Explicit construction of a family of profiles with $\Theta(m^4)$ distinct preferences

Proof. We set $c_1 = (0, 0)$ and $c_2 = (1, 2)$. According to the classification of hypersurfaces, $H(c_1, c_2)$ is horizontal (more precisely of type H^-). We then place c_3 in such a way that both $H(c_1, c_3)$ and $H(c_2, c_3)$ are vertical. To do so, we need to fix the values of x_3 and y_3 (coordinates of c_3) such that:

$$\begin{aligned} |x_1 - x_3| &> |y_1 - y_3| \\ \text{and } |x_2 - x_3| &> |y_2 - y_3|. \end{aligned}$$

This can be done by setting, for instance,

$$y_3 = \frac{y_1 + y_2}{2}$$

and

$$x_3 = \max\{x_1, x_2\} + 2|y_1 - y_2|$$

We check that, indeed, for $i \in \{1, 2\}$, we have

$$\begin{aligned} |x_3 - x_i| &= |\max\{x_1, x_2\} - x_i + 2|y_1 - y_2|| \\ &\geq 2|y_1 - y_2| > \frac{1}{2}|y_1 - y_2| \geq \left| \frac{y_1 + y_2}{2} - y_i \right| = |y_3 - y_i| \end{aligned}$$

where the strict inequality follows from the fact that $y_1 \neq y_2$. Geometrically, choosing y_3 between y_1 and y_2 ensures that $|y_3 - y_i|$ is upper bounded by $|y_1 - y_2|$. To guarantee that $|x_i - x_3| > |y_i - y_3|$, it is then sufficient that x_3 is taken large enough - here, the distance from $\max\{x_1, x_2\}$ (and so in particular from both x_1 and x_2) to x_3 is greater than the above mentioned upper bound $|y_1 - y_2|$.

We will now generalize the idea: we want $H(c_{2k}, c_i)$ to be horizontal for all $k \geq 1, i < 2k$, and $H(c_{2k+1}, c_i)$ to be vertical for $k \geq 1, i < 2k+1$. Let us detail only the case of horizontal hypersurfaces (the case of vertical ones being symmetric).

We set

$$x_{2k} = \frac{\max_{i < 2k} \{x_i\} + \min_{i < 2k} \{x_i\}}{2}$$

and

$$y_{2k} = \max_{i < 2k} \{y_i\} + 2 \left(\max_{i < 2k} \{x_i\} - \min_{i < 2k} \{x_i\} \right).$$

The geometrical intuition remains the same - as we need, for all $i < 2k$, $|x_{2k} - x_i| < |y_{2k} - y_i|$, we choose the value of x_{2k} so that $|x_{2k} - x_i|$ is upper bounded by $\max_{i < 2k} \{x_i\} - \min_{i < 2k} \{x_i\}$,

and we chose then y_{2k} in such a way that $|y_{2k} - y_i|$ is greater than this upper bound. Formally, we have:

$$\begin{aligned}
 |y_{2k} - y_i| &= \left| \max_{i < 2k} \{y_i\} + 2 \left(\max_{i < 2k} \{x_i\} - \min_{i < 2k} \{x_i\} \right) - y_i \right| \\
 &= \left| \max_{i < 2k} \{y_i\} - y_i + 2 \left(\max_{i < 2k} \{x_i\} - \min_{i < 2k} \{x_i\} \right) \right| \\
 &\geq 2 \left(\max_{i < 2k} \{x_i\} - \min_{i < 2k} \{x_i\} \right) \\
 &> \frac{1}{2} \left(\max_{i < 2k} \{x_i\} - \min_{i < 2k} \{x_i\} \right) \\
 &\geq \left| \frac{\max_{i < 2k} \{x_i\} + \min_{i < 2k} \{x_i\}}{2} - x_i \right| = |x_{2k} - x_i|
 \end{aligned}$$

Therefore we have $|y_{2k} - y_i| > |x_{2k} - x_i|$, so $H(c_{2k}, c_i)$ is horizontal. Analogously, we set

$$y_{2k+1} = \frac{\max_{i < 2k+1} \{y_i\} + \min_{i < 2k+1} \{y_i\}}{2}$$

and

$$x_{2k+1} = \max_{i < 2k+1} \{x_i\} + 2 \left(\max_{i < 2k+1} \{y_i\} - \min_{i < 2k+1} \{y_i\} \right).$$

We prove as above (just by swapping the roles of x and y) that, in this case, the hypersurfaces $H(c_{2k+1}, c_i)$ are all vertical.

For $k \leq m$, we denote by H_k (resp. V_k) the number of horizontal (resp. vertical) hypersurfaces after adding the k -th candidate.

As all horizontal hypersurfaces intersect all vertical hypersurfaces, these intersections already define $(H_k + 1)(V_k + 1)$ different areas (with distinct preferences). Hence, denoting by A_m the number of areas after adding the k -th candidate, we have $A_m \geq (H_m + 1)(V_m + 1)$.

Each time we add a candidate c_k , we obtain $k - 1$ new hypersurfaces $H(c_1, c_k), \dots, H(c_{k-1}, c_k)$, all horizontal if k is even, or all vertical if k is odd. Consequently:

- if k is even, $H_k = H_{k-1} + (k - 1)$ and $V_k = V_{k-1}$;
- if k is odd, $H_k = H_{k-1}$ and $V_k = V_{k-1} + k - 1$.

We can deduce that $H_m \in \Theta(m^2)$ and $V_m \in \Theta(m^2)$, and thus $A_m \in \Omega(m^4)$. \square



Final thoughts

Chapter 8

Conclusion

All (good) things come to an end, so does this thesis. But it is well-known that every end is a new beginning, and every solved problem raises new questions.

We started this thesis with a basic opening problem - there was a group of friends discussing who they are going to vote for in the upcoming parliamentary election.¹ This example brought us to the social choice theory, a field which studies, among other things, the problems connected to the decision making where a group of individuals have to choose together among a set of available alternatives.

We have seen that even though some questions and problems tackled by social choice can seem very natural and even “basic”, the solution is often not obvious to find from a theoretical or computational point of view. To circumvent computational difficulties or impossibility results and paradoxes, an assumption of a common underlying structure in the preferences is often made. In this regard, we have presented some of the most common preference structures (namely, single-peaked preferences, single-crossing preferences and 1-Euclidean preferences), as well as some of their most important theoretical and computational interests.

Before the final goodbye, let us recall the main results of this thesis and discuss the research directions that open to us. This conclusion is organized as follows: there are three sections, one for each part of the thesis. We recall *very briefly* the background of each part, as well as our main contributions, at the beginning of the corresponding section. Our aim here is to help the reader to get back to the swing of things by refreshing her/his memory. If needed, more details can be found in the corresponding parts and chapters. On the other hand, the reader is free to skip these two paragraphs if (s)he remember perfectly the corresponding part.

¹By the way, I’m writing these lines nearly 10 years later, and about a week before the Czech presidential elections; my dear friends from the opening example are spread all over Europe and... I still do not know who I am going to vote for!

The rest of each section is then dedicated to some possible research directions and open questions arising from our works.

8.1 Looking back to Part I

Reminder of the content Part I is an opening part. Chapter 1 is purely introductory - its aim is to introduce the notion and notations used in the remainder of the thesis. Also, it introduces formally the domains of single-peaked, single-crossing and 1-Euclidean preferences, and provides a state of the art.

Chapter 2 is there to deepen the state of art from the previous chapter by focusing on the relations between the three domain restrictions mentioned above. The starting point of this chapter is the fact that the 1-Euclidean domain is undoubtedly less understood (from the structural point of view) than the single-peaked domain and the single-crossing domain. Several works shed light on the structural comprehension of 1-Euclidean domain. In particular, it is well-known that 1-Euclidean preferences are both single-peaked and single-crossing. Therefore, the understanding of the domain of preferences that are both single-peaked and single-crossing (denoted by SPSC) seems essential for a better understanding of the 1-Euclidean domain. Indeed, the SPSC domain is quite well understood. In particular, Elkind et al. (2020) give a characterization of the SPSC domain. By combining the results of Elkind et al. (2020) with the technique of Bredereck et al. (2013) (based on works of Abello (1991) and Galambos and Reiner (2008)) describing how to enumerate all single-crossing profiles, it is in particular possible to enumerate all maximal SPSC profiles w.r.t. a given single-peaked axis A .

8.1.1 1D-Euclidean preferences: towards a characterization

At the end of Chapter 2, we have enumerated numerically all SPSC profiles (up to renaming of candidates) on at most 8 candidates. We have then checked (by using linear programming) for each of these profiles whether or not it was 1-Euclidean. These results seem to reveal that a very small proportion of SPSC profiles are actually 1-Euclidean. For completeness, and to avoid the reader to return back to Chapter 2, we recall these results in Table 8.1.1.

m	4	5	6	7	8
SP-SC	2	12	286	33 592	23 178 480
Euclidean	2	12	168 (59 %)	4680 (14%)	229 468 (1%)

Table 8.1.1: The number of maximal SPSC, resp. 1-Euclidean, profiles on m candidates (w.r.t. a given axis A).

We recall that Chen et al. (2017) proved that the 1-Euclidean domain cannot be characterized by finitely many forbidden minors, and asked whether or not there is a *good* characterization of this domain, i.e., a polynomial time algorithm that would identify

whether or not a given preference profile contains a forbidden minor.

We have investigated this question, and we came with a conjecture stating that the problem of deciding whether or not a given *maximal* preference profile is 1-Euclidean can be reduced (in polynomial time) to a problem of cycle detection in a directed graph.²

A very natural, short-term future work, would be to prove (or disprove) the conjecture. We explain here how the conjecture was found, and we provide some numerical results supporting it.

The starting point comes from the paper by Doignon and Falmagne (1994). They have noticed that given a maximal SPSC profile \mathcal{P} single-peaked w.r.t. an axis $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$, any mapping f providing a 1-Euclidean representation of \mathcal{P} , if it exists, verifies:

$$f(c_1) < f(c_2) < \dots < f(c_m).$$

In other words, the candidates are placed on the line in the single-peaked order. Any voter v such that $f(v) < \frac{f(c_1)+f(c_2)}{2}$ has the preference $>_v = (c_1, c_2, \dots, c_m)$. Moving towards the candidate c_2 , the voter crosses at some point the bisector between the two candidates c_1 and c_2 (denoted here by $B(c_1, c_2)$). As a consequence, her preference becomes $(c_2, c_1, c_3, \dots, c_m)$ - the candidates c_1 and c_2 are swapped. Moving more and more to the right, another bisector will be crossed and hence another pair of candidates swapped - see Figure 8.1.1 for an illustration (with $m = 3$). Therefore, deciding whether a profile is (maximal) 1-Euclidean boils down to finding an order on the candidates (i.e., an axis) as well as a single-crossing order on the preferences, and then determining if there are possible values $f(c_1), f(c_2), \dots, f(c_m)$ inducing the order of bisectors corresponding to the single-crossing order.

Indeed, the single-crossing order $<_O$ gives us an order on consecutive swaps of candidates. If a pair of candidates $\{c_i, c_j\}$ is swapped before $\{c_k, c_l\}$ in $<_O$, then the bisector $B(c_i, c_j)$ must lie on the left of the bisector $B(c_k, c_l)$. More formally, for two pairs of candidates $\{c_i, c_j\}$ and $\{c_k, c_l\}$, we have the following constraint:

$$f(c_i) + f(c_j) < f(c_k) + f(c_l) \text{ iff } \{c_i, c_j\} <_O \{c_k, c_l\}.$$

Doignon and Falmagne (1994) show that the profile is 1-Euclidean iff the system of inequalities induced by all couples of pairs $\{c_i, c_j\}$ and $\{c_k, c_l\}$ (together with the constraint $f(c_1) < f(c_2) < \dots < f(c_m)$) has a solution. Scott (1964) provides a general mathematical (namely, linear-algebraic) technique allowing to decide whether or not such a system has a solution. We will now investigate the question from a “structural” point of view.

²Note that the restriction to maximal profiles is possible thanks to the results of Doignon and Falmagne (1994) and their direct corollaries.

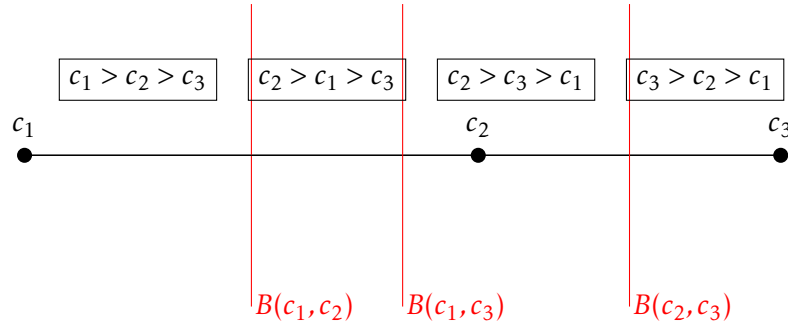


Figure 8.1.1: An example of a Euclidean embedding with respect to the axis $A = c_1 \triangleleft c_2 \triangleleft c_3$.

Towards the conjecture Let us consider a maximal SPSC profile \mathcal{P} (w.r.t. A) with the following order of swaps (induced by the single-crossing order):

$$O = (c_1 c_2, c_1 c_3, c_2 c_3, c_1 c_4, c_2 c_4, c_3 c_4, c_1 c_5, c_2 c_5, c_1 c_6, c_3 c_5, c_4 c_5, c_2 c_6, c_3 c_6, c_4 c_6, c_5 c_6) \quad (8.1)$$

The first preference of this order is $>_{o_1} = (c_1, c_2, \dots, c_m)$. The second preference is $>_{o_2} = (c_2, c_1, c_3, \dots, c_m)$ as c_1 and c_2 are swapped. The third preference is $>_{o_3} = (c_2, c_3, c_1, c_4, \dots, c_m)$ as c_1 and c_3 are swapped, and so on to the last preference $>_{o_m} = (c_m, c_{m-1}, \dots, c_1)$.

The profile \mathcal{P} is not 1-Euclidean, as the system of inequalities given by Doignon and Falmagne (1994) has no solution. Indeed, let us suppose for contradiction that there exists a 1-Euclidean representation f of \mathcal{P} . In particular, it verifies the three following inequalities (the corresponding arcs are in red in Figure 8.1.2):

$$f(c_2) + f(c_3) < f(c_1) + f(c_4), \text{ as } B(c_2, c_3) \text{ is on the left of } B(c_1, c_4). \quad (8.2)$$

$$f(c_1) + f(c_6) < f(c_3) + f(c_5), \text{ as } B(c_1, c_6) \text{ is on the left of } B(c_3, c_5). \quad (8.3)$$

$$f(c_4) + f(c_5) < f(c_2) + f(c_6), \text{ as } B(c_4, c_5) \text{ is on the left of } B(c_2, c_6). \quad (8.4)$$

If we sum these inequalities, we obtain $0 < 0$. In other words, they cannot all hold together, so the profile is not 1-Euclidean. Let us now define a graph denoted by $C_m(A, O)$ (an example relative to the order O given in 8.1 and axis $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$ is given in Figure 8.1.2; the arcs corresponding to the inequalities 8.2 - 8.4 are given in red):

Definition 8.1.1: Graph $C_m(A, O)$

Let m be the number of candidates, O a single-crossing order and $A = c_{i_1} \triangleleft c_{i_2} \triangleleft \dots \triangleleft c_{i_m}$ (with $\{i_1, i_2, \dots, i_m\} = \{1, 2, \dots, m\}$) an axis. We define a graph $C_m(A, O)$ as follows:

- There is one vertex for each couple of candidates c_{i_k}, c_{i_j} such that $c_{i_j} \triangleleft c_{i_k}$. This vertex is denoted by $i_k i_j$.
- There is an arc $(i_k i_j, i_{k+1} i_j)$ for any two vertices $i_k i_j$ and $i_{k+1} i_j$.
- There is an arc $(i_k i_{j+1}, i_k i_j)$ for any two vertices $i_k i_{j+1}$ and $i_k i_j$.
- Each couple of swaps $c_{i_j} c_{i_k}$ and $c_{i_p} c_{i_q}$ (with $c_{i_j} \triangleleft c_{i_p} \triangleleft c_{i_q} \triangleleft c_{i_k}$ (resp. $c_{i_p} \triangleleft c_{i_j} \triangleleft c_{i_k} \triangleleft c_{i_q}$) such that $c_{i_j} c_{i_k}$ is before $c_{i_p} c_{i_q}$ in O , there are two arcs $(i_k i_q, i_p i_j)$ and $(i_k i_p, i_q i_j)$ (resp. $(i_j i_p, i_q i_k)$ and $(i_k i_p, i_q i_j)$) in $C_m(A, O)$.

A Curious Tortoise Intervention



The Definition 8.1.1 is given for an arbitrary axis on the set of candidates $\{c_1, c_2, \dots, c_m\}$. However, we have assumed that $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$, which simplifies the notations: the definition rewrites as follows:

- There is one vertex for each couple of candidates c_j, c_i such that $i < j$. This vertex is denoted by ji .
- There is an arc $(ji, (j+1)i)$ for each $i < m-1$ and $j > i$.
- There is an arc $(j(i+1), ji)$ for each $i < m-1$ and $j > i+1$.
- Each couple of swaps $c_i c_j$ and $c_k c_l$ with $i < k < l < j$ (resp. $k < i < j < l$) such that $c_i c_j$ is before $c_k c_l$ in O , there are two arcs (jl, ki) and (jk, li) (resp. (ik, lj) and (jk, li)) in $C_m(A, O)$.

This graph can be interpreted as follows: each vertex ji corresponds to the quantity $f(c_j) - f(c_i)$ for any possible representation function f . There is an arc from ji to $j'i'$ if $f(c_j) - f(c_i) < f(c_{j'}) - f(c_{i'})$.

Indeed, with this interpretation (and still under the assumption that $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$):

- The first set of arcs simply states that for any $j > i$, $f(c_j) - f(c_i) < f(c_{j+1}) - f(c_i)$, in other words, $f(c_j) < f(c_{j+1})$.
- The second set of arcs is also implied by the (set of) condition(s)

$$f(c_1) < f(c_2) < \dots < f(c_m)$$

It actually states that for any $j > i + 1$, $f(c_j) - f(c_{i+1}) < f(c_j) - f(c_i)$; in other words, $f(c_i) < f(c_{i+1})$.

- Finally, if the swap $c_i c_j$ is before $c_k c_l$ in O , we have the bisector $B(c_i, c_j)$ on the left of $B(c_k, c_l)$. This is written $c_i + c_j < c_k + c_l$. If $i < k < l < j$, this can be rewritten as $c_j - c_m < c_k - c_i$ and $c_j - c_k < c_l - c_i$, which yields two arcs from the last point of Definition 8.1.1 (the case of $k < i < j < l$ is treated similarly).

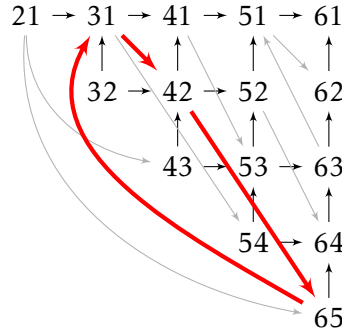


Figure 8.1.2: O is a single-crossing order given in 8.1, the axis $A = c_1 \triangleleft c_2 \triangleleft \dots \triangleleft c_m$. $C_6(A, O)$ contains a circuit. We note that for “diagonal” arcs, only the arcs corresponding to consecutive couples of candidates in O are given - the others can be obtained by transitive closure. For more readability, diagonal arcs not important for this example are greyed.

We can now go back to the profile \mathcal{P} and build its corresponding graph $C_m(A, O)$ (see Figure 8.1.2). If we focus on the three “problematic” constraints 8.2-8.4, we notice that they are together responsible for a circuit in $C_m(P)$. Indeed, it is not difficult to see that if there is a circuit in $C_m(A, O)$, then the order of bisectors induced by O is not feasible (as a circuit in $C_m(A, O)$ can be interpreted as a circuit in constraints needed to be verified by any representation function f), and hence the profile is only SPSC but not 1-Euclidean. The question is now whether this results is just a new, stronger necessary condition for a profile to be Euclidean, or if this condition is actually sufficient.³ We conjecture that this condition is sufficient:

Conjecture 8.1.1

Let \mathcal{P} be a maximal SPSC profile on m candidates w.r.t. an axis A and an order of swaps (induced by the single-crossing order) O . This profile is Euclidean if and only if the corresponding graph $C_m(A, O)$ is acyclic.

³Indeed, intuitively, one can imagine that the order of bisector might involve a more complicated constraints that can not be expressed by a simple circuit.

Numerical results We performed numerical experiments in order to support our conjecture, or to find a counter-example. To do so, we reused our simple code designed in Chapter 2 to enumerate all SPSC profiles.

For $m \leq 7$, we have listed all SPSC profiles. For each of them, we have performed the cycle detection in $C_m(A, O)$ as well as a linear programming resolution. Therefore, we have proved experimentally that the conjecture holds for $m \leq 7$.

For $m = 8$, although we were able to list all SPSC profiles (we recall that there are 23 178 480 such profiles), the circuit detection approach was very time consuming. Our code was originally designed as a basic tool to help us understand the relation between the single-peaked domain, the SPSC domain and the 1-Euclidean domain. Therefore, it was not optimized to perform serious numerical experiments.

A Curious Tortoise Intervention



In other words, the code is as slow as a tortoise... But I have nothing to do with it !

While the code can be clearly optimized to check if the conjecture also holds for $m = 8$, it seems quite tricky to have analogous experimental results for $m > 8$. To get more insights, we have performed more experiments on randomly generated SPSC profiles on m candidates with m varying between 8 and 13. We have proceeded as follows: We first set m at random, uniformly between 9 and 13. Then we built a random maximal SPSC profile on m candidates, and the corresponding graph $C_m(A, O)$. Finally, we have checked if $C_m(A, O)$ contained a circuit, and we compared with the result of linear programming. We have performed 1 000 000 tests and observed that the conjectured held for all of them. Even though 1 000 000 tests is clearly not enough (let us recall that it represents about 4% of all SPSC profiles on 8 candidates), the results are encouraging, and it seems that the conjecture is worth further study.

A Curious Tortoise Intervention



Before ending this section, let us discuss some ideas about how to prove the conjecture. A natural idea is to make a proof by construction. Let us consider a SPSC profile with respect to an axis A and an order of swaps (induced by the single-crossing order) O . If $C_m(A, O)$ does not contain a circuit, we can find a topological order on its vertices. Quite a straightforward idea would be to order the quantities $f(c_j) - f(c_i)$ where $j > i$ (i.e., the vertices of $C_m(A, O)$), and then somehow define the values of f (without using linear programming) in such a way that all the quantities fit together while respecting the fixed order. However, the topological order on vertices of $C_m(A, O)$ is not necessarily unique. The following example shows that we cannot just take any of them - actually, as we will see, the



topological order adds more constraints on the values of f , and these may not be feasible. To this purpose, let us consider a preference profile \mathcal{P} that is maximal single-peaked single-crossing and induces the following order O of swaps:

$$O = (c_1 c_2, c_1 c_3, c_2 c_3, c_1 c_4, c_2 c_4, c_3 c_4, c_1 c_5, c_1 c_6, c_2 c_5, c_2 c_6, c_3 c_5, c_3 c_6, c_4 c_5, c_4 c_6, c_5 c_6)$$

The corresponding graph $C_m(A, O)$ does not contain a circuit, so it is Euclidean (we recall that we have proved the conjecture numerically for $m \leq 7$). Let us build a topological order on vertices of $C_m(A, O)$. For instance, we can take the following order T :

$$T = (65, 32, 21, 31, 54, 43, 42, 41, 64, 53, 63, 52, 62, 51, 61)$$

However, there is no representation f respecting this order. In fact, we have:

$$f(c_3) - f(c_2) > f(c_6) - f(c_5)$$

and

$$f(c_4) - f(c_3) > f(c_5) - f(c_4).$$

If we sum up these two conditions, we obtain

$$f(c_4) - f(c_2) > f(c_6) - f(c_4).$$

But the vertex 64 is ordered after the vertex 42 in T . Thus, T is not feasible. We can check that there is another feasible topological order T' that consists simply in putting 64 just before 42:

$$T' = (65, 32, 21, 31, 54, 43, 64, 42, 41, 53, 63, 52, 62, 51, 61)$$

Let us fix the values of f in the most naive and straightforward way (that obviously does not work all the time, as we will discuss later): The first element of T' is 65 - we can hence fix the value $f(c_6) - f(c_5)$ arbitrarily, let say, $f(c_6) - f(c_5) = 1$. Then we have $f(c_3) - f(c_2) > 1$, so we can choose, for instance, $f(c_3) - f(c_2) = 2$. Using the same strategy, we fix $f(c_2) - f(c_1) = 3$. The next vertex is 31, and from the values fixed so far, we have $f(c_3) - f(c_1) = f(c_3) - f(c_2) + f(c_2) - f(c_1) = 5$. We fix then $f(c_5) - f(c_4) = 6$, and $f(c_4) - f(c_3) = 7$. However, this last choice is problematic - the next vertex is $f(c_6) - f(c_4)$, and from previously fixed values, we have $f(c_6) - f(c_4) = 7$. We need hence to come back and change the value of $f(c_4) - f(c_3)$ to, for instance, 6.5. We see then that all remaining values can be inferred from previous ones, and that the order is respected. Moreover, if we put $f(c_1) = 0$, we obtain the following values:



$$f(c_1) = 0, f(c_2) = 3, f(c_3) = 5, f(c_4) = 11.5, f(c_5) = 17.5, f(c_6) = 18.5.$$

Note that from there, all bisectors can be drawn up, which will divide the real line into $\frac{m(m-1)}{2} - 1$ segments and two half-lines. For each voter v , we can then identify her corresponding segment (resp. half-line), and fix $f(v)$ within it.

This example raises two (main) questions:

1. Can we characterize feasible (resp. non-feasible) topological orders?
2. Once the topological order is fixed, how to obtain values of f on the set of candidates? Actually, in the example of T' , we were very lucky to fix the values (nearly) iteratively. However, there was one case where the value of $f(c_4) - f(c_3)$ was conditioned by the value $f(c_6) - f(c_4)$; this second value came after the first one in the topological order, but was already determined by the values $f(c_6) - f(c_5)$ and $f(c_5) - f(c_4)$ (coming both before 43 in T'). We were forced to make a step back to correct the value of $f(c_4) - f(c_3)$ in function of $f(c_6) - f(c_4)$. Can we bound the number of such steps back? Or, better, can we prevent them?

Note that, in particular, a positive answer to both these questions would yield a combinatorial algorithm for recognizing 1-Euclidean preferences.

8.2 Looking back to Part II

Reminder of the content Even though single-peaked preferences guarantee many good theoretical and computational properties, they have an important drawback: they are very unlikely to appear in real-world applications if the preferences are not for numerical values. That is why we try to make the structure less restrictive. Even though the majority of desirable properties are lost, the knowledge of an underlying structure is still interesting, in particular for data analysis purposes.

In this part, we proposed two different extensions of single-peaked preferences. In Chapter 3, it is the structure of the axis itself that is relaxed: an axis is seen as a graph (more precisely a path), and we introduce the notion of the single-peakedness with respect to a graph as follows: a preference profile is single-peaked with respect to a given graph \mathcal{G} if each of its preferences is a traversal of \mathcal{G} . Trivially, any profile is single-peaked with respect to the complete graph. However, such a result is useless for data analysis purposes. Indeed, the sparser the graph is, the more structural information it provides. Two notions of sparsity were introduced in this chapter, and we looked for solutions minimizing (one of) them.

In Chapter 4, we studied the so-called nearly single-peakedness. This time, we kept the axis structure, but we allowed the preferences to slightly deviate from it if necessary. A distance measure needs then to be introduced to evaluate the distance of a profile from a given axis. In particular, given a preference profile, we looked for an axis that minimizes this distance. Several measures of single-peakedness of a profile exists in the literature. We proposed a new measure of single-peakedness called *k-forbidden triple single-peakedness (FT)* that directly follows from the very definition of Black's single-peakedness condition.

Our contribution In Chapter 3, we proposed Integer Linear Programming formulations (ILP) finding a compatible graph that minimizes the number of edges (resp. the maximum degree of a vertex). We show that both corresponding problems are NP-hard. Nevertheless, if an optimal solution is a tree (resp. a path or a cycle for the second ILP), we prove the integrality of any optimal basic solution of the linear program obtained by relaxing the integrality constraints. This provides an alternative polynomial time recognition algorithm for preferences single-peaked on a tree (resp. a path or a cycle). We also provide a polynomial time recognition algorithm for preferences single-peaked on a pseudotree. We performed numerical experiments on both real-world and synthetic elections, in order to evaluate the relevance and efficiency of these methods in practice.

As said above, a new measure of single-peakedness is proposed in Chapter 4. In order to compare it to existing measures of nearly single-peakedness, we undertook an axiomatic study to identify some theoretical properties that are (or not) guaranteed by this measure. In particular, it allowed us to formalize some differences between this measure, *k-voter deletion single-peakedness* and *k-global swap single-peakedness*.

Regarding computational aspects, we proved that, as for most of the proposed measures in the literature, it is NP-hard to find an axis minimizing the *k-forbidden triples* measure. Nevertheless, we proposed an ILP formulation of this problem which reveals efficient in practice. We performed various numerical tests on both real-world and synthetic data to evaluate the efficiency of the *FT* measure in practice, and also to provide a comparison with some other notions of nearly single-peakedness.

8.2.1 Single-peaked preferences on an arbitrary graph: combining minimization criteria, the likelihood of solutions, and even more

The approach introduced in Chapter 3 raises several questions. Most of them are related to the interpretability of the results - indeed, we have emphasized that identifying the underlying structure of a given profile may be useful for data analysis. Let us now discuss some natural issues, and propose therefore several possible, mainly short-term/middle-term, research directions:

How do we choose the minimization criterion?

We propose in Chapter 3 two different measures of graph sparsity that we would like to minimize - namely, the number of edges and the maximum vertex degree. However, we cannot simply state that we will minimize both of them at the same time, as these criteria can be antagonist. Mathilde has already noticed in Chapter 3 that, unsurprisingly, different minimization criteria yield different solutions, and that is not always evident which minimization criterion to prefer.

For instance, one can argue that minimizing the number of edges should be preferred, as it is a very natural and easy to interpret measure. However, one may also argue that the maximum degree minimization somehow takes into account the minimization of number of edges. Indeed, the sum of degrees of all vertices equals twice the number of edges. In particular, if we have m vertices and a maximal degree equal to k , the number of edges is upper bounded by $\frac{km}{2}$. On the other hand, minimizing the number of edges does not provide any guarantee on the maximal vertex degree. From this point of view, the maximum degree minimization might seem a better criterion.

From another point of view, the minimization of the number of edges guarantee to return an acyclic solution if it exists. This can be a strong argument to promote this criterion, as several desirable theoretical and computational properties are guaranteed for acyclic graphs - for instance, there will always be a Condorcet winner.

In practice, there is not a clear answer as to which criterion is better. For instance, we can imagine a profile compatible with a star graph, i.e., a tree $T = (V_T, E_T)$ for which there exists a vertex $c \in V_T$ such that $\{c, c'\} \in E_T$ for each $c' \in V_T$. We have already seen that the only information provided by this graph was that the star center (i.e., the vertex c) was ranked first or second by any voter. However, we do not have any information on relations between the others candidates. If we allow to add some more edges, we can obtain a more “decentralized” graph providing structural information on more candidates. More generally, the minimization of the maximum degree will tend to give more “equitable” solutions (in the sense that each vertex will have approximately the same degree), while the minimization of the number of edges may tend to identify some “hub” vertices with a strong degree that will ensure the connection of many (possibly “isolated”) vertices.

Depending on the context, one or the other solution can be preferred. For instance, the minimizing the maximum degree may be more convenient for a recommendation system, while minimizing the number of edges can be suitable for an election context (as, typically, the “central” vertices will correspond to very popular candidates).

It would be interesting to deepen this question and to study if there are some problems for which one of the criteria seems more suitable. Alternatively, we could try to

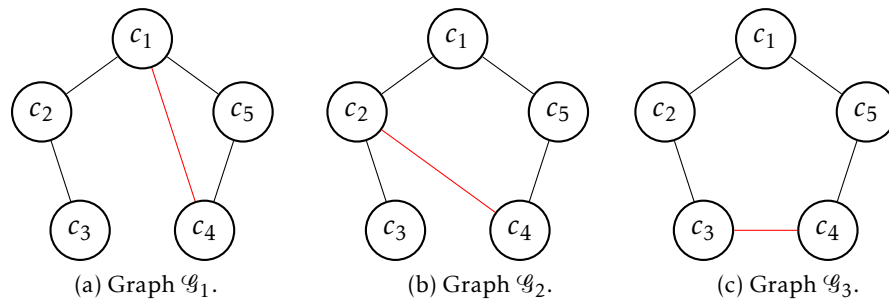


Figure 8.2.1: There are three graphs minimizing the number of edges compatible with the profile \mathcal{P} .

combine both criteria, which would yield multidimensional optimization problems, or minimizing the number of edges with an upper bound on the degree.

How to decide which solution is “the best” if several solutions exist?

Let us suppose that we have a profile of preferences, and we want to find a graph with respect to which this profile would be single-peaked. Assume that we have answered the previous question and chose the most suitable criterion for our problem. For instance, let us assume that we minimize the number of edges, and that we have the following profile \mathcal{P} of 4 preferences over 5 candidates:

$$>_1: (c_1, c_2, c_3, c_4, c_5)$$

$$>_2: (c_5, c_1, c_2, c_3, c_4)$$

$$>_3: (c_5, c_4, c_1, c_2, c_3)$$

$$>_4: (c_3, c_2, c_1, c_5, c_4)$$

Any graph compatible with \mathcal{P} contains the following necessary edges: $\{c_1, c_2\}$, $\{c_1, c_5\}$, $\{c_2, c_3\}$ and $\{c_4, c_5\}$. Moreover, the vertex c_4 needs to be connected to one of the vertices c_1, c_2 or c_3 so that the definition of single-peakedness on a graph is satisfied. Put together, there are 3 possible graphs $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ with respect to which \mathcal{P} is single-peaked - see Figure 8.2.1.

But each of these graph provides a different structural information. Our preference profile can actually be seen as a statistical sampling of a population sharing a common hidden structure. More visually, we can imagine that there is a “hidden” graph (here $\mathcal{G}_1, \mathcal{G}_2$ or \mathcal{G}_3) from which each preference ranking was generated as a random traversal. But we only have access to the resulting profile \mathcal{P} , and we would like to learn the structure behind, i.e., to identify the most likely graph from which \mathcal{P} was generated.⁴ This problem raises in particular two questions that yield an interesting research direc-

⁴Note that a similar problem was recently studied by Sliwinski and Elkind (2019) for the profiles single-peaked w.r.t. trees.

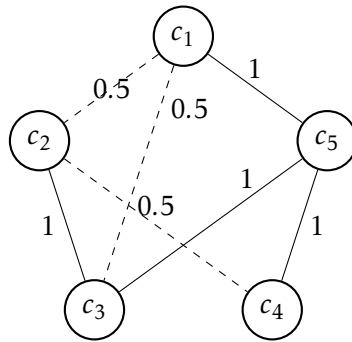


Figure 8.2.2: An illustration of a graph with fractional edges. The dashed edges have a capacity 0.5, while the full edges have a capacity 1.

tion: first, we need to determine the set of *all* possible solutions. For instance, we have seen in the above example that there were three possible graphs minimizing the number of edges. Second, we need to evaluate, for each of these solutions, the likelihood that this graph has generated a given profile. In other words, the “true” solution would be a graph \mathcal{G} that maximizes the probability to obtain \mathcal{P} if we generate randomly the preferences single-peaked with respect to \mathcal{G} .

Continuous relaxation of ILP formulations: interpretation of fractional edges

We have defined the single-peakedness on an arbitrary graph by using the notion of traversals of graphs. In particular, given a preference $(c_{i_1}, c_{i_2}, \dots, c_{i_m})$, then for each c_{i_k} , there must be an edge connecting it to one of its predecessors $c_{i_1}, c_{i_2}, \dots, c_{i_{k-1}}$. The edges are seen here as binary objects - each edge is or is not in the graph.

However, it seems possible to allow the presence of fractional edges by giving them a suitable interpretation. In a continuous version, each edge has a capacity between 0 and 1, and $(c_{i_1}, c_{i_2}, \dots, c_{i_m})$ is a traversal of the graph if for each c_{i_k} , the sum of the capacities of edges $\{c_{i_j}, c_{i_k}\}$ with $j < k$ is greater than or equal to 1. For instance, let us consider the graph \mathcal{G} given in Figure 8.2.2. The preference $(c_1, c_5, c_4, c_2, c_3)$ is a traversal of \mathcal{G} in a fractional sense: even if the candidate c_2 is not connected to any of its predecessors (namely, c_1, c_5, c_4) by an integral edge, we have $x_{12} = x_{24} = 0.5$, so $x_{12} + x_{24} + x_{25} \geq 1$ - the sum of capacities of edges $\{c_1, c_2\}$, $\{c_2, c_4\}$ and $\{c_2, c_5\}$ is greater than or equal to 1.

In other words, there can be synergies between several edges. In this example, c_2 cannot be directly connected to either c_1 or c_4 . However, having both c_1 and c_4 in the preference prefix allows us to connect c_2 . A natural question that is worth investigating is whether there is any social choice context in which this interpretation could be useful?

Note also another particularity of fractional edges. They somehow add a kind of “one-way” neighbourhood relationship: for instance, if the first and second most preferred

candidates of a given voter v are c_2 and c_3 , her third most preferred candidate can be c_1 . However, conversely, if the most preferred candidate of another voter v' is c_1 , her second most preferred candidate can be neither c_2 nor c_3 . We can once more ask ourselves the question of whether this interpretation may reveal useful in some context.

An alternative interpretation of fractional edges consists in assuming that the probability of using an edge somehow involves its capacity. This could be used, for instance, by a recommendation system. For illustration, let us consider that the preference structure is given by the graph \mathcal{G} in Figure 8.2.2. Let us assume that there is a voter whose most preferred candidate is c_2 . Strictly speaking, we should only recommend her the candidate c_3 , as any (possibly fractional) traversal of \mathcal{G} starting by c_2 must have c_3 in second position. However, we could consider a variant in which c_1 or c_4 can also be proposed to the voter, just with less probability than c_3 - typically, we could recommend c_3 with probability 0.5, or also c_1 or c_4 , each with the same probability 0.25. Maybe this kind of interpretation might provide a generalisation where the preference structure is probabilistic and thus less restrictive?

What about partial preferences /preferences with ties?

The assumption that the preferences are strict complete orders is actually quite strong. In practice, voters are rarely asked to give a complete ranking on the whole set of candidates. Worse, voters often do not know their complete preference - for instance, we often have a precise idea about who is our most preferred candidate, and, on the other hand, which candidate would be unacceptable to us. However, there may be several candidates (typically in the middle of our ranking) who leave us indifferent. It may be very difficult to rank them exactly, and if we are asked to do so, we may be tempted to rank them more or less randomly, which will introduce noise disturbing the structure recognition.

There are several possibilities to circumvent this issue. We propose here two possibilities, and provide a brief discussion for each of them:

- We can ask the voters to only provide the top-orders: each voter will give a strict incomplete ranking $(c_{i_1}, c_{i_2}, \dots, c_{i_k})$ on the set of candidates $\{c_1, c_2, \dots, c_m\}$. Each candidate c' that does not appear in this top order is less preferred than any of the candidates $c_{i_1}, c_{i_2}, \dots, c_{i_k}$. However, the candidates that do not appear in the top order are not comparable between them.

There are several issues to deal with. One of the most important issues is the following one: while we can of course generalize the single-peakedness to subsets of candidates by requiring each top order $(c_{i_1}, c_{i_2}, \dots, c_{i_k})$ to be a traversal of a subgraph restricted to the set of vertices $\{c_{i_1}, c_{i_2}, \dots, c_{i_k}\}$, the resulting graph on the whole set of candidates may not be connected. An interpretation of such a solution would be required. Note that the less candidates the voters rank in their top orders, the more challenging it will be to build a graph providing a relevant

information - for instance, if each voter gives only her most preferred candidate, an empty graph will be compatible with the profile, but would not provide any structural information.

- We can ask the voters to rank all candidates by allowing ties. Given a preference $>_v$ with ties, we can enumerate a set of preferences $>$ compatible with $>_v$ defined as follows: $c > c'$ whenever $c >_v c'$, and we can fix arbitrary $c > c'$ or $c' > c$ for any tie $c \sim_v c'$. For instance, there are two preferences $>_1$ and $>_2$ compatible with the preference $c_1 >_v c_2 \sim_v c_3$, namely $c_1 >_1 c_2 >_1 c_3$ and $c_1 >_2 c_3 >_2 c_2$. The notion of single-peakedness on a graph can then be generalized as follows: a preference $>_v$ (possibly with ties) is single-peaked with respect to a given graph \mathcal{G} if there exist a strict preference $>$ (i.e., without tie) compatible with $>_v$ that is single-peaked with respect to \mathcal{G} .

Given a preference profile $\mathcal{P} = \{>_1, >_2, \dots, >_m\}$, let us denote by P_i the set of preferences compatible with $>_i$. A preference profile \mathcal{P} is single-peaked with respect to a given graph \mathcal{G} if there exists a *compatible profile* $\mathcal{P}' = \{>'_1, >'_2, \dots, >'_m\}$ with $>'_i \in P_i$ single-peaked with respect to \mathcal{G} .

Obviously, the straightforward algorithm consisting in checking all compatible profiles would be very computationally consuming: for instance, if we have a profile of n preferences $>_1, >_2, \dots$, each of them containing a unique tie between a pair of candidates, there are two compatible strict preferences for each $>_i$. This yields 2^n compatible profiles to check. Clearly, a more sophisticated method would be needed to check whether a profile with ties is single-peaked with respect to a given graph. While it seems to us that such a method could be found quite easily, this questions becomes a little bit harder when it comes to finding *an optimal* graph (in terms of a given sparsity criterion).

8.2.2 Nearly single-peaked preferences on an arbitrary graph

The main motivation for introducing the notion of nearly single-peakedness can be summarized as follows: the main problem of (generalized) single-peaked preferences is that each voter must be *perfectly* compatible with the graph. Therefore, the voters with very specific preferences force the presence of edges that are not necessary for the majority of the population. In Chapter 4, we have tried to circumvent this problem by introducing the notion of nearly single-peakedness. We have assumed that there is a true hidden underlying structure, and that the preferences can be seen as its noisy observations. More precisely, we have assumed that the underlying structure is an axis.

This idea can be generalized by supposing that the underlying structure can be an arbitrary graph. In such a case, we tackle a multi-objective optimization problem, as our aim is to minimize both the sparsity of graph (for instance, in terms of the number of edges or the maximum vertex degree) and the distance of the profile from this graph.

Obviously, these two criteria are in conflict: it will be necessary to make a trade-off between the graph sparsity (i.e., the amount of structural information provided by a solution) and the distance of the profile from the graph (i.e., the level of relevance of the structure for a given profile).

This problem, combining both approaches studied in Part II, represents a promising research direction which could possibly be useful, especially in real-world applications (as, for instance, recommendation systems). Indeed, the assumptions that a profile is single-peaked with respect to a very restrictive structure (as, for instance, an axis), and that each preference is perfectly compatible with it, are too strong and hence unlikely to be verified by a real-world preference profile.

8.3 Looking back to Part III

Reminder of the content In this part, we present another way to make classical domain restrictions “richer” in terms of expressivity. It consists in adding more dimensions to the structure. We focus in this part on Euclidean preferences - more precisely, we study multidimensional Euclidean preferences with respect to norms ℓ_1, ℓ_2 and ℓ_∞ , mainly in the 2-dimensional case.

Our contribution In Chapter 5, we answered the open question of the complexity of the Kemeny ranking problem. More generally, we showed that every weighted (bipartite) tournament can be seen as the (weighted) majority graph of a 2-Euclidean profile. Thus, essentially, hardness results for computational choice problems that can be formulated on the (weighted) majority tournament are still true if preferences are 2-Euclidean.

In Chapter 6, we proposed a heuristic algorithm for recognizing 2-Euclidean preferences with respect to the norm ℓ_2 . We recall that Peters (2017) showed that this problem is $\exists\mathbb{R}$ -hard. Several algorithms providing an approximate 2-Euclidean embedding exist, but they often encounter the following problem: they tend to provide a *degenerate* solution.

Our heuristic works in two steps. First, we check if necessary conditions to be 2-Euclidean are fulfilled. If not, the profile is not 2-Euclidean, and we stop here. Otherwise, it might be 2-Euclidean, and we try to “guess” a 2-Euclidean representation of the profile. If a representation is not found within a given timeout, the profile remains undecided.

The idea of this heuristic is obviously straightforward, and it is with no surprise that its naive implementation provides very poor performance. Our main contribution consists in developing an optimized version of this heuristic that can reasonably deal with profiles up to 8 candidates. We could also get some theoretical insights into the 2-

Euclidean domain.

In Chapter 7, we studied 2-Euclidean preferences with respect to the norm ℓ_1 . We recall that for $d = 2$, the notions of Euclidean profile with respect to the norm ℓ_1 and with respect to the norm ℓ_∞ are equivalent. We then focused on geometrical properties of 2-Euclidean representations with respect to the norm ℓ_1 , which yield several structural properties of this domain. Namely, we proved that any 2-Euclidean profile w.r.t. the norm ℓ_1 on 4 candidates contains at most 19 pairwise distinct preferences, and we actually conjecture that there is a unique maximal profile on 4 candidates. We proved that at most 4 candidates can be ranked last in a 2-Euclidean profile w.r.t. the norm ℓ_1 . We generalized this result for $d \geq 2$ - in such a case, there are at most $2d$ candidates ranked last if the norm ℓ_∞ is used, and 2^d candidates ranked last if the norm ℓ_1 is used. We also pointed out some similarities and differences between 2-Euclidean preferences w.r.t. the norm ℓ_1 and 2-Euclidean preferences w.r.t. the norm ℓ_2 .

8.3.1 A heuristic algorithm for recognizing 2-Euclidean preferences with respect to the norm ℓ_1 : further studies

The heuristic for recognizing 2-Euclidean preferences (with respect to the norm ℓ_2) studied in Chapter 6 could provide quite promising results. Several questions and possible improvements arise from this study. We now discuss some of them in more details.

Improving the NO-test of the heuristic

Our experiments show that improving the NO-certificates seems to be the key point to address. Indeed, although we cannot prove it formally, we believe that most of the profiles for which the algorithm cannot decide are actually not 2-Euclidean. Providing tighter necessary conditions on being 2-Euclidean would help to identify these profiles.

In its present form, the heuristic uses two NO-certificates. The first one uses the well-known formula (Bennett and Hays, 1960) giving the maximal number of pairwise distinct preferences in a 2-Euclidean profile in function of the number of candidates m . Any profile containing more pairwise distinct preferences cannot be 2-Euclidean. In practice, this NO-certificate is outperformed by the second NO-certificate based on the characterization of 2-Euclidean profiles on 4 candidates (Kamiya et al., 2011). There are actually 3 maximal 2-Euclidean profiles (w.r.t. the norm ℓ_2) on 4 candidates; hence, given a preference profile \mathcal{P} , any restriction of \mathcal{P} on four candidates must be a subprofile of one of the three maximal profiles.

A natural extension would be to use a characterization of 2-Euclidean profiles on 5 candidates. However, such a characterization does not yet exist in the literature. We could experimentally conjecture that there are 543 maximal 2-Euclidean profiles on 5 candidates. Proving this conjecture seems to be a challenging question. Moreover, it might be of limited interest for our heuristic: indeed, we have carried an additional experiment where we have assumed that our experimentally found characterization held,

turning it into a new NO-certificate (note that some 2-Euclidean input instances could then possibly be wrongly identified as non 2-Euclidean, as some maximal profiles are possibly missing in our list), and we did not observe any significant improvements of the recognition rate.

Similarly, we could identify some forbidden minors on 6 candidates (we could prove that these profiles on 6 candidates were not 2-Euclidean by using geometrical arguments showing that there was no possible representation mapping). Several dozens of such minors were turned into NO-certificates, but this did not improve significantly the recognition rate as well.

Put together, it seems that the NO-certificates based on forbidden minors (or on the characterizations of profiles with small number of candidates) are not efficient enough. One should probably identify another type of (more “structural”) conditions to obtain more tight certificates. This question seems difficult, although very stimulating.

Improving the YES-test of the heuristic

Although the YES-test of the heuristic was quite optimized (and seems indeed to provide decent performances, especially for profiles with up to 8 candidates), there is still room for improvement.

From the theoretical point of view, one could reiterate the idea that for any representation function f and any candidate c_i ranked last at least once, $f(c_i)$ is a vertex of the convex hull of the set $\{f(c_1), f(c_2), \dots, f(c_m)\}$. In practice, the positions of the candidates ranked last at least once are chosen to form a polygon, and the remaining positions are picked at random inside this polygon.

But if we denote by C_L the set of candidates ranked last at least once, the same idea applies on the profile \mathcal{P}_L restricted on $C \setminus C_L$. Therefore, we can do better than picking the positions of candidates in $C \setminus C_L$ at random - we can reiterate the process and determine the set C_{L_2} of candidates ranked last at least once in \mathcal{P}_L . The positions of the candidates of C_{L_2} will then be chosen to form a polygon (inside the “first” polygon with vertices corresponding to the positions of the candidates of C_L), and we will consider a new profile restriction \mathcal{P}_{L_2} on the set $C_L \setminus C_{L_2}$.

The idea of an “iterative” convex-hull seems promising, because we could see that the heuristic performed very badly if all positions were picked at random, without taking into account the size of convex-hull.

More generally, our code was developed and optimized to deal with profiles involving at most 9 candidates. Significant code changes would be required if we want to make it operational for profiles involving more candidates.

Improving the experimental studies and the interpretability of the results

We have performed numerical experiments on data sets available in Preflib (Mattei and Walsh, 2013). Unsurprisingly, none of these instance has been recognized as 2-Euclidean⁵. More precisely, the NO-test was in general sufficient to conclude that a given instance were not 2-Euclidean. Consequently, these experiments do not allow us to conclude about the performance of the YES-test on real-world instances. To improve our study, we could consider as input subprofiles of Preflib instances. Indeed, with smaller profiles, we have more chance to identify 2-Euclidean profiles. Moreover, we have seen that our algorithm seems to struggle on profiles with about 10 voters and 10-20 candidates. Providing more instances of this size could possibly support this observation even more, or identify more precisely the range of values m and n for which the algorithm struggles. Note that another main difficulty we encounter is that we do not know if an input (real-world) profile is 2-Euclidean or not. Hence, if such a profile remains undecided, there is no way to know if it is because of the NO-test that could not be able to detect a non-Euclidean profile, or because of the YES-test that was not able to find a convenient 2-Euclidean representation. Therefore, it is quite challenging to independently evaluate the performances of the NO-test and the performances of the YES-test.

Also, it would be meaningful to better position our heuristic with respect to multi-dimensional unfolding algorithms (see for instance Busing et al. (2005)). Although this issue may seem straightforward, it is not clear how it should be undertaken. Actually, there are several algorithms that, given a preference profile, provide an approximate 2-Euclidean representation of it. Note that the starting point is not the same: actually, these algorithms do not tackle the decision question of whether or not a given profile is 2-Euclidean. In particular, they do not allow to distinguish 2-Euclidean instances from non-Euclidean ones (they are not guaranteed to return an exact representation when the profile is 2-Euclidean), so we cannot define in a simple way a proper recognition rate for them, and compare it with our (exact) method.

Last but not least, the question of interpretability of the representation returned by the heuristic deserves attention. Given a 2-Euclidean profile, the heuristic returns a *random* representation of it. In other words, different executions of the algorithm yield different representations of the same profile. A natural question is whether all these representations share some common structural aspects. For instance, we know that in any representation f , each candidate ranked last at least once will lie on the convex hull of the set $\{f(c_1), f(c_2), \dots, f(c_3)\}$. Are there other theoretical/geometrical properties verified by any possible representation? This challenging question should definitely be studied: any element of answer would shed light into the structural comprehension of the 2-Euclidean domain.

⁵except those that were trivially 2-Euclidean.

8.3.2 d -Euclidean preferences with respect to the norm ℓ_1 and ℓ_∞

In Chapter 7, we studied 2-Euclidean preferences with respect to the norm ℓ_1 and ℓ_∞ . So far, this domain restriction has not received much attention in the literature. The novelty of the topic opens hence several exciting, mostly middle-term or even long-term research directions.

Characterization of profiles on 4 candidates: maximal profile size

A first natural step would be to prove the conjecture that there is a unique maximal profile on 4 candidates under the norm ℓ_1 (containing 19 pairwise distinct preferences). We have conducted a preliminary study of this question. At first sight, the conjecture could be proven numerically - one could enumerate all possible profiles of 19 (pairwise distinct) preferences, and check (by using an ILP formulation) if each of this profile is 2-Euclidean with respect to the norm ℓ_1 . Putting aside computational aspects of this method, there is a serious theoretical issue that makes this approach insufficient.

Actually, there is no reason that all (inclusion-wise) maximal 2-Euclidean profiles contain the same number of pairwise distinct preferences. We recall that, for instance, in case of single-peaked preferences, it is well-known that any maximal profile⁶ on m candidates contains 2^{m-1} pairwise distinct preferences. Any maximal single-crossing (resp. 1-Euclidean) profile on m candidates contains $\frac{m(m-1)}{2} + 1$ pairwise distinct preferences. Last but not least, Bennett and Hays (1960) provides a formula for the size of any maximal 2-Euclidean profile with respect to the norm ℓ_2 . However, an analogous result is not known for 2-Euclidean preferences with respect to the norm ℓ_1 .

This question seems quite challenging. In the case of the ℓ_2 norm, we could enumerate all 2-intersections and 3-intersections of boundary hypersurfaces. We could then use the Euler's formula to obtain the number of different preference areas (see Chapter 7 for more details). In the case of the ℓ_1 norm, it is not easy to establish the (maximal) number of 2-intersections and 3-intersections; we have seen in Chapter 7 that this task was already quite complex with only 4 candidates. Note also that these numbers may vary from one representation to another, while there was always the same number of intersections for any non-degenerate representation in the sense of ℓ_2 .

There is another issue to deal with in the case of the ℓ_1 norm. Assume that we have somehow enumerated all 2-intersections and 3-intersections of boundary hypersurfaces. We can then count the number of different areas into which the plane is divided. However, contrary to the case of the ℓ_2 norm, two different areas can correspond to the same preference - see for instance Figure 8.3.1 where there are two different areas corresponding to the same preference (c_4, c_3, c_1, c_2) . In other words, the number of areas provides an upper bound on the number of pairwise distinct preferences, but this

⁶Actually, there is a unique maximal profile, up to a renaming of candidates.

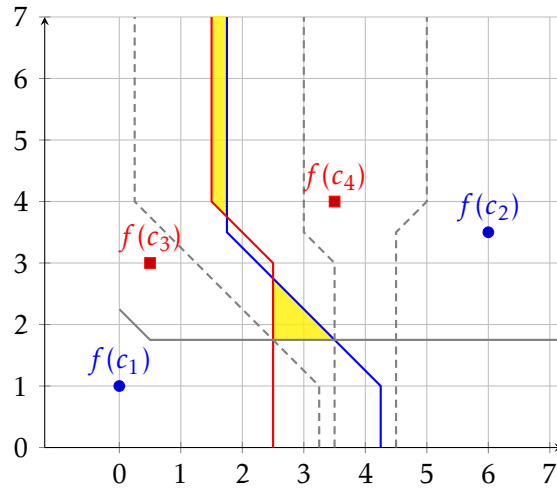


Figure 8.3.1: A 2-Euclidean representation f with 4 candidates. Both yellow areas correspond to the same preference (c_4, c_3, c_1, c_2) . The red hypersurface corresponds to $H(c_3, c_4)$ and the blue one to $H(c_1, c_2)$. For a simpler and less crowded look, the other hypersurfaces are not labeled.

bound may not be reached.

We started with the question of whether or not all maximal profiles contain the same number of pairwise distinct preferences, and we have identified several related questions. This example illustrates that the domain of 2-Euclidean preferences with respect to the norm ℓ_1 remains quite unexplored, and offers a very beautiful and challenging research direction that is definitely worth being followed.

Characterization of profiles on 4 candidates: the 2-Euclidean domain w.r.t. ℓ_1 and the permutohedron

We have actually investigated the question of characterization of profiles on 4 candidates. The maximal profile containing 19 pairwise preferences corresponds to the subgraph of permutohedron that is in red in Figure 8.3.2 (up to the renaming of candidates). We denote this subgraph by \mathcal{G}^* in which follows.

To prove that this profile is the unique maximal 2-Euclidean profile w.r.t. ℓ_1 on 4 candidates, it would be sufficient to prove that any 2-Euclidean profile corresponds to a subgraph of \mathcal{G}_0^* . More generally, it would be interesting to characterize the 2-Euclidean domain with respect to the norm ℓ_1 as a set of subgraphs of the permutohedron. We recall that analogous problems were studied for instance for single-peaked or single-crossing preferences.

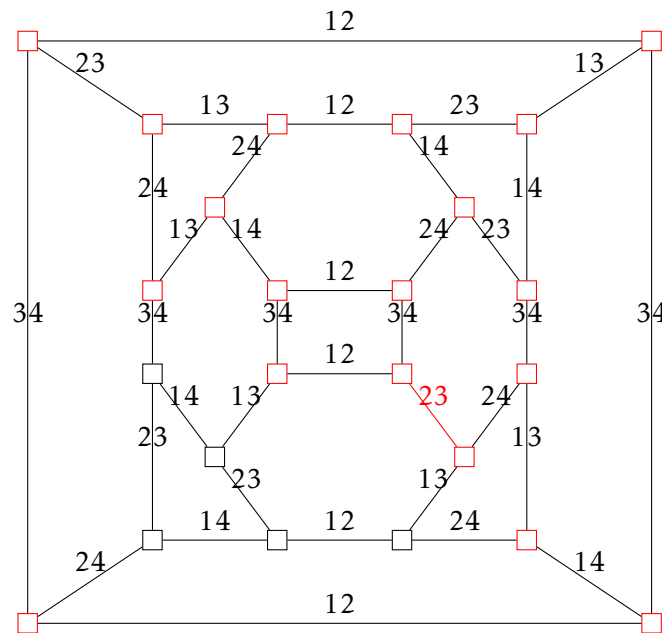


Figure 8.3.2: The maximal 2-Euclidean profile \mathcal{P}_0^* as a subgraph \mathcal{G}_0^* (red vertices) of the permutohedron.

Recognition of 2-Euclidean preferences with respect to the norm ℓ_1

While Peters (2017) showed that the recognition problem for 2-Euclidean preferences w.r.t. ℓ_1 is in NP, the question of whether or not the problem is polytime solvable remains open. Answering this question is another interesting research topic. In particular, it could be insightful (from the structural point of view) to develop a, hopefully efficient, recognition algorithm for this domain restriction. We note that a parallel with the single-peaked domain can be found: there are at most 2 candidates ranked last in a single-peaked profile. One of the underlying ideas of the recognition algorithm proposed by Doignon and Falmagne (1994) for single-peaked preferences consists in identifying these candidates corresponding to the extremities of a single-peaked axis. We have seen that there are at most 4 candidates ranked last in any 2-Euclidean profile w.r.t. the norm ℓ_1 . Maybe this information could be used to design a recognition algorithm.

d -Euclidean preferences with respect to the norm ℓ_1 and ℓ_∞

Finally, a very natural, broad research direction is to study d -Euclidean preferences for $d > 2$. For instance, we have seen that there were profiles that are d -Euclidean with respect to the norm ℓ_1 but not with respect to the norm ℓ_∞ . Does the opposite hold? Answering this question would require a deeper structural comprehension of both d -

Euclidean preferences w.r.t. to norm ℓ_1 and w.r.t. the norm ℓ_∞ . This questions seems very challenging, as we have seen that already for $d = 2$, the encountered problems were anything but easy.

Bibliography

- James Abello. The weak bruhat order of S_n , consistent sets, and catalan numbers. *SIAM Journal on Discrete Mathematics*, 4(1):1–16, 1991.
- Ravindra K Ahuja, Thomas L Magnanti, and James B Orlin. *Network flows*. Alfred P. Sloan School of Management, Massachusetts Institute of Technology, 1988.
- Kenneth Arrow. *Social Choice and Individual Values*. John Wiley & Sons, Chapman & Hall, 1951.
- Giorgio Ausiello, Alessandro D’Atri, and Marco Protasi. Structure preserving reductions among convex optimization problems. *J. Comput. Syst. Sci.*, 21(1):136–153, 1980.
- Miguel A Ballester and Guillaume Haeringer. A characterization of the single-peaked domain. *Social Choice and Welfare*, 36(2):305–322, February 2011.
- Salvador Barberà, Faruk Gul, and Ennio Stacchetti. Generalized median voter schemes and committees. *Journal of Economic Theory*, 61(2):262–289, 1993.
- Salvador Barberà and Bernardo Moreno. Top monotonicity: A common root for single peakedness, single crossing and the median voter result. *Games and Economic Behavior*, 73(2):345–359, 2011. ISSN 0899-8256.
- John Bartholdi, Craig A Tovey, and Michael A Trick. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6(2):157–165, 1989.
- John Bartholdi III and Michael A Trick. Stable matching with preferences derived from a psychological model. *Operations Research Letters*, 5(4):165–169, 1986.
- Joseph F. Bennett and William L. Hays. Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, 25(1):27–43, Mar 1960.
- D. Black. *The Theory of Committees and Elections*. Cambridge University Press, 1958.
- Duncan Black. On the rationale of group decision-making. *The Journal of Political Economy*, 56(1):23–34, 1948.

- Anna Bogomolnaia and Jean-Francois Laslier. Euclidean preferences. *HAL, Working Papers*, 01 2004.
- Kellogg S. Booth and George S. Lueker. Testing for the consecutive ones property, interval graphs, and graph planarity using pq-tree algorithms. *Journal of Computer and System Sciences*, 13(3):335–379, 1976. ISSN 0022-0000.
- Ingwer Borg and Patrick Groenen. *Modern Multidimensional Scaling: Theory and Applications (Springer Series in Statistics)*. 08 2005.
- Sylvain Bouveret, Renaud Blanch, Antoinette Baujard, François Durand, Herrade Igersheim, Jérôme Lang, Annick Laruelle, Jean-François Laslier, Isabelle Lebon, and Vincent Merlin. Voter autrement 2017 - online experiment, July 2018.
- Felix Brandt, Markus Brill, Edith Hemaspaandra, and Lane A Hemaspaandra. Bypassing combinatorial protections: Polynomial-time algorithms for single-peaked electorates. *J. of Art. Intelligence Research*, 53:439–496, 2015.
- Robert Brederick, Jiehua Chen, and Gerhard J. Woeginger. A characterization of the single-crossing domain. *Social Choice and Welfare*, 41(4):989–998, 2013. ISSN 01761714, 1432217X.
- Robert Brederick, Jiehua Chen, and Gerhard J Woeginger. Are there any nicely structured preference profiles nearby? *Mathematical Social Sciences*, 79:61–73, 2016.
- Laurent Bulteau and Jiehua Chen. 2-dimensional Euclidean preferences, 2022.
- Frank Busing, Patrick Groenen, and Willem Heiser. Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, 70: 587–587, 09 2005.
- Timothy M Chan and Mihai Pătraşcu. Counting inversions, offline orthogonal range counting, and related problems. In *Proceedings of the 21st annual ACM-SIAM symposium on Discrete Algorithms (SIAM)*, pages 161–173, 2010.
- Jiehua Chen and Sanjukta Roy. Multi-dimensional stable roommates in 2-dimensional Euclidean space. In *Proceedings of the 30th Annual European Symposium on Algorithms (ESA)*, 2022.
- Jiehua Chen, Kirk Pruhs, and Gerhard J. Woeginger. The one-dimensional Euclidean domain: Finitely many obstructions are not enough. *Social Choice and Welfare*, 48(2): 409–432, 2017.
- Jiehua Chen, Martin Nöllenburg, Sofia Simola, Anaïs Villedieu, and Markus Wallinger. Multidimensional Manhattan preferences, 2022.

- Adam Clearwater, Clemens Puppe, and Arkadii Slinko. Generalizing the single-crossing property on lines and trees to intermediate preferences on median graphs. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJ-CAI)*, 2015.
- Vincent Conitzer. Computing slater rankings using similarities among candidates. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence (AAAI)*, pages 613–619, 2006.
- Vincent Conitzer. Eliciting single-peaked preferences using comparison queries. *J. of Art. Intelligence Research*, 35:161–191, 2009.
- Vincent Conitzer and Tuomas Sandholm. Communication complexity of common voting rules. EC '05, page 78–87, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 1595930493.
- C.H. Coombs. *A Theory of Data*. Wiley, 1964.
- Clyde H Coombs. Psychological scaling without a unit of measurement. *Psychological Review*, 57(3):145–158, 1950.
- Bernard Debord. Caractérisation des matrices des préférences nettes et méthodes d'agrégation associées. *Mathématiques et sciences humaines*, 97:5–17, 1987.
- Gabrielle Demange. Single-peaked orders on a tree. *Math. Social Sciences*, 3(4):389–396, 1982.
- J.P. Doignon and J.C. Falmagne. A polynomial time algorithm for unidimensional unfolding representations. *Journal of Algorithms*, 16(2):218–233, 1994. ISSN 0196-6774.
- Anthony Downs. *An economic theory of democracy*. Harper & Row New York, 1957.
- Edith Elkind and Piotr Faliszewski. Recognizing 1-euclidean preferences: An alternative approach. In *International Symposium on Algorithmic Game Theory*, pages 146–157. Springer, 2014.
- Edith Elkind and Martin Lackner. On detecting nearly structured preference profiles. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014.
- Edith Elkind, Piotr Faliszewski, and Arkadii Slinko. Clone structures in voters' preferences. In *Proceedings of the 13th ACM conference on electronic commerce*, pages 496–513, 2012.
- Edith Elkind, Martin Lackner, and Dominik Peters. Structured preferences. In *Trends in computational social choice*, chapter 10, pages 187–207. AI Access, 2017.
- Edith Elkind, Piotr Faliszewski, and Piotr Skowron. A characterization of the single-peaked single-crossing domain. *Social Choice and Welfare*, 54(1):167–181, Jan 2020. ISSN 1432-217X.

- Edith Elkind, Martin Lackner, and Dominik Peters. Preference restrictions in computational social choice: A survey, 2022.
- Gábor Erdélyi, Martin Lackner, and Andreas Pfandler. Computational aspects of nearly single-peaked electorates. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI)*, pages 283–289, 2013. An extended version appeared in *J. of AI Research* in 2017.
- Gábor Erdélyi, Martin Lackner, and Andreas Pfandler. Computational aspects of nearly single-peaked electorates. *J. of Art. Intelligence Research*, 58:297–337, 2017.
- Bruno Escoffier, Jérôme Lang, and Meltem Öztürk. Single-peaked consistency and its complexity. In *ECAI*, volume 8, pages 366–370, 2008.
- Bruno Escoffier, Olivier Spanjaard, and Magdaléna Tydrichová. Recognizing single-peaked preferences on an arbitrary graph: Complexity and algorithms. In *Proceedings of the 13th International Symposium on Algorithmic Game Theory (SAGT)*, pages 291–306, 2020.
- Bruno Escoffier, Olivier Spanjaard, and Magdaléna Tydrichová. Measuring Nearly Single-Peakedness of an Electorate: Some New Insights. In *Proceedings of the 7th International Conference on Algorithmic Decision Theory (ADT)*, pages 19–34, November 2021.
- Bruno Escoffier, Olivier Spanjaard, and Magdaléna Tydrichová. Euclidean preferences in the plane under ℓ_1 , ℓ_2 and ℓ_∞ norms, 2022a.
- Bruno Escoffier, Olivier Spanjaard, and Magdaléna Tydrichová. Weighted majority tournaments and Kemeny ranking with 2-dimensional Euclidean preferences. *Discrete Applied Mathematics*, 318:6–12, 09 2022b.
- Piotr Faliszewski, Edith Hemaspaandra, and Lane A Hemaspaandra. The complexity of manipulative attacks in nearly single-peaked electorates. In *TARK*, pages 228–237. ACM, 2011. an extended version appeared in *Artif. Intel.* in 2014.
- Scott L Feld and Bernard Grofman. Research note partial single-peakedness: An extension and clarification. *Public Choice*, 51(1):71–80, 1986.
- Ádám Galambos and Victor Reiner. Acyclic sets of linear orders via the bruhat orders. *Social Choice and Welfare*, 30(2):245–264, Feb 2008. ISSN 1432-217X.
- Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., NY, USA, 1979. ISBN 0716710447.
- Diego Garzia and Stefan Marschall. *Voting advice applications*. Oxford University Press, 2019.
- Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4): 587–601, 1973. ISSN 00129682, 14680262.

- Michał T. Godziszewski, Paweł Batko, Piotr Skowron, and Piotr Faliszewski. An analysis of approval-based committee rules for 2d-euclidean elections. *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 35(6):5448–5455, May 2021.
- IJ Good and TN Tideman. Stirling numbers and a geometric structure from voting theory. *Journal of Combinatorial Theory, Series A*, 23(1):34–45, 1977.
- Jean-Michel Grandmont. Intermediate preferences and the majority rule. *Econometrica*, pages 317–330, 1978.
- GT Guilbaud and P Rosenstiehl. Analyse algébrique d’un scrutin. *Math. Sci. Hum*, 4: 9–33, 1963.
- William L. Hays and Joseph F. Bennett. Multidimensional unfolding: Determining configuration from complete rank order preference data. *Psychometrika*, 26(2):221–238, Jun 1961.
- Harold Hotelling. Stability in competition. *Economic Journal*, 39:41–57, 1929.
- Veikko Isotalo. Designing voting advice applications: The finnish case. Master’s thesis, 2020.
- Dejan Jovanović and Leonardo de Moura. Solving non-linear arithmetic. In Bernhard Gramlich, Dale Miller, and Uli Sattler, editors, *Automated Reasoning*, pages 339–354, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-31365-3.
- Hidehiko Kamiya, Akimichi Takemura, and Hiroaki Terao. Ranking patterns of unfolding models of codimension one. *Advances in Applied Mathematics*, 47(2):379–400, 2011.
- John G Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.
- Vicki Knoblauch. Recognizing one-dimensional Euclidean preference profiles. *Journal of Mathematical Economics*, 46:1–5, 01 2010.
- Jozef Kratica, Marija Milanovic, and Djordje Dugošija. A mixed integer linear programming formulation of the maximum betweenness problem. *European Journal of Operational Research*, 206:522–527, 11 2010.
- J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, Mar 1964.
- D. T. Lee and C. K. Wong. Voronoi diagrams in $l_1(l_\infty)$ metrics with 2-dimensional storage applications. *SIAM Journal on Computing*, 9(1):200–211, 1980.
- Nicholas Mattei and Toby Walsh. Preflib: A library of preference data [HTTP://PREFLIB.ORG](http://preflib.org). In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT 2013)*, pages 259–270, 2013.

- David C McGarvey. A theorem on the construction of voting paradoxes. *Econometrica*, pages 608–610, 1953.
- J. A. Mirrlees. An Exploration in the Theory of Optimum Income Taxation¹². *The Review of Economic Studies*, 38(2):175–208, 04 1971. ISSN 0034-6527.
- Bernard Monjardet. *Acyclic Domains of Linear Orders: A Survey*, pages 139–160. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-79128-7.
- Guillermo Romero Moreno, Javier Padilla, and Enrique Chueca. Learning VAA: A new method for matching users to parties in voting advice applications. *Journal of Elections, Public Opinion and Parties*, 32(2):339–357, 2022. ISSN 1745-7289.
- H. Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980. ISSN 00485829, 15737101.
- Klaus Nehring and Clemens Puppe. The structure of strategy-proof social choice—part i: General characterization and possibility results on median spaces. *J. of Economic Theory*, 135(1):269–305, 2007.
- C.C. Nembua, P.R.), J.P. Barthélemy, M. do Rosário Rolfsen Salles, and B. Monjardet. *Permutação et choix social*. 1989.
- Richard G Niemi. Majority decision-making with partial unidimensionality. *American Political Science Review*, 63(2):488–497, 1969.
- Dominik Peters. Recognising multidimensional Euclidean preferences. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pages 642–648, 2017.
- Dominik Peters. Single-peakedness and total unimodularity: New polynomial-time algorithms for multi-winner elections. In *Proceedings of the 32nd Conference on Artificial Intelligence (AAAI)*, pages 1169–1176, 2018.
- Dominik Peters and Edith Elkind. Preferences single-peaked on nice trees. In *AAAI*, pages 594–600, 2016.
- Dominik Peters and Martin Lackner. Preferences single-peaked on a circle. In *Proceedings of the 31st Conference on Artificial Intelligence (AAAI)*, pages 649–655, 2017.
- Dominik Peters and Martin Lackner. Preferences single-peaked on a circle. *Journal of Artificial Intelligence Research*, 68:463–502, 2020.
- Clemens Puppe. The single-peaked domain revisited: A simple global characterization. *Journal of Economic Theory*, 176:55–80, 2018. ISSN 0022-0531.
- Clemens Puppe and Arkadii Slinko. Condorcet domains, median graphs and the single crossing property. *Economic Theory*, 67, 02 2019.

-
- Kevin WS Roberts. Voting over income tax schedules. *Journal of Public Economics*, 8(3): 329–340, 1977.
- Paul Rothstein. Representative voter theorems. *Public Choice*, 72(2):193–212, 1991.
- Mark Allen Satterthwaite. Strategy-proofness and arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975. ISSN 0022-0531.
- Dana Scott. Measurement structures and linear inequalities. *Journal of Mathematical Psychology*, 1(2):233–247, 1964. ISSN 0022-2496.
- Patrick Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48(3/4): 303–312, 1961.
- J Sliwinski and E Elkind. Preferences single-peaked on a tree: Sampling and tree recognition. pages 580–586. Association for the Advancement of Artificial Intelligence, 2019.
- Richard P Stanley et al. An introduction to hyperplane arrangements. *Geometric combinatorics*, 13(389-496):24, 2004.
- Xin Sui, Alex Francois-Nienaber, and Craig Boutilier. Multi-dimensional single-peaked consistency and its approximations. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 375–382, 2013.
- Michael A Trick. Recognizing single-peaked preferences on a tree. *Mathematical Social Sciences*, 17(3):329–334, 1989.
- P. Valtr. Probability thatn random points are in convex position. *Discrete & Computational Geometry*, 13(3):637–643, Jun 1995. ISSN 1432-0444.
- H. Peyton Young. Condorcet’s theory of voting. *American Political science review*, 82(4): 1231–1244, 1988.
- Lan Yu, Hau Chan, and Edith Elkind. Multiwinner elections under preferences that are single-peaked on a tree. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 425–431, 2013.

... lopussa kiitos seisoo. (Finnish proverb)