Methods for analyzing routing games

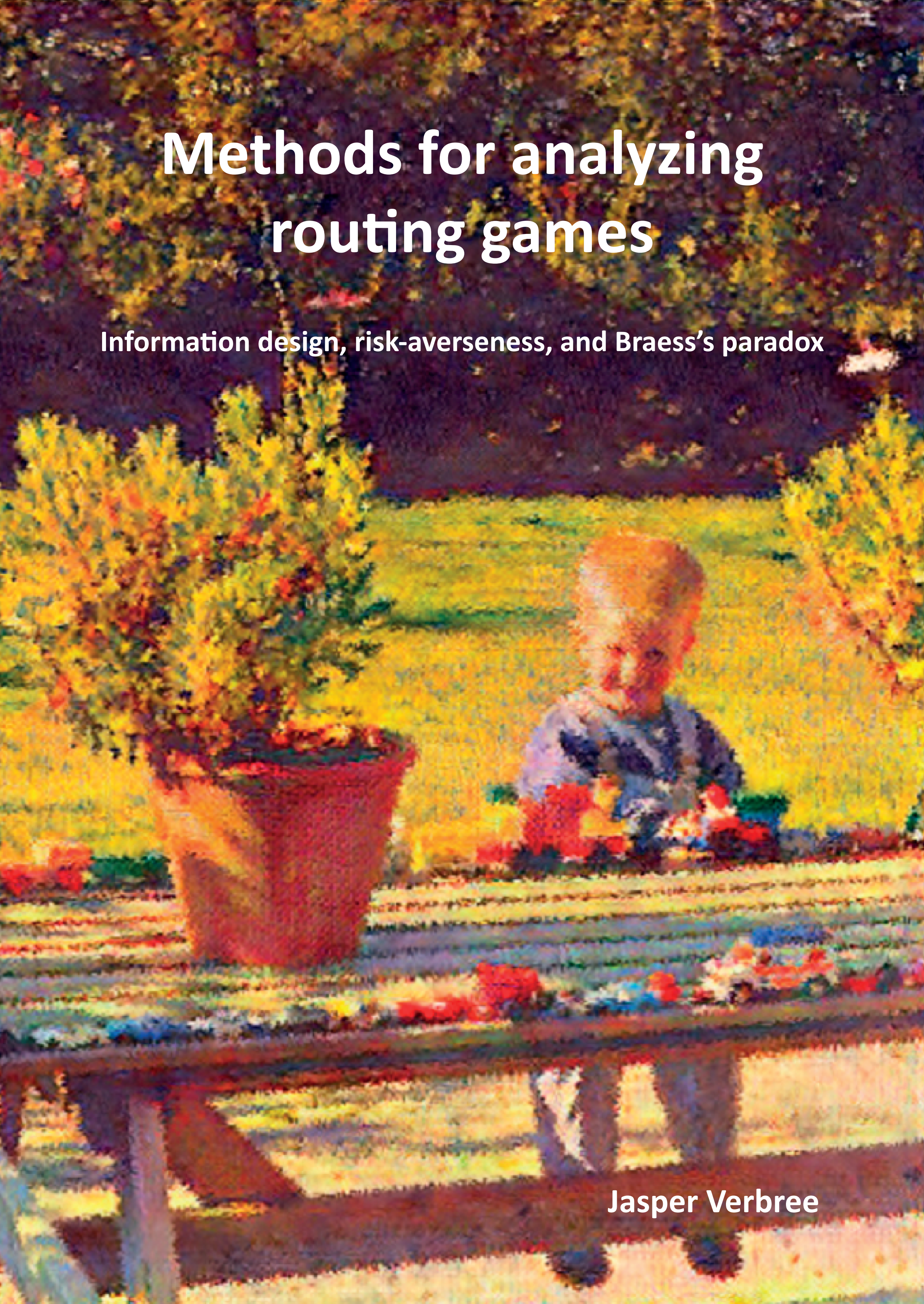
Verbree, Jasper

DOI:
10.33612/diss.872396326

Publication date:
2024

Methods for analyzing routing games

Information design, risk-averseness, and Braess's paradox



Jasper Verbree

Book cover: The author routing traffic in his youth.

ISBN/EAN: 978-94-6473-355-6

Published by Ipskamp Printing

The research described in this thesis has been carried out at the Faculty of Science and Engineering, University of Groningen, the Netherlands. This thesis was completed using the thesis LaTeXtemplate by Drs. K. Bunte, A. Taghribi, and E. Talavera Martinez

Methods for analyzing routing games

Information design, risk-averseness, and Braess's paradox

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. ir. J.M.A. Scherpen
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

maandag 5 februari 2024 om 11.00 uur

door

Jasper Verbree

geboren op 3 februari 1995

Promotor
Prof. dr. D. Bauso

Copromotor
Dr. A.K. Cherukuri

Beoordelingscommissie
Prof. dr. J.G. Peypouquet
Prof. dr. M. Cao
Prof. dr. B. Gharesifard

Contents

Acknowledgments

"The end is near." These are the first words that go through my mind as I start writing this last, final part of my thesis. Perhaps it is overly dramatic, but I've always liked to put a bit more drama in my writing than strictly necessary, or even wise. At least it is true. The end is near. The end of my time as a Ph.D. student here in Groningen. The end of my project, which has been such a major part of my life for these last four years. The end of a chapter. Endings have their good sides, of course. Pursuing a Ph.D can be like an island, a tiny region of the world of research that you explore in excruciating detail. The chance to now expand my view, look at a broader picture, and start something new is an exciting prospect. There were also times that the work was hard, when I was exhausted and plodding on seemed too daunting a task. To put the finishing touches on a project I once doubted I could complete is very satisfying. However, when I look back at my time as a Ph.D. student, the good times outnumber the bad by quite a margin, and so I would like to take some time to thank all of those who helped make the last few years so enjoyable.

First of all there is my daily supervisor, Ashish Cherukuri. Thank you, Ashish, for all the time you have given me. At the start of my project my weeks had a clear pattern to them. During the week I would slow down, as I encountered new problems and challenges, until we had our weekly meeting and I would leave energized and full of new ideas. Thank you for that. Later on, when weekly meetings were not required anymore, you were still always available when I needed you. You gave me the freedom to work on my own ideas and projects, which I appreciate immensely, and which is one of the main reasons I so enjoyed my Ph.D. project.

I also want to thank my promotor, Dario Bauso. You gave me some important pointers on communication, which gave me some useful insights into myself.

Next, I want to express my gratitude toward the members of the reading committee: Ming Cao, Bahman Gharesifard, and Juan Peypouquet. Thank you for your time and feedback!

Then there are my office mates. Wouter Baar, Emmanuel de la Cruz Piña, Bo Jin and Amin Maghami. The corona period was not the ideal time to bond in the office, but for what little time we had, I enjoyed your company a lot. In the same line of thought I want to thank my friends, Henk and Tessa van Waarde and Lammert and Charlotte Wiersma, who so generously loaned me office space in their homes when the university was closed, and stood in for the office mates that I therefore had to miss. Thank you guys, I would probably not have finished this project without you.

I also want to thank my family, my father and mother, my brothers and sisters. In law or otherwise. You kept me stable throughout these times, lifted me up when I was down, and I am thankful that I got to do this project here in Groningen, with all of you so (relatively) close.

And then there are my colleagues. So many that I am sure to fail to name them all here, but let me try anyway. Agung, Alireza, Arijit, Auriane, Azin, Bahadir, Bangguo, Brenda, Emin, Frederika, Hayden, Ivett, Juan, Julien, Kathinka, Linda, Lorenzo, Luis, Marco, Matthijs, Mireny, Najmeh, Ping, Rafael, Rhido, Rory, Saeed, Santiago, Sepide, Simon, Stefanny, Taraneh, Vaibhav, Yifan, Yvonne, Zaki, and probably a lot more. Thank you all for making the office lively. A special mention goes to Nikolai Charalampadis, Anne-men Huijser, Jasper Koning, Brayan Shali, and Paul Wijnbergen, without whom I may never have started this project at all. I also want to say thanks, again, to Emin and Anne-men, my paranymphs, who doubtlessly will get me through the final defense without too much trouble.

Finally, my greatest thanks go to my wife, Joanne, who I am quite sure is the loveliest person alive, and my son Adam, who is working hard to change that. You are the ground beneath my feet and the sun in my sky. I would be lost without you.

Jasper Verbree
Groningen
January 4, 2024

Chapter 1

Introduction

Networks are a more and more prevalent part of our daily life. Think, for instance, of traffic networks, power grids, water and gas distribution systems, and communication networks such as the internet. All of these have become an integral part of our world. As our dependency on these systems grows, so do the networks themselves. They increase in size and complexity, and as a result the need for efficient ways of designing and regulating these networks grows.

One way in which the efficient usage of these networks is hindered is through the use of competing agents. Multiple parties want to transport their product over the network for the smallest cost possible, not taking into account the extra costs their usage imposes on the other users of the network. We are all too familiar with the additional travel time caused by congestion on the road, a phenomenon that could sometimes be alleviated by proper coordination of the traffic flow. This perspective of 'selfish' agents making use of a shared network to facilitate their transport objectives, while aiming to minimize their own costs of transportation gives rise to the subject of *routing games*. The inefficiencies caused by the competition among agents create the potential for a social planner or controller to step in and be of benefit to all users by regulating the design, usage, and information provision of these systems. In this thesis, we investigate several questions related to the efficient design and control of routing games. Since all chapters share this overarching theme, Chapter 2 is devoted to giving a thorough introduction to the subject, including some notation, definitions, and preliminary results necessary for the rest of this thesis.

1.1 Braess's paradox

One type of inefficiency that can occur in routing games is called *Braess's paradox*, named after its discoverer Dietrich Braess, who first presented it in 1968 [1](see [2] for an English translation). The phenomenon occurs when the cost that all agents experience when traversing a network can be reduced by removing one or more connections in the network. It is a fascinating phenomenon that has inspired a large body of research since its discovery. However, the problem of avoiding the paradox when designing networks, and indeed even that of detecting the presence of the paradox in a given routing game has proven to be difficult. It is this difficulty, and the hope to alleviate it, that motivates the material presented in Chapters 3 and 4. In Chapter 3 we give a rigorous analysis of how the choices of agents using a network change as the amount of *demand* that is routed over the network changes. Although results in Chapter 3 were initially derived in the search for methods for alleviating Braess's paradox, there are some surprising observations on how the usage of the network can be easily derived for very large levels of demand. These results are interesting in their own right and warrant their own presentation, and thus we have chosen to present Chapter 3 separately from Chapter 4, where we delve into the subject of Braess's paradox proper. In Chapter 4 we make use of the results of Chapter 3 to derive a number of conditions, which are computationally feasible to check, and which reveal the presence of Braess's paradox. We also show the necessary and sufficient condition for the presence of Braess's paradox that underlies all of these conditions, which is less useful in practice but may be valuable for further analysis. Finally we show how the obtained insights change the perspective on Braess's paradox. Although often presented as a problematic phenomenon, the paradox is fundamentally a matter of balance. Removing a link from a network may improve efficiency at one level of demand, but only if it decreases efficiency for another level of demand. Whether the presence of a link is beneficial or detrimental then depends on how the demand on the network varies, and on what measure one uses to determine the 'value' of a link, as we demonstrate in the last part of Chapter 4.

1.2 Inferring the prior

The phenomenon of Braess's paradox shows that sometimes a network becomes more efficient after a link is removed. Similarly, sometimes all participants of a routing game are better off when some *information* is withheld from them. This creates the opportunity for *information design*, where a central authority called a *travel information system*(TIS), who has more information about the network than the agents, can release this information strategically in order to promote efficient usage of the network. However, it is reasonable to assume that the behaviour of the participants does not only depend on the information supplied by the TIS but also on their own beliefs and preferences, which the TIS does not necessarily have access to. To optimize information provision, the TIS thus needs a way to learn about these private parameters of the participants.

This is the issue under consideration in Chapter 5. In it, we study routing games subject to *uncertainty*. Costs associated with the usage of a link now include a stochastic term, and as a consequence, participants can only make estimations about what the cost of a particular route through the network is. This estimation depends on a prior belief the participants hold on the state that the network is in, and on the information supplied by a TIS. The situation is meant to model a routing game over a traffic network, where costs can be subject to uncertainty (think of weather conditions or accidents on the road), and drivers make use of navigation software to inform them about the state of the network. The TIS would like to supply information in such a way that the average travel time of all participants is minimized, but is hindered by the fact that the choices of the drivers also depend on their own beliefs, called the *prior belief*, which the TIS does not necessarily have access to. We study how the choices of the participants can be used by the TIS to infer information about the prior belief they hold. We investigate under what conditions information provision strategies can be designed that allow the TIS to fully identify this belief, and subsequently show how to design these strategies.

1.3 Risk-based variational inequalities

Like Chapter 5, Chapter 6 is motivated by routing games which are subject to uncertainty. Fundamental in the study of routing games are *variational inequalities*(VI), which can be used to compute equilibrium solutions to routing games, in which no users can decrease their experience cost by changing their strategy. In Chapter 6 we investigate the case where participants try to minimize their uncertain cost in a risk-averse manner, where they use the *conditional-value-at-risk*(CVaR) as a risk measure. The result is a *CVaR-based* VI, involving the CVaR of random costs. To properly design and regulate games involving these CVaR-based VIs, a planner would need a way to solve these VIs. However, finding a solution of such a VI is hindered by the fact that no *unbiased* estimators of the CVaR of a random variable are available. Thus we need to investigate the potential of some standard methods for solving VI, or stochastic VI, when applied in the context of biased estimators. We study three different schemes for finding solutions of CVaR-based VI and show that as long as enough samples of the uncertainty are taken, a solution can be obtained up to any level of desired accuracy. We also provide an explicit relation between the number of samples taken, and the achieved accuracy of the solution, and compare the performance of the different methods using an example of a routing game.

1.4 Notation

Here we introduce part of the general notation that we use throughout this thesis. The concept of *routing games* and related notation is introduced in Chapter 2. More specific instances of notation, that are less common, or only relevant to one or two chapters, will be defined when they are first encountered.

The set of natural and real numbers are denoted by $\mathbb{N}$ and $\mathbb{R}$ respectively. We also use $\mathbb{R}_{\geq 0}$ to denote the set of non-negative real numbers, and for the empty set we write $\emptyset$. For a given $n \in \mathbb{N}$, we write $[n] = \{1, 2, \cdots, n\}$ and $[n]_0 = \{0, 1, 2, \cdots, n\}$. For a finite set $\mathcal{S}$, the number of elements in $\mathcal{S}$ is $|\mathcal{S}|$. For a subset $\mathcal{S} \subseteq \mathcal{P}$ the complement of $\mathcal{S}$ in $\mathcal{P}$ is denoted $\mathcal{S}^c$ or $\mathcal{P} \setminus \mathcal{S}$; i.e. $\mathcal{S}^c := \mathcal{P} \setminus \mathcal{S} := \{p \in \mathcal{P} \mid p \notin \mathcal{S}\}$. The closure of a set $\mathcal{S} \subseteq \mathbb{R}^n$ is denoted by $\mathrm{cl}(\mathcal{S})$. The normal cone to a given set $\mathcal{X} \subseteq \mathbb{R}^n$ at $x \in \mathcal{X}$ is defined as $\mathcal{N}_{\mathcal{X}}(x) := \{y \in \mathbb{R}^n \mid y^\top(z - x) \leq 0 \ \forall z \in \mathcal{X}\}$. The set $\mathcal{T}_{\mathcal{X}}(x) := \mathrm{cl}\big(\cup_{y \in \mathcal{X}} \cup_{\lambda > 0} \lambda(y - x)\big)$ is referred to as the tangent cone to $\mathcal{X}$ at $x \in \mathcal{X}$. The simplex of all vectors $f \in \mathbb{R}^n$ for which the elements sum to D is denoted $\mathcal{H}_D := \{f \in \mathbb{R}^n \mid \sum_{i \in [n]} f_p = D\}$. When talking about probability distributions $q \in \mathbb{R}^m$, we also sometimes write $\Delta_1^m := \{q \in \mathbb{R}^m \mid \sum_{i \in [m]} q_i = 1\}$.

We use $\mathbb{0}$ and $\mathbb{1}$ to denote the vector of all zeros and all ones respectively, where the dimension should be clear from the context. For a vector $f \in \mathbb{R}^n$, we let f_i denote the i-th element. Similarly, given a set $\mathcal{S} \subset [n]$, we use $f_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$ for the vector constructed by removing all elements whose indices are not in $\mathcal{S}$. For a matrix $A \in \mathbb{R}^{n \times m}$ we write $A_i \in \mathbb{R}^{1 \times m}$ for the i-th row of the matrix, and A_{ij} for the element of A in the i-th row and the j-th column. For a square matrix $A \in \mathbb{R}^{n \times n}$ and a set $\mathcal{S} \subset [n]$ we write $A_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{S}|}$ for the matrix constructed by removing all rows and columns of A whose indices are not in $\mathcal{S}$. We also use $A^\top$ to denote the transpose of the matrix $A \in \mathbb{R}^{n \times m}$; that is $A^\top \in \mathbb{R}^{m \times n}$ is defined by $(A^\top)_{ij} := A_{ji}$. The kernel of a matrix $A \in \mathbb{R}^{n \times m}$ is denoted $\ker(A) := \{f \in \mathbb{R}^m \mid Af = \mathbb{0}\}$. The space of $n \times m$ column stochastic matrices is $\mathsf{CS}(n, m) := \{A \in \mathbb{R}_{\geq 0}^{n \times m} \mid \sum_{i \in [n]} A_{ij} = 1 \text{ for all } j \in [m]\}$.

Given $f, \breve{f} \in \mathbb{R}^n$ and $\mu \in [0, 1]$ we use $\mathrm{coco}_\mu(f, \breve{f}) := \mu f + (1 - \mu)\breve{f}$ to denote the convex combination of f and $\breve{f}$. The Euclidean norm of a vector $f \in \mathbb{R}^n$ is given by $\|f\| := \sqrt{\sum_{i \in [n]} f_i^2}$. The Euclidean projection of f onto the set $\mathcal{F}$ is then denoted $\Pi_{\mathcal{F}}(f) := \arg\min_{\breve{f} \in \mathcal{F}} \|f - \breve{f}\|$. The ϵ-neighbourhood of f is defined as $\mathcal{C}_\epsilon(f) := \{\breve{f} \in \mathbb{R}^n \mid \|f - \breve{f}\| < \epsilon\}$.

For a function $g : \mathbb{R}^n \to \mathbb{R}$, $f \mapsto g(f)$, we write $\nabla g(f)$ for the gradient at f. For a function $g : \mathbb{R}^n \to \mathbb{R}^m$, $f \mapsto g(f)$ we use $Dg(f)$ to denote the Jacobian at f. For a function $g : \mathbb{R} \to \mathbb{R}^n$, $f \mapsto g(f)$, we write $\frac{\partial^+}{\partial f} g(f) := \lim_{h \to 0^+} \frac{g(f+h) - g(f)}{h}$ for the right-hand derivative of g at f. Similarly, $\frac{\partial^-}{\partial f} g(f) := \lim_{h \to 0^-} \frac{g(f+h) - g(f)}{h}$ denotes the left-hand derivative of g at f.

1.5 Publications and origin of chapters

Preprints

1. J. Verbree, A. Cherukuri, "Wardrop equilibrium and Braess's paradox for varying demand", *submitted to Mathematical Programming,* 2023 (**Ch. 3 & 4**). Available online at arxiv.org/pdf/2310.04256.
2. J. Verbree, A.Cherukuri, "Inferring the prior in routing games using public signaling", *under preparation,* 2023 (**Ch. 5**). Available online at arxiv.org/pdf/2109.05895.
3. J. Verbree, A. Cherukuri, "CVaR-based variational inequalities: stochastic approximation approaches using computationally-efficient projections", *submitted to Systems & Control Letters,* 2023 (**Ch. 6**).

Conference publications

1. J. Verbree, A. Cherukuri, "Stochastic approximation for CVaR-based variational inequalities", *In Proceedings of the IEEE Conference on Decision and Control*, pp. 2216-2221, 2020 (**Ch. 6**).

Chapter 2

Routing games

In this chapter, we take the time to introduce the concept of a *routing game*, which is the shared theme of all material presented in this thesis. Routing games are a class of games in which participants, or agents, aim to traverse a network, choosing a route in such a way that their own travel cost is minimized. However, the cost of traversing the network depends not only on which route an agent takes, but also on the extent to which this route overlaps with the routes of other agents. Parts of the network that are heavily used can become congested and therefore increase in cost. Our use of terms like "routes" and "congestion" already hints at the inspiration for these types of games: traffic networks. The archetypical example of a routing game is a network of roads that services a number of traffic participants. Modelling and control of traffic remains one of the main areas in which the subject of routing games finds application, although other uses of these types of games exist, for instance in communication networks [3] or in mechanical, electrical, and hydraulic systems [4]. In this chapter we will give a rigorous mathematical description of the specific type of routing game that we consider, introduce fundamental related concepts, and discuss some of the assumptions that hold throughout this thesis as well as basic results needed for the subsequent chapters. For a more general introduction to the subject of routing games, see [5].

2.1 The model

Our starting point for the introduction of routing games is a network, an example of which can be seen in Figure 2.1. A network consists of a set of points, called *nodes* or *vertices*, and a set of connections, called *edges*, between these points. Thus a network is completely defined by the associated *graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = [N]$, $N \in \mathbb{N}$ is the set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. In this work we only consider *directed* graphs, meaning that an edge $e_k \in \mathcal{E}$ consists of an *ordered* pair of vertices $(v_k^{\text{in}}, v_k^{\text{out}})$, where $v_k^{\text{in}}, v_k^{\text{out}} \in \mathcal{V}$, indicating that there is a connection from v_k^{in} to v_k^{out}. Using these connections, we can start at a vertex v_o and move along the edges in $\mathcal{E}$ until we reach some vertex v_d to find a *path* from v_o to v_d. Thus a path p from v_o

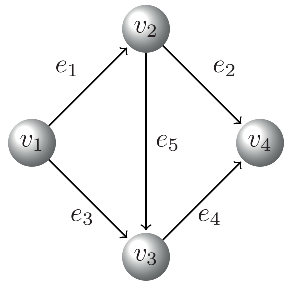


Figure 2.1: The Wheatstone network.

to v_d is an ordered set of edges $(e_{p_1}, \cdots, e_{p_l})$ such that $v_{p_1}^{\text{in}} = v_o$ and $v_{p_l}^{\text{out}} = v_d$ and $v_{p_k}^{\text{out}} = v_{p_{k+1}}^{\text{in}}$ for all $k \in [l-1]$. Paths are also defined to be *acyclic*, meaning that $v_{p_i}^{\text{in}} \neq v_{p_j}^{\text{out}}$ for all $1 \leq i \leq j \leq l$. The next step in constructing a routing game is now to associate an origin $v_o \in \mathcal{V}$ and a destination $v_d \in \mathcal{V}$ to the network. The paths relevant to the the routing game are all those starting at v_o and ending at v_d and the set of all these paths is denoted $\mathcal{P}$. It is also possible to define a routing game using multiple origin-destination pairs, however, in this thesis we mainly consider *single origin-destination* networks. In this case, there is only one origin and one destination in the network, and all traffic needs to be routed from this origin to the destination over the paths in $\mathcal{P}$. The next element of a routing game is the total amount of traffic that needs to be routed in this way, which we refer to as the *demand* and is denoted $D \geq 0$. The way in which this demand is divided among the paths in $\mathcal{P}$ gives rise to a vector of path flows $f \in \mathbb{R}^n_{\geq 0}$, where f_p denotes the amount of flow routed over path p, and $n = |\mathcal{P}|$. We assume the flow to be *non-atomic*, meaning that f is a continuous variable. The set of all *feasible* flows f is therefore given by

$$\mathcal{F}_D := \Big\{ f \in \mathbb{R}^n_{\geq 0} \mid \sum_{p \in \mathcal{P}} f_p = D \Big\}. \tag{2.1}$$

Based on the flow over the paths in $\mathcal{P}$, the flow over an edge $e_k \in \mathcal{E}$ of the network is given by the sum of the flows over all paths that contain e_k, i.e.

$$f_{e_k} := \sum_{p \ni e_k} f_p. \tag{2.2}$$

For any traffic participant, traversing one of the edges in the network incurs a cost that depends on the amount of flow on that edge. Specifically, each edge e_k of the network has a cost function $C_{e_k} : \mathbb{R} \to \mathbb{R}$ associated to it that maps the edge-flow f_{e_k} to the edge-cost $C_{e_k}(f_{e_k})$. The functions $\{C_{e_k}\}_{e_k \in \mathcal{E}}$ are assumed to be *continuous*, and *non-negative* and *non-decreasing* on $\mathbb{R}_{\geq 0}$. The cost of traversing the path p is then equal to the sum of the costs of all edges in that path; i.e.,

$$C_p(f) := \sum_{e_k \in p} C_{e_k}(f_{e_k}).$$

For the sake of convenience, we define the following notation:

$$\begin{aligned} C_e(f) &:= \big(C_{e_1}(f_{e_1}), \cdots, C_{e_m}(f_{e_m})\big)^\top, \\ C(f) &:= \big(C_1(f), \cdots, C_n(f)\big)^\top, \\ \mathcal{C} &:= \{C_{e_k}\}_{e_k \in \mathcal{E}}, \end{aligned} \tag{2.3}$$

where $m := |\mathcal{E}|$. Throughout this thesis we make the assumption that the functions in $\mathcal{C}$ are continuous, and non-negative and non-decreasing on $\mathbb{R}_{\geq 0}$ explicit by writing $\mathcal{C} \subset \mathcal{K}$. Here $\mathcal{K}$ is thus the set of all functions that are continuous, and non-negative and non-decreasing on $\mathbb{R}_{\geq 0}$. With the feasible set and the related costs defined, the final ingredient of a routing game is the way in which drivers make routing choices. For this we assume that all traffic participants want to minimize their own travel cost and choose a path accordingly. The resulting flow is therefore assumed to be a *Wardrop equilibrium*(WE):

Definition 2.1.1. *Given a set of paths $\mathcal{P}$, associated cost functions $\mathcal{C} \subset \mathcal{K}$, and a demand $D \geq 0$, a flow $f^D \in \mathbb{R}^n_{\geq 0}$ is called a Wardrop equilibrium if $f^D \in \mathcal{F}_D$ and for every $p \in \mathcal{P}$ such that $f^D_p > 0$ we have*

$$C_p(f^D) \leq C_r(f^D) \quad \textit{for all } r \in \mathcal{P}. \tag{2.4}$$

We will denote the set of all Wardrop equilibria as $\mathcal{W}_D$. •

Remark 2.1.2. In the hope that it provides the reader with a clear intuition for what a routing game is, and for what applications it may be useful, we have introduced the concept using a network represented by a graph. However, the definition of Wardrop equilibrium requires only the existence of the set $\mathcal{P}$, the set $\mathcal{C}$, and the demand D. In fact, a routing game is completely defined by only the set of paths $\mathcal{P}$, the set of cost functions $\mathcal{C}$, and the demand D, but this general formulation of routing games can result in situations for which there no longer exists a graph that represents the structure of the game, (see Appendix A for an example). We will encounter this possibility later in this thesis when we consider routing games in which the flow on some paths is constrained to be zero. However, by simply removing any path whose flow is constrained to zero from consideration, and defining a routing game using only the set of paths $\mathcal{P}$, the edge-costs in $\mathcal{C}$, and the demand D, it is immediately clear that all results that apply to classical routing games with a graph representation also apply to these games. For this reason we do not impose the existence of a graph that represents the routing game in Definition 2.1.1 or any future definitions and results, but instead only consider the sets $\mathcal{P}$ and $\mathcal{C}$ as given. •

In the next section we discuss preliminaries on the subject of Wardrop equilibria.

2.2 Wardrop equilibria

The intuitive motivation behind the concept of Wardrop equilibrium is that individual drivers want to minimize their own travel costs. Therefore they would change their routing choices if a path with lower cost was available to them. Thus, for the flow to be in equilibrium it is required that every driver experiences the same cost, and this cost must be minimal among all paths. This leads naturally to the introduction of the set of WE as given in Definition 2.1.1. However, for the purpose of analysis and computation, alternative characterizations of the set of WE can be more useful. For instance, the set of WE can equivalently be defined as the set of solutions to a *variational inequality*(VI).

Definition 2.2.1. *(Variational inequalities (VIs)): Given a map $C : \mathbb{R}^n \to \mathbb{R}^n$ and a set $\mathcal{F} \subset \mathbb{R}^n$, the associated variational inequality problem, denoted $\mathrm{VI}(\mathcal{F}, C)$ is to find $f^* \in \mathcal{F}$ such that the following holds:*

$$C(f^*)^\top (f - f^*) \geq 0, \text{ for all } f \in \mathcal{F}. \tag{2.5}$$

The set of solutions $f^* \in \mathcal{F}$ satisfying the above property is denoted as $\mathrm{SOL}(\mathcal{F}, C)$. •

Based on the feasible set $\mathcal{F}_D$ and the cost function C as given by (2.1) and (2.3), we have the following relation between WE and VIs.

Proposition 2.2.2. *(WE as the solution of a VI [6]): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}$, and $D \geq 0$ be given. Then, f^D is a Wardrop equilibrium if and only if $f^D \in \mathrm{SOL}(\mathcal{F}_D, C)$.*

We note that when $C(f) = Af$ for some matrix $A \in \mathbb{R}^{n \times n}$, then we use the notation $\mathrm{SOL}(\mathcal{F}, A) := \mathrm{SOL}(\mathcal{F}, C)$. Alternatively, the set of WE can be defined as the solution set of a convex optimization problem, a characterization that is more useful from a computational perspective.

Proposition 2.2.3. *(WE as the solution of an optimization problem [7]): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}$, and $D \geq 0$ be given, and let the map $V : \mathbb{R} \to \mathbb{R}$ be given by*

$$V(D) := \min_{f \in \mathcal{F}_D} \sum_{e_k \in \mathcal{E}} \int_0^{f_{e_k}} C_{e_k}(z) dz. \tag{2.6}$$

A flow f^D is a Wardrop equilibrium if and only if it is an optimal solution of the above minimization problem; that is, if and only if $f^D \in \mathcal{F}_D$ and

$$\sum_{e_k \in \mathcal{E}} \int_0^{f_{e_k}^D} C_{e_k}(z) dz = V(D).$$

The repeated reference to the *set of* WE may already have made clear that a WE is in general not unique. However, the set of WE does have a property sometimes referred to as *essential uniqueness* [8], meaning that for all WE $f^D \in \mathcal{W}_D$ the cost on all edges is the same.

Proposition 2.2.4. *(All WE induce the same edge costs [9]): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}$, and $D \geq 0$ be given. We have*

$$C_{e_k}(f_{e_k}^D) = C_{e_k}(\check{f}_{e_k}^D) \quad \text{for all } f^D, \check{f}^D \in \mathcal{W}_D \text{ and } e_k \in \mathcal{E}.$$

If we strengthen the assumption that the functions C_{e_k} are non-decreasing on $\mathbb{R}_{\geq 0}$ by imposing that these functions are instead strictly increasing on $\mathbb{R}_{\geq 0}$, the above proposition implies the following corollary on the uniqueness of WE.

Corollary 2.2.5. *Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}$, and $D \geq 0$ be given, where all functions in $\mathcal{C}$ are strictly increasing. We have*

$$f_{e_k}^D = \check{f}_{e_k}^D \quad \text{for all } f^D, \check{f}^D \in \mathcal{W}_D \text{ and } e_k \in \mathcal{E}.$$

Chapter 3

Varying demand

A routing game is defined by a set of paths, a set of cost functions on the edges in those paths, and finally by the amount of traffic that traverses the network, often referred to as the *demand*. It follows naturally that an important subject in the study of routing games is how changes demand affect how flow is divided among the paths. In other words: “how does the set of WE change as the demand changes?” Of course this is a question that has been studied before. An early investigation into the relation between WE and demand establishes useful properties such as continuity and potential non-differentiability of the WE relative to changes in demand [10]. In [11] the authors show how in routing games with multiple origin-destination pairs the relation between demand and the experienced travel costs can be counterintuitive, to the extent that an increase in demand can decrease travel costs. Later, in [12], an insightful characterization of the ’directional derivatives’ of WE is provided for a broad class of routing games. This offers a comprehensive understanding of how WE evolve in response to changes in demand. More recently [9] studies the relation between demand and *price of anarchy*, which is a measure of the inefficiency caused by the ’selfish’ routing choices of drivers. We take a moment to highlight the contribution of [12], which gives quite a complete picture of how changes in certain parameters, including the demand, affect WE of a broad class of routing games.

Our work in this chapter studies a question closely related to the work in [12]. Specifically we look at single origin-destination routing games where the functions in $\mathcal{C}$ are affine and characterize the set of directions in which the set of WE can change as the demand increases. The reason results in [12] are not directly applicable to our case is that they depend on the assumption that WE have unique edge flows, which is not valid in our case. In addition the argumentation in [12] is given in terms of the flows on edges rather than in terms of flows on paths as we do, though both perspectives are eventually equivalent. However, the main motivation for writing this chapter instead of relying on the results in [12] is that many of the intermediary results are required for Chapter 4, which investigates the phenomenon of Braess’s paradox, where the removal of one or more edges from a network can decrease the travel cost under WE. In a way this chapter and Chapter 4 are two parts of one story, where the former is necessary for the latter, and the latter motivates the former.

Nonetheless, the results of this chapter are interesting in their own right and are therefore presented separately.

Organization

In broad terms the structure of this chapter is as follows. After introducing some required definitions, notations, and other preliminaries we start with an illustrative example, showcasing a lot of the properties of the evolution of WE which we establish formally in later parts of the chapter. The example demonstrates that the set of WE moves continuously with respect to the demand, and what is more, that the evolution of the WE is piecewise affine[1]. The first step of our investigation then considers the evolution of two important sets: the set of paths with minimal cost, also called the *active* set, and the set of paths that for some WE carry a positive amount of flow, also referred to as the *used* set. These sets remain constant on the intervals of demand in which the evolution of the WE is affine, and we investigate how these sets can change at the endpoints of these intervals. Next we turn our attention to the evolution of the costs of the paths under WE. This too evolves in a piecewise affine manner, and we show that the points of non-differentiability lie exactly at the endpoints of the intervals on which the evolution of the WE is affine. With all of the preceding results established, we can address the main topic of the chapter, and obtain a characterization of the evolution of the set of WE as the set of solutions to a variational inequality. We finish by discussing some interesting consequences of this main result among which is the possibility of constructing a direct method for finding all WE above a sufficiently high level of demand. Throughout the chapter we illustrate our findings using examples.

3.1.1 Definitions, notation and preliminaries

To ease the exposition in this chapter we introduce some concepts and notation related to the cost of paths under WE. First of all, here and in Chapter 4 we will assume that the cost functions for the edges are, in addition to continuous, non-negative and non-decreasing, also *affine*. That is, for each edge $e_k \in \mathcal{E}$ there exist parameters $\alpha_{e_k}, \beta_{e_k} \geq 0$ such that

$$C_{e_k}(f_{e_k}) := \alpha_{e_k} f_{e_k} + \beta_{e_k}. \tag{3.1}$$

This assumption allows us to write the path-cost function concisely as

$$C(f) = Af + \beta, \tag{3.2}$$

[1]Since the WE is not unique, a piecewise affine evolution here means that we can divide the range of the demand into intervals on which the set of directions in which the set of WE evolves remains constant

where $\beta = (\beta_p)_{p\in\mathcal{P}}$ is the vector with entries $\beta_p = \sum_{e_k\in p}\beta_{e_k}$ and $A \in \mathbb{R}^{n\times n}_{\geq 0}$ is a matrix with the (p, r)-th entry given by $A_{pr} = \sum_{e_k\in p\cap r}\alpha_{e_k}$. Therefore A is symmetric and in fact positive semi-definite. This can be seen by noting that $A = B^T QB$, where $B \in \mathbb{R}^{m\times n}$ is defined by $B_{k,i} = 1$ if edge e_k is in path p_i and $B_{k,i} = 0$ otherwise, and $Q \in \mathbb{R}^{m\times m}$ is a diagonal matrix with $Q_{k,k} = \alpha_{e_k} \geq 0$. Throughout this chapter and the next we will make the assumption that the functions in $\mathcal{C}$ are affine explicit by writing $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$. Here $\mathcal{K}_{\text{aff}}$ is thus the set of all non-decreasing affine functions which are non-negative on $\mathbb{R}_{\geq 0}$.

We recall from Proposition 2.2.4 that the costs of the edges are the same for all WE in the set $\mathcal{W}_D$, which has several useful consequences. First of all, in combination with (3.1) this gives a representation of $\mathcal{W}_D$ as a set satisfying a number of affine constraints, showing that the set $\mathcal{W}_D$ is convex. In addition we note that since all edge costs are the same for all WE in $\mathcal{W}_D$, the same holds for the costs of the paths and therefore we can define a map from demand to a vector of path costs as follows:

$$\lambda^{\text{vec}}(D) := C(f^D), \quad f^D \in \mathcal{W}_D.$$

We thus have that the p-th element of $\lambda^{\text{vec}}(D)$ is the cost of path p under any WE in the set $\mathcal{W}_D$. There are two important sets related to this path-cost under WE that are essential to the analysis in this chapter. The first of these sets we call the *active* set. Given a demand D, the active set is the set of all paths that have minimal cost under WE. We denote this set as

$$\mathcal{R}_D^{\text{act}} := \{p \in \mathcal{P} \mid \lambda_p^{\text{vec}}(D) \leq \lambda_r^{\text{vec}}(D), \text{ for all } r \in \mathcal{P}\}. \tag{3.3}$$

We recall that for any WE in $\mathcal{W}_D$, a positive amount of flow is routed onto path p only if p has minimal cost among all paths. It follows that when in WE all traffic experiences the same cost, and this cost is exactly that of the paths that are in the active set. We write

$$\lambda^{\text{WE}}(D) := \lambda_p^{\text{vec}}(D), \quad p \in \mathcal{R}_D^{\text{act}}. \tag{3.4}$$

The other important set we call the *used* set. It is the set of all paths for which there exists at least one WE in $\mathcal{W}_D$ which assigns a positive amount of flow to that path. Formally we have

$$\mathcal{R}_D^{\text{use}} := \{p \in \mathcal{P} \mid \text{there exists an } f^D \in \mathcal{W}_D \text{ such that } f_p^D > 0\}. \tag{3.5}$$

Since a WE only routes flow onto paths with minimal cost we have $\mathcal{R}_D^{\text{use}} \subseteq \mathcal{R}_D^{\text{act}}$.

The final concept we want to introduce is that of the set of *breakpoints* of a routing game. It turns out that we can divide the non-negative real line into a finite number of intervals, and for any two levels of demand in one such interval, the corresponding active sets will be the same, as will the corresponding used sets. We use the term *breakpoints* to refer to the endpoints of these intervals. The set of all these breakpoints is denoted $\mathcal{D}$. We formally introduce this set in the following result.

Corollary 3.1.1. *(Piecewise constant evolution of active and used sets): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. There exists a set $\mathcal{D} := (D_0, D_1, \cdots, D_M, D_{M+1}) \subset \mathbb{R}_{\geq 0} \cup \{+\infty\}$ of finitely many points, with $D_0 = 0$, $D_{M+1} = \infty$ and $D_j > D_{j-1}$ for all $j \in [M+1]$, and associated sets of subsets of $\mathcal{P}$ denoted $\{\mathcal{J}_0^{\text{act}}, \mathcal{J}_1^{\text{act}}, \cdots, \mathcal{J}_M^{\text{act}}\}$ and $\{\mathcal{J}_0^{\text{use}}, \mathcal{J}_1^{\text{use}}, \cdots, \mathcal{J}_M^{\text{use}}\}$, such that, for all $i \in [M]_0$ and $D \in (D_i, D_{i+1})$, we have*

$$\mathcal{R}_D^{\text{act}} = \mathcal{J}_i^{\text{act}}, \quad \mathcal{R}_D^{\text{use}} = \mathcal{J}_i^{\text{use}}.$$

Furthermore, $\mathcal{J}_i^{\text{act}} \neq \mathcal{J}_j^{\text{act}}$ and $\mathcal{J}_i^{\text{use}} \neq \mathcal{J}_j^{\text{use}}$ for all $i \neq j$.

The parts of the above concerning the active set are established in [9, Section 4]. The arguments mainly rely on the upcoming Lemma, also from [9], which establishes properties of convex combinations of WE. For ease of reference we recall that $\text{coco}_\mu(f, \widetilde{f}) = \mu f + (1-\mu)\widetilde{f}$.

Lemma 3.1.2. *(Evolution of the active set [9]): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. For any $0 \leq D^- \leq D^+$ that satisfy $\mathcal{R}_{D^-}^{\text{act}} = \mathcal{R}_{D^+}^{\text{act}}$, the following hold*

1. *For all $D \in [D^-, D^+]$, we have $\mathcal{R}_D^{\text{act}} = \mathcal{R}_{D^-}^{\text{act}} = \mathcal{R}_{D^+}^{\text{act}}$.*
2. *If $f^{D^-} \in \mathcal{W}_{D^-}$, $f^{D^+} \in \mathcal{W}_{D^+}$ and $\mu \in [0,1]$, then $\text{coco}_\mu(f^{D^-}, f^{D^+}) \in \mathcal{W}_T$, where $T = \text{coco}_\mu(D^-, D^+)$.*

This Lemma, in combination with the affine form of C and the already established parts of Corollary 3.1.1 related to the active set, can also be used to prove the parts of Corollary 3.1.1 concerning the used set. For the sake of rigour a proof is given in Appendix B.

3.2 The evolution of WE

We start our investigation into the evolution of the set of WE with a set of examples, which will provide useful intuition and context for the subsequent analysis. Throughout the chapter we revisit these examples to demonstrate obtained results.

Example 3.2.1. *(Evolution of WE):* We discuss here two simple networks:

(Case-a) For the first example, consider the network depicted in Figure 3.1, often called the Wheatstone network, where the edge-cost functions are given by

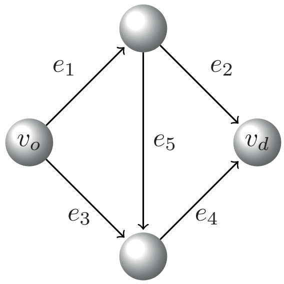


Figure 3.1: The Wheatstone network.

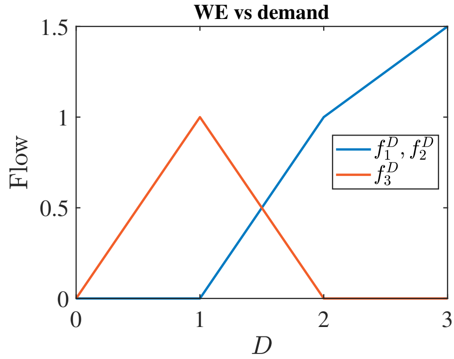


Figure 3.2: The WE at different demands for the routing game defined by the Wheatstone network (Figure 3.1) and costs (3.7).

$$\begin{aligned} C_{e_1}(f_{e_1}) &= f_{e_1}, & C_{e_2}(f_{e_2}) &= 1, \\ C_{e_3}(f_{e_3}) &= 1, & C_{e_4}(f_{e_4}) &= f_{e_4}, \\ & C_{e_5}(f_{e_5}) = 0. \end{aligned} \tag{3.6}$$

There are three paths from the origin to the destination in this network, namely $p_1 = (e_1, e_2)$, $p_2 = (e_3, e_4)$ and $p_3 = (e_1, e_5, e_4)$, with cost functions given respectively by

$$\begin{aligned} C_1(f) &= f_1 + f_3 + 1, \\ C_2(f) &= f_2 + f_3 + 1, \\ C_3(f) &= f_1 + f_2 + 2f_3. \end{aligned} \tag{3.7}$$

Using this, we can find the following explicit expression for the WE as a function of the demand D.

$$f^D = \begin{cases} \begin{pmatrix} 0 & 0 & D \end{pmatrix}^\top & \text{for } D \in [0,1], \\ \begin{pmatrix} D-1 & D-1 & 2-D \end{pmatrix}^\top & \text{for } D \in [1,2], \\ \begin{pmatrix} \frac{D}{2} & \frac{D}{2} & 0 \end{pmatrix}^\top & \text{for } D \in [2,\infty). \end{cases} \tag{3.8}$$

The first important observation is that f^D changes continuously, and what is more, it evolves in a piecewise affine manner. The points at which the evolution changes from one affine piece to the next will later turn out to be exactly the breakpoints in $\mathcal{D}$ defined in Corollary 3.1.1. Notice that in this example the WE is unique for any demand and the evolution of the WE as demand increases is therefore fully characterized by the right-hand derivative of the map $D \mapsto f^D$:

$$f^\delta(D) := \frac{\partial^+}{\partial D} f^D = \begin{cases} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^\top & \text{for } D \in [0,1), \\ \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}^\top & \text{for } D \in [1,2), \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}^\top & \text{for } D \in [2,\infty). \end{cases}$$

Also note that on the intervals where the map $D \mapsto f^D$ is affine, naturally the direction in which the WE moves is constant.

(Case-b) When the WE are not unique, the situation can become more complicated. To illustrate this, we slightly modify our example. Instead of connecting the top and bottom nodes of the Wheatstone network (Figure 3.1) with the edge e_5, we merge them into one node. The resulting network is depicted in Figure 3.3.

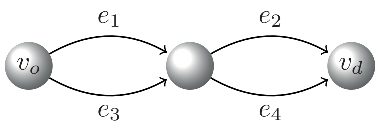


Figure 3.3: The Wheatstone network after merging the top and bottom nodes.

The edge-cost functions are still given by (3.6), except that edge e_5 no longer exists. There are now four paths in $\mathcal{P}$, namely $p_1 = (e_1, e_2)$, $p_2 = (e_3, e_4)$, $p_3 = (e_1, e_4)$ and $p_4 = (e_3, e_2)$, and the path-cost function is given by

$$C(f) = Af + b = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} f + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}. \tag{3.9}$$

Note that $Af^0 = \mathbb{0}$ for $f^0 = (-1,-1,1,1)^\top$ and so, for any flow f and any value $\epsilon \in \mathbb{R}$, we have $C(f + \epsilon f^0) = C(f)$. In other words, re-routing equal amounts of flow from p_1 to p_2 and from p_3 to p_4, or vice versa, does not change the path-cost. As a consequence, the WE are not always unique, and we instead find a set of WE given by

$$\mathcal{W}_D := \begin{cases} \left\{ \begin{pmatrix} 0 & 0 & D & 0 \end{pmatrix}^\top \right\} & \text{for } D \in [0,1], \\ \{ f \in \mathcal{F}_D \mid f_1 + f_3 = 1, f_2 + f_3 = 1 \} & \text{for } D \in [1,\infty). \end{cases} \tag{3.10}$$

Using this, we obtain the directions along which the WE can move as the demand increases as:

$$\Gamma_D := \begin{cases} \{f^\delta \in \mathcal{H}_1 \mid f_1^\delta, f_2^\delta, f_4^\delta = 0, f_3^\delta = 1\}, & D \in [0,1), \\ \{f^\delta \in \mathcal{H}_1 \mid f_1^\delta, f_2^\delta, f_4^\delta \geq 0, f_1^\delta = f_2^\delta = -f_3^\delta\}, & D = 1, \\ \{f^\delta \in \mathcal{H}_1 \mid f_1^\delta + f_3^\delta = 0, \;\; f_2^\delta + f_3^\delta = 0\}, & D \in [1,\infty), \end{cases} \tag{3.11}$$

where we recall that $\mathcal{H}_T := \{f \in \mathbb{R}^n \mid \sum_{i\in[n]} f_p = T\}$. Instead of a single direction, we find a set of directions along which the set of WE evolves. That is, for every D and every $f^\delta \in \Gamma_D$, there exist $f^D \in \mathcal{W}_D$ and $\bar{\epsilon} > 0$ such that for all $\epsilon \in [0, \bar{\epsilon})$ the flow $f^D + \epsilon f^\delta$ is a WE for the demand $D + \epsilon$. Note that, as in the previous example, the non-negative real line is divided into intervals such that the set of directions Γ_D is constant in each interval. •

In the above example we observed that the set of WE changes continuously with respect to the demand. Our first goal is to establish that this holds for any routing game. When equilibrium flows over the edges are unique, such a result is already established in [10]. We now show that the same holds when considering sets of WE.

Lemma 3.2.2. *(Continuity of the map $D \mapsto \mathcal{W}_D$): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. The (set-valued) map $D \mapsto \mathcal{W}_D$ is continuous; that is, for every $D \geq 0$ and $\epsilon > 0$ the following hold:*

1. *There exists a $\delta > 0$ such that for all $f^D \in \mathcal{W}_D$ and $T \geq 0$ satisfying $|T - D| < \delta$ there exists $f^T \in \mathcal{W}_T$ such that $\|f^D - f^T\| < \epsilon$.*
2. *There exists a $\delta > 0$ such that for all $T \geq 0$ satisfying $|T - D| < \delta$ and all $f^T \in \mathcal{W}_T$ there exists $f^D \in \mathcal{W}_D$ such that $\|f^D - f^T\| < \epsilon$.*

The first and second parts of the above definition are known as lower and upper semicontinuity respectively [13], or as inner and outer semicontinuity, respectively [14]. The result follows from [15, Theorem 4.2] and noting that since A is positive semi-definite, and symmetric, it is *cocoercive* [15, Theorem 3.3].

With continuity properly established, we first turn our attention to the evolution of the active and used sets. Our next result sheds light on how these two sets evolve on the interval $[D_i, D_{i+1}]$, where $D_i, D_{i+1} \in \mathcal{D}$. Recall that $\mathcal{D}$ is the set of breakpoints of the routing game; that is, $\mathcal{D}$ contains the points where the active and used sets change (see Corollary 3.1.1).

Lemma 3.2.3. *(Relationship between active and used sets over an interval): For a given $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$, let $D_i, D_{i+1} \in \mathcal{D}$ and let $\mathcal{J}_i^{\text{act}} \subseteq \mathcal{P}$ and $\mathcal{J}_i^{\text{use}} \subseteq \mathcal{P}$ be the associated active and used sets on (D_i, D_{i+1}), respectively. Then, we have*

$$\mathcal{R}_{D_i}^{\text{use}} \subseteq \mathcal{J}_i^{\text{use}} \subseteq \mathcal{J}_i^{\text{act}} \subseteq \mathcal{R}_{D_i}^{\text{act}}, \tag{3.12a}$$

$$\mathcal{R}^{\text{use}}_{D_{i+1}} \subseteq \mathcal{J}^{\text{use}}_{i} \subseteq \mathcal{J}^{\text{act}}_{i} \subseteq \mathcal{R}^{\text{act}}_{D_{i+1}}. \tag{3.12b}$$

Proof. We will first establish (3.12a). Let $p \in \mathcal{R}^{\text{act}}_{D_i}$ and $r \in (\mathcal{R}^{\text{act}}_{D_i})^c$. It follows that $\lambda^{\text{vec}}_p(D_i) < \lambda^{\text{vec}}_r(D_i)$. Continuity of the map $D \mapsto \mathcal{W}_D$, as proven in Lemma 3.2.2, implies continuity of the map $\lambda^{\text{vec}}(\cdot)$. Therefore, it follows that for small enough $\epsilon > 0$ we have $\lambda^{\text{vec}}_p(T) < \lambda^{\text{vec}}_r(T)$ for all $T \in [D_i, D_i + \epsilon)$. It follows that $r \in (\mathcal{R}^{\text{act}}_T)^c$ for all $T \in [D_i, D_i + \epsilon)$, which shows that $r \in (\mathcal{J}^{\text{act}}_i)^c$. Thus we have $\mathcal{J}^{\text{act}}_i \subseteq \mathcal{R}^{\text{act}}_{D_i}$. Similarly, it follows from Lemma 3.2.2 that for small enough $\epsilon > 0$ there exist $f^T \in \mathcal{W}_T$ such that $f^T_p > 0$ for all $p \in \mathcal{R}^{\text{use}}_{D_i}$ and $T \in [D_i, D_i + \epsilon)$. This shows that $\mathcal{R}^{\text{use}}_{D_i} \subseteq \mathcal{R}^{\text{use}}_T$ for all $T \in [D_i, D_i + \epsilon)$. Thus we have $\mathcal{R}^{\text{use}}_{D_i} \subseteq \mathcal{J}^{\text{use}}_i$. By definition of the active and used sets, we also have $\mathcal{J}^{\text{use}}_i \subseteq \mathcal{J}^{\text{act}}_i$, proving (3.12a). The result (3.12b) concerning $\mathcal{R}^{\text{use}}_{D_{i+1}}$ and $\mathcal{R}^{\text{act}}_{D_{i+1}}$ follows by the same arguments considering the interval $(D_{i+1} - \epsilon, D_{i+1}]$. □

The implication of the above is that when the demand D moves from the point D_i into the interval (D_i, D_{i+1}), the used set $\mathcal{R}^{\text{use}}_D$ can only gain elements, while the active set $\mathcal{R}^{\text{act}}_D$ can only lose elements. When the demand D then moves from the interval (D_i, D_{i+1}) to the point D_{i+1} the situation is reversed. That is, $\mathcal{R}^{\text{use}}_D$ can only lose elements, while $\mathcal{R}^{\text{act}}_D$ can only gain elements. Also note that since $\mathcal{J}^{\text{use}}_i \neq \mathcal{J}^{\text{use}}_j$ and $\mathcal{J}^{\text{act}}_i \neq \mathcal{J}^{\text{act}}_j$ for all $i \neq j$, both the active and the used set must change as D moves from (D_{i-1}, D_i) to (D_i, D_{i+1}). Turning our attention back to Example 3.2.1a, using (3.7) and (3.8), we derive $\mathcal{R}^{\text{act}}_D$ and $\mathcal{R}^{\text{use}}_D$ as:

$$(\mathcal{R}^{\text{act}}_D, \mathcal{R}^{\text{use}}_D) = \begin{cases} (\{p_3\}, \emptyset) & \text{for } D = 0, \\ (\{p_3\}, \{p_3\}) & \text{for } D \in (0,1), \\ (\{p_1,p_2,p_3\}, \{p_3\}) & \text{for } D = 1, \\ (\{p_1,p_2,p_3\}, \{p_1,p_2,p_3\}) & \text{for } D \in (1,2), \\ (\{p_1,p_2,p_3\}, \{p_1,p_2\}) & \text{for } D = 2, \\ (\{p_1,p_2\}, \{p_1,p_2\}) & \text{for } D \in (2,\infty). \end{cases} \tag{3.13}$$

We see that the evolution of the active and used sets indeed adheres to the result in Lemma 3.2.3. We can also use Lemma 3.2.3 to establish the following minor but useful extension of Lemma 3.1.2:

Corollary 3.2.4. *(Convex combinations of WE in $[D_i, D_{i+1}]$): For a given $\mathcal{P}$, $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ and $D_i, D_{i+1} \in \mathcal{D}$, let $D, T \in [D_i, D_{i+1}]$. Then, for any $f^D \in \mathcal{W}_D$, $f^T \in \mathcal{W}_T$, and $\mu \in [0,1]$, we have $\operatorname{coco}_\mu(f^D, f^T) \in \mathcal{W}_{\operatorname{coco}_\mu(D,T)}$.*

Proof. The result for the case $D, T \in (D_i, D_{i+1})$ is already stated in Lemma 3.1.2. Now consider the case $D = D_i$ and $T = D_{i+1}$. For given $\mu \in [0,1]$, $f^D \in \mathcal{W}_D$, and $f^T \in \mathcal{W}_T$, denote $f^\mu := \operatorname{coco}_\mu(f^D, f^T)$. Since $f^D, f^T \geq 0$, we have $f^\mu \geq 0$, and it follows in a straightforward manner that $f^\mu \in \mathcal{F}_{T_\mu}$, where $T_\mu := \operatorname{coco}_\mu(D, T)$.

Now let $p \in \mathcal{P}$ be a path such that $f_p^\mu > 0$. It follows that either $f_p^D > 0$ or $f_p^T > 0$. Since $D = D_i$ and $T = D_{i+1}$, we obtain either $p \in \mathcal{R}_{D_i}^{\text{use}}$ or $p \in \mathcal{R}_{D_{i+1}}^{\text{use}}$. In both cases, Lemma 3.2.3 implies that $p \in \mathcal{J}_i^{\text{use}}$, and subsequently the same result implies that $p \in \mathcal{R}_{D_i}^{\text{act}} \cap \mathcal{R}_{D_{i+1}}^{\text{act}}$. Thus, we have

$$\begin{aligned} C_p(f^D) &\leq C_r(f^D), \quad \text{for all } r \in \mathcal{P}, \\ C_p(f^T) &\leq C_r(f^T), \quad \text{for all } r \in \mathcal{P}. \end{aligned}$$

Since the function C is affine and f^μ is a convex combination of f^D and f^T, we get $C_p(f^\mu) \leq C_r(f^\mu)$ for all $r \in \mathcal{P}$. This establishes the WE condition (2.4) and thus we have shown that $f^\mu \in \mathcal{W}_{T_\mu}$. The cases $D = D_i$, $T \in (D_i, D_{i+1})$ and $D \in (D_i, D_{i+1})$, $T = D_{i+1}$ follow using similar arguments. □

The main goal of this chapter is to characterize the evolution of the set of WE. We already noted in Example 3.2.1 that there exists a (set of) directions f^δ along which the WE moves when the demand increases in the intervals between the breakpoints in $\mathcal{D}$. We refer to such a direction as a *direction of increase*. The collection of all such vectors, denoted Γ_D, is referred to as the *set of directions of increase*. Our aim is to characterize this set Γ_D, which is formally defined as follows:

Definition 3.2.5. *(Set of directions of increase): Let $\mathcal{P}$, $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$, and $D \geq 0$ be given. The set of* directions of increase Γ_D *is the set of all directions $f^\delta \in \mathcal{H}_1$ in which the flow can be increased, starting from some flow in $\mathcal{W}_D$, such that the new flow is a WE as long as the increase is small enough. That is,*

$$\Gamma_D := \{f^\delta \in \mathcal{H}_1 \mid \exists f^D \in \mathcal{W}_D,\ \ \bar{\epsilon} > 0 \text{ such that } f^D + \epsilon f^\delta \in \mathcal{W}_{D+\epsilon} \text{ for all } \epsilon \in [0, \bar{\epsilon}]\}.$$

Before we address this evolution of the WE, we first focus on the evolution of the WE-cost λ^{WE}. The following result relates the evolution of λ^{vec} to the set Γ_D.

Proposition 3.2.6. *(The evolution of λ^{vec}): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. For any $i \in [M]_0$ there exists a vector δC^i such that the following hold:*

1. *for all $D \in [D_i, D_{i+1})$ and $f^\delta \in \Gamma_D$*

$$Af^\delta = \delta C^i, \tag{3.14}$$

2. *for all $T \in [D_i, D_{i+1}]$*

$$\lambda^{\text{vec}}(T) = \lambda^{\text{vec}}(D_i) + (T - D_i)\delta C^i, \tag{3.15}$$

3. $\delta C^i \neq \delta C^{i+1}$.

Proof. We start with the second claim, which is a consequence of the affine form of C, given in (3.2), and the convexity result in Corollary 3.2.4. Let $f^{D_i} \in \mathcal{W}_{D_i}$ and $f^{D_{i+1}} \in \mathcal{W}_{D_{i+1}}$ and define

$$f^{\delta_0} := (D_{i+1} - D_i)^{-1}(f^{D_{i+1}} - f^{D_i}).$$

By Corollary 3.2.4 any convex combination of f^{D_i} and $f^{D_{i+1}}$ is a WE. To be specific, pick some $\mu \in [0,1]$ and let $f^\mu := \mathrm{coco}_\mu(f^{D_i}, f^{D_{i+1}})$. Then, $f^\mu \in \mathcal{W}_{T_\mu}$, where $T_\mu = \mathrm{coco}_\mu(D_i, D_{i+1})$. Furthermore, we have

$$\begin{aligned} f^\mu &= f^{D_i} + (1-\mu)(f^{D_{i+1}} - f^{D_i}) \\ &= f^{D_i} + (T_\mu - D_i)\frac{f^{D_{i+1}} - f^{D_i}}{D_{i+1} - D_i} \\ &= f^{D_i} + (T_\mu - D_i)f^{\delta_0}. \end{aligned}$$

Using the fact that C is affine and the definition of λ^{vec} we derive

$$\begin{aligned} \lambda^{\mathrm{vec}}(T_\mu) &= C(f^\mu) \\ &= C\big(f^{D_i} + (T_\mu - D_i)f^{\delta_0}\big) \\ &= C(f^{D_i}) + (T_\mu - D_i)Af^{\delta_0} \\ &= \lambda^{\mathrm{vec}}(D_i) + (T_\mu - D_i)Af^{\delta_0}. \end{aligned} \tag{3.16}$$

Since the above holds for all $\mu \in [0,1]$ we have $\lambda^{\mathrm{vec}}(T) = \lambda^{\mathrm{WE}}(D_i) + (T_\mu - D_i)Af^{\delta_0}$ for all $T \in [D_i, D_{i+1}]$. Setting $\delta C^i := Af^{\delta_0}$ the second statement is proven.

To show the first statement, let $D^- \in [D_i, D_{i+1})$ and let $f^\delta \in \Gamma_{D^-}$. It follows that there exist $f^{D^-} \in \mathcal{W}_{D^-}$, $D^+ \in (D^-, D_{i+1}]$ and $f^{D^+} \in \mathcal{W}_{D^+}$ such that

$$f^{D^+} = f^{D^-} + (D^+ - D^-)f^\delta.$$

Using the same derivation as in (3.16) we find

$$\lambda^{\mathrm{vec}}(D^+) = \lambda^{\mathrm{vec}}(D^-) + (D^+ - D^-)Af^\delta. \tag{3.17}$$

However, from (3.15) we have

$$\begin{aligned} \lambda^{\mathrm{vec}}(D^+) &= \lambda^{\mathrm{vec}}(D_i) + (D^- - D_i)\delta C^i + (D^+ - D^-)\delta C^i \\ &= \lambda^{\mathrm{vec}}(D^-) + (D^+ - D^-)\delta C^i, \end{aligned} \tag{3.18}$$

where we have again used $\lambda^{\mathrm{vec}}(D^-) = \lambda^{\mathrm{vec}}(D_i) + (D^- - D_i)\delta C^i$. Comparing (3.17) and (3.18) we get $Af^\delta = \delta C^i$, proving the first statement.

The third statement we show by contradiction. Therefore, assume $\delta C^i = \delta C^{i+1}$ holds for some $i \in [M]_0$ and let $p, r \in \mathcal{J}_i^{\mathrm{act}}$. By definition of the active set it follows that $\lambda_p^{\text{vec}}(D) = \lambda_r^{\text{vec}}(D)$ for all $D \in (D_i, D_{i+1})$. Combined with (3.15) this implies $\delta C_p^i = \delta C_r^i$. Under the assumption that $\delta C^i = \delta C^{i+1}$ we then have, by (3.15), that $\lambda_p^{\text{vec}}(D) = \lambda_r^{\text{vec}}(D)$ for all $D \in (D_i, D_{i+2})$ and $p, r \in \mathcal{J}_i^{\text{act}}$. As a consequence either all of the paths in $\mathcal{J}_i^{\text{act}}$ are in the active set on the interval (D_{i+1}, D_{i+2}) or none of them are. That is, one of the following holds: $\mathcal{J}_i^{\text{act}} \cap \mathcal{J}_{i+1}^{\text{act}} = \emptyset$ or $\mathcal{J}_i^{\text{act}} \subseteq \mathcal{J}_{i+1}^{\text{act}}$. However, note that from Lemma 3.2.3 we have $\mathcal{R}_{D_{i+1}}^{\text{use}} \subseteq \mathcal{J}_i^{\text{use}} \subseteq \mathcal{J}_i^{\text{act}}$ and $\mathcal{R}_{D_{i+1}}^{\text{use}} \subseteq \mathcal{J}_{i+1}^{\text{use}} \subseteq \mathcal{J}_{i+1}^{\text{act}}$. Since $D_{i+1} > 0$, we have $\mathcal{R}_{D_{i+1}}^{\text{use}} \neq \emptyset$ and it follows that $\mathcal{J}_i^{\text{act}} \cap \mathcal{J}_{i+1}^{\text{act}} \neq \emptyset$. Thus we must have $\mathcal{J}_i^{\text{act}} \subseteq \mathcal{J}_{i+1}^{\text{act}}$.

Now since $\mathcal{J}_i^{\text{act}} \neq \mathcal{J}_{i+1}^{\text{act}}$, there must exist some path $r' \in (\mathcal{J}_i^{\text{act}})^c \cap \mathcal{J}_{i+1}^{\text{act}}$. Since r' is not in the active set $\mathcal{J}_i^{\text{act}}$, one can find a path $p' \in \mathcal{J}_i^{\text{act}}$ and a demand $D \in (D_i, D_{i+1})$ such that $\lambda_{p'}^{\text{vec}}(D) < \lambda_{r'}^{\text{vec}}(D)$. On the other hand, by Lemma 3.2.3, $r' \in \mathcal{J}_{i+1}^{\text{act}}$ implies $r' \in \mathcal{R}_{D_{i+1}}^{\text{act}}$ and so, $\lambda_{r'}^{\text{vec}}(D_{i+1}) \leq \lambda_{p'}^{\text{vec}}(D_{i+1})$. Combining the above two facts with the affine form (3.15), we deduce that $\delta C_{p'}^i > \delta C_{r'}^i$. Since $\delta C^{i+1} = \delta C^i$ it then follows from (3.15) and $\lambda_{r'}^{\text{vec}}(D_{i+1}) \leq \lambda_{p'}^{\text{vec}}(D_{i+1})$ that $\lambda_{p'}^{\text{vec}}(D) > \lambda_{r'}^{\text{vec}}(D)$ for all $D \in (D_{i+1}, D_{i+2})$. However, this contradicts the fact that $p' \in \mathcal{J}_i^{\text{act}} \subseteq \mathcal{J}_{i+1}^{\text{act}}$. We see that we arrive at a contradiction, and therefore the premise must be false. We conclude that $\delta C^{i+1} \neq \delta C^i$. □

Proposition 3.2.6 shows that the map λ^{vec} is affine on the intervals between the breakpoints in $\mathcal{D}$ and non-differentiable at the points in $\mathcal{D}$. A similar result can be obtained for the evolution of λ^{WE}.

Corollary 3.2.7. *(Evolution of λ^{WE}): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. For any $i \in [M]_0$ there exists a value $\delta\lambda^i \geq 0$ such that for all $T \in [D_i, D_{i+1}]$*

$$\lambda^{\text{WE}}(T) = \lambda^{\text{WE}}(D_i) + (T - D_i)\delta\lambda^i.$$

Furthermore, $\delta\lambda^i = \min_{r \in \mathcal{J}_i^{\text{act}}} \delta C_r^i$.

Proof. The result follows from (3.15) and from the fact that for any $i \in [M]_0$ and $D \in (D_i, D_{i+1})$ we have $\lambda^{\text{WE}}(D) = \lambda_p^{\text{vec}}(D)$ for any $p \in \mathcal{J}_i^{\text{act}}$. □

Note that we do not state that $\delta\lambda^i \neq \delta\lambda^{i+1}$, as surprisingly, this does not necessarily hold. We show this fact, and illustrate Proposition 3.2.6 and Corollary 3.2.7 as well as some of the complexities of the evolution of the cost with an example:

Example 3.2.8. *(λ^{WE} can be differentiable at points in $\mathcal{D}$):*

(a) First we show that, even though we know $\delta C^i \neq \delta C^{i+1}$, we can have $\delta C^i = \delta C^j$ when $j \notin \{i-1, i, i+1\}$. Consider the network depicted in Figure 3.4, which is the Wheatstone network with the additional edge e_6. For e_1, e_2, e_3, e_4, and e_5 we use the

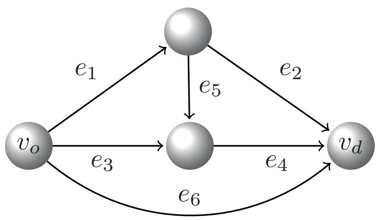


Figure 3.4: The Wheatstone network with an added parallel path.

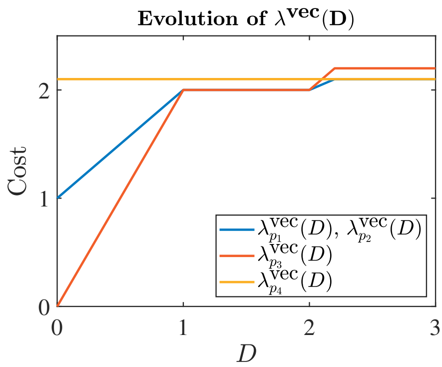


Figure 3.5: The evolution of $\lambda^{\text{vec}}(D)$ for the routing game discussed in Example 3.2.8a.

cost functions given in (3.6), and for the new edge e_6 we use

$$C_{e_6}(f_{e_6}) := 2.1.$$

There are four paths through this network, given by $p_1 := (e_1, e_2)$, $p_2 := (e_3, e_4)$, $p_3 := (e_1, e_5, e_4)$ and $p_4 := (e_6)$. The resulting path-cost function is given by

$$C(f) = Af + b = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} f + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2.1 \end{pmatrix}.$$

As long as the WE-cost of this game is lower than the constant cost of path p_4, the WE will be the same as that of the game in Example 3.2.1a, with the additional element $f_{p_4} = 0$. Furthermore, after the WE-cost has reached the constant cost of path p_4, all subsequent flow will be routed onto path p_4. That is, we have

$$f^D = \begin{cases} \begin{pmatrix} 0 & 0 & D & 0 \end{pmatrix}^\top & \text{for } D \in [0,1], \\ \begin{pmatrix} D-1 & D-1 & 2-D & 0 \end{pmatrix}^\top & \text{for } D \in [1,2], \\ \begin{pmatrix} \frac{D}{2} & \frac{D}{2} & 0 & 0 \end{pmatrix}^\top & \text{for } D \in [2,2.2], \\ \begin{pmatrix} 1.1 & 1.1 & 0 & D-2.2 \end{pmatrix}^\top & \text{for } D \in [2.2,\infty). \end{cases}$$

Therefore we obtain

$$\lambda^{\text{vec}}(D) = \begin{cases} \begin{pmatrix} 1+D & 1+D & 2D & 2.1 \end{pmatrix}^\top & \text{for } D \in [0,1], \\ \begin{pmatrix} 2 & 2 & 2 & 2.1 \end{pmatrix}^\top & \text{for } D \in [1,2], \\ \begin{pmatrix} 1+\frac{D}{2} & 1+\frac{D}{2} & D & 2.1 \end{pmatrix}^\top & \text{for } D \in [2,2.2], \\ \begin{pmatrix} 2.1 & 2.1 & 2.2 & 2.1 \end{pmatrix}^\top & \text{for } D \in [2.2,\infty). \end{cases}$$

In Figure 3.5 we see the above illustrated, and it is immediately apparent that on the intervals $D \in (1,2)$ and $D \in (2.2,\infty)$ the costs of all paths remain constant. In other words, we have $\delta C^1 = \delta C^3 = \mathbb{0}$.

(b) Using the same network we can also show that even though $\lambda^{\text{vec}}(\cdot)$ is necessarily not differentiable at the points in $\mathcal{D}$, the same does not hold for $\lambda^{\text{WE}}(\cdot)$. To see this, again consider the network in Figure 3.4, where as before the cost functions of the edges e_1, e_2, e_3, and e_4 are given by (3.6), but for the edges e_5 and e_6 we set

$$C_{e_5}(f_{e_5}) := f_{e_5}, \quad C_{e_6}(f_{e_6}) := 2 + f_{e_6}.$$

The resulting path-cost function is given by

$$C(f) = Af + b = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} f + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}$$

and we get the following expression for the WE:

$$f^D = \begin{cases} \begin{pmatrix} 0 & 0 & D & 0 \end{pmatrix}^\top & \text{for } D \in [0,\frac{1}{2}], \\ \frac{1}{3}\begin{pmatrix} 2D-1 & 2D-1 & 2-D & 0 \end{pmatrix}^\top & \text{for } D \in [\frac{1}{2},2], \\ \frac{1}{3}\begin{pmatrix} D+1 & D+1 & 0 & D-2 \end{pmatrix}^\top & \text{for } D \in [2,\infty). \end{cases}$$

For $D \in (\frac{1}{2},2)$ and $D \in (2,\infty)$ we have $\mathcal{R}_D^{\text{act}} = \{p_1,p_2,p_3\}$ and $\mathcal{R}_D^{\text{act}} = \{p_1,p_2,p_4\}$, respectively. Therefore $\lambda^{\text{vec}}(\cdot)$ should not be differentiable at $D = 2$, which is verified by noting that

$$\lambda_{p_4}^{\text{vec}}(D) = \begin{cases} 2 & \text{for } D \in [0,2], \\ \frac{1}{3}D + \frac{4}{3} & \text{for } D \in [2,\infty). \end{cases}$$

However, $\lambda^{\text{WE}}(\cdot)$ is given by

$$\lambda^{\text{WE}}(D) = \begin{cases} 3D & \text{for } D \in [0,\frac{1}{2}], \\ \frac{1}{3}D + \frac{4}{3} & \text{for } D \in [\frac{1}{2},\infty). \end{cases}$$

The full evolution of $\lambda^{\mathrm{vec}}(D)$ and λ^{WE} is depicted in Figure 3.6. We see that $\lambda^{\mathrm{WE}}(\cdot)$

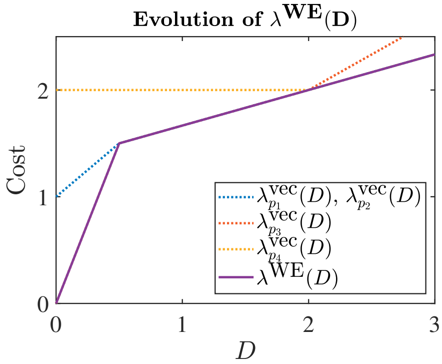


Figure 3.6: The evolution of $\lambda^{\mathrm{WE}}(D)$ and $\lambda^{\mathrm{vec}}_{p_4}(D)$ for the routing game discussed in Example 3.2.8b.

is differentiable on $(\frac{1}{2}, \infty)$, even though $\lambda^{\mathrm{vec}}_{p_3}(\cdot)$ and $\lambda^{\mathrm{vec}}_{p_4}(\cdot)$ clearly show a breakpoint at $D = 2$. •

With results on the evolution of λ^{vec} and λ^{WE} in place, our next goal is to characterize Γ_D in a comprehensive way. For ease of exposition we first define the set of *directions of feasibility*.

Definition 3.2.9. *(Direction of feasibility): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}$, and $D \geq 0$ be given. The set of* directions of feasibility *M_D is the set of all directions $f^\delta \in \mathcal{H}_1$ in which the flow can be increased such that no flow is assigned to or taken from paths that are inactive under WE, and a non-negative flow is assigned to paths that are unused under WE. That is,*

$$\mathcal{M}_D := \{f^\delta \in \mathcal{H}_1 \mid f^\delta_{\mathcal{R}^{\mathrm{act}}_D \setminus \mathcal{R}^{\mathrm{use}}_D} \geq 0,\ \ f^\delta_{(\mathcal{R}^{\mathrm{act}}_D)^c} = 0\}, \tag{3.19}$$

where $(\mathcal{R}^{\mathrm{act}}_D)^c = \mathcal{P} \setminus \mathcal{R}^{\mathrm{act}}_D$.

The main result of this chapter is that Γ_D can be obtained as the set of solutions of a VI problem, where the feasible set is given by $\mathcal{M}_D$, and the map is given by $f \mapsto Af$. Specifically, we will show the following:

Theorem 3.2.10. *(Directions of increase as solutions to a VI): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$, and $D \geq 0$ be given. Then*

$$\Gamma_D = \mathrm{SOL}(\mathcal{M}_D, A).$$

Working towards a proof of the above, we first need to establish two intermediate statements. We start by showing that the set of directions of increase is contained in the set of directions of feasibility.

Lemma 3.2.11. *(Γ_D is a subset of $\mathcal{M}_D$): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ and $D \geq 0$ be given. Then $\Gamma_D \subseteq \mathcal{M}_D$.*

Proof. Let $f^\delta \in \Gamma_D$, and let $D_i, D_{i+1} \in \mathcal{D}$ satisfy $D \in [D_i, D_{i+1})$. Then there exist $f^D \in \mathcal{W}_D$, $T \in (D, D_{i+1}]$, and $f^T \in \mathcal{W}_T$ such that $f^T = f^D + (T-D)f^\delta$. Thus we have $f^\delta = (T-D)^{-1}(f^T - f^D)$. Note that since $D \in [D_i, D_{i+1})$ and $T \in (D, D_{i+1}]$ it follows from Lemma 3.2.3 that $\mathcal{R}_T^{\mathrm{use}} \subseteq \mathcal{R}_D^{\mathrm{act}}$. Therefore $(\mathcal{R}_D^{\mathrm{act}})^c \subseteq (\mathcal{R}_T^{\mathrm{use}})^c$. Since paths that are not in the used set do not carry any flow under WE, this inclusion implies $f^T_{(\mathcal{R}_D^{\mathrm{act}})^c} = 0$. Similarly, paths that are not in the active set are not in the used set either, and therefore do not carry any flow under WE, and thus we have $f^D_{(\mathcal{R}_D^{\mathrm{act}})^c} = 0$. Combining all of the above we find $f^\delta_{(\mathcal{R}_D^{\mathrm{act}})^c} = 0$.

Once again, since paths not in the used set do not carry flow under WE, we have $f^D_{\mathcal{R}_D^{\mathrm{act}} \setminus \mathcal{R}_D^{\mathrm{use}}} = 0$. By feasibility of the WE we also have $f^T \geq 0$. It follows that $f^\delta_{\mathcal{R}_D^{\mathrm{act}} \setminus \mathcal{R}_D^{\mathrm{use}}} \geq 0$, which completes the proof. □

Next is a more technical result, that establishes some useful properties of the set $\mathrm{SOL}(\mathcal{M}, A)$, where $\mathcal{M}$ is of the same form as the set $\mathcal{M}_D$.

Proposition 3.2.12. *(Properties of $\mathrm{SOL}(\mathcal{M}, A)$): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$, $T \in \mathbb{R}$, and $\mathcal{R}, \mathcal{Q} \subseteq \mathcal{P}$ satisfying $\mathcal{R} \subseteq \mathcal{Q}$ be given. In addition, let*

$$\mathcal{M} := \{f^\delta \in \mathcal{H}_T \mid f^\delta_{\mathcal{Q}\setminus\mathcal{R}} \geq 0, \quad f^\delta_{\mathcal{Q}^c} = 0\}.$$

We then have the following:

1. *$\mathrm{SOL}(\mathcal{M}, A)$ is non-empty,*
2. *There exists a vector $\delta C \in \mathbb{R}^n$ such that $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$ if and only if $f^\delta \in \mathcal{M}$ and $Af^\delta = \delta C$,*
3. *If $f^\delta_p > 0$ for some $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$ and $p \in \mathcal{P}$, then $A_p f^\delta = \min_{r \in \mathcal{Q}} A_r f^\delta$,*
4. *If $p \in \mathcal{R}$ and $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$, then we have $A_p f^\delta = \min_{r\in\mathcal{Q}} A_r f^\delta$.*

Proof. Part 1: The first claim follows by noting that $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$ if and only if it solves [16, Section 1.3.1]

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} f^\top A f \\ \text{subject to} \quad & f \in \mathcal{M}. \end{aligned} \tag{3.20}$$

Since A is positive semi-definite and $\mathcal{M}$ is closed and convex, this minimization problem has a non-empty solution set, showing that $\mathrm{SOL}(\mathcal{M}, A)$ is non-empty.

Part 2: To prove the "only if" side, let $f^{\delta_1}, f^{\delta_2} \in \mathrm{SOL}(\mathcal{M}, A)$. By definition of the set of solutions of a VI (Definition 2.2.1) we have

$$\begin{aligned}(Af^{\delta_1})^\top(f^{\delta_2} - f^{\delta_1}) \geq 0,\\ (Af^{\delta_2})^\top(f^{\delta_1} - f^{\delta_2}) \geq 0.\end{aligned} \tag{3.21}$$

Since A is symmetric and positive semidefinite we also have

$$\begin{aligned}0 &\leq (f^{\delta_1} - f^{\delta_2})^\top A(f^{\delta_1} - f^{\delta_2})\\ &= (Af^{\delta_1})^\top(f^{\delta_1} - f^{\delta_2}) - (Af^{\delta_2})^\top(f^{\delta_1} - f^{\delta_2}) \leq 0.\end{aligned}$$

where the last inequality follows from (3.21). Thus, we have $A(f^{\delta_1} - f^{\delta_2}) = 0$. This shows that there exists a $\delta C \in \mathbb{R}^n$ such that $Af^\delta = \delta C$ for all $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$. Now we show the "if" part of the claim. Let $f^{\delta_0} \in \mathcal{M}$ be a flow satisfying $Af^{\delta_0} = \delta C$. We wish to show that $f^{\delta_0} \in \mathrm{SOL}(\mathcal{M}, A)$. Since this set is non-empty, consider some $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$. From what we have shown above, we obtain $Af^{\delta_0} = Af^\delta$. Therefore, using this equality, the following derivation holds for any $f \in \mathcal{M}$:

$$\begin{aligned}(Af^{\delta_0})^\top(f - f^{\delta_0}) &= (Af^\delta)^\top(f - f^{\delta_0})\\ &= (Af^\delta)^\top f - (f^\delta)^\top Af^{\delta_0}\\ &= (Af^\delta)^\top f - (f^\delta)^\top Af^\delta\\ &= (Af^\delta)^\top(f - f^\delta) \geq 0.\end{aligned}$$

Here the final inequality holds since $f \in \mathcal{M}$ and $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$. We see that $(Af^{\delta_0})^\top(f - f^{\delta_0}) \geq 0$ for all $f \in \mathcal{M}$ and therefore $f^{\delta_0} \in \mathrm{SOL}(\mathcal{M}, A)$. This shows that the second claim holds.

Part 3: Let $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$ and assume for the sake of contradiction that there exist $p \in \mathcal{P}$ and $r \in \mathcal{Q}$ such that $f_p^\delta > 0$ and $A_p f^\delta > A_r f^\delta$. Since $f_p^\delta > 0$ and $r \in \mathcal{Q}$ it follows by definition of $\mathcal{M}$, that $f := f^\delta - \epsilon(E^p - E^r) \in \mathcal{M}$ for small enough $\epsilon > 0$. Here E^i is the vector defined by $(E^i)_i = 1$ and $(E^i)_j = 0$ for all $j \neq i$. We then have

$$\begin{aligned}(Af^\delta)^\top(f - f^\delta) &= \epsilon(f^\delta)^\top A(E^p - E^r)\\ &= \epsilon(A_p f^\delta - A_r f^\delta) < 0.\end{aligned}$$

This contradicts $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$. Therefore the premise must be false, and we conclude that $f_p^\delta \neq 0$ implies $A_p f^\delta = \min_{r \in \mathcal{Q}} A_r f^\delta$, proving the third claim. Similar arguments can be used to show the last claim. □

The above result is not explicitly about routing games, since the feasible set of the VI allows for negative flows, and is unbounded. Nevertheless, it is quite closely related to the subject. Note for instance that when the set $\mathcal{R}$ is empty, the first three statements of the proposition are equivalent to well-known results on the existence and essential uniqueness of WE, and properties of the cost under WE, which can be proven in much the same way. To be specific, when $\mathcal{R}$ is empty, the set $\mathcal{M}$ is simply the feasible set for the routing game over the set of paths in $\mathcal{Q}$, with a path-cost function given by $C(f) = A_{\mathcal{Q}}f$ and a demand equal to T, where we recall that $A_{\mathcal{Q}} \in \mathbb{R}^{|\mathcal{Q}|\times|\mathcal{Q}|}$ is the matrix A with the rows and columns associated with paths in $\mathcal{Q}^c$ removed. The first two claims in the above result then simply state the existence and essential uniqueness of WE and the third claim states that any used path has minimal cost among all paths, which is the WE condition. Using this perspective, we get a nice intuitive interpretation of the third and fourth claims when $\mathcal{Q} = \mathcal{R}_D^{\text{act}}$ and $\mathcal{R} = \mathcal{R}_D^{\text{use}}$. Note that in this case we have $\mathcal{M} = \mathcal{M}_D$. The third claim then shows that for any $f^\delta \in \text{SOL}(\mathcal{M}_D, A)$, any path in the active set that gains flow when moving in the direction of f^δ must have the smallest increase in cost among all active paths. Similarly, the fourth claim shows that when moving in the direction of that same f^δ any path in the used set must have the smallest increase in cost among all active paths, regardless of how the flow on that path changes. Despite the similarities to results for routing games, here the set $\mathcal{M}$ is not bounded and therefore the cases are not identical. For the sake of rigor, we have therefore included the proof here.

With the previous two results established, we are ready to prove the main result of this chapter: Theorem 3.2.10.

Proof. First we show that $\text{SOL}(\mathcal{M}_D, A) \subseteq \Gamma_D$, and to do this we rely heavily on the conclusions of Proposition 3.2.12. Note the parallel between the set $\mathcal{M}$ given there and the set $\mathcal{M}_D$ used here; that is $\mathcal{M} = \mathcal{M}_D$ when we set $\mathcal{Q} := \mathcal{R}_D^{\text{act}}$ and $\mathcal{R} := \mathcal{R}_D^{\text{use}}$. We start by showing that there exists an $f^D \in \mathcal{W}_D$ satisfying $f^D_{\mathcal{R}_D^{\text{use}}} > 0$, a fact which holds because the set $\mathcal{W}_D$ is convex. To see why, consider the following argument: By definition of the used set, there exists a WE $f^p \in \mathcal{W}_D$ satisfying $f^p_p > 0$, and this is true for any $p \in \mathcal{R}_D^{\text{use}}$. Setting $f^D = \sum_{r\in\mathcal{R}_D^{\text{use}}} \mu_r f^r$, where $\mu_r \in (0,1)$ for all $r \in \mathcal{R}_D^{\text{use}}$ and $\sum_{r\in\mathcal{R}_D^{\text{use}}} \mu_r = 1$ we clearly have $f^D_{\mathcal{R}_D^{\text{use}}} > 0$. From the convexity of $\mathcal{W}_D$ we also have $f^D \in \mathcal{W}_D$. Thus there always exists a flow $f^D \in \mathcal{W}_D$ satisfying $f^D_{\mathcal{R}_D^{\text{use}}} > 0$.

Now let $f^D \in \mathcal{W}_D$ be such a WE satisfying $f^D_{\mathcal{R}_D^{\text{use}}} > 0$. For a given direction $f^\delta \in \text{SOL}(\mathcal{M}_D, A)$ and $\epsilon > 0$ we write

$$f^{D+\epsilon} = f^D + \epsilon f^\delta.$$

Note that since $f^D_{\mathcal{R}_D^{\text{use}}} > 0$ and $f^\delta \in \mathcal{M}_D$ we have $f^{D+\epsilon} \geq 0$ as long as $\epsilon > 0$ is small enough. We will show that for such a small enough ϵ we have $f^{D+\epsilon} \in \mathcal{W}_{D+\epsilon}$ which then implies $f^\delta \in \Gamma_D$. Let $p \in \mathcal{P}$ be a path such that $f_p^{D+\epsilon} > 0$. This implies that either (a) $f_p^D > 0$ or (b) $f_p^\delta > 0$. If $f_p^D > 0$ it follows that $p \in \mathcal{R}_D^{\text{use}}$. In addition, Proposition 3.2.12-4 then tells us that $A_p f^\delta = \min_{r\in\mathcal{R}_D^{\text{act}}} A_r f^\delta$. Similarly, when $f_p^\delta > 0$, the definition of $\mathcal{M}_D$ shows that $p \in \mathcal{R}_D^{\text{act}}$ and from Proposition 3.2.12-3 we obtain $A_p f^\delta = \min_{r \in \mathcal{R}_D^{\text{act}}} A_r f^\delta$. Also for both cases (a) and (b), we have $p \in \mathcal{R}_D^{\text{act}}$ and so $C_p(f^D) = \min_{r \in \mathcal{P}} C_r(f^D)$. Using these properties and Proposition 3.2.6, it follows that, as long as $\epsilon > 0$ is small enough,

$$\begin{aligned} C_p(f^{D+\epsilon}) &= C_p(f^D) + \epsilon A_p f^\delta \\ &\leq \min_{r \in \mathcal{P}} \big(C_r(f^D) + \epsilon A_r f^\delta \big) \\ &= \min_{r \in \mathcal{P}} C_r(f^{D+\epsilon}). \end{aligned}$$

Thus $f_p^{D+\epsilon} > 0$ implies $C_p(f^{D+\epsilon}) = \min_{r \in \mathcal{P}} C_r(f^{D+\epsilon})$ which means that condition (2.4) is satisfied, and it follows that $f^{D+\epsilon} \in \mathcal{W}_{D+\epsilon}$. Therefore $\mathrm{SOL}(\mathcal{M}_D, A) \subseteq \Gamma_D$.

To show $\Gamma_D \subseteq \mathrm{SOL}(\mathcal{M}_D, A)$, let $f^\delta \in \Gamma_D$. Since $\mathrm{SOL}(\mathcal{M}_D, A) \subseteq \Gamma_D$ we then know that there exists an $f^{\delta_0} \in \Gamma_D \cap \mathrm{SOL}(\mathcal{M}_D, A)$. By Lemma 3.2.11 we have $f^\delta \in \mathcal{M}_D$, and by Proposition 3.2.6 we have $Af^\delta = Af^{\delta_0}$. Therefore it follows from Proposition 3.2.12 that $f^\delta \in \mathrm{SOL}(\mathcal{M}_D, A)$. This establishes that $\Gamma_D \subseteq \mathrm{SOL}(\mathcal{M}_D, A)$. In conclusion, we find $\Gamma_D = \mathrm{SOL}(\mathcal{M}_D, A)$. □

To illustrate Theorem 3.2.10, we look at the routing game associated to Figure 3.3 as discussed in Example 3.2.1b, for the demand $D = 1$. Using (3.9) and (3.10) we find $\mathcal{R}_D^{\text{act}} = \{p_1, p_2, p_3, p_4\}$, $\mathcal{R}_D^{\text{use}} = \{p_3\}$. Thus $\mathcal{M}_D = \{f^\delta \in \mathcal{H}_1 \mid f_1^\delta, f_2^\delta, f_4^\delta \geq 0\}$. We also have $\ker(A) = \{f \in \mathbb{R} \mid f_1 = f_2 = -f_3\}$, which shows that $\ker(A) \cap \mathcal{M}_D$ is non-empty. Therefore, let $f \in \ker(A) \cap \mathcal{M}_D$. Since A is positive semi-definite we find

$$(Af^\delta)^\top (f - f^\delta) \leq 0$$

for any $f^\delta \in \mathcal{M}_D$. Using (2.5) in the definition of $\mathrm{SOL}(\mathcal{M}_D, A)$ it follows that in order for $f^\delta \in \mathrm{SOL}(\mathcal{M}_D, A)$ to hold, we need $f^\delta \in \ker(A) \cap \mathcal{M}_D$. Furthermore, when $f^\delta \in \ker(A) \cap \mathcal{M}_D$, we have $(Af^\delta)^\top (f - f^\delta) = 0$ for any $f \in \mathcal{M}_D$. In other words, $\mathrm{SOL}(\mathcal{M}_D, A) = \ker(A) \cap \mathcal{M}_D$. From (3.11) we then have $\Gamma_D = \mathrm{SOL}(\mathcal{M}_D, A)$, as claimed.

We finish this section by establishing some basic properties of the set Γ_D, which are now straightforward consequences of Theorem 3.2.10. First we have that Γ_D is closed, which follows from the fact that $\mathcal{M}_D$ is closed [16, Section 1.1], and that it is convex, which follows from [16, Theorem 2.3.5].

Corollary 3.2.13. *Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\text{aff}}$, and $D \geq 0$ be given. Then, Γ_D is non-empty, closed, and convex.*

In Example 3.2.1 we already noted that in the intervals in-between the points of $\mathcal{D}$ the set directions of increase Γ_D remains constant. Using Theorem 3.2.10 it is now easy to prove that this holds for any routing game.

Lemma 3.2.14. *Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\text{aff}}$ and $D_i, D_{i+1} \in \mathcal{D}$ be given. There exists a set $\Gamma^i \subset \mathcal{H}_1$ such that for all $D \in (D_i, D_{i+1})$ we have $\Gamma_D = \Gamma^i$. Furthermore $\Gamma^{i+1} \cap \Gamma^i = \emptyset$ and $\Gamma_{D_i} \subseteq \Gamma^i$.*

Proof. The existence of Γ^i such that $\Gamma_D = \Gamma^i$ for all $D \in (D_i, D_{i+1})$ follows directly from Theorem 3.2.10 and observing that $\mathcal{M}_D$ remains constant on the interval (D_i, D_{i+1}), which follows from Corollary 3.1.1. To show $\Gamma_{D_i} \subseteq \Gamma^i$, let $f^\delta \in \Gamma_{D_i}$ and pick $T \in (D_i, D_{i+1}]$, $f^{D_i} \in \mathcal{W}_{D_i}$, and $f^T \in \mathcal{W}_T$ such that

$$f^T = f^{D_i} + (T - D_i) f^\delta.$$

Pick $\mu \in (0,1)$ and let $T_\mu = \text{coco}_\mu(D_i, T)$ and $f^\mu = \text{coco}_\mu(f^{D_i}, f^T)$. Note that $T_\mu \in (D_i, D_{i+1})$, $T \in (T_\mu, D_{i+1})$, and that from Corollary 3.2.4 we have $f^\mu \in \mathcal{W}_{T_\mu}$. Moreover, $f^T = f^\mu + (T - T_\mu) f^\delta$, and this holds for any $\mu \in (0,1)$. Therefore $f^\delta \in \Gamma_{T_\mu} = \Gamma^i$, showing that $\Gamma_{D_i} \subseteq \Gamma^i$. For the last statement, note that if there exists $f^\delta \in \Gamma^i \cap \Gamma^{i+1}$ then the first statement of Proposition 3.2.6 implies $\delta C^{i+1} = \delta C^i$, contradicting the third statement of Proposition 3.2.6. Therefore $\Gamma^{i+1} \cap \Gamma^i = \emptyset$. □

3.3 Finding the final breaking point D_M

An interesting consequence of the results in Section 3.2, and most importantly Theorem 3.2.10, is that it enables us to design a method to directly calculate the final breaking point D_M of the set of breakpoints $\mathcal{D}$, as well as the WE-cost λ^{WE} and λ^{vec} for all $D \geq D_M$, and at least one WE f^D for all $D \geq D_M$. In other words, this method allows us to fully characterize the properties of the WE on the interval $[D_M, \infty)$. Having access to this characterization will prove to be surprisingly useful in the next chapter on Braess's paradox.

We start the exposition on how to obtain D_M and related quantities by showing an interesting relationship between the sets Γ^M and $\text{SOL}(\mathcal{F}_1, A)$. Namely, these two sets overlap, and this allows us to obtain δC^M by solving $\text{VI}(\mathcal{F}_1, A)$.

Theorem 3.3.1. *(Finding δC_M by solving $\text{VI}(\mathcal{F}_1, A)$): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\text{aff}}$, and a demand $D \geq D_M = \max\left(\mathcal{D} \setminus \{\infty\}\right)$ be given. Then we have:*

1. *$\Gamma_D \cap \text{SOL}(\mathcal{F}_1, A)$ is non-empty,*
2. *$Af^\delta = \delta C^M$ for all $f^\delta \in \text{SOL}(\mathcal{F}_1, A)$,*
3. *$\delta\lambda^M = \min_{r \in \mathcal{P}} \delta C_p^M$.*

Proof. We note that the second statement follows from the first. To see this, let $f^\delta \in \Gamma_D \cap \text{SOL}(\mathcal{F}_1, A)$. From Proposition 3.2.6 we have $Af^\delta = \delta C^M$, and Proposition

3.2.12 then gives us $A\widetilde{f}^\delta = \delta C^M$ for all $\widetilde{f}^\delta \in \mathrm{SOL}(\mathcal{F}_1, A)$. Thus, to prove the first two claims it is enough to show that $\Gamma_D \cap \mathrm{SOL}(\mathcal{F}_1, A)$ is non-empty, which we now do.

First we show that $\Gamma_D \cap \mathcal{F}_1$ is non-empty. For any $f^\delta \in \Gamma_D$ we already have $f^\delta \in \mathcal{H}_1$. Therefore to establish that $\Gamma_D \cap \mathcal{F}_1$ is non-empty it is enough to prove the existence of $f^\delta \in \Gamma_D$ such that $f^\delta \geq 0$. This we will do by constructing a sequence $\{f^{\delta,i}\}_{i\in\mathbb{N}} \subset \Gamma_D$ that converges to an f^δ satisfying $f^\delta \geq 0$. Then, using the fact that Γ_D is closed (as stated in Corollary 3.2.13), we conclude that $f^\delta \in \mathcal{F}_1 \cap \Gamma_D$.

Let $f^D \in \mathcal{W}_D$, and let $\{f^{T_i}\}_{i\in\mathbb{N}}$ be a sequence of WE satisfying $f^{T_i} \in \mathcal{W}_{T_i}$, where $D_M \leq D < T_1$, $T_i < T_{i+1}$ for all i, and $\lim_{i\to\infty} T_i = \infty$. We then define the sequence $\{f^{\delta,i}\}_{i\in\mathbb{N}}$ by setting

$$f^{\delta,i} := (T_i - D)^{-1}(f^{T_i} - f^D).$$

Since $D \geq D_M$ and $T_i > D$ for all $i \in \mathbb{N}$, it follows from Corollary 3.2.4 that any convex combination of f^D and f^{T_i} is a WE; that is, for any $\epsilon \in [0, (T_i - D)^{-1})$ we find that $f^D + \epsilon f^{\delta,i}$ is a WE. By Definition 3.2.5, we deduce that $f^{\delta,i} \in \Gamma_D$ for all $i \in \mathbb{N}$. In addition, the sequence $\{f^{\delta,i}\}_{i\in\mathbb{N}}$ is bounded. To show this, let $t := \max_{r\in\mathcal{P}} f_r^D$. Since $f^{T_i} \geq 0$ for all $i \in \mathbb{N}$ we obtain $(f_r^{T_i} - f_r^D) \geq -t$ for all $r \in \mathcal{P}$ and all $i \in \mathbb{N}$. Therefore, $f_r^{\delta,i} \geq -t(T_i - D)^{-1}$ for all $r \in \mathcal{P}$ and $i \in \mathbb{N}$. Letting $\mathcal{R}_i := \{r \in \mathcal{P} \mid f_r^{\delta,i} < 0\}$ it follows that

$$\sum_{r\in\mathcal{R}_i} f_r^\delta \geq -nt(T_i - D)^{-1}.$$

Since $f^{\delta,i} \in \mathcal{H}_1$ the above implies that $f_r^{\delta,i} \leq nt(T_i - D)^{-1} + 1$ for all $r \in \mathcal{P}$. Since $T_1 \leq T_i$ for all $i \in \mathbb{N}$ we get $-t(T_1 - D)^{-1} \leq f_r^{\delta,i} \leq -nt(T_1 - D)^{-1} + 1$ for all $r \in \mathcal{P}$ and $i \in \mathbb{N}$. In other words, $\{f^{\delta,i}\}_{i\in\mathbb{N}}$ is bounded. Therefore, the sequence contains a subsequence, denoted $\{f^{\delta,i_k}\}_{k\in\mathbb{N}}$, that converges. For this subsequence, since $\lim_{k\to\infty} T_{i_k} = \infty$ it follows that $\lim_{k\to\infty} f_r^{\delta,i_k} \geq 0$ for all $r \in \mathcal{P}$. We see that

$$f^\delta := \lim_{k\to\infty} f^{\delta,i_k} \geq 0.$$

Since Γ_D is closed, it follows that $f^\delta \in \Gamma_D$. Thus, there exists $f^\delta \in \Gamma_D \cap \mathcal{F}_1$.

The final step is to show that $f^\delta \in \Gamma_D \cap \mathcal{F}_1$ implies $f^\delta \in \mathrm{SOL}(\mathcal{F}_1, A)$. This we do by showing that f^δ is a WE of the routing game with the path cost function $\widetilde{C}(f) = Af$ and the feasible set $\mathcal{F}_1$.

Given an $f^\delta \in \Gamma_D \cap \mathcal{F}_1$, let $p \in \mathcal{P}$ satisfy $f_p^\delta > 0$. By definition of the set Γ_D we then know that for some $T > D$ there exist WE $f^T \in \mathcal{W}_T$ satisfying $f_p^T > 0$. Since $D \geq D_M$ this implies that the path p is used, and therefore active, in the interval (D_M, ∞). In other words, $p \in \mathcal{J}_M^{\mathrm{act}}$. For the sake of contradiction assume that $r \in \mathcal{P}$ satisfies $A_r f^\delta < A_p f^\delta$. Since $f^\delta \in \Gamma_D$ we have $Af^\delta = \delta C^M$, which implies $\lambda^{\mathrm{vec}}(T) = \lambda^{\mathrm{vec}}(D) + (T - D)Af^\delta$ for any $T \in (D, \infty)$. It follows that for large enough T we have $\lambda_r^{\mathrm{vec}}(T) < \lambda_p^{\mathrm{vec}}(T)$, which implies that $p \notin \mathcal{R}_T^{\mathrm{act}}$. Since $T > D$, and

$D \geq D_M$ this implies that for some $T \in (D_M, \infty)$, the path p is not active. In other words $p \notin \mathcal{J}_M^{\text{act}}$. We have arrived at a contradiction, and therefore we conclude that $A_p f^\delta = \min_{r \in \mathcal{P}} A_r f^\delta$. In other words, f^δ is a WE of the routing game with the path cost function $\widetilde{C}(f) = Af$ and the feasible set $\mathcal{F}_1$. By Proposition 2.2.2 this implies $f^\delta \in \text{SOL}(\mathcal{F}_1, A)$, which concludes the proof.

For the third statement, note that when $\delta C_r^M < \delta\lambda^M$ for some $r \in P$, this implies that in the interval (D_M, ∞), the cost of path r under WE increases at a slower rate than the WE-cost. Since the cost of all paths in this interval increases at a constant rate, this implies that at some demand $T > D_M$, $\lambda_r^{\text{vec}}(T) < \lambda^{\text{WE}}(T)$. However, the WE-cost is the minimal cost of all paths under WE, and therefore this is not possible. Thus we have $\delta\lambda^M = \min_{r \in \mathcal{P}} \delta C_r^M$. This concludes the proof. □

The intuition behind the above result is as follows. Consider a demand D in the "final" interval $[D_M, \infty)$, for which we have $\Gamma_D \subseteq \Gamma^M$ by Lemma 3.2.14, and consider the set of directions in which the WE moves as demand increases from D; that is, consider the set Γ_D. Since the interval $[D_M, \infty)$ stretches to infinity without encountering any other breakpoint, there must be some direction in Γ_D along which we can move indefinitely. For any direction f^δ which takes flow from some path, that is the vector has some negative component, one can only move a finite amount in that direction as sooner or later the flow on that path then becomes zero, and we can no longer move in that direction. Thus, there must be some $f^\delta \in \Gamma_D \cap \mathcal{F}_1$. In addition, when moving in such a "non-negative" direction of increase the cost of all paths receiving flow must remain minimal to satisfy the WE conditions. Therefore the cost of paths receiving flow must show the minimal increase among all paths. This implies that in fact $f^\delta \in \text{SOL}(\mathcal{F}_1, A)$, which in turn shows the above result. Note that these observations do not imply $\text{SOL}(\mathcal{M}_D, A) = \text{SOL}(\mathcal{F}_1, A)$, as for instance demonstrated in Example 3.2.1b. There we see that for any $D \geq D_M = 1$ there exist $f^\delta \in \Gamma_D$ and $p \in \mathcal{P}$ such that $f_p^\delta < 0$. However, once we have obtained δC^M by solving $\text{SOL}(\mathcal{F}_1, A)$ we can give a full characterization of $\Gamma^M \cap \text{SOL}(\mathcal{F}_1, A)$. Note that the result considers Γ_D for some $D \geq D_M$ instead of Γ^M, but whenever $D > D_M$ we have $\Gamma_D = \Gamma^M$.

Proposition 3.3.2. *Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\text{aff}}$, and $D \geq D_M = \max\big(\mathcal{D} \setminus \{\infty\}\big)$ be given. The set $\Gamma_D \cap \text{SOL}(\mathcal{F}_1, A)$ is equal to the set of solutions of the following minimization problem:*

$$\begin{aligned} &\text{minimize} && \beta^\top f^\delta \\ &\text{subject to} && Af^\delta = \delta C^M \\ &&& f^\delta \in \mathcal{F}_1. \end{aligned} \tag{3.22}$$

Proof. We start with the observation that taken together Proposition 3.2.12 and Theorem 3.3.1 imply $f^\delta \in \text{SOL}(\mathcal{F}_1, A)$ if and only if $f^\delta \in \mathcal{F}_1$ and $Af^\delta = \delta C^M$. In other words, the feasible set of (3.22) is equal to $\mathrm{SOL}(\mathcal{F}_1, A)$. Our first aim is therefore to show that f^δ is a solution to (3.22) if and only if it is feasible and $f^\delta \in \Gamma^M$, which will prove the result for all $D > D_M$. The same arguments are also sufficient for the case $D = D_M$.

Our first step is to show that any $f^\delta \in \Gamma^M$ that is feasible for (3.22) is a solution to the given minimization problem. To do so, let $f^{D_M} \in \mathcal{W}_{D_M}$ and consider the minimization problem

$$\begin{aligned} \text{minimize} \quad & (f^\delta)^\top (Af^{D_M} + \beta) \\ \text{subject to} \quad & Af^\delta = \delta C^M \\ & f^\delta \in \mathcal{F}_1. \end{aligned} \tag{3.23}$$

This minimization problem is actually equivalent to (3.22), meaning that f^δ is a solution to (3.22) if an only if it is a solution to (3.23). To see this, note that since $f^{D_M} \in \mathcal{W}_{D_M}$ we have $Af^{D_M} + \beta = \lambda^{\mathrm{vec}}(D_M)$. Therefore, if $p \in \mathcal{P}$ satisfies $f_p^{D_M} > 0$, then it follows that $p \in \mathcal{R}_{D_M}^{\mathrm{use}}$, and by Lemma 3.2.3 we then have $p \in \mathcal{J}_M^{\mathrm{use}}$. We see that p is in the used set on the interval $[D_M, \infty)$, and therefore it must maintain minimal cost under WE among all paths as D increases to infinity. Consequently $\delta C_p^M = \min_{r \in \mathcal{P}} \delta C_p^M = \delta \lambda^M$. We also have, by Proposition 3.2.6, $\delta C^M = Af^\delta$ for any $f^\delta \in \Gamma_D$. Thus we have $A_p f^\delta = \delta \lambda^M$ whenever $f_p^{D_M} > 0$. Since $f^{D_M} \in \mathcal{F}_{D_M}$ this implies

$$(f^\delta)^\top A f^{D_M} = D_M \delta \lambda^M \quad \text{for all } f^\delta \in \mathrm{SOL}(\mathcal{F}_1, A). \tag{3.24}$$

We see that the term $(f^\delta)^\top A f^{D_M}$ is constant over the feasible set of (3.23). Therefore f^δ solves (3.23) if and only if it minimizes the term $\beta^\top f^\delta$ over the feasible set. This is exactly the objective function of (3.22), and since the feasible sets of (3.22) and (3.23) are the same, this shows that (3.22) and (3.23) have the same solution set.

Now we find a lower bound on (3.23). We have $\lambda^{\mathrm{vec}}(D_M) = Af^{D_M} + \beta$, and in addition $\lambda^{\mathrm{WE}}(D_M) = \min_{r \in \mathcal{P}} \lambda_r^{\mathrm{vec}}(D_M)$. In other words, $\lambda^{\mathrm{vec}}(D_M) \geq \lambda^{\mathrm{WE}}(D_M)\mathbb{1}$. For any $f^\delta \in \mathcal{F}_1$ all elements of f^δ are non-negative, and sum to one, and therefore it follows that for all $f^\delta \in \mathrm{SOL}(\mathcal{F}_1, A)$ we have

$$\begin{aligned} (f^\delta)^\top (Af^{D_M} + \beta) &= (f^\delta)^\top \lambda^{\mathrm{vec}}(D_M) \\ &\geq \lambda^{\mathrm{WE}}(D_M). \end{aligned}$$

Next we show that any $f^\delta \in \Gamma_D \cap \mathrm{SOL}(\mathcal{F}_1, A)$ achieves this lower bound. We know from Theorem 3.3.1 that $\Gamma_D \cap \mathrm{SOL}(\mathcal{F}_1, A)$ is non-empty, so we can pick $\widetilde{f}^\delta \in \Gamma_D \cap \mathrm{SOL}(\mathcal{F}_1, A)$. By Theorem 3.2.10 we then have $\widetilde{f}^\delta \in \mathrm{SOL}(\mathcal{M}_D, A)$ for any $D > D_M$. Now let $p \in \mathcal{P}$ satisfy $\widetilde{f}_p^\delta > 0$. From the definition of Γ_D it follows that $p \in \mathcal{R}_T^{\mathrm{use}} = \mathcal{J}_M^{\mathrm{use}}$ for some $T > D$, and by Lemma 3.2.3 this implies $p \in \mathcal{R}_{D_M}^{\mathrm{act}}$.

Therefore $\lambda_p^{\text{vec}}(D_M) = \lambda^{\text{WE}}(D_M)$. Since $\widetilde{f}^\delta \in \mathcal{F}_1$ it follows that

$$(\widetilde{f}^\delta)^\top \lambda^{\text{vec}}(D_M) = \lambda^{\text{WE}}(D_M), \tag{3.25}$$

showing that $\widetilde{f}^\delta$ achieves the lower bound we established for (3.23). In other words, $\widetilde{f}^\delta$ is a solution of (3.23), and is therefore also a solution to (3.22). This shows that $\Gamma_D \cap \text{SOL}(\mathcal{F}_1, A)$ is contained within the set of solutions of (3.22).

Now for the other inclusion let f^δ be a solution of (3.22), and therefore also a solution of (3.23). We now know that this implies $(f^\delta)^\top \lambda^{\text{vec}}(D_M) = \lambda^{\text{WE}}(D_M)$. Since $\lambda^{\text{vec}}(D_M) \geq \lambda^{\text{WE}}(D_M)\mathbb{1}$ and $f^\delta \in \mathcal{F}_1$ this gives us

$$\lambda_p^{\text{vec}}(D_M) = \lambda^{\text{WE}}(D_M) \quad \text{ for all } p \in \mathcal{P} \text{ satisfying } f_p^\delta > 0.$$

In other words, $f_p^\delta > 0$ implies $p \in \mathcal{R}_{D_M}^{\text{act}}$. Together with $f^\delta \in \mathcal{F}_1$ this shows that $f^\delta \in \mathcal{M}_{D_M}$. Thus we have $f^\delta \in \mathcal{M}_{D_M}$ and $Af^\delta = \delta C^M$. From Theorem 3.2.10 we have $\Gamma_{D_M} = \text{SOL}(\mathcal{M}_{D_M}, A)$, from Proposition 3.2.6 we have $A\widehat{f}^\delta = \delta C^M$ for all $\widehat{f}^\delta \in \Gamma_{D_M}$ and from Proposition 3.2.12 it then follows that $\widehat{f}^\delta \in \Gamma_{D_M}$ if and only if $\widehat{f}^\delta \in \mathcal{M}_{D_M}$ and $A\widehat{f}^\delta = \delta C^M$. Thus we see that $f^\delta \in \Gamma_{D_M}$. From Corollary 3.2.14 we have $\Gamma_{D_M} \subseteq \Gamma^M$, and thus we conclude that $f^\delta \in \Gamma_D$ for any $D \geq D_M$. This completes the proof. □

Making use of the above, we can also obtain an expression for $\lambda^{\text{WE}}(D)$ for any $D \geq D_M$, as shown in the next result.

Lemma 3.3.3. *(Obtaining $\lambda^{\text{WE}}(D)$ for $D \geq D_M$): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. In addition, let $D \geq D_M = \max(\mathcal{D} \setminus \{\infty\})$ and $\beta^M = \beta^\top f^\delta$, where f^δ is a solution of* (3.22). *We have*

$$\lambda^{\text{WE}}(D) = \delta\lambda^M D + \beta^M.$$

Proof. First pick $D > D_M$, and let f^δ be a solution of (3.22). From Proposition 3.3.2 $f^\delta \in \Gamma^M \cap \text{SOL}(\mathcal{F}_1, A)$ follows. Additionally $\lambda^{\text{WE}}(D) = \lambda^{\text{WE}}(D_M) + (D - D_M)\delta\lambda^M$ holds. Now consider the expression

$$(f^\delta)^\top \Big(A\big(f^{D_M} + (D - D_M)f^\delta\big) + \beta \Big),$$

where $f^{D_M} \in \mathcal{W}_{D_M}$. Since $f^{D_M} \geq 0$ and $f^\delta \geq 0$ it follows that $f^{D_M} + (D - D_M)f^\delta \geq 0$. In combination with $f^\delta \in \Gamma^M$ this shows that $f^{D_M} + (D - D_M)f^\delta \in \mathcal{W}_D$. Thus we have $A\big(f^{D_M} + (D - D_M)f^\delta\big) + \beta = \lambda^{\text{vec}}(D)$. By the same arguments used to derive (3.25) we obtain $(f^\delta)^\top \lambda^{\text{vec}}(D) = \lambda^{\text{WE}}(D)$. Thus we have

$$\begin{aligned} (f^\delta)^\top \Big(A\big(f^{D_M} + (D - D_M)f^\delta\big) + \beta \Big) &= (f^\delta)^\top \lambda^{\text{vec}}(D) \\ &= \lambda^{\text{WE}}(D). \end{aligned} \tag{3.26}$$

We also have the following:

$$(f^\delta)^\top \Big(A\big(f^{D_M} + (D - D_M)f^\delta\big) + \beta\Big) = (f^\delta)^\top A f^{D_M} + (D - D_M)(f^\delta)^\top A f^\delta + \beta^\top f^\delta. \tag{3.27}$$

From (3.24) we have $(f^\delta)^\top A f^{D_M} = D_M \delta\lambda^M$. From arguments similar to those for (3.25) we also find $(D - D_M)(f^\delta)^\top A f^\delta = (D - D_M)\delta\lambda^M$. These facts combined with (3.26) and (3.27) yield

$$\begin{aligned}\lambda^{\mathrm{WE}}(D) &= (f^\delta)^\top \Big(A\big(f^{D_M} + (D - D_M)f^\delta\big) + \beta\Big)\\ &= D\delta\lambda^M + \beta^\top f^\delta.\end{aligned}$$

Since f^δ is a solution of (3.22) we therefore have

$$\lambda^{\mathrm{WE}}(D) = D\delta\lambda^M + \beta^M.$$

The above holds for any $D > D_M$, and it follows from Corollary 3.2.7 that it then also holds for $D = D_M$, which concludes the proof. □

Now that we have access to an expression for λ^{WE} on the interval $[D_M, \infty)$, our next result shows how we can use this information to obtain $\mathcal{J}_M^{\mathrm{act}}$.

Lemma 3.3.4. *(Obtaining $\mathcal{J}_M^{\mathrm{act}}$): For a given $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ and $D \in \mathbb{R}$ let f^δ be a solution to* (3.22). *Consider the following index sets:*

$$\begin{aligned}\mathcal{I}_1 &= \{r \in \mathcal{P} \mid f_r^\delta > 0\},\\ \mathcal{I}_2 &= \{r \in \mathcal{P} \mid f_r^\delta = 0,\ \delta C_r^M > \delta\lambda^M\},\\ \mathcal{I}_3 &= \{r \in \mathcal{P} \mid f_r^\delta = 0,\ \delta C_r^M = \delta\lambda^M\}.\end{aligned}$$

Let f^* be a solution of the following convex minimization problem:

$$\begin{aligned}
&\text{minimize} && f^\top A f + f^\top \beta && && \text{(3.28a)}\\
&\text{subject to} && f_r = 0 && \forall r \in \mathcal{I}_2, && \text{(3.28b)}\\
& && f_r \geq 0 && \forall r \in \mathcal{I}_3, && \text{(3.28c)}\\
& && C_r(f) \geq \delta\lambda^M D + \beta^M && \forall r \in \mathcal{I}_3, && \text{(3.28d)}\\
& && C_r(f) = \delta\lambda^M D + \beta^M && \forall r \in \mathcal{I}_1, && \text{(3.28e)}\\
& && \mathbb{1}^\top f = D. && && \text{(3.28f)}
\end{aligned}$$

We then have that $p \in \mathcal{J}_M^{\mathrm{act}}$ if and only if $\delta C_p^M = \delta\lambda^M$ and $C_p(f^*) = \delta\lambda^M D + \beta^M$.

The following is a long proof, and for this reason we first provide some intuition to clarify the underlying ideas. The optimization problem (3.28) is designed to identify the set $\mathcal{J}_M^{\text{act}}$. Note that due to the previous results we can already obtain δC^M and $\delta\lambda^M$ as well as an element f^δ of Γ^M, which provide quite some information about $\mathcal{J}_M^{\text{act}}$. For instance, on the interval (D_M, ∞), the cost under WE of any path $r \in \mathcal{I}_2$ increases faster than $\delta\lambda^M$, where the latter is the minimum over all paths of the increase in cost under WE. Thus, the cost of these paths can not remain minimal on this interval, which immediately shows that $\mathcal{J}_M^{\text{act}} \cap \mathcal{I}_2 = \emptyset$. Similarly, since $f^\delta \in \Gamma^M$ we know that any path in $\mathcal{I}_1$ is used under WE in the interval (D_M, ∞), which shows that $\mathcal{I}_1 \subseteq \mathcal{J}_M^{\text{act}}$. What remains is to find which paths in $\mathcal{I}_3$ are in $\mathcal{J}_M^{\text{act}}$, which can be done by solving (3.28). The reason that this works is that (3.28) is designed in such a way that for any solution f^* the flow $f^* + \epsilon f^\delta$ is a WE as long as $\epsilon > 0$ is large enough. Appropriately picking ϵ will then ensure that the constructed flow is a WE with demand in (D_M, ∞). Consequently the active set $\mathcal{J}_M^{\text{act}}$ is given by the paths that have minimal cost given the flow $f^* + \epsilon f^\delta$. Using the conditions imposed by the constraints we can show that these are exactly the paths for which $\delta C_p^M = \delta\lambda^M$ and $C_p(f^*) = \delta\lambda^M D + \bar{\beta}$, completing the argument. Before we start the proof, we explicitly note that the result holds for *any* $D \in \mathbb{R}$. Therefore it can be used without prior knowledge of D_M.

Proof of Lemma 3.3.4. We start by noting that from (3.22) we have $f^\delta \geq 0$ and from Theorem 3.3.1 we have $\delta\lambda^M = \min_{r\in\mathcal{P}} \delta C^M$. This implies that $\mathcal{P} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3$.

The first part of the proof is now to show that

$$(f^*)^\top A f^* + (f^*)^\top \beta = D(\delta\lambda^M D + \beta^M)$$

holds for any optimizer f^* of (3.28). To prove that this is true, we construct a specific optimal solution f' of (3.28) in the following way. Let $f^T \in \mathcal{W}_T$ for some $T > D_M$, and let

$$f' := f^T + (D - T) f^\delta.$$

To show that f' is an optimal solution of (3.28) our first step is to show that it satisfies all the constraints. Then we obtain the objective function value $(f')^\top A f' + (f')^\top \beta$ and finally we show that this value is optimal, proving that f' is a solution of (3.28).

Step 1: f' is feasible: We start by considering the constraints on the paths in $\mathcal{I}_1$ given in (3.28e). Note that since f^δ is a solution of (3.22), Proposition 3.3.2 gives $f^\delta \in \Gamma_T$, and Theorem 3.2.10 then implies $f^\delta \in \text{SOL}(\mathcal{M}_T, A)$. Also note that we have

$$\begin{aligned} C(f') &= A f' + \beta, \\ &= A\big(f^T + (D - T) f^\delta\big) + \beta, \\ &= A f^T + \beta + (D - T) A f^\delta, \\ &= C(f^T) + (D - T) A f^\delta. \end{aligned} \tag{3.29}$$

Now let $p \in \mathcal{I}_1$, which gives $f_p^\delta > 0$, and since $f^\delta \in \Gamma_T$ is a direction of increase, this implies that $p \in \mathcal{R}_{T+}^{\text{use}}$ for some $T^+ > T$. However, since $T \in (D_M, \infty)$, we obtain $p \in \mathcal{J}_M^{\text{use}} \subseteq \mathcal{J}_M^{\text{act}}$. Therefore, we have $p \in \mathcal{R}_T^{\text{act}}$ and so, $C_p(f^T) = \lambda^{\text{WE}}(T)$. Consequently, Lemma 3.3.3 then tells us that $C_p(f^T) = \delta\lambda^M T + \beta^M$. Furthermore, the fact that $p \in \mathcal{J}_M^{\text{use}}$, in combination with Proposition 3.2.12, implies $A_p f^\delta = \min_{r \in \mathcal{J}_M^{\text{act}}} A_r f^\delta$. It then follows from Corollary 3.2.7 that $A_p f^\delta = \delta\lambda^M$. Collecting these deduced facts that $C_p(f^T) = \delta\lambda^M T + \beta^M$ and $A_p f^\delta = \delta\lambda^M$ and employing them in (3.29) then gives us

$$\begin{aligned} C_p(f') &= \delta\lambda^M T + \beta^M + (D - T)\delta\lambda^M, \\ &= \delta\lambda^M D + \beta^M. \end{aligned}$$

Thus, f' satisfies the constraint (3.28e). Similar arguments can be used to show that any path $p \in \mathcal{I}_3$ satisfies $C_p(f^T) \geq \delta\lambda^M T + \beta^M$ and $A_p f^\delta = \delta\lambda^M$, leading to the conclusion that $C_p(f') \geq \delta\lambda^M D + \beta^M$. That is, f' satisfies (3.28c).

To show that the constraint on paths in $\mathcal{I}_2$ holds, let $p \in \mathcal{I}_2$, which by definition means $\delta C_p^M > \delta\lambda^M$. Since $f^\delta \in \Gamma_T$, from Proposition 3.2.12, we have $\delta C^M = A f^\delta$. In combination with $\delta\lambda^M = \min_{r \in \mathcal{J}_M^{\text{act}}} A_r f^\delta$ we see that $\delta C_p^M > \min_{r \in \mathcal{P}} \delta C_r^M$. In other words, the cost under WE of path p can not remain minimal on the entire interval (D_M, ∞) and therefore $p \notin \mathcal{J}_M^{\text{act}}$. Since $T \in (D_M, \infty)$, this implies $p \notin \mathcal{R}_T^{\text{act}}$, which gives $p \notin \mathcal{R}_T^{\text{use}}$. Therefore, $f_p^T = 0$. Furthermore, since $p \notin \mathcal{R}_T^{\text{act}}$ and $f^\delta \in \text{SOL}(\mathcal{M}_T, A)$ it follows from the definition of $\mathcal{M}_T$ that $f_p^\delta = 0$. Consequently, $f'_p = 0$, and so, f' satisfies the constraint (3.28b).

For the constraint (3.28c) on the paths in $\mathcal{I}_3$, note that since $f^T \in \mathcal{F}_T$, we have $f^T \geq 0$. For any p with $f_p^\delta = 0$ it then follows that $f'_p \geq 0$, as required. That the final constraint (3.28f) holds follows from the definition of f'. Thus, in summary, f' satisfies all constraints in (3.28), and is therefore feasible.

Step 2: Obtaining an expression for $(f')^\top A f' + (f')^\top$*:* The next step required for showing f' is an optimal solution of (3.28) is to further derive $(f')^\top A f' + (f')^\top \beta$, which will later be shown to be the lower bound of the objective function of (3.28) over the feasible set. To get this expression, we show that if $f'_p \neq 0$ for some p, then $C_p(f') = \delta\lambda^M D + \beta^M$. This fact is consequently used to show that we have $(f')^\top A f' + (f')^\top \beta = (f')^\top C(f') = D(\delta\lambda^M D + \beta^M)$.

Let p be a path such that $f'_p \neq 0$. It follows that either $f_p^\delta > 0$ or $f_T^\delta > 0$ (note that both vectors are non-negative, so values less than zero are not possible). For the first case, $f_p^\delta > 0$, we have $p \in \mathcal{I}_1$ which we have already shown implies $C_p(f') = \delta\lambda^M D + \beta^M$. For the second case, $f_p^T > 0$, we have $p \in \mathcal{R}_T^{\text{use}} \subseteq \mathcal{R}_T^{\text{act}}$, and therefore we have $C_p(f^T) = \lambda^{\text{WE}}(T) = \delta\lambda^M T + \beta^M$. Setting $\mathcal{M}_T = \mathcal{M}$ in Proposition 3.2.12 yields $A_p f^\delta = \min_{r \in \mathcal{R}_T^{\text{act}}} A_r f^\delta$, and Corollary 3.2.7 then implies $A_p f^\delta = \delta\lambda^M$. Using these conclusions in (3.29), we obtain $C_p(f') = \delta\lambda^M D + \beta^M$. In summary, $f'_p \neq 0$ implies $C_p(f') = \delta\lambda^M D + \beta^M$ and we then have

$$\begin{aligned}(f')^\top A f' + (f')^\top \beta &= (f')^\top C(f), \\ &= D(\delta\lambda^M D + \beta^M).\end{aligned} \tag{3.30}$$

Step 3: f' is optimal: To finish proving that f' is an optimal solution of (3.28), we show that $D(\delta\lambda^M D + \beta^M)$ is in fact a lower bound on $f^\top A f + f^\top \beta$ for any f satisfying the constraints of (3.28). Therefore, let f be an element of the feasible set of (3.28). Consider p such that $f_p \neq 0$. Since $\mathcal{P} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \mathcal{I}_3$, the constraints in (3.28) then imply $p \in \mathcal{I}_1 \cup \mathcal{I}_3$. From constraint (3.28e), if $p \in \mathcal{I}_1$, then we have $C_p(f) = \delta\lambda^M D + \beta^M$. On the other hand, if $p \in \mathcal{I}_3$, then we obtain $C_p(f) \geq \delta\lambda^M D + \beta^M$ from (3.28d).

Since $\mathbb{1}^\top f = D$, the same derivation as in (3.30) then gives

$$f^\top A f + f^\top \beta \geq D(\delta\lambda^M D + \beta^M D). \tag{3.31}$$

We see that $D(\delta\lambda^M D + \beta^M)$ is a lower bound on the objective function value of (3.28), and f' achieves this lower bound. Since f' is also feasible for this minimization problem, it follows that it is an optimal solution of (3.28).

From the above we draw the conclusion that any optimizer f^* of (3.28) satisfies $(f^*)^\top A f^* + (f^*)^\top \beta = D(\delta\lambda^M D + \beta^M)$. As shown in the derivation of (3.31), any feasible f and path p satisfying $f_p \neq 0$ satisfy $C_p(f) \geq \delta\lambda^M D + \beta^M$. Consequently we have the following for any optimizer f^* of (3.28):

$$C_p(f^*) = \delta\lambda^M D + \beta^M \quad \text{for all } p \text{ such that } f^*_p \neq 0. \tag{3.32}$$

The next part of the proof is to establish that for any optimizer f^* of (3.28), there exists $D^+ > 0$ such that the following holds:

$$f^{D^+} := f^* + (D^+ - D) f^\delta \in \mathcal{W}_{D^+}.$$

Step 4: f^{D^+} is a WE: We start by noting that if $f^*_p < 0$, then the constraint (3.28e) along with the definition of the set $\mathcal{I}_1$ imply $f^\delta_p > 0$. Thus, there exists a large enough D^+ such that $f^{D^+} \geq 0$, and therefore, $f^{D^+} \in \mathcal{F}_{D^+}$. Next, a similar derivation as in (3.29) gives us

$$C(f^{D^+}) = C(f^*) + (D^+ - D) A f^\delta. \tag{3.33}$$

Now choose p such that $f^{D^+}_p > 0$. This implies that either $f^\delta_p > 0$ or $f^*_p > 0$. For the first case, $f^\delta_p > 0$, we have $p \in \mathcal{I}_1$ and the constraint (3.28e) gives us $C_p(f^*) = \delta\lambda^M D + \beta^M$, and we have already shown in Step 1 that we then also have $A_p f^\delta = \delta\lambda^M$. Thus, we get

$$C_p(f^{D^+}) = \delta\lambda^M D^+ + \beta^M. \tag{3.34}$$

Now consider the second case $f_p^* > 0$. Using (3.32), we deduce $C_p(f^*) = \delta\lambda^M D + \beta^M$. Furthermore, constraint (3.28b) implies that $p \in \mathcal{I}_1 \cup \mathcal{I}_3$. From Step 1, we get that for $p \in \mathcal{I}_1 \cup \mathcal{I}_3$, the expression $\delta C_p^M = \delta\lambda^M$ holds. In combination with (3.33) this shows that (3.34) holds.

To establish that f^{D^+} is a WE, all that remains to be shown is that we have $C_p(f^{D^+}) \geq \delta\lambda^M D^+ + \beta^M$ whenever $f_p^{D^+} = 0$. Therefore, let p be a path such that $f_p^{D^+} = 0$. This can occur when $f_p^* < 0$ and $f_p^\delta > 0$, however, in this case the previous arguments already show that (3.34) holds. The only other way in which $f_p^{D^+} = 0$ is when $f_p^* = 0$ and $f_p^\delta = 0$. We split this scenario into two cases. First we consider $\delta C_p^M > \delta\lambda^M$. In this case it follows from $\delta C^M = Af^\delta$ in combination with (3.33) that for a large enough D^+ we get

$$C_p(f^{D^+}) \geq \delta\lambda^M D^+ + \beta^M. \tag{3.35}$$

The second case is $\delta C_p^M = \delta\lambda^M$ (In Step 3 we already argued that $\delta C_p^M = \delta\lambda^M$ for all $p \in \mathcal{I}_1$ and from the definition of $\mathcal{I}_2$ and $\mathcal{I}_3$ it follows that $\delta C_p^M < \delta\lambda^M$ is not possible.) For this case (3.28d) gives $C_p(f^*) \geq \delta\lambda^M D + \beta^M$. Once again, using $\delta C^M = Af^\delta$ in combination with (3.33) we find that (3.35) holds. In conclusion, for a large enough D^+ we have $f^{D^+} \geq 0$ and $f_p^{D^+} = 0$ implies (3.35), and $f_p^{D^+} > 0$ implies (3.34). In other words, as long as D^+ is large enough, f^{D^+} is a WE. It follows that we can pick D^+ such that $D^+ \in (D_M, \infty)$ and $f^{D^+} \in \mathcal{W}_{D^+}$.

To finish the proof we now have $\mathcal{J}_M^{\text{act}} = \mathcal{R}_{D^+}^{\text{act}}$. Since the cost under WE is unique, it follows that $p \in \mathcal{J}_M^{\text{act}}$ if and only if it has minimal cost among all paths for the flow f^{D^+}. In Step 4, we have shown that the paths with minimal cost are exactly those for which $\delta C_p^M = \delta\lambda^M$ and $C_p(f^*) = \delta\lambda^M D + \beta^M$ hold and therefore, this concludes the proof. $\square$

Now that we can derive $\mathcal{J}_M^{\text{act}}$, we finish this chapter by showing that we can also find D_M, as well as an associated WE f^{D_M}.

Corollary 3.3.5. *Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given, and consider the following minimization problem:*

$$\begin{aligned} \text{minimize} \quad & \mathbb{1}^\top f \\ \text{subject to} \quad & C_p(f) \leq C_r(f) && \textit{for all } p \in \mathcal{J}_M^{\text{act}} \\ & f_r = 0 && \textit{for all } r \in (\mathcal{J}_M^{\text{act}})^c \\ & f \geq 0. \end{aligned}$$

for any solution f^* of the above we have $\mathbb{1}^\top f^* = D_M$ and $f^* \in \mathcal{W}_{D_M}$.

Proof. Since any WE for a demand in (D_M, ∞) satisfies the constraints, we see that the feasible set is non-empty. Also note that due to the affine and non-strict nature of the constraints, this implies that any WE in $\mathcal{W}_{D_M}$ is also feasible. It is then easy to prove that any solution f^* to the given minimization problem must be a WE, and also that any flow of the form $f^* + \epsilon f^\delta$, where $f^\delta \in \Gamma_{D_M} \cap \mathcal{F}_1$ and $\epsilon > 0$, is a WE as well. This shows that f^* is a WE in the interval $[D_M, \infty)$, and it follows that $\mathbb{1}^\top f^* = D_M$, which implies $f^* \in f^{D_M}$. □

3.4 Conclusions

In this chapter we looked at how the set of Wardrop equilibria of a routing game with affine cost functions on the edges evolves as the demand increases, without imposing that the WE are unique with respect to their induced flow on the edges. We have obtained useful results on the evolution of the active and used sets, the vector of costs under WE, and given a specific demand have given a full characterization of the set of directions of increase, that is the set of directions in which the set of WE moves as the demand increases. We have then shown how these results can be used to fully characterize the WE, WE-cost, and the evolution of these in the "final" interval in which these values evolve in an affine manner.

Chapter 4

Braess's paradox

Braess's paradox(BP) is an intriguing phenomenon that occurs in routing games when the removal of a path from a network decreases the cost under Wardrop equilibrium(WE) of the associated routing game. It is a counter-intuitive and fascinating subject, that is best introduced using an example. Thus, for those readers who are unfamiliar with BP, we recommend looking at Example 4.1.1 first, to see the phenomenon illustrated, before reading on. Be advised that the exposition in Example 4.1.1 assumes the reader is already familiar with the concepts of routing games and Wardrop equilibrium, as introduced in Chapter 2. It is also worth mentioning that most literature on BP considers the phenomenon as related to the removal of a set of *edges* from a network. In this chapter we consider the slightly more general perspective of BP related to the removal of a set of *paths* from the network. The two perspectives are similar, and in fact any BP caused by the removal of a set of edges is also caused by the removal of a set of paths, but the converse is not true, as we show in Example 4.1.1b. We also feel analysis is more transparent from the path perspective, and some of our obtained results are more natural to present from this point of view. However, the number of paths in a network can grow exponentially with the number of nodes, so for practical implementation one should convert the statements to an edge-based formulation. All results we present here are straightforward to modify in this way when required.

Literature review

We give a short overview of some of the literature on Braess's paradox that is most relevant to this chapter. For an extensive overview of work published on this subject see [17]. The counter-intuitive Braess's "paradox" was first discovered in 1968 by Dietrich Braess, who presented it in [1](see [2] for an English translation), and has been extensively studied ever since [18–23]. One of the essential results obtained in the literature is that BP, or more precisely BP with respect to the removal of an *edge* from a network, can occur if and only if the considered network is not series parallel. The result is formally proven in [24], but the observation was already stated in [18]. We note that a network is series parallel if and only if it does not have the

Wheatstone network (see Figure 4.1) embedded in its structure, and it is for this reason that this network can be seen as the archetypical example of a network subject to BP. In a way the relation between the occurrence of BP and network structure is completely captured by this one fact: if the network contains an instance of the Wheatstone network, BP can occur, and if it does not, it can not. As such, other literature has focused on how the presence or absence of BP can be revealed based on other parameters of a routing game, such as the level of demand or the specifics of the cost functions [20, 22, 23]. There has also been interesting work on the the likelihood of BP occurring in a routing game [21,25], indicating that it is quite likely that BP occurs in real life traffic networks, and that the phenomenon may in fact be quite prevalent. Further supporting this line of thought are some famous real life examples of the paradox revealed by the closure of roads [26].

Of course, since BP increases travel costs, the ultimate goal is to prevent the paradox from occurring in traffic networks as much as possible. To bring obtained results closer to practical implementation various approaches have been suggested [27–29]. However, the general problem of finding the optimal subset of *edges* in a network that minimizes the travel cost under WE has proven to be very difficult, and has in fact been shown to be NP-hard [30]. Despite this established difficulty, detecting BP is exactly the aim of this chapter. In the work presented here, we study the relationship between varying demand and the occurrence of BP, and how knowledge on the evolution of the set of WE as demand varies can be used to reveal the presence of BP. For this reason the results in this chapter heavily rely on the observations and conclusions obtained in Chapter 3 on the evolution of the set of WE. In this way, Chapter 3 and the current chapter are two parts of one story, where the former is required for the latter, and the latter motivates the former.

Organization

The structure of this chapter is as follows. To familiarize the reader with Braess's paradox we first provide two illustrative examples, which also highlight one of the advantages of the path-based perspective over the edge-based perspective. Then, before we start the analysis proper, we introduce some required definitions, notation, and preliminaries. Next we use results from Chapter 3 to explore the relation between the evolution of the cost under WE and the evolution of the active and used sets, and a new set called the *necessary* set, consisting of all sets of paths that can not be removed from the game without changing the WE. We then use the obtained insight to construct *upper bounds* on the WE-cost of the routing game, meaning that whenever the WE-cost exceeds any of these upper bounds, it is certain that BP occurs. It turns out that the results in Chapter 3 on the "final" evolution of the set of WE is very useful for constructing these upper bounds. We also show how the same line of reasoning can be used to find a necessary and sufficient condition for the occurrence of BP, though the practical value of this is limited since it requires investigation of a large number of subsets of all paths. Finally we discuss some surprising consequences of our results for the effects of BP when it occurs. We show that, depending on which measure one uses, the removal of a path that causes BP may not be a good idea, since the detrimental effects the path's presence has at one level of demand are often compensated by the beneficial effects it has at other levels of demand.

4.1 Introductory examples and preliminaries

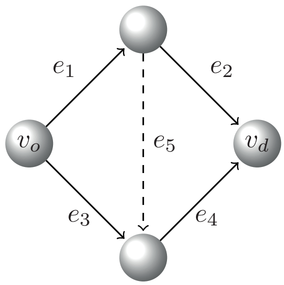


Figure 4.1: The Wheatstone network.

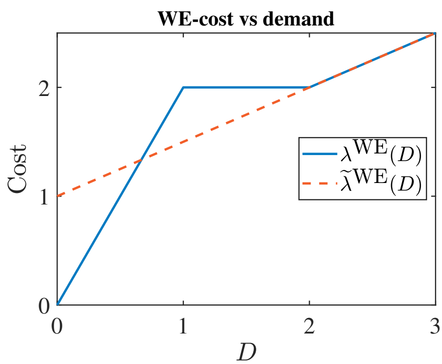


Figure 4.2: The cost under Wardrop equilibrium at different demands for the routing game defined by the Wheatstone network (Figure 4.1) and costs (4.1).

To familiarize the reader with the concept of Braess's paradox, and to showcase the difference between BP caused by edges and BP caused by paths, we start with an illustrative example. Note that we assume familiarity of the reader with the subject of routing games, as introduced in Chapter 2. For our example we consider the archetypical instance of BP that occurs in the Wheatstone network, with the edge cost functions used in Example 3.2.1a discussed in Chapter 3. For ease of reference we recall this example here.

Example 4.1.1. *(Evolution of BP):*

(a) Consider the network in Figure 4.1, with edge-cost functions given by

$$\begin{aligned} C_{e_1}(f_{e_1}) &= f_{e_1}, & C_{e_2}(f_{e_2}) &= 1, \\ C_{e_3}(f_{e_3}) &= 1, & C_{e_4}(f_{e_4}) &= f_{e_4}, \\ & & C_{e_5}(f_{e_5}) = 0. \end{aligned} \tag{4.1}$$

The set of paths is then given by $\mathcal{P} = \{p_1, p_2, p_3\}$, where $p_1 = (e_1, e_2)$, $p_2 = (e_3, e_4)$ and $p_3 = (e_1, e_5, e_4)$. The resulting path-cost function is

$$C(f) = Af + b = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} f + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

For this example the Wardrop equilibria are unique for each level of demand, and given by

$$f^D = \begin{cases} \begin{pmatrix} 0 & 0 & D \end{pmatrix}^\top & \text{for } D \in [0,1], \\ \begin{pmatrix} D-1 & D-1 & 2-D \end{pmatrix}^\top & \text{for } D \in [1,2], \\ \begin{pmatrix} \frac{D}{2} & \frac{D}{2} & 0 \end{pmatrix}^\top & \text{for } D \in [2,\infty). \end{cases}$$

This allows us to derive that the cost under WE is given by

$$\lambda^{\mathrm{WE}}(D) = \begin{cases} 2D & \text{if } 0 \le D \le 1, \\ 2 & \text{if } 1 \le D \le 2, \\ \frac{D}{2} + 1 & \text{if } 2 \le D. \end{cases}$$

Next we consider the same network, with the same edge-cost functions, but with edge e_5 removed. This effectively removes the path p_3 from consideration, and leaves us with a routing game over only the paths p_1 and p_2, where the path-cost function is given by

$$\widetilde{C}(\widetilde{f}) = \widetilde{A}\widetilde{f} + \widetilde{b} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \widetilde{f} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Once again, the WE is unique for each level of demand. In this case it is given by

$$\widetilde{f}^D = \begin{pmatrix} \frac{D}{2} & \frac{D}{2} \end{pmatrix}^\top \quad \text{for all } D \ge 0,$$

which gives us the following expression for the cost under WE:

$$\widetilde{\lambda}^{\mathrm{WE}}(D) = \frac{D}{2} + 1 \quad \text{for all } D \ge 0.$$

In Figure 4.2 we compare the WE-cost λ^{WE} of the original game with $\widetilde{\lambda}^{\mathrm{WE}}$, that is the cost of the modified game with edge e_5 removed. Note that for a level of demand between $\frac{2}{3}$ and 2, the modified game achieves a lower cost than the original game. This is the phenomenon referred to as Braess's paradox. Counter-intuitively, the removal of a part of the network has decreased the travel cost of all users.

However, there is more to this story, as can already be deduced from Figure 4.2. We see that the presence of edge e_5 is detrimental when $\frac{2}{3} < D < 2$, but also that its presence is beneficial when $0 \leq D < \frac{2}{3}$, and neutral when $D \in \{\frac{2}{3}\} \cup [2, \infty)$. Clearly, Braess's paradox is demand dependent, and the occurrence of the paradox at one level of demand does not necessarily imply that the presence of the relevant set of edges is detrimental overall. The situation warrants further investigation, and this is the subject of this chapter. However, before moving on we discuss one more example which highlights the advantage of considering BP from the perspective of removing paths rather than edges.

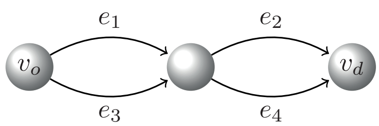


Figure 4.3: The Wheatstone network after merging the top and bottom nodes.

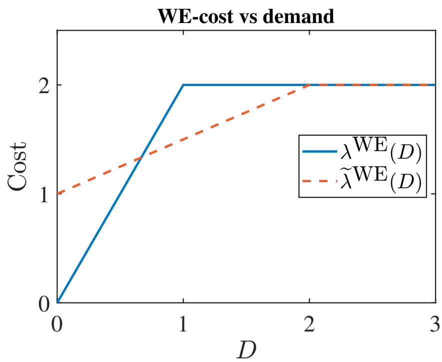


Figure 4.4: An illustration of BP comparing the cost of the routing games over the network Figure 4.3, with and without the path $p_4 = (e_1, e_4)$ present. The costs are defined by (4.1).

(b) Classically, Braess' paradox refers to the situation where removal of an *edge* (or a set of edges) leads to a lesser cost for all participants, but in this chapter we will consider a slightly more generalized form of the paradox, where removal of a *path*, or set of paths, from the network leads to a lesser cost for all participants. The previous example shows an instant where these two cases are the same. However, it is possible that a BP is present in a routing game that only emerges as the result of removing a (set of) path(s), but not as the result of removing a (set of) edge(s). To show this, we revisit Example 3.2.1b. That is, we consider a routing game over the network in Figure 4.3, with the edge-costs given by (4.1), taking into account that edge e_5 no longer exists. Setting $p_1 = (e_1, e_2)$, $p_2 = (e_3, e_4)$, $p_3 = (e_3, e_2)$ and $p_4 = (e_1, e_4)$, the path-cost function becomes

$$C(f) = Af + b = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 2 \end{pmatrix} f + \begin{pmatrix} 1 \\ 1 \\ 2 \\ 0 \end{pmatrix}.$$

We obtain the following expression for the set of WE

$$\mathcal{W}_D := \begin{cases} \left\{ \begin{pmatrix} 0 & 0 & 0 & D \end{pmatrix}^\top \right\} & \text{for } D \in [0,1], \\ \{ f \in \mathcal{F}_D \mid f_1 + f_4 = 1, f_2 + f_4 = 1 \} & \text{for } D \in [1, \infty), \end{cases} \tag{4.2}$$

and therefore the WE cost is given by

$$\lambda^{\mathrm{WE}}(D) = \begin{cases} 2D & \text{for } D \in [0,1], \\ 2 & \text{for } D \in [1, \infty). \end{cases}$$

If we instead consider the same routing game, but with the path $p_4 = (e_1, e_4)$ removed, the path-cost function is given by

$$\widetilde{C}(\widetilde{f}) = \widetilde{A}\widetilde{f} + \widetilde{b} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \widetilde{f} + \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix},$$

and we obtain the following expression for the WE

$$\widetilde{f}^D := \begin{cases} \begin{pmatrix} \frac{D}{2} & \frac{D}{2} & 0 \end{pmatrix}^\top & \text{for } D \in [0,2], \\ \begin{pmatrix} 1 & 1 & D-2 \end{pmatrix}^\top & \text{for } D \in [2, \infty). \end{cases}$$

The resulting WE-cost is given by

$$\widetilde{\lambda}^{\mathrm{WE}}(D) = \begin{cases} 1 + \frac{D}{2} & \text{for } D \in [0,2], \\ 2 & \text{for } D \in [2, \infty). \end{cases}$$

Figure 4.4 shows the comparison between the cost of the original game λ^{WE} and $\widetilde{\lambda}^{\mathrm{WE}}$, that is the cost of the modified game with path p_4 removed. Note that the case is almost identical to that of Example 4.1.1a. The same BP occurs on the interval $D \in (\frac{2}{3}, 2)$, but in this case there is no edge that is responsible for the inefficiency. Instead it is the presence of the path $p_4 = (e_1, e_4)$ that causes the BP. Also note that the network considered in this last example is series parallel, and thus we see that the result that BP does not occur in series parallel networks no longer holds when considering BP from a path perspective, which is an interesting observation in itself.•

4.1.1 Notation, Facts and Definitions

In this section we take the time to introduce and recall some concepts required for the exposition of the rest of this chapter. We note that some of these preliminaries have already been introduced in Chapter 3. For ease of reference, and to make this chapter as self-contained as possible, we recall this material here, but for a more detailed discussion we refer the reader to Section 3.1.1. We also assume that the reader is familiar with the general preliminaries on routing games given in Chapter 2.

Affine costs, WE-costs, and active and used sets

In this chapter, as in the previous, we assume that the edge-cost functions are of the form

$$C_{e_k}(f_{e_k}) := \alpha_{e_k} f_{e_k} + \beta_{e_k},$$

where $\alpha_{e_k}, \beta_{e_k} \geq 0$. We recall that in this case the path-cost function can be written as

$$C(f) = Af + \beta,$$

where $\beta = (\beta_p)_{p \in \mathcal{P}}$ is the vector with entries $\beta_p = \sum_{e_k \in p} \beta_{e_k}$ and $A \in \mathbb{R}^{n \times n}_{\geq 0}$ is a symmetric and positive semidefinite matrix with the (p, r)-th entry given by $A_{pr} = \sum_{e_k \in p \cap r} \alpha_{e_k}$. Throughout this and the previous chapter we make the assumption that the edge-cost functions are affine explicit by writing $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$. We also recall from Section 3.1.1 the definitions of the vector of path costs under WE $\lambda^{\text{vec}}(D)$, the cost under WE $\lambda^{\text{WE}}(D)$, the active set $\mathcal{R}_D^{\text{act}}$ and the used set $\mathcal{R}_D^{\text{use}}$, respectively given by

$$\begin{aligned}
\lambda^{\text{vec}}(D) &:= C(f^D), \quad f^D \in \mathcal{W}_D, \\
\lambda^{\text{WE}}(D) &:= \lambda_p^{\text{vec}}(D), \quad p \in \mathcal{R}_D^{\text{act}}, \\
\mathcal{R}_D^{\text{act}} &:= \{p \in \mathcal{P} \mid \lambda_p^{\text{vec}}(D) \leq \lambda_r^{\text{vec}}(D), \text{ for all } r \in \mathcal{P}\} \\
\mathcal{R}_D^{\text{use}} &:= \{p \in \mathcal{P} \mid \text{There exists an } f^D \in \mathcal{W}_D \text{ such that } f_p^D > 0\},
\end{aligned}$$

where $\mathcal{W}_D$ denotes the set of WE of the routing game at demand D. Finally we recall that the set $\mathcal{D}$ denotes the "breakpoints" of the routing game; that is, in between the points in $\mathcal{D}$ the active and used sets remain constant, and the points in $\mathcal{D}$ are exactly the points where the piecewise affine function λ^{vec} is not differentiable.

Necessary sets

In addition to the active and used sets, we now define the closely related concept of the *necessary* set. Intuitively the necessary set at demand D, denoted $\mathcal{N}_D$, is the set of all sets of paths that can not be removed from the game without changing the WE at the demand D. Formally we have the following definition:

Definition 4.1.2. *(Necessary sets): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. we say that a set $\mathcal{S}^{nec} \subseteq \mathcal{P}$ is* necessary *at demand D if*

$$f^D_{\mathcal{S}^{nec}} \neq 0 \textit{ for all } f^D \in \mathcal{W}_D.$$

That is, for every WE f^D there exists at least one path in the set $\mathcal{S}^{nec}$ that takes non-zero flow. We use $\mathcal{N}_D$ to denote the set of all necessary sets at D; that is,

$$\mathcal{N}_D := \{\mathcal{S}^{nec} \subseteq \mathcal{P} \mid f^D_{\mathcal{S}^{nec}} \neq 0 \textit{ for all } f^D \in \mathcal{W}_D\}.$$

When $\mathcal{S} \notin \mathcal{N}_D$ we say that $\mathcal{S}$ is unnecessary *at demand D.* •

Note that for any $\mathcal{S} \subset \mathcal{P}$, if $\mathcal{S} \cap \mathcal{R}^{\text{use}}_D = \emptyset$ then $\mathcal{S} \notin \mathcal{N}_D$. Due to this close relation to the used set, one might expect that the dependency of the necessary set on the demand is similar to that of the the used set. Therefore we find it important to note that unlike the used set, the necessary set is not guaranteed to stay constant in the intervals between the points of $\mathcal{D}$. For instance, in Example 4.1.1b we have $\mathcal{D} = \{0, 1, \infty\}$. However, we can deduce from the expression given for $\mathcal{W}_D$ in (4.2) that $\{p_4\} \in \mathcal{N}_D$ for $D \in (0, 2)$, but $\{p_4\} \notin \mathcal{N}_D$ for $D \geq 2$, showing that the necessary set changes at $D = 2$.

Modified games

Next we introduce the concept of a *modified game*. Braess's paradox is related to the difference in WE-cost of a routing game and a modified version of that routing game, where a set of paths has been removed from consideration. For this reason we find it useful to introduce notation for the sets of paths, sets of WE, WE-costs, etc. of this modified game. Thus, given a routing game defined by a set of paths $\mathcal{P}$ and cost functions $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$, a modified game is constructed by *removing* a set $\mathcal{S}^{\text{rem}} \subset \mathcal{P}$ from consideration; that is, we replace the feasible set $\mathcal{F}_D$ of the original game with the set

$$\widetilde{\mathcal{F}}_D := \{\widetilde{f} \in \mathbb{R}^n_{\geq 0} \mid \sum_{i \in \mathcal{P}} \widetilde{f}_i = D, \;\; \widetilde{f}_{\mathcal{S}^{\text{rem}}} = 0\}.$$

We introduce the following notation:

$$\widetilde{\mathcal{P}} := \mathcal{P} \setminus \mathcal{S}^{\texttt{rem}},$$
$$\widetilde{\mathcal{H}}_D := \{\widetilde{f} \in \mathbb{R}^n \mid \sum_{i \in \widetilde{\mathcal{P}}} \widetilde{f}_i = D, \;\; \widetilde{f}_{\mathcal{S}^{\texttt{rem}}} = 0\}.$$

For notational convenience the dimension of the flows $\widetilde{f} \in \widetilde{\mathcal{F}}_D$ of the modified game is kept equal to $n = |\mathcal{P}|$. For a modified game, we say that $\widetilde{f}^D$ is a WE when $\widetilde{f}^D \in \widetilde{\mathcal{F}}_D$ and for all $p \in \widetilde{\mathcal{P}}$ such that $\widetilde{f}_p > 0$ we have

$$C_p(\widetilde{f}^D) \leq C_r(\widetilde{f}^D) \quad \text{for all } r \in \widetilde{\mathcal{P}}.$$

Analogously we use this $\widetilde{(\cdot)}$ notation for other concepts related to the modified game. Specifically, we have the following:

$$\begin{aligned}
\widetilde{\mathcal{W}}_D &:= \mathrm{SOL}(\widetilde{\mathcal{F}}_D, C) \\
\widetilde{\lambda}^{\mathrm{vec}}(D) &:= C(\widetilde{f}^D) \text{ for any } \widetilde{f}^D \in \widetilde{\mathcal{W}}_D, \\
\widetilde{\mathcal{R}}_D^{\mathrm{act}} &:= \{p \in \widetilde{\mathcal{P}} \mid \widetilde{\lambda}_p^{\mathrm{vec}}(D) \leq \widetilde{\lambda}_r^{\mathrm{vec}}(D) \text{ for all } r \in \widetilde{\mathcal{P}}\}, \\
\widetilde{\mathcal{R}}_D^{\mathrm{use}} &:= \{p \in \widetilde{\mathcal{P}} \mid \text{There exists an } \widetilde{f}^D \in \widetilde{\mathcal{W}}_D \text{ such that } \widetilde{f}_p^D > 0\}, \\
\widetilde{\lambda}^{\mathrm{WE}}(D) &:= \widetilde{\lambda}^{\mathrm{vec}}(D) \text{ for any } p \in \widetilde{\mathcal{R}}_D^{\mathrm{act}} \\
\widetilde{\mathcal{M}}_D &:= \{\widetilde{f}^\delta \in \widetilde{\mathcal{H}}_1 \mid \widetilde{f}^\delta_{\widetilde{\mathcal{R}}_D^{\mathrm{act}} \setminus \widetilde{\mathcal{R}}_D^{\mathrm{use}}} \geq 0, \;\; \widetilde{f}^\delta_{(\widetilde{\mathcal{R}}_D^{\mathrm{act}})^c} = 0\}, \\
\widetilde{\Gamma}_D &:= \mathrm{SOL}(\widetilde{\mathcal{M}}_D, A).
\end{aligned}$$

Similarly we use $\widetilde{\mathcal{D}}$ to denote the set of breakpoints of a modified game and use $\widetilde{D}_i$ to denote the i-th breakpoint of $\widetilde{\mathcal{D}}$. We write $\widetilde{M}$ for the index of the greatest finite valued breakpoint $\widetilde{D}_{\widetilde{M}}$ in $\widetilde{\mathcal{D}}$ and use $\delta\widetilde{\lambda}^i$ and $\delta\widetilde{C}^i$ to denote the directions in which respectively $\widetilde{\lambda}^{\mathrm{WE}}$ and $\widetilde{\lambda}^{\mathrm{vec}}$ evolve on the interval between $\widetilde{D}_i$ and $\widetilde{D}_{i+1}$. When we need to discuss multiple modified games simultaneously, as is sometimes the case in a proof, we use a similar notation for concepts related to these modified games, replacing $\widetilde{(\cdot)}$ with $\breve{(\cdot)}$ or with $(\cdot)'$, $(\cdot)''$ or $(\cdot)'''$. (e.g. we use $\breve{\mathcal{P}}$, and $\mathcal{P}'$, $\mathcal{P}''$ and $\mathcal{P}'''$ to denote the related sets of paths, and similarly for the feasible set, WE-cost etc.).

We note that modified games, as presented here, are technically not routing games of the same form as those presented in Chapter 2, because of the additional constraints on the feasible set, and because the introduced concept of WE for these modified games does not take into account the costs of paths in the set $\mathcal{S}^{\texttt{rem}}$. Of course, instead of imposing $\widetilde{f}_{\mathcal{S}}^{\texttt{rem}} = 0$, we could simply drop this set of paths $\mathcal{S}^{\texttt{rem}}$ from consideration, and let $\widetilde{f} \in \mathbb{R}_{\geq 0}^{|\widetilde{\mathcal{P}}|}$. In this case we could define $\widetilde{\mathcal{F}}_D$ without imposing additional restrictions, and in this representation the modified game is of the form presented in Chapter 2. However, removing a set of paths from a routing game like this can result in a situation that is no longer representable by a graph. That is, in that case there does not exist a graph, an associated origin-destination pair and a set of cost functions such that the resulting routing game over all paths from the origin to the destination has the same cost function as that of the modified game (see Appendix A for an example). For this reason, and because of some notational conveniences we have chosen to define modified games in the presented way. Since there exists a straightforward way to equivalently represent a modified game in a form aligning with the presentation of routing games in Chapter 2, all results that we have established and will establish for routing games also hold for modified games. We highlight that this implies that $\widetilde{\mathcal{W}}_D$ and $\widetilde{\Gamma}_D$, as defined above, are indeed the set of WE of the modified game and the set of directions of increase of the modified game respectively, as the notation suggests.

Directions of decrease

The results presented in Chapter 3 consider the set of directions of *increase*, which is the set of directions in which the set of WE evolves as the demand increases. For our investigation into BP it will be helpful to also have these results formulated in terms of *directions of decrease*.

Definition 4.1.3. *(Set of directions of decrease): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\text{aff}}$ and $D > 0$ be given. The set of* directions of decrease Γ_D^- *is the set of all directions $f^\delta \in \mathcal{H}_{-1}$ in which the flow can be decreased, starting from some flow in $\mathcal{W}_D$, such that the new flow is a WE as long as the decrease is small enough. That is,*

$$\Gamma_D^- := \{f^\delta \in \mathcal{H}_{-1} \mid \exists f^D \in \mathcal{W}_D,\ \ \bar{\epsilon} > 0 \ \text{ such that } f^D + \epsilon f^\delta \in \mathcal{W}_{D-\epsilon} \ \text{ for all } \epsilon \in [0, \bar{\epsilon}]\}.$$

Similar to the definition of the set of directions of feasibility (Definition 3.2.9), we define the set $\mathcal{M}_D^-$ of feasible descent directions as

$$\mathcal{M}_D^- := \{f^\delta \in \mathcal{H}_{-1} \mid f^\delta_{\mathcal{R}_D^{\text{act}} \setminus \mathcal{R}_D^{\text{use}}} \geq 0,\ \ f^\delta_{(\mathcal{R}_D^{\text{act}})^c} = 0\}.$$

The arguments made in Chapter 3 can then be repeated to obtain the following modified version of Theorem 3.2.10.

Lemma 4.1.4. *(Directions of decrease as solutions to a VI): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\text{aff}}$, and $D > 0$ be given. Then*

$$\Gamma_D^- = \text{SOL}(\mathcal{M}_D^-, A).$$

Properties of V

Finally we recall from Chapter 2 the definition of the function V:

$$V(D) := \min_{f \in \mathcal{F}_D} \sum_{e_k \in \mathcal{E}} \int_0^{f_{e_k}} C_{e_k}(z)dz.$$

A useful result from [9] shows that V is differentiable on the positive real line, as well as convex, and that the derivative of V at D is given by $\lambda^{\mathrm{WE}}(D)$.

Proposition 4.1.5. *(Properties of WE cost λ^{WE} [9]): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ be given. The function V is differentiable, and for any $D > 0$ we have*

$$\frac{\partial}{\partial D} V(D) = \lambda^{\mathrm{WE}}(D).$$

In addition λ^{WE} is non-decreasing, and consequently, V is convex.

For a modified game we define

$$\widetilde{V}(D) := \min_{\widetilde{f} \in \widetilde{\mathcal{F}}_D} \sum_{e_k \in \mathcal{E}} \int_0^{\widetilde{f}_{e_k}} \widetilde{C}_{e_k}(z)dz. \tag{4.3}$$

We then have the following straightforward observations on the relationship between routing games, modified games and necessary sets.

Lemma 4.1.6. *(Relations between the original and the modified game): For given $\mathcal{P}$, $\mathcal{S}^{rem} \subset \mathcal{P}$, $\mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ and D, let $V(D)$ and $\widetilde{V}(D)$ be as defined in* (2.6) *and* (4.3) *respectively. The following then hold:*

- $V(D) \leq \widetilde{V}(D)$,
- $V(D) = \widetilde{V}(D)$ *if and only if* $\mathcal{S}^{rem} \notin \mathcal{N}_D$,
- *if* $\mathcal{S}^{rem} \notin \mathcal{N}_D$, *then* $\widetilde{f}^D \in \widetilde{\mathcal{W}}_D$ *if and only if* $\widetilde{f}^D_{\mathcal{S}^{rem}} = 0$ *and* $\widetilde{f}^D \in \mathcal{W}_D$. *As a consequence we then have* $\lambda^{\mathrm{WE}}(D) = \widetilde{\lambda}^{\mathrm{WE}}(D)$.

The above Lemma is illustrated in Figure 4.5, which shows the functions $V(D)$ and $\widetilde{V}(D)$ for the games considered in Example 4.1.1a. For demand in the interval $(0, 2)$, where the path $p_3 = (e_1, e_5, e_4)$ is a necessary set, we see that $V(D) < \widetilde{V}(D)$, and exactly at the point $D = 2$, where this path is no longer necessary, we have $V(D) = \widetilde{V}(D)$.

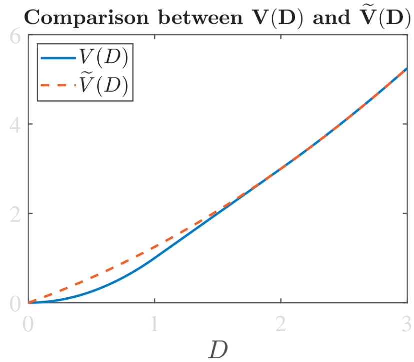


Figure 4.5: A comparison between $V(D)$ and $\widetilde{V}(D)$ for the Wheatstone network in Figure 4.1, with and without edge e_5 present and with edge-costs defined by (4.1).

4.2 The evolution of WE-costs in routing games

Our investigation into the relation between varying demand and Braess's paradox begins with an analysis of how changes in the active and used sets affect the evolution of the WE-cost. In particular we look at how the slope of λ^{WE} changes at the breakpoints in $\mathcal{D}$. To ease the exposition of results on this subject we define the following notation for left- and right-hand derivatives of the function λ^{WE}:

$$\delta\lambda^+(D) := \frac{\partial^+}{\partial D}\lambda^{\mathrm{WE}}(D), \quad \delta\lambda^-(D) := \frac{\partial^-}{\partial D}\lambda^{\mathrm{WE}}(D).$$

Note that from Corollary 3.2.7 we know that λ^{WE} is piecewise affine, and differentiable outside of the set $\mathcal{D}$, where we thus have $\delta\lambda^+(D) = \delta\lambda^-(D)$. Now using Proposition 3.2.12 with $\mathcal{M} = \mathcal{M}_D$ and $\mathcal{M} = \mathcal{M}_D^-$ we have the following observation:

Corollary 4.2.1. *(Relation between* $\mathrm{SOL}(\mathcal{M}_D, A)$ *and slope of* λ^{WE}*): Let* $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$*, and* $D \geq 0$ *be given. We have*

$$\delta\lambda^+(D) = \min_{r \in \mathcal{R}_D^{\mathrm{act}}} A_r f^\delta \qquad \textit{for all } f^\delta \in \mathrm{SOL}(\mathcal{M}_D, A),$$

$$\delta\lambda^-(D) = -\min_{r \in \mathcal{R}_D^{\mathrm{act}}} A_r f^\delta \quad \textit{for all } f^\delta \in \mathrm{SOL}(\mathcal{M}_D^-, A).$$

We see that sets of the form $\mathrm{SOL}(\mathcal{M}_D, A)$ are important when studying the evolution of $\delta\lambda^+(D)$ and $\delta\lambda^-$(D). The next result concerns $\mathrm{SOL}(\mathcal{M}, A)$, where the set $\mathcal{M}$ is of a form similar to that of $\mathcal{M}_D$ given in Definition 3.2.9. As discussed after Proposition 3.2.12, sets of the form $\mathrm{SOL}(\mathcal{M}, A)$ of this form are not explicitly about calculating WE of a routing game, but are quite closely related to the subject.

Lemma 4.2.2. *(Properties of* $\mathrm{SOL}(\mathcal{M}, A)$*): Let* $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ *and sets* $\mathcal{Q}, \mathcal{R}, \widetilde{\mathcal{Q}}, \widetilde{\mathcal{R}} \subseteq \mathcal{P}$ *satisfying* $\mathcal{R} \subseteq \mathcal{Q}$ *and* $\widetilde{\mathcal{R}} \subseteq \widetilde{\mathcal{Q}}$ *be given. In addition let*

$$\begin{aligned}
\mathcal{M} &:= \{f^\delta \in \mathcal{H}_1 \mid f^\delta_{\mathcal{Q}\setminus\mathcal{R}} \geq 0,\ \ f^\delta_{\mathcal{Q}^c} = 0\},\\
\widetilde{\mathcal{M}} &:= \{f^\delta \in \mathcal{H}_1 \mid f^\delta_{\widetilde{\mathcal{Q}}\setminus\widetilde{\mathcal{R}}} \geq 0,\ \ f^\delta_{\widetilde{\mathcal{Q}}^c} = 0\},\\
\mathcal{M}^- &:= \{f^\delta \in \mathcal{H}_{-1} \mid f^\delta_{\mathcal{Q}\setminus\mathcal{R}} \geq 0,\ \ f^\delta_{\mathcal{Q}^c} = 0\},\\
\widetilde{\mathcal{M}}^- &:= \{f^\delta \in \mathcal{H}_{-1} \mid f^\delta_{\widetilde{\mathcal{Q}}\setminus\widetilde{\mathcal{R}}} \geq 0,\ \ f^\delta_{\widetilde{\mathcal{Q}}^c} = 0\},
\end{aligned}$$

and let $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$, $\widetilde{f}^\delta \in \mathrm{SOL}(\widetilde{\mathcal{M}}, A)$*. If* $\widetilde{\mathcal{M}} \subseteq \mathcal{M}$ *then*

$$\min_{r\in\widetilde{\mathcal{Q}}} A_r \widetilde{f}^\delta \geq \min_{r\in\mathcal{Q}} A_r f^\delta.$$

Similarly, if $f^{\delta-} \in \mathrm{SOL}(\mathcal{M}^-, A)$, $\widetilde{f}^{\delta-} \in \mathrm{SOL}(\widetilde{\mathcal{M}}^-, A)$ *and* $\widetilde{\mathcal{M}}^- \subseteq \mathcal{M}^-$ *then*

$$\min_{r\in\widetilde{\mathcal{Q}}} A_r \widetilde{f}^{\delta-} \leq \min_{r\in\mathcal{Q}} A_r f^{\delta-}.$$

Proof. We first prove the statement concerning the case $\widetilde{\mathcal{M}} \subseteq \mathcal{M}$. For $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$, let $p \in \mathcal{P}$ be a path satisfying $f^\delta_p \neq 0$. Then, either $f^\delta_p > 0$ or $p \in \mathcal{R}$. For both these cases Proposition 3.2.12 tells us that $A_p f^\delta = \min_{r\in\mathcal{Q}} A_r f^\delta$. Using this and the fact that $f^\delta \in \mathcal{H}_1$ and $f^\delta_{\mathcal{Q}^c} = 0$, we obtain $(f^\delta)^\top A f^\delta = \min_{r\in\mathcal{Q}} A_r f^\delta$. Similarly we find $(\widetilde{f}^\delta)^\top A \widetilde{f}^\delta = \min_{r\in\widetilde{\mathcal{Q}}} A_r \widetilde{f}^\delta$. Now, assume for the sake of contradiction that

$$\min_{r\in\widetilde{\mathcal{Q}}} A_r \widetilde{f}^\delta < \min_{r\in\mathcal{Q}} A_r f^\delta. \tag{4.4}$$

Since $f^\delta \in \mathrm{SOL}(\mathcal{M}, A)$ and $\widetilde{f}^\delta \in \widetilde{\mathcal{M}} \subseteq \mathcal{M}$, we have

$$(f^\delta)^\top A(\widetilde{f}^\delta - f^\delta) \geq 0, \tag{4.5}$$

which gives us the following derivation:

$$\begin{aligned}
(\widetilde{f}^\delta)^\top A \widetilde{f}^\delta = \min_{r\in\widetilde{\mathcal{Q}}} A_r \widetilde{f}^\delta < \min_{r\in\mathcal{Q}} A_r f^\delta &= (f^\delta)^\top A f^\delta\\
&\leq (f^\delta)^\top A \widetilde{f}^\delta,
\end{aligned}$$

where the first inequality is due to (4.4) and the second inequality follows from (4.5). The above implies

$$(\widetilde{f}^\delta)^\top A(f^\delta - \widetilde{f}^\delta) > 0. \tag{4.6}$$

On the other hand, since A is positive semi-definite

$$(\widetilde{f}^\delta - f^\delta)^\top A(f^\delta - \widetilde{f}^\delta) \leq 0. \tag{4.7}$$

Expanding the left-hand side of this inequality and combining it with (4.5), we obtain $(\widetilde{f}^\delta)^\top A(f^\delta - \widetilde{f}^\delta) \leq 0$, which contradicts (4.6). Therefore the premise is false, and the proof is complete.

For the other case, $\widetilde{\mathcal{M}}^- \subseteq \mathcal{M}^-$, the result can be proven using the same arguments, where the direction of the inequality reverses, since f^δ is in $\mathcal{H}_{-1}$ rather than $\mathcal{H}_1$, and thus $(f^\delta)^\top A f^\delta = -\min_{r\in\mathcal{Q}} A_r f^\delta$, and similarly for $\widetilde{f}^\delta$. □

The above result allows us to specify how the slope of λ^{WE} changes at the breakpoints $D_i \in \mathcal{D}$, in the special case where the used or active set at D_i is the same as in the interval (D_{i-1}, D_i). We have the following result:

Lemma 4.2.3. *(Slope of λ^{WE} for constant active and used set): For a given $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$, let $D_i \in \mathcal{D}$, where $i \geq 1$. We have the following:*

$$\begin{aligned} \mathcal{J}_{i-1}^{\mathrm{use}} = \mathcal{R}_{D_i}^{\mathrm{use}} &\Rightarrow \delta\lambda^{i-1} > \delta\lambda^i, \qquad & \mathcal{J}_{i}^{\mathrm{use}} = \mathcal{R}_{D_i}^{\mathrm{use}} &\Rightarrow \delta\lambda^{i-1} < \delta\lambda^i,\\ \mathcal{J}_{i-1}^{\mathrm{act}} = \mathcal{R}_{D_i}^{\mathrm{act}} &\Rightarrow \delta\lambda^{i-1} < \delta\lambda^i, \qquad & \mathcal{J}_{i}^{\mathrm{act}} = \mathcal{R}_{D_i}^{\mathrm{act}} &\Rightarrow \delta\lambda^{i-1} > \delta\lambda^i. \end{aligned}$$

Proof. We start with the claim for the case $\mathcal{J}_{i-1}^{\mathrm{use}} = \mathcal{R}_{D_i}^{\mathrm{use}}$. From Corollary 3.2.7 we have $\delta\lambda^{i-1} = \min_{r\in\mathcal{J}_{i-1}^{\mathrm{act}}} \delta C_r^{i-1}$. Now let $D \in (D_{i-1}, D_i)$. From Proposition 3.2.6 we then have $\delta C_r^{i-1} = A f^\delta$ for any $f^\delta \in \Gamma_D$. Furthermore, Theorem 3.2.10 gives $\Gamma_D = \mathrm{SOL}(\mathcal{M}_D, A)$ and it follows from the definition of $\mathcal{M}_D$ and Corollary 3.1.1 that

$$\mathcal{M}_D = \{f^\delta \in \mathcal{H}_1 \mid f^\delta_{\mathcal{J}_{i-1}^{\mathrm{act}}\setminus\mathcal{J}_{i-1}^{\mathrm{use}}} \geq 0,\;\; f^\delta_{(\mathcal{J}_{i-1}^{\mathrm{act}})^c} = 0\}.$$

In summary, we found that $\delta\lambda^{i-1} = \min_{r\in\mathcal{J}_{i-1}^{\mathrm{act}}} A_r f^\delta$, where $f^\delta \in \mathrm{SOL}(\mathcal{M}_D, A)$, and $\mathcal{M}_D$ is given by the above equality.

We can obtain a similar result for $\delta\lambda^i$. From Proposition 3.2.6 we find that $\lambda^{\mathrm{vec}}(T) = \lambda^{\mathrm{vec}}(D_i) + \delta C^i$ for any $T \in [D_i, D_{i+1}]$, and from Corollary 3.2.7 we have $\lambda^{\mathrm{WE}}(T) = \lambda^{\mathrm{WE}}(D_i) + (T - D_i)\delta\lambda^i$. Since the WE $\lambda^{\mathrm{WE}}(T)$ is the minimum of all $\lambda^{\mathrm{vec}}(T)$ it follows that it is equal to the costs of those paths that have minimum cost under WE at demand D_i, and that keep minimal cost as λ^{vec} moves in the direction of δC^i. In other words, $\delta\lambda^i = \min_{r\in\mathcal{R}_{D_i}^{\mathrm{act}}} \delta C^i$. Proposition 3.2.6 also gives $\delta C^i = A f^\delta$ for any $f^\delta \in \Gamma_{D_i}$, and Theorem 3.2.10 gives $\Gamma_{D_i} = \mathrm{SOL}(\mathcal{M}_{D_i}, A)$. In summary, we found that $\delta\lambda^i = \min_{r\in\mathcal{R}_{D_i}^{\mathrm{act}}} A_r f^\delta$, where $f^\delta \in \mathrm{SOL}(\mathcal{M}_{D_i}, A)$, and $\mathcal{M}_{D_i}$ is given by

$$\mathcal{M}_{D_i} = \{f^\delta \in \mathcal{H}_1 \mid f^\delta_{\mathcal{R}_{D_i}^{\mathrm{act}}\setminus\mathcal{R}_{D_i}^{\mathrm{use}}} \geq 0,\;\; f^\delta_{(\mathcal{R}_{D_i}^{\mathrm{act}})^c} = 0\}.$$

By assumption we have $\mathcal{J}_{i-1}^{\mathrm{use}} = \mathcal{R}_{D_i}^{\mathrm{use}}$, and from Lemma 3.2.3 we have $\mathcal{J}_{i-1}^{\mathrm{act}} \subseteq \mathcal{R}_{D_i}^{\mathrm{act}}$. Therefore we have $\mathcal{M}_D \subseteq \mathcal{M}_{D_i}$. The rest of the proof follows in the same manner as the proof of Lemma 4.2.2, with $\mathcal{M} = \mathcal{M}_{D_i}$ and $\widetilde{\mathcal{M}} = \mathcal{M}_D$. The only difference is that the inequalities (4.4) and (4.6) are non-strict while the inequality in (4.7) is strict, therefore preserving the contradiction. To see that (4.7) holds strictly, note that it holds with equality only if $Af^\delta = A\widetilde{f}^\delta$. However, in our case we have $Af^\delta = \delta C^i$ and $A\widetilde{f}^\delta = \delta C^{i-1}$. Thus equality of (4.7) would imply $\delta C^i = \delta C^{i-1}$, which contradicts Proposition 3.2.6.

This completes the proof for the case $\mathcal{J}_{i-1}^{\text{use}} = \mathcal{R}_{D_i}^{\text{use}}$. The claim for the case $\mathcal{J}_{i-1}^{\text{act}} = \mathcal{R}_{D_i}^{\text{act}}$ follows by the same arguments, where we find $\mathcal{M}_{D_i} \subseteq \mathcal{M}_D$ instead of $\mathcal{M}_D \subseteq \mathcal{M}_{D_i}$. The other claims can be proven similarly, where we consider $\mathcal{M}_D^-$ and $\mathcal{M}_{D_i}^-$ instead of $\mathcal{M}_D$ and $\mathcal{M}_{D_i}$. □

In light of Theorem 3.2.10, the above is a very intuitive result. For instance, when $\mathcal{J}_{i-1}^{\text{use}} = \mathcal{R}_{D_i}^{\text{use}}$, in order for $\Gamma_{D_i} \neq \Gamma_D$ to hold we must have $\mathcal{J}_{i-1}^{\text{act}} \neq \mathcal{R}_{D_i}^{\text{act}}$. However, as discussed before, Lemma 3.2.3 shows that the active set can only gain elements as D moves from (D_{i-1}, D_i) to D_i. Thus we find $\mathcal{M}_D \subset \mathcal{M}_{D_i}$, which shows that the set of directions to which Γ_D is restricted has grown strictly larger. In other words, there are new paths that are now feasible for carrying flow. Since the division of flow is a result of the traffic participants trying to minimize their own travel time, it seems natural that an increase in options will decrease the 'speed' at which the WE-cost grows as the demand increases. Similarly, $\mathcal{J}_{i-1}^{\text{act}} = \mathcal{R}_{D_i}^{\text{act}}$ implies a scenario in which there are fewer options for the evolution of the flow, and as such, the 'speed' at which the WE-cost grows as the demand increases becomes larger. We can also see this in Example 4.1.1a. We recall from (3.13) the evolution of the active and used set for this example:

$$(\mathcal{R}_D^{\text{act}}, \mathcal{R}_D^{\text{use}}) = \begin{cases} (\{p_3\}, \emptyset) & \text{for } D = 0, \\ (\{p_3\}, \{p_3\}) & \text{for } D \in (0,1), \\ (\{p_1, p_2, p_3\}, \{p_3\}) & \text{for } D = 1, \\ (\{p_1, p_2, p_3\}, \{p_1, p_2, p_3\}) & \text{for } D \in (1,2), \\ (\{p_1, p_2, p_3\}, \{p_1, p_2\}) & \text{for } D = 2, \\ (\{p_1, p_2\}, \{p_1, p_2\}) & \text{for } D \in (2, \infty). \end{cases}$$

Comparing this to Figure 4.1, we see that at $D = 1$, where p_1 and p_2 become active, the slope of the cost decreases. Conversely, at $D = 2$, where path p_3 leaves the used set, it is no longer possible to take flow away from that path, and as a consequence, the slope of the cost increases. Note that this intuition is exactly the one that is defied by Braess's paradox, where we see that more options increase travel time. Thus, even though this intuition does not hold for the WE-cost, it does hold for the slope of the WE-cost. It is also interesting to note that using this perspective, Lemma 4.2.2 implies that BP does not occur in a network in which $\beta_{e_k} = 0$ for all $e_k \in \mathcal{E}$.

Lemma 4.2.3 also sheds light on Example 3.2.8b, in which the slope of $\lambda^{\text{WE}}(\cdot)$ does not change at $D = 2$, despite this being a breakpoint. Lemma 4.2.3 tells us that this can only happen when both the active and used set change simultaneously at the considered breakpoint. That is, there needs to be both a path that loses all flow, and there must be a previously inactive path that becomes active. The slope of $\lambda^{\mathrm{WE}}(\cdot)$ then remains constant when the effects on the evolution of $\lambda^{\mathrm{WE}}(\cdot)$ of these changes in the used and active set exactly cancel each other out.

A useful observation for the upcoming analysis is the following consequence of Lemma 4.2.3.

Corollary 4.2.4. *(Slope of λ^{WE} when all paths are active): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ and $D \geq 0$ be given. In addition let $i \in [M]$ satisfy $\mathcal{J}_i^{\mathrm{act}} = \mathcal{P}$. Then $\delta\lambda^{i-1} > \delta\lambda^i$.*

Proof. From Lemma 3.2.3 we know that $\mathcal{J}_i^{\mathrm{act}} \subseteq \mathcal{R}_{D_i}^{\mathrm{act}}$. Since $\mathcal{J}_i^{\mathrm{act}} = \mathcal{P}$ and $\mathcal{R}_{D_i}^{\mathrm{act}} \subseteq \mathcal{P}$ this implies $\mathcal{R}_{D_i}^{\mathrm{act}} = \mathcal{P}$. Thus we have $\mathcal{J}_i^{\mathrm{act}} = \mathcal{R}_{D_i}^{\mathrm{act}}$, and thus result follows from Lemma 4.2.3. □

The final result of this section connects the previous analysis of the evolution of λ^{WE} to the concepts of modified games and necessary sets. It shows that whenever a set $\mathcal{S}^{\mathrm{rem}}$ is not necessary, the slope of the WE-cost of the associated modified game is at least as steep as the slope of the WE-cost of the original game.

Lemma 4.2.5. *(Necessary sets and the slope of λ^{WE}): Let $\mathcal{P}, \mathcal{S}^{rem} \subset \mathcal{P}, \mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ and $D > 0$ be given. If $S \notin \mathcal{N}_D$, the following hold:*

$$\delta\widetilde{\lambda}^{+}(D) \geq \delta\lambda^{+}(D),$$
$$\delta\widetilde{\lambda}^{-}(D) \geq \delta\lambda^{-}(D).$$

Proof. We start by proving the first inequality. From Corollary 4.2.1 we obtain

$$\delta\lambda^{+}(D) = \min_{r \in \mathcal{R}_D^{\mathrm{act}}} A_r f^{\delta}, \quad \delta\widetilde{\lambda}^{+}(D) = \min_{r \in \widetilde{\mathcal{R}}_D^{\mathrm{act}}} A_r \widetilde{f}^{\delta}.$$

where $f^{\delta} \in \mathrm{SOL}(\mathcal{M}_D, A)$ and $\widetilde{f}^{\delta} \in \mathrm{SOL}(\widetilde{\mathcal{M}}_D, A)$. Here

$$\widetilde{\mathcal{M}}_D := \{\widetilde{f}^{\delta} \in \widetilde{\mathcal{H}}_1 \mid \widetilde{f}^{\delta}_{\widetilde{\mathcal{R}}_D^{\mathrm{act}} \setminus \widetilde{\mathcal{R}}_D^{\mathrm{use}}} \geq 0, \;\; \widetilde{f}^{\delta}_{(\widetilde{\mathcal{R}}_D^{\mathrm{act}})^c} = 0\}.$$

Now, since $\mathcal{S}^{\mathrm{rem}} \notin \mathcal{N}_D$ it follows from Lemma 4.1.6 that $\widetilde{f}^D \in \widetilde{\mathcal{W}}_D$ if and only if $\widetilde{f}^D \in \mathcal{W}_D \cap \widetilde{\mathcal{F}}_D$. Consequently $\widetilde{\mathcal{R}}_D^{\mathrm{use}} \subseteq \mathcal{R}_D^{\mathrm{use}} \cap (\mathcal{S}^{\mathrm{rem}})^c$ and in addition we have $\widetilde{\lambda}^{\mathrm{vec}}(D) = \lambda^{\mathrm{vec}}(D)$, which implies $\widetilde{\mathcal{R}}_D^{\mathrm{act}} = \mathcal{R}_D^{\mathrm{act}} \cap (\mathcal{S}^{\mathrm{rem}})^c$. These facts collectively imply $\widetilde{\mathcal{M}}_D \subseteq \mathcal{M}_D$ and it follows from Lemma 4.2.2 that

$$\min_{r \in \widetilde{\mathcal{R}}_D^{\mathrm{act}}} A\widetilde{f}^{\delta} \geq \min_{r \in \mathcal{R}_D^{\mathrm{act}}} Af^{\delta}.$$

Thus we find

$$\delta\widetilde{\lambda}^+(D) \geq \delta\lambda^+(D).$$

This shows that the first inequality holds. The second inequality follows by similar arguments, but considering sets of feasible *descent* directions $\mathcal{M}_D^-$, rather than sets of feasible *ascent* directions $\mathcal{M}_D$. □

4.3 Conditions revealing Braess's paradox

Now we are ready to give our first result on Braess's paradox. It states that if a set $\mathcal{S}^{\text{rem}}$ is not necessary at D, then it is either not necessary for all lower levels of demand, or it causes BP at some lower levels of demand.

Theorem 4.3.1. *(Sets not in $\mathcal{N}_D$ are "non-essential" for lower demands): Let $\mathcal{P}$, $\mathcal{S}^{rem} \subset \mathcal{P}$, $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ and D be given. If $\mathcal{S}^{rem} \notin \mathcal{N}_D$ then exactly one of the following holds:*

- *$\mathcal{S}^{rem} \notin \mathcal{N}_T$ for all $T \in [0, D]$,*
- *There exists D^-, D^+ satisfying $0 < D^- < D^+ \leq D$ and*

$$\lambda^{\text{WE}}(T) > \widetilde{\lambda}^{\text{WE}}(T) \quad \textit{for all } T \in (D^-, D^+). \tag{4.8}$$

Proof. First we recall from Proposition 4.1.5 that $V(\cdot)$ and $\widetilde{V}(\cdot)$ are continuously differentiable, with $\frac{d}{dT}V(T) = \lambda^{\text{WE}}(T)$ and $\frac{d}{dT}\widetilde{V}(T) = \widetilde{\lambda}^{\text{WE}}(T)$. Furthermore, since $\mathcal{S}^{\text{rem}} \notin \mathcal{N}_D$ Lemma 4.1.6 tells us that $V(D) = \widetilde{V}(D)$ and $\lambda^{\text{WE}}(D) = \widetilde{\lambda}^{\text{WE}}(D)$. We define $D^+ \leq D$ as the smallest value such that $\lambda^{\text{WE}}(T) = \widetilde{\lambda}^{\text{WE}}(T)$ for all $T \in [D^+, D]$. In other words, the derivatives of $V(\cdot)$ and $\widetilde{V}(\cdot)$ are equal on the interval $[D^+, D]$. Since $V(D) = \widetilde{V}(D)$ this implies $V(T) = \widetilde{V}(T)$ for all $T \in [D^+, D]$. By Lemma 4.1.6 it follows that $\mathcal{S}^{\text{rem}} \notin \mathcal{N}_T$ for all $T \in [D^+, D]$. If $D^+ = 0$ this gives us $\mathcal{S}^{\text{rem}} \notin \mathcal{N}_T$ for all $T \in [0, D]$, which corresponds to the first scenario.

Alternatively, if $D^+ > 0$, it remains to be shown that there exists a $D^- \in (0, D^+)$ such that (4.8) holds. First, we note that using the fact $\mathcal{S}^{\text{rem}} \notin \mathcal{N}_{D^+}$ in Lemma 4.2.5 gives us $\delta\widetilde{\lambda}^-(D^+) \geq \delta\lambda^-(D^+)$. That is, we have either $\delta\widetilde{\lambda}^-(D^+) = \delta\lambda^-(D^+)$ or $\delta\widetilde{\lambda}^-(D^+) > \delta\lambda^-(D^+)$. The former of these is not possible. To see this, recall that by definition of D^+, we have $\lambda^{\text{WE}}(D^+) = \widetilde{\lambda}^{\text{WE}}(D^+)$. Since λ^{WE} and $\widetilde{\lambda}^{\text{WE}}$ are continuous, piecewise affine functions with only finitely many points in which the functions are not differentiable, it follows that if $\delta\widetilde{\lambda}^-(D^+) = \delta\lambda^-(D^+)$, then there exists some $\epsilon > 0$ such that $\lambda^{\text{WE}}(T) = \widetilde{\lambda}^{\text{WE}}(T)$ for all $T \in (D^+ - \epsilon, D^+]$. This however contradicts the definition of D^+ as the smallest value such that $\lambda^{\text{WE}}(T) = \widetilde{\lambda}^{\text{WE}}(T)$ for all $T \in [D^+, D]$. Therefore, $\delta\widetilde{\lambda}^-(D^+) = \delta\lambda^-(D^+)$ is not possible and we have $\delta\widetilde{\lambda}^-(D^+) > \delta\lambda^-(D^+)$. This then implies that there exists some $D^- > 0$ such that $D^- < D^+$ and (4.8) holds, completing the proof. □

Of course, as a method for detecting BP Theorem 4.3.1 leaves much to be desired, not only because it does not specify at which level of demand the considered set of paths $\mathcal{S}^{\mathtt{rem}}$ causes a BP, but also because it leaves the option open that the set does not cause BP at all, but is instead not a necessary set for all lower levels of demand. However, it is still a potentially useful result. To see this, recall that in Chapter 3 we gave a straightforward method to obtain the "final" active set $\mathcal{J}_M^{\mathrm{act}}$, as well as f^{D_M} and an $f^{\delta} \in \Gamma_{D_M}$. This gives us an easy method for finding sets of paths that are not used, and therefore not necessary, in the interval (D_M, ∞). In light of Theorem 4.3.1 the usefulness of any set of paths with this property then becomes rather suspect, because it shows that this set is either responsible for BP at some levels of demand in the interval $(0, D_M)$, or it can be completely removed from the network without affecting the WE-cost of the game at any level of demand. Since we can easily obtain $(\mathcal{J}_M^{\mathrm{act}})^c$, this supplies a social planner with a quick method for identifying sets of paths in the network whose benefits to the network warrant further investigation, and which are potential candidates for removal.

Though Theorem 4.3.1 has its uses, it would still be preferable if we could also find a condition that more explicitly guarantees the presence of BP, and does so for a specific level of demand. Our next result gives the essential observation that will allow us to find more efficient ways of detecting BP. To ease its exposition we define the following function:

Definition 4.3.2. *(Affine extension functions): Let $\mathcal{P}$, $\mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ and $\vec{\mathcal{P}} \subset \mathcal{P}$ be given, and consider the game $(\vec{\mathcal{P}}, \mathcal{C})$. For $i \in [\vec{\mathcal{M}}]_0$, we define the following function:*

$$u_{\vec{\mathcal{P}},i}(T) := \vec{\lambda}^{\mathrm{WE}}(\vec{D}_i) + (D - \vec{D}_i)\delta\vec{\lambda}^i,$$

Note that $u_{\vec{\mathcal{P}},i}$ is simply the affine function that describes $\vec{\lambda}^{\mathrm{WE}}$ on the interval $[\vec{D}_i, \vec{D}_{i+1})$, but extended to the whole of $\mathbb{R}_{\geq 0}$. We call $u_{\vec{\mathcal{P}},i}$ an affine extension. •

where $\widetilde{D}_i \in \widetilde{\mathcal{D}}$. Note that the function $u_{\widetilde{\mathcal{P}},i}(\cdot)$ is simply the affine function that describes $\widetilde{\lambda}^{\mathrm{WE}}(D)$ on the interval $[\widetilde{D}_i, \widetilde{D}_{i+1}]$, but extended to the whole of $\mathbb{R}_{\geq 0}$. Our upcoming result shows that for any affine extension $u_{\widetilde{\mathcal{P}},i}$ of any modified game $(\widetilde{\mathcal{P}}, \mathcal{C})$, there exists another modified game $(\check{\mathcal{P}}, \mathcal{C})$ that achieves a WE-cost lower than $u_{\widetilde{\mathcal{P}},i}(D)$ for all $D \leq \widetilde{D}_{i+1}$.

Lemma 4.3.3. *(An upper bound on minimum WE-cost of all modified games): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$ be given. For any $\widetilde{\mathcal{P}} \subseteq \mathcal{P}$, $i \in [\widetilde{M}]_0$ and $D \leq \widetilde{D}_{i+1}$, there exists a set $\mathcal{S}_{i,D}^{rem} \subset \mathcal{P}$ such that for the modified game over the set $\check{\mathcal{P}}$, where $\check{\mathcal{P}} = \mathcal{P} \setminus \mathcal{S}_{i,D}^{rem}$ the associated WE-cost, given by $\check{\lambda}_{i,D}^{\mathrm{WE}}(\cdot)$, satisfies*

$$\check{\lambda}_{i,D}^{\mathrm{WE}}(D) \leq u_{\widetilde{\mathcal{P}},i}(D).$$

Proof. For notational convenience, we prove the result for the case $\widetilde{\mathcal{P}} = \mathcal{P}$. The case $\widetilde{\mathcal{P}} \subset \mathcal{P}$ can be proven using the same arguments. Let $\mathcal{J}_i^{\text{use}}$ be the used-set for the routing game defined by $\mathcal{P}$ and $\mathcal{C}$ on the interval (D_i, D_{i+1}), and let $\mathcal{S}' = (\mathcal{J}_i^{\text{use}})^c$. We consider the modified game associated with $\mathcal{P}'$ where $\mathcal{P}' := \mathcal{P} \setminus \mathcal{S}'$. Clearly, $(\mathcal{J}_i^{\text{use}})^c$ is unnecessary with respect to the original game on the interval (D_i, D_{i+1}); i.e., $(\mathcal{J}_i^{\text{use}})^c \notin \mathcal{N}_D$ for all $D \in (D_i, D_{i+1})$. By Lemma 4.1.6, the WE-cost of the modified game, denoted by $\lambda'^{\text{WE}}(\cdot)$, then satisfies

$$\lambda'^{\text{WE}}(D) = \lambda^{\text{WE}}(D), \quad \text{for all } D \in [D_i, D_{i+1}]. \tag{4.9}$$

This already shows that $\lambda'^{\text{WE}}(D) = u_{\mathcal{P},i}(D)$ holds on the interval $[D_i, D_{i+1})$. Therefore, if $D_i = 0$, the proof would be complete.

Now, assume $D_i > 0$ and note that $\mathcal{P}' = \mathcal{R}'^{\text{use}}_D = \mathcal{R}'^{\text{use}}_T$ for all $D, T \in (D_i, D_{i+1})$. Corollary 3.1.1 then tells us that there exist $D'_j, D'_{j+1} \in \mathcal{D}'$ for which we have $(D_i, D_{i+1}) \subseteq (D'_j, D'_{j+1})$ and $\mathcal{J}'^{\text{use}}_j = \mathcal{P}'$. Therefore, using this inclusion and (4.9), we have $\delta\lambda'^j = \delta\lambda^i$. This fact in combination with $\lambda'^{\text{WE}}(D_i) = \lambda^{\text{WE}}(D_i)$ shows that $\lambda'^{\text{WE}}(D) = u_{\mathcal{P},i}(D)$ for all $D \in [D'_j, D'_{j+1})$. As before, if $D'_j = 0$, the proof would be complete.

When $D'_j > 0$, note that $\mathcal{J}'^{\text{act}}_j = \mathcal{J}'^{\text{use}}_j = \mathcal{P}'$. From Lemma 3.2.3 we then have $\mathcal{R}'^{\text{act}}_{D'_j} = \mathcal{J}'^{\text{act}}_j$, and it follows from Lemma 4.2.3 that $\delta\lambda'^{j-1} > \delta\lambda'^j$. This in combination with $\delta\lambda'^j = \delta\lambda^i$ and $\lambda'^{\text{WE}}(D'_j) = u_{\mathcal{P},i}(D'_j)$ shows that $\lambda'^{\text{WE}}(D) < u_{\mathcal{P},i}(D)$ for all $D \in [D'_{j-1}, D'_j)$. As before, if $D'_{j-1} = 0$, the proof would be complete. Note that the conclusions up until this point also give $u_{\mathcal{P}',j-1}(D) < u_{\mathcal{P},i}(D)$ for all $D < D'_j$.

If $D'_{j-1} > 0$, we can define $\mathcal{S}'' = (\mathcal{J}'^{\text{use}}_{j-1})^c$ and consider the modified game associated to $\mathcal{P}''$, where $\mathcal{P}'' := \mathcal{P} \setminus \mathcal{S}''$. First we show that $\mathcal{P}''$ is non-empty. To see this, note that $\mathcal{P}'' = \emptyset$ implies $\mathcal{J}'^{\text{use}}_{j-1} = \emptyset$, which in turn implies that all for all demands $D \in (D'_{j-1}, D'_j)$ and all WE $f'^D \in \mathcal{W}_D$, $f'^D_p = 0$. This however means that the $f^D \in \mathcal{F}_0$, which contradicts the assumption that $D'_{j-1} > 0$. We conclude that $\mathcal{P}''$ is nonempty.

We also have $\mathcal{J}'^{\text{use}}_{j-1} \subseteq \mathcal{P}'$, and Corollary 3.1.1 gives $\mathcal{J}'^{\text{use}}_{j-1} \neq \mathcal{J}'^{\text{use}}_j$. Since $\mathcal{J}'^{\text{use}}_j = \mathcal{P}'$ we see that $\mathcal{J}'^{\text{use}}_{j-1} \subset \mathcal{P}'$. Therefore $\mathcal{P}'' \subset \mathcal{P}'$.

We can now apply the arguments made for comparing λ'^{WE} with $u_{\mathcal{P},i}$ to compare λ''^{WE} with $u_{\mathcal{P}',j-1}$, which gives

$$\lambda''^{\text{WE}}(D) = u_{\mathcal{P}',j-1} \quad \text{for all } D \in [D''_k, D''_{k+1}), \tag{4.10}$$

where k is such that $[D'_{j-1}, D'_j) \subseteq [D''_k, D''_{k+1})$ for some $D''_k, D''_{k+1} \in \mathcal{D}''$. If $D''_k = 0$, the proof is complete, and if $D''_k > 0$, the same arguments as before give

$$\begin{aligned} \lambda''^{\text{WE}}(D) &< u_{\mathcal{P}',j-1} \quad \text{for all } D \in [D''_{k-1}, D''_k), \\ u_{\mathcal{P}'',k-1} &< u_{\mathcal{P}',j-1} \quad \text{for all } D \leq D''_k, \end{aligned}$$

where $D''_{k-1} \in \mathcal{D}''$ satisfies $D''_{k-1} < D''_k \leq D'_{j-1}$. This then shows that the result holds on the interval $[D''_{k-1}, D_{i+1})$. If $D''_{k-1} = 0$, the proof is complete. If $D''_{k-1} > 0$ we can again repeat our arguments, to extend the interval on which the statement is shown to hold. Each time we repeat the arguments the set of paths under consideration is a strict subset of the previously considered set (e.g. $\mathcal{P}'' \subset \mathcal{P}' \subset \mathcal{P}$), since there are only finitely many paths, we can only repeat the arguments finitely many times, but we can always repeat the argument as long as the used set at the lowest value in the interval where the statement is shown to hold is non-empty. We conclude that after a finite number of repetitions the used set at the lowest value of the interval where the statement is shown to hold is empty, which implies that demand at this point is zero. Therefore, the proof is complete. □

Lemma 4.3.3 is an essential part of the upcoming results for detecting BP. For this reason we highlight the implications of this result by revisiting Example 4.1.1. At the start of this chapter we showed that the routing game discussed in Example 4.1.1a over the network in Figure 4.1 experiences BP when $D \in (\frac{2}{3}, 2)$. In Figure 4.2 we see the comparison between the WE-cost of the original game and that of the modified game with edge e_5 removed. However, note that $\widetilde{\lambda}^{\mathrm{WE}}(D)$ is simply the extension to $\mathbb{R}_{\geq 0}$ of the line piece describing $\lambda^{\mathrm{WE}}(D)$ on the interval $(D_2, D_3) = (2, \infty)$. That is $\widetilde{\lambda}^{\mathrm{WE}}(D) = u_{\mathcal{P},2}(D)$. Thus we see that $\lambda^{\mathrm{WE}}(D) > u_{\mathcal{P},2}(D)$ for $D \in (\frac{2}{3}, 2)$, and therefore Lemma 4.3.3 implies that the network is subject to BP on this interval. Using this approach we could have known that BP was present in this interval even without analysing a modified game. Note that we can not always count on BP to be revealed by comparing λ^{WE} with functions of the form $u_{\mathcal{P},i}$. For instance, the BP present in Example 4.1.1b, shown in Figure 4.4, is not revealed in this way, but instead by comparing λ^{WE} with $u_{\vec{\mathcal{P}},0}$, where $\vec{\mathcal{P}} = \mathcal{P} \setminus \{(e_1, e_4)\}$. The full potential of Lemma 4.3.3 for use in detecting BP is revealed later in Theorem 4.3.7. However, before we can establish that result, we need to give some intermediate statements, which themselves give additional useful methods for detecting BP. The first of these statements is that as a consequence of Lemma 4.3.3, any increase in the slope of λ^{WE} reveals the presence of BP.

Corollary 4.3.4. *(BP revealed by increase of slope of λ^{WE}): For a given $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\mathrm{aff}}$, let $D_i \in \mathcal{D}$ with $D_i > 0$. If $\delta\lambda^{i-1} < \delta\lambda^i$ then there exists, for all $T \in [D_{i-1}, D_i)$, a set $\mathcal{S}_T^{rem} \subset \mathcal{P}$ such that $\widetilde{\lambda}^{\mathrm{WE}}(T) < \lambda^{\mathrm{WE}}(T)$, where $\widetilde{\lambda}^{\mathrm{WE}}(T)$ is the WE-cost at demand T for the modified game formed by removing paths in $\mathcal{S}_T^{rem}$.*

Proof. The result follows from Lemma 4.3.3 after noting that $\lambda^{\mathrm{WE}}(D_i) = u_{\mathcal{P},i}(D_i)$ and that for $D \in (D_{i-1}, D_i)$ we have

$$\lambda^{\mathrm{WE}}(D) = \lambda^{\mathrm{WE}}(D_i) + (D - D_i)\delta\lambda^{i-1},$$

$$u_{\mathcal{P},i}(D) = u_{\mathcal{P},i}(D_i) + (D - D_i)\delta\lambda^i.$$

□

Lemma 4.3.3 and the above corollary provide potentially useful ways of detecting BP. However, they do require investigating the WE-cost of a routing game for multiple levels of demand, which increases the computational intensity. The following result instead gives us a sufficient condition for the presence of BP at one level of demand D, which requires no investigation of modified games, or of the same game at multiple levels of demand.

Theorem 4.3.5. *(Paths losing flow reveals BP): Let $\mathcal{P}$, $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ and D be given. If $\Gamma_D \cap \mathcal{F}_1 = \emptyset$, then there exists a set $\mathcal{S}^{rem}$ such that $\widetilde{\lambda}^{\text{WE}}(D) < \lambda^{\text{WE}}(D)$.*

Proof. Let D be such that $\Gamma_D \cap \mathcal{F}_1 = \emptyset$. In addition let $D_i, D_{i+1} \in \mathcal{D}$ satisfy $D \in [D_i, D_{i+1})$ and let $\mathcal{S}^{\texttt{rem}} := (\mathcal{J}_i^{\text{use}})^c$. We consider the modified game over the set of paths $\widetilde{\mathcal{P}} := \mathcal{J}_i^{\text{use}}$. Note that $\mathcal{S}^{\texttt{rem}} \notin \mathcal{N}_T$ for all $T \in [D_i, D_{i+1})$ and by Lemma 4.1.6 we therefore have $\lambda^{\text{WE}}(T) = \widetilde{\lambda}^{\text{WE}}(T)$ for all $T \in [D_i, D_{i+1}]$. Consequently, for proving the result, it suffices to show that there exists a set $\widehat{\mathcal{P}} \subset \widetilde{\mathcal{P}}$ such that $\widehat{\lambda}^{\text{WE}}(D) < \widetilde{\lambda}^{\text{WE}}(D)$.

Our first aim is to show that for the game defined over the set of paths $\widetilde{\mathcal{P}} := \mathcal{J}_i^{\text{use}}$ we have $D < \widetilde{D}_{\widetilde{M}}$; that is, D does not lie in the 'final' interval of the modified game (see Theorem 3.3.1). This we do by proving that $\widetilde{\Gamma}_D \cap \mathcal{F}_1 = \emptyset$. Indeed this is enough as from Theorem 3.3.1, if $D \geq \widetilde{D}_{\widetilde{M}}$, then $\widetilde{\Gamma}_D \cap \text{SOL}(\mathcal{F}_1, A)$ is non-empty.

Note that we have $\mathcal{S}^{\texttt{rem}} \notin \mathcal{N}_D$ from above. Consequently, by Lemma 4.1.6, we have $\widetilde{f}^D \in \widetilde{\mathcal{W}}_D$ if and only if $\widetilde{f}^D \in \mathcal{W}_D$ and $\widetilde{f}^D_{\mathcal{S}^{\texttt{rem}}} = 0$. Now pick $\widetilde{f}^D \in \widetilde{\mathcal{W}}_D$ and $\widetilde{f}^\delta \in \widetilde{\Gamma}_D$, such that $\widetilde{f}^D + \epsilon \widetilde{f}^\delta \in \widetilde{\mathcal{W}}_{D+\epsilon}$ as long as ϵ is small enough. For any $\epsilon > 0$ that then also satisfies $D + \epsilon \in (D, D_{i+1})$ we thus obtain a WE of the modified game over the set $\mathcal{J}_i^{\text{use}}$ at demand $D + \epsilon$. Note that we have $(\mathcal{J}_i^{\text{use}})^c \notin \mathcal{N}_D$ and $(\mathcal{J}_i^{\text{use}})^c \notin \mathcal{N}_{D+\epsilon}$. It follows from Lemma 4.1.6 that $\widetilde{f}^D \in \mathcal{W}_D$, and as long as ϵ is small enough we also have $\widetilde{f}^D + \epsilon \widetilde{f}^\delta \in \mathcal{W}_{D+\epsilon}$, which shows that $\widetilde{f}^\delta \in \Gamma_D$. Thus we see that $\widetilde{\Gamma}_D \subseteq \Gamma_D$, which implies $\widetilde{\Gamma}_D \cap \mathcal{F}_1 = \emptyset$ and so, we have $D < \widetilde{D}_{\widetilde{M}}$.

Now let $\widetilde{D}_j, \widetilde{D}_{j+1} \in \widetilde{D}$ satisfy $D \in [\widetilde{D}_j, \widetilde{D}_{j+1})$. Since $\widetilde{\mathcal{R}}_T^{\text{act}} = \widetilde{\mathcal{J}}_j^{\text{act}} = \widetilde{\mathcal{P}}$ for all $T \in (\widetilde{D}_j, \widetilde{D}_{j+1})$ it follows from Lemma 3.2.3 that $\widetilde{\mathcal{J}}_j^{\text{act}} = \widetilde{\mathcal{R}}^{\text{act}}_{\widetilde{D}_{j+1}}$. Lemma 4.2.3 therefore gives us $\delta\widetilde{\lambda}^j < \delta\widetilde{\lambda}^{j+1}$. The statement then follows from Corollary 4.3.4. □

The previous three results supply us with useful, practically implementable methods for detecting the presence of BP, but the statements themselves provide little direction on how one can find a set of paths $\mathcal{S}^{\texttt{rem}}$ whose removal alleviates the detected paradox. However, based on the proofs of these results there is some guidance we can offer on how to find $\mathcal{S}^{\texttt{rem}}$, as we illustrate in our next example.

Example 4.3.6. First note that the proof of Theorem 4.3.5 depends on Corollary 4.3.4, which in turn depends on Lemma 4.3.3. Finally Lemma 4.3.3 is established by looking at the evolution of the costs of modified games where attention is constrained only to the used set, and this is part that can help in the search for a set $\mathcal{S}^{\texttt{rem}}$ that causes BP. For instance, we see that in the proof of Lemma 4.3.3, whenever the considered used set loses one or more paths as we decrease the level of demand, the game is again modified by dropping the now unused paths from consideration. We see that the arrived-at subsets of paths that achieve lower WE-costs are found by limiting our attention to the used set, and then decreasing the flow continuously, dropping a path from consideration whenever it leaves the used set. Similarly, in the proof of Theorem 4.3.5 we also limit our attention to only the used set, and we consider the demand $\widetilde{D}_{i+1}$, which is the smallest level of demand higher than D such that the used set of this modified game loses one or more paths. If we write $\mathcal{S}' := \widetilde{\mathcal{P}} \setminus \widetilde{\mathcal{R}}^{\text{use}}_{D_{i+1}}$ for the set of paths that is no longer used at $\widetilde{D}_{i+1}$, following the discussed procedure of Lemma 4.3.3, the first step to reveal BP on the interval $(\widetilde{D}_i, \widetilde{D}_{i+1})$ is then to drop this set $\mathcal{S}'$. In other words, what we know is that there exists a set $\mathcal{S}^{\texttt{rem}}$ that contains $\mathcal{S}'$ and which causes the detected BP. We note that we only have $\mathcal{S}' \subset \mathcal{S}^{\texttt{rem}}$ and not necessarily $\mathcal{S}' = \mathcal{S}^{\texttt{rem}}$. To see this, consider the network given in Figure 4.6. This is

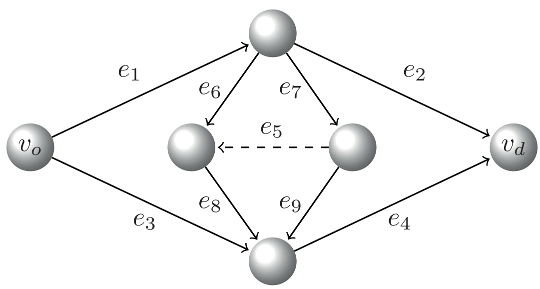


Figure 4.6: The nested Wheatstone network, in which the middle edge of the Wheatstone network has been replaced with an instance of the Wheatstone network.

the Wheatstone network, modified by replacing the middle edge by another instance of the Wheatstone network. For this reason this is called a *nested Wheatstone network*. This network has five paths, namely $p_1 = (e_1, e_2)$, $p_2 = (e_3, e_4)$, $p_3 = (e_1, e_7, e_9, e_4)$, $p_4 = (e_1, e_6, e_8, e_4)$, and $p_5 = (e_1, e_7, e_5, e_8, e_4)$.

Let the cost functions of the edges be given by

$$
\begin{aligned}
C_{e_1}(f_{e_1}) &= f_{e_1}, & C_{e_2}(f_{e_2}) &= 7, \\
C_{e_3}(f_{e_3}) &= 7, & C_{e_4}(f_{e_4}) &= f_{e_4}, \\
C_{e_6}(f_{e_6}) &= 1, & C_{e_7}(f_{e_7}) &= f_{e_7}, \\
C_{e_8}(f_{e_8}) &= f_{e_8}, & C_{e_7}(f_{e_7}) &= 1, \\
& & C_{e_5}(f_{e_5}) = 0. &
\end{aligned}
$$

The resulting WE, which is unique in this case, is shown in Figure 4.7.

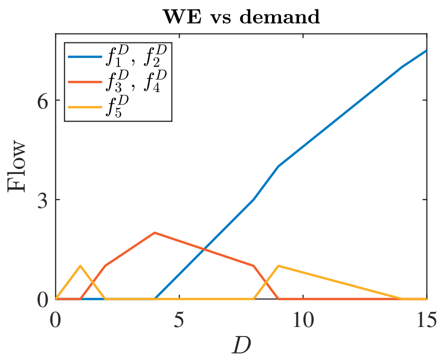


Figure 4.7: The evolution of f^D for the routing game over the network in Figure 4.6, as discussed in Example 4.3.6.

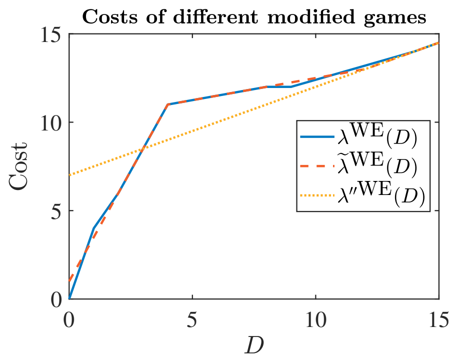


Figure 4.8: The evolution of the WE-cost, for the three games discussed in Example 4.3.6.

Full expressions for the WE and WE-costs of the games considered in this example are given in Appendix C. In Figure 4.8 we see a comparison between the WE-cost of the original game, the game over the paths $\widetilde{\mathcal{P}} = \mathcal{P} \setminus \{p_5\}$ and the game over the paths $\mathcal{P}'' = \mathcal{P} \setminus \{p_3, p_4, p_5\}$. Note that on the interval $D \in (4, 8)$, the flow over paths p_3 and p_4 decreases, which by Theorem 4.3.5 indicates that a BP is present. At $D_4 = 8$ the used set is given by $\mathcal{R}^{\mathrm{use}}_{D_4} = \{p_1, p_2, p_3, p_4\}$ and thus we have $\mathcal{P} \setminus \mathcal{R}^{\mathrm{use}}_{D_4} = \{p_5\}$. However, as we see in Figure 4.8, it is not the set $\{p_5\}$ which causes the BP in the interval $(4, 8)$, but rather the sets $\{p_3, p_4\}$ or $\{p_3, p_4, p_5\}$. This shows that indeed there exists a set $\mathcal{S}^{\mathtt{rem}}$ causing BP that satisfies $\mathcal{P} \setminus \mathcal{R}^{\mathrm{use}}_{D_4} \subset \mathcal{S}^{\mathtt{rem}}$, but it is not equal to $\mathcal{P} \setminus \mathcal{R}^{\mathrm{use}}_{D_4}$.

We also note that when a BP is detected using Theorem 4.3.5, it is not necessarily the subset of paths that is losing flow that causes BP, but there does exist a set containing the subset of paths that loses flow, that causes a paradox. To see this, note that in our example the path p_5 loses flow on the interval $D \in (9, 14)$. Indeed, on the interval $D \in (\frac{32}{3}, 14)$ we find confirmation that the removal of path p_5 decreases the WE-cost. However, on the interval $D \in (9, \frac{32}{3})$, removal of p_5 increases WE cost, despite p_5 being the only path that loses flow in this interval. We see that in this case the BP is associated with the removal of the set $\{p_3, p_4, p_5\}$. •

Theorem 4.3.5 gives us a feasible way of detecting Braess's paradox in a network at one specific level of demand. The downside is that the given condition is only sufficient; that is, Braess's paradox can still be present in the network even when the given condition is not satisfied. For instance, considering the routing game defined by (3.7) in Example 4.1.1a and looking at the evolution of the associated WE in Figure 4.2, we see that $\Gamma_D \cap \mathcal{F}_1 = \emptyset$ for $D \in [1, 2)$, revealing BP in this range of demands, while the condition is not satisfied for $D \in (\frac{2}{3}, 1)$ despite the presence of BP there. What is more, in Example 4.1.1b we have that $\Gamma_D \cap \mathcal{F}_1 = \emptyset$ does not hold for any level of demand, completely missing the BP that is present in that example. Our next result shows how the obtained results can be used to give a necessary and sufficient condition for the existence of Braess's paradox, though the practical usefulness of this result is limited, since applying it requires searching the space of all subsets of the set of paths. We recall that M is used to denote the index of the last finite valued breaking point $D_M \in \mathcal{D}$ of a routing game.

Theorem 4.3.7. *(Final cost evolution of modified games reveals all BPs): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. The routing game is subject to a Braess's paradox at demand D if and only if there exists a set $\mathcal{S}^{rem} \subset \mathcal{P}$ such that $\widetilde{\Gamma}_D \cap \mathcal{F}_1 \neq \emptyset$ and*

$$u_{\widetilde{\mathcal{P}}, \widetilde{M}}(D) < \lambda^{\text{WE}}(D). \tag{4.11}$$

Proof. First assume that $(\mathcal{P}, \mathcal{C})$ is subject to Braess's paradox at demand D. That is, there exists a set $\mathcal{S}'$ such that $\lambda'^{\text{WE}}(D) < \lambda^{\text{WE}}(D)$, where λ'^{WE} stands for the WE-cost for the game $(\mathcal{P} \setminus \mathcal{S}', \mathcal{C})$. To reiterate, we need to show that there exists $\mathcal{S}^{\text{rem}}$ such that $\widetilde{\Gamma}_D \cap \mathcal{F}_1 \neq \emptyset$ and (4.11) holds. Consider therefore the case when $\Gamma'_D \cap \mathcal{F}_1 = \emptyset$. It follows from Theorem 4.3.5 that the routing game $(\mathcal{P} \setminus \mathcal{S}', \mathcal{C})$ is also subject to Braess's paradox at demand D. That is, there exists a set $\mathcal{S}'' \subset \mathcal{P}'$ such that $\lambda''^{\text{WE}}(D) < \lambda'^{\text{WE}}(D)$. With this we deduce that since $\mathcal{P}$ is subject to BP caused by the set $\mathcal{S}'$ and $\mathcal{P}'$ is subject to BP by removing the paths $\mathcal{S}''$, then $\mathcal{P}$ is subject to BP caused by the set $\mathcal{S}' \cup \mathcal{S}''$. We can repeat this argument until we find a set $\widehat{\mathcal{S}}$ such that the modified game $(\mathcal{P} \setminus \widehat{\mathcal{S}})$ satisfies $\widehat{\Gamma}_D \cap \mathcal{F}_1 \neq \emptyset$. Let $\widehat{\mathcal{S}}' = (\widehat{\mathcal{R}}_D^{\text{use}})^c$. We then consider the game $(\widehat{\mathcal{P}} \setminus \widehat{\mathcal{S}}', \mathcal{C})$. Once again, if $\widehat{\Gamma}'_D \cap \mathcal{F}_1 = \emptyset$, this means that the game $(\widehat{\mathcal{P}} \setminus \widehat{\mathcal{S}}', \mathcal{C})$ is subject to BP at demand D, and therefore there exists a set $\widehat{\mathcal{S}}'' \subset \widehat{\mathcal{P}}'$ such that for the game over $(\widehat{\mathcal{P}}' \setminus \widehat{\mathcal{S}}'', \mathcal{C})$ we find $\widehat{\lambda}''^{\text{WE}}(D) < \widehat{\lambda}'^{\text{WE}}(D)$. We can repeat the above arguments until we find a set $\mathcal{S}^{\text{rem}} \subset \mathcal{P}$ that satisfies $\widetilde{\mathcal{P}} := \mathcal{P} \setminus \mathcal{S}^{\text{rem}}$, $\widetilde{\mathcal{R}}_D^{\text{use}} = \widetilde{\mathcal{P}}$, $\widetilde{\Gamma}_D \cap \mathcal{F}_1 \neq \emptyset$, and $\widetilde{\lambda}^{\text{WE}}(D) < \lambda^{\text{WE}}(D)$.

Next, let $\widetilde{f}^\delta \in \widetilde{\Gamma}_D \cap \mathcal{F}_1$. Since $\widetilde{\mathcal{R}}_D^{\text{use}} = \widetilde{\mathcal{P}}$ we have $\widetilde{\lambda}_p^{\text{WE}}(D) = \widetilde{\lambda}_r^{\text{WE}}(D)$ for all $p, r \in \widetilde{\mathcal{P}}$. In addition we have

$$\widetilde{\mathcal{M}}_D = \{f^\delta \in \mathcal{H}_1 \mid f^\delta_{\widetilde{\mathcal{P}}^c} = 0\},$$

and from Theorem 3.2.10 we know that $\widetilde{f}^\delta \in \text{SOL}(\widetilde{\mathcal{M}}_D, A)$. It follows from Propositions 3.2.12 that $A_p\widetilde{f}^\delta = A_r\widetilde{f}^\delta$ holds for all $p, r \in \widetilde{\mathcal{P}}$. In other words, as we move in the direction of $\widetilde{f}^\delta$, the costs of all paths remain equal. Since $\widetilde{f}^\delta \in \mathcal{F}_1$ it also follows that for any $\widetilde{f}^D \in \widetilde{\mathcal{W}}_D$ and any $\epsilon > 0$ we have $\widetilde{f}^D + \epsilon\widetilde{f}^\delta \geq 0$. Consequently, $\widetilde{f}^D + \epsilon\widetilde{f}^\delta$ is a WE for any $\epsilon > 0$. From this it follows that $\widetilde{f}^\delta \in \widetilde{\Gamma}_T$ for all $T \geq D$. Since we know from Lemma 3.2.14 that $\widetilde{\Gamma}^i \neq \widetilde{\Gamma}^{i+1}$ for all $i \in [\widetilde{M}]$, it follows that $D \in [\widetilde{D}_{\widetilde{M}}, \infty)$. As a consequence, using the definition of the affine extension function $u_{\widetilde{\mathcal{P}},\widetilde{M}}$ we arrive at $\widetilde{\lambda}(D) = u_{\widetilde{\mathcal{P}},\widetilde{M}}(D)$. And since $\widetilde{\lambda}(D) < \lambda^{\text{WE}}(D)$, we conclude that $u_{\widetilde{\mathcal{P}},\widetilde{M}}(D) < \lambda^{\text{WE}}(D)$. This shows one direction of the implication.

For the other direction, assume that there exists some $\mathcal{S}^{\texttt{rem}} \subset \mathcal{P}$ such that $u_{\widetilde{\mathcal{P}},\widetilde{M}}(D) < \lambda^{\text{WE}}(D)$. If $D < \widetilde{D}_{\widetilde{M}}$, then it follows from Lemma 4.3.3 that the game $(\mathcal{P}, \mathcal{C})$ is subject to BP at demand D. If $D \geq \widetilde{D}_{\widetilde{M}}$, then we have $\widetilde{\lambda}^{\text{WE}}(D) = u_{\widetilde{\mathcal{P}},\widetilde{M}}(D)$ and therefore $\widetilde{\lambda}^{\text{WE}}(D) < \lambda^{\text{WE}}(D)$, which shows that the game is subject to BP at demand D. This completes the proof. □

Though it is computationally expensive to use the above proposition for detection of BP, it does have one advantage worth mentioning. Namely, that one does not have to check for BP separately for different levels of demand. Instead there is one piecewise affine function that serves as a bound on the WE-cost. Simply checking whether the value of WE-cost exceeds the value of this function reveals the presence or absence of BP. However, constructing this function requires finding a set of subsets of $\mathcal{P}$ that define it, and it can be computationally infeasible to check all subsets of $\mathcal{P}$ to know which ones to use. Despite this limitation, it is still a useful result. For instance, we observed earlier that for Example 4.1.1a, comparing $\lambda^{\text{WE}}(D)$ to $u_{\mathcal{P},2}(D)$ reveals BP on the interval $D \in (\frac{2}{3}, 1)$ and we can now generalize this observation with the following Corollary:

Corollary 4.3.8. *(Easily attainable upper bound on achievable WE-cost): Let $\mathcal{P}, \mathcal{C} \subset \mathcal{K}$ and D be given. If $u_{\mathcal{P},M}(D) < \lambda(D)$, then the network suffers from Braess's paradox at demand D.*

Once the evolution of λ^{WE} has been mapped out, the above provides an easily obtained 'upper bound' on the cost $\lambda^{\text{WE}}(D)$, such that whenever $\lambda^{\text{WE}}(D)$ exceeds this bound, the network necessarily suffers from Braess's paradox.

Before finishing our exposition on BP in this section we use the obtained results to show that for any network BP can only occur on a finite interval of demand. Though it seems rather surprising, to the best of our knowledge this result has not been proven before in full generality, though versions limited to the Wheatstone network (Figure 4.1) have been obtained [23,31].

Theorem 4.3.9. *(Braess's paradox occurs on finite interval): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ be given. There exists a value $D^{BP} \geq 0$ such that $\lambda^{\text{WE}}(D) \leq \widetilde{\lambda}^{\text{WE}}(D)$ for all $D \geq D^{\text{BP}}$ and all $\mathcal{S}^{rem} \subset \mathcal{P}$.*

Proof. Let $D_M \in \mathcal{D}$ and $\widetilde{D}_{\widetilde{M}} \in \widetilde{\mathcal{D}}$ be the largest finite valued breakpoints of the original and modified game respectively. It follows that there exist $\delta\lambda^M$ and $\delta\widetilde{\lambda}^{\widetilde{M}}$ such that

$$\begin{aligned}\lambda^{\text{WE}}(T) &= \lambda^{\text{WE}}(D_M) + (T - D_M)\delta\lambda^M, \\ \widetilde{\lambda}^{\text{WE}}(T) &= \widetilde{\lambda}^{\text{WE}}(\widetilde{D}_{\widetilde{M}}) + (T - \widetilde{D}_{\widetilde{M}})\delta\widetilde{\lambda}^{\widetilde{M}}\end{aligned}$$

for all $T \geq \max(D_M, \widetilde{D}_{\widetilde{M}})$. For the sake of contradiction, assume that $\delta\widetilde{\lambda}^{\widetilde{M}} < \delta\lambda^M$. From Proposition 4.1.5 we have

$$\begin{aligned}\frac{\partial}{\partial D}\big(\widetilde{V}(T) - V(T)\big) &= \widetilde{\lambda}^{\text{WE}}(T) - \lambda^{\text{WE}}(T) \\ &= \lambda^{\text{WE}}(D_M) - \widetilde{\lambda}^{\text{WE}}(\widetilde{D}_{\widetilde{M}}) \\ &\quad + (T - D_M)\delta\lambda^M - (T - \widetilde{D}_{\widetilde{M}})\delta\widetilde{\lambda}^{\widetilde{M}}\end{aligned} \tag{4.12}$$

for any $T \geq \max(D_M, \widetilde{D}_{\widetilde{M}})$. Since we assume $\delta\widetilde{\lambda}^{\widetilde{M}} < \delta\lambda^M$, the above relation implies that for large enough T, we get $\widetilde{V}(T) < V(T)$ which contradicts Lemma 4.1.6. Therefore, we obtain $\delta\widetilde{\lambda}^{\widetilde{M}} \geq \delta\lambda^M$. Now consider two cases: (a) $\delta\widetilde{\lambda}^{\widetilde{M}} = \delta\lambda^M$ and (b) $\delta\widetilde{\lambda}^{\widetilde{M}} > \delta\lambda^M$. For (a) note that if $\widetilde{\lambda}^{\text{WE}}(T) < \lambda^{\text{WE}}(T)$ for any $T \geq \max(D_M, \widetilde{D}_{\widetilde{M}})$, then we arrive at a similar contradiction with Lemma 4.1.6 as before. Thus, for case (a), we must have $\widetilde{\lambda}^{\text{WE}}(T) \geq \lambda^{\text{WE}}(T)$ for all $T \geq \max(D_M, \widetilde{D}_{\widetilde{M}})$. For case (b), from (4.12), for all large values of T, we have $\widetilde{\lambda}^{\text{WE}}(T) \geq \lambda^{\text{WE}}(T)$. Hence, combining the reasoning of both cases, we find that there exists some value D^{BP} such that $\widetilde{\lambda}^{\text{WE}}(T) \geq \lambda^{\text{WE}}(T)$ for all $T \geq D^{\text{BP}}$. This completes the proof. □

Note that D^{BP} can be strictly larger than D_M, as is the case in Example 4.1.1b, where $D_M = 1$ while Figure 4.4 shows that BP occurs on the interval $D \in (\frac{2}{3}, 2)$.

4.4 The benefits of Braess's paradox

In the final part of this chapter we want to discuss how the results we have obtained on BP show that even when BP is detected in a network, one should be careful in drawing the conclusion that a set of paths is better removed. We start from Theorem 4.3.1 which, as discussed, hints at the possibility that a set $\mathcal{S}^{\texttt{rem}}$ that is not necessary at some demand D is not useful to the network. If this set of paths had not been present, either the WE-cost would have stayed the same for all lower levels of demand, or better yet, would have decreased for some of these demands.

Of course this tells us nothing about what happens for higher demands. A path that is unnecessary for one level of demand may be very important when demand is higher. It can even be the case that a path is necessary at some level of demand, becomes unnecessary at a higher level of demand, and finally becomes a necessary part of the "final" set of used paths $\mathcal{J}_M^{\text{use}}$. This phenomenon is showcased in the following example:

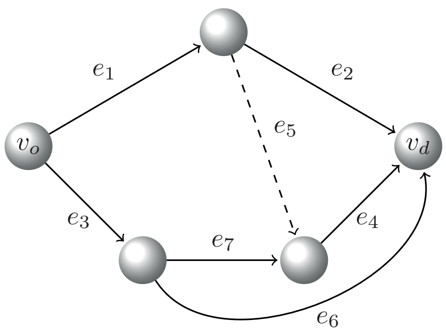


Figure 4.9: Modification of the Wheatstone network, in which there is one additional path.

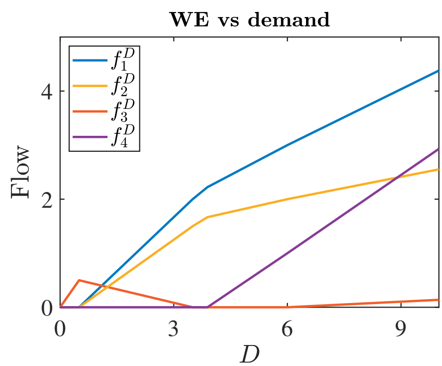


Figure 4.10: The evolution of f^D for the routing game over the network in Figure 4.9 defined by the costs (4.13).

Example 4.4.1. Consider the network in Figure 4.9, and let the cost functions of the edges be given by

$$\begin{aligned} C_{e_1}(f_{e_1}) &= 2f_{e_1}, & C_{e_2}(f_{e_2}) &= f_{e_2} + 1, \\ C_{e_3}(f_{e_3}) &= f_{e_3} + 1, & C_{e_4}(f_{e_4}) &= 2f_{e_4}, \\ C_{e_5}(f_{e_5}) &= 0, & C_{e_6}(f_{e_6}) &= f_{e_6} + 5, \\ & & C_{e_7}(f_{e_7}) &= f_{e_7}. \end{aligned}$$

For this network, the set $\mathcal{P}$ contains four paths, namely $p_1 = (e_1, e_2)$, $p_2 = (e_3, e_7, e_4)$,

$p_3 = (e_1, e_5, e_4)$ and $p_4 = (e_3, e_6)$. The path-cost function is then given by

$$C(f) = Af + b = \begin{pmatrix} 3 & 0 & 2 & 0 \\ 0 & 4 & 2 & 1 \\ 2 & 2 & 4 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix} f + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 6 \end{pmatrix}. \tag{4.13}$$

We see the evolution of the WE, which is unique in this case, in Figure 4.10. An explicit expression for the WE is provided in Appendix D. Note that path p_3 first carries all flow, and then loses all flow as it rerouted onto the paths p_1 and p_2. Finally p_3 becomes necessary again, as we see that $\mathcal{J}_M^{\text{use}} = \mathcal{P}$ in this case. The intuition for why this happens is that before path p_4 becomes active, the routing game is essentially a variant of that over the Wheatstone network discussed in Example 4.1.1a. As was the case there, The path over edge e_5 takes all flow at low levels of demand, but eventually becomes unnecessary for higher demands as paths p_1 and p_2 can carry the flow more efficiently. However, for even higher levels of demand, path p_4 becomes active, and this changes the evolution of the WE. Specifically, the availability of path p_4 means that less flow is routed onto the edges e_7 and e_5, and this reduces the increase in cost of path p_3, which then becomes a viable alternative to the paths p_1 and p_2 again.

Note that we can control when path p_4 becomes active by tuning the free flow cost $C_{e_6}(0) = \beta_{e_6}$ of edge e_6, which can be used to illustrate the ways in which BP can remain hidden or be revealed. When $\beta_{e_6} \in (0, 3.5)$, path p_4 becomes active before p_3 loses all flow, which prevents p_3 from ever losing all flow. In this case p_3 never becomes unnecessary, and Theorem 4.3.1 does not reveal the presence of BP. We can then still appeal to Theorem 4.3.5 to reveal the BP, since there remains a range of demands on which the flow over p_3 decreases. Setting $\beta_{e_6} = 0$ will prevent this as well, resulting in a scenario where p_4 becomes active before p_3 starts losing flow, preventing p_3 from ever decreasing in flow. The BP is still present, but the only result presented in this text that can help us reveal it is now Theorem 4.3.7. •

The above example already urges caution when it comes to dismissing a set of paths that becomes unnecessary, and thus causes BP at some level of demand, as useless or detrimental overall, since we do not know what happens at higher levels of demand. However, even for lower levels of demand the conclusions are not that straightforward. The following result shows that when a set $\mathcal{S}^{\text{rem}}$ causes BP at some demand D, its presence must have been strictly beneficial at some lower level of demand.

Theorem 4.4.2. *(Paths causing BP are useful at lower demands): let $\mathcal{P}$, $\mathcal{S}^{\text{rem}} \subset \mathcal{P}$, $\mathcal{C} \subset \mathcal{K}_{\text{aff}}$ and D be given. If $\lambda^{\text{WE}}(D) > \widetilde{\lambda}^{\text{WE}}(D)$ then there exist a D^-, D^+ such that*

$0 < D^- < D^+ < D$ and

$$\lambda^{\mathrm{WE}}(T) < \widetilde{\lambda}^{\mathrm{WE}}(T) \quad \text{for all } T \in (D^-, D^+).$$

Proof. The arguments for proving this are similar to those for Theorem 4.3.1. We note that $\lambda^{\mathrm{WE}}(\cdot)$ and $\widetilde{\lambda}^{\mathrm{WE}}(\cdot)$ are the derivatives of $V(D)$ and $\widetilde{V}(D)$ respectively, and are in addition continuous and piecewise affine with only finitely many points at which they are not differentiable. From Lemma 4.1.6 we know that $V(T) \leq \widetilde{V}(T)$ for all $T \in \mathbb{R}_{\geq 0}$. It follows that $\lambda^{\mathrm{WE}}(D) > \widetilde{\lambda}^{\mathrm{WE}}(D)$ implies $V(D) < \widetilde{V}(D)$. Since $V(0) = \widetilde{V}(0) = 0$ this implies that for some range of demands (D^-, D^+) with $0 \leq D^- < D^+ < D$ we must have $\lambda^{\mathrm{WE}}(T) < \widetilde{\lambda}^{\mathrm{WE}}(T)$ for all $T \in (D^-, D^+)$. □

The above result lends us a different perspective on Braess' paradox. We have already seen that the phenomenon is highly dependent on the demand, but this corollary shows that even though addition of a set of paths may increase the travel time of all participants at one level of demand, looking at a more complete picture, we see that the same set of paths must have decreased travel time for some lower level of demand. When deciding to keep or remove a path from a network it would thus be helpful to consider the effect of that path on the network for the entire range of demands in which the network functions.

This leads naturally to the question how to quantify the value of a path to the network while considering a range of demands. An initial, perhaps naive, method may be to consider the function

$$J(D) = \int_0^D \widetilde{\lambda}^{\mathrm{WE}}(z) - \lambda^{\mathrm{WE}}(z) dz$$

as a measure of the value of a set of paths $\mathcal{S}^{\mathrm{rem}}$ to the network on the range of demands from zero to D. For this we have the following result:

Proposition 4.4.3. *(Benefits of a path using a simple measure): let $\mathcal{P}$, $\mathcal{S}^{rem} \subset \mathcal{P}$, $\mathcal{C} \subset \mathcal{K}$ and $D \geq 0$ be given. We have*

$$J(D) \geq 0 \text{ for all } D \in [0, \infty),$$

with $J(D) = 0$ if and only if $\mathcal{S}^{rem} \notin \mathcal{N}_D$.

Proof. Since $\lambda^{\mathrm{WE}}(D) = \frac{d}{dD} V(D)$ and $\widetilde{\lambda}^{\mathrm{WE}}(D) = \frac{d}{dD} \widetilde{V}(D)$ we get

$$\begin{aligned}\int_0^D \widetilde{\lambda}^{\mathrm{WE}}(z) - \lambda^{\mathrm{WE}}(z) dz &= \widetilde{V}(z)|_0^D - V(z)|_0^D \\ &= \widetilde{V}(D) - V(D)\end{aligned}$$

$$\geq 0,$$

with equality holding if and only if $\widetilde{V}(D) = V(D)$. Using Lemma 4.1.6 the proof is finished. □

Using this measure, we see that in the worst-case scenario the set $\mathcal{S}^{\texttt{rem}}$ is "neutral" to the network, which only occurs when $\mathcal{S}^{\texttt{rem}}$ is unnecessary at demand D. If this is not the case, then the presence of $\mathcal{S}^{\texttt{rem}}$ is strictly beneficial to some degree.

As mentioned, J may be a fairly naive measure. Even when not knowing anything about the levels of demand that a network is likely to carry, we may weigh certain levels more than others. A reasonable approach may be to weigh each level of demand by that amount of demand, as it is more important for the system to perform well when demand is high than when demand is low. In this case we have the measure

$$W(D) = \int_0^D z\big(\widetilde{\lambda}^{\mathrm{WE}}(z) - \lambda^{\mathrm{WE}}(z)\big)dz.$$

For which we have the following result:

Proposition 4.4.4. *(Detriments of a path using a measure weighed by demand): let $\mathcal{P}$, $\mathcal{S}^{\texttt{rem}} \subset \mathcal{P}$, $\mathcal{C} \subset \mathcal{K}$ and $D > 0$ be given. If $\mathcal{S}^{\texttt{rem}} \notin \mathcal{N}_D$ then*

$$W(D) \leq 0$$

with equality holding if and only if $\mathcal{S}^{\texttt{rem}} \notin \mathcal{N}_T$ for all $T \in (0, D)$.

Proof. The result can be obtained by integration by parts:

$$\begin{aligned} W(D) &= \int_0^D z\big(\widetilde{\lambda}^{\mathrm{WE}}(z) - \lambda^{\mathrm{WE}}(z)\big)dz \\ &= z\big(\widetilde{V}(z) - V(z)\big)\big|_0^D - \int_0^D \widetilde{V}(z) - V(z)dz \\ &= \int_0^D V(z) - \widetilde{V}(z)dz. \end{aligned}$$

Here the last equality follows since $\mathcal{S}^{\texttt{rem}} \notin \mathcal{N}_D$ and therefore we have, by Lemma 4.1.6, that $V(D) = \widetilde{V}(D)$. From the same Lemma we know that $V(T) \leq \widetilde{V}(T)$ for all $T \in [0, D]$ and it follows that we end up with an integral over a non-positive function, which is only equal to the zero function if $\widetilde{V}(T) = V(T)$ for all $T \in (0, D)$. By Lemma 4.1.6, this happens if and only if $\mathcal{S}^{\texttt{rem}}$ is not necessary for all $T \in (0, D)$. This completes the proof. □

Thus, if we use this measure, and look at a range of demands from zero to a demand at which the set $\mathcal{S}^{\mathtt{rem}}$ is unnecessary, then it follows that $\mathcal{S}^{\mathtt{rem}}$ is at best "neutral" to the network performance, and this best case scenario only occurs when $\mathcal{S}^{\mathtt{rem}}$ is "useless" in that for the whole range of considered demands the set is unnecessary.

4.5 Conclusion

In this chapter we have studied the relation between the demand of a routing game with affine cost functions on the edges and the occurrence of Braess's paradox. First we have given a rigorous analysis of the evolution of the WE-cost of such routing games and subsequently we have used the gained insight to give sufficient conditions for the presence of BP which are computationally feasible to check. In addition we have provided a very efficient way to find a set of paths in the network that are potential candidates for causing BP, or are otherwise not beneficial to the network. We have also given a necessary and sufficient condition for the occurrence of Braess's paradox that can be usefully employed, but is computationally intractable to check in full. Finally we have shown that any set of paths responsible for BP at some level of demand must at other levels of demand strictly reduce the WE-cost. Based on this observation we have constructed two measures on the value of a set of paths, and have shown that even when a set of paths is observed to cause BP at some demand, removal of that set from the network could still be detrimental to the performance of the network overall, depending on which measure one uses.

Chapter 5

Inferring the prior

Recent years have seen increased utilization of *traffic information systems* (TISs) such as Google Maps and Waze by users of traffic networks. These TISs supply the drivers with information about some uncertain states of a network that they can not otherwise obtain. On one hand the increased utilization of TISs can cause problems such as congestion, and pose various challenges for traffic management [32], but on the other hand they also create the opportunity for *information design*, where information about the state of the network is strategically revealed in order to minimize congestion. To motivate this idea, we note that sometimes the travel cost of all drivers can be reduced when information about certain routes is withheld [33].

A fitting framework for studying the effects of information on decision making is *Bayesian persuasion* [34]. In the context of routing games, applying this framework entails assuming there are several states that the network can inhabit, representing for instance the presence or absence of road congestion, accidents, or weather events, and the participants are assumed to have a prior belief about the probability of each state occurring. The TIS releases information about the state using a set of messages or signals and in this way influences the posterior belief formed by the participants. Subsequently, participants select routes that minimize the expected travel cost in a selfish manner under the posterior belief, i.e., they route according to a *Wardrop equilibrium*. Note that we consider the case where all participants receive the same signal, commonly known as *public signalling*. In the above explained framework, the TIS can influence the flow by carefully designing the map from states to messages, also known as the signalling scheme. The effects of such a design naturally depend on the prior of the participants. However, the TIS may not know this prior in advance, presenting a problem for the implementation of this method. For instance, we will show that when aiming to minimize the total travel time of all participants, any error in the estimation of the prior by the TIS can result in decreased performance.

The aim of this chapter is to address this problem by studying how the prior of a population influences the Wardrop equilibrium, and how information about the prior can be inferred from observing the equilibrium flows under a signalling scheme.

Literature review

Adaptions of the Bayesian persuasion framework [34] for information design in the context of routing games have been studied in several recent works. In [35] the potential of information design to reduce travel times is show-cased for two common examples, in [36] the cost-performance of incentive-compatible signalling schemes are studied in comparison to socially optimal solutions, and in [37, 38] the relative performance of different strategies of information design, such as public and private signalling, are obtained. Instead of assuming that all users participate in persuasion, the works [39, 40] determine optimal information provision for heterogeneous populations, where a part of the users do not "trust" the TIS. Closer to the subject of our work, [41] also studies the effects of a mismatch between the actual distribution and the prior belief of a population concerning some parameters of a congestion game. In particular it introduces a type of routing game called a 'subjective Bayesian congestion game' which considers information that users have about the signals other users receive. Recent works also investigate the possible pitfalls of information provision by TISs. For example, [42, 43] explore inefficiencies caused by competing TISs; [33] highlights how knowing more routes can cause more congestion, revealing informational Braess's paradox; and [44] demonstrates oscillating traffic behaviour when information about travel times is available in real time. An analysis of how the benefits and detriments of revealing information to the population relate to the specifics of the cost functions and structure of the uncertainty is given in [45].

The viewpoint adopted in this chapter of learning about private parameters, such as the prior, of users in a routing setup is similar in spirit to [46] and [47]. In the former, the problem of estimating the learning rate of the population that employs a mirror descent algorithm to adapt route choices is considered. In the latter, learning of the cost functions of paths is studied. In a broader context, [48] investigates incentive design for a set of non-cooperative agents by learning the cost functions that govern their decisions. Our work is partly related to learning in routing games, where a lot of focus is on learning from the perspective of participants, see [49, 50] and references therein. The work [51] looks at a Bayesian framework and explores how participants learn about the state of the network in repeated play. It is worth noting that none of the works consider learning preferences or biases inherently present in the decisions of users in the context of information design.

Finally, we note that a popular alternative to information design for influencing flows in a traffic network is *incentive design*. For routing games, this area focuses on how tolls and subsidies can be used to influence the behaviour of traffic participants, see [52] and references therein for an overview, and [53] for an investigation on the potential of using incentive and information design in tandem.

Organization

We start this chapter by introducing the model of a routing game that can be in several states, and for which information about the state is communicated to the population by a TIS. Then we discuss an example that shows how a mismatch between the prior of the population and the estimate thereof made by the TIS can lead to a higher average travel cost. Our first point of study is then to find when a signalling scheme exists that allows the TIS to fully identify the prior. We show that under mild conditions, such a scheme exists, employing as many signals as there are states. Next we provide an iterative procedure for constructing such a signalling scheme, which in each iteration uses observations of the equilibrium flow incurred by the current signalling scheme. Next we discuss the challenges and potential encountered when generalizing these results to the case where the population is divided into fractions, each adhering to their own prior. We also show that a subclass of the signalling schemes we consider has some robustness properties, meaning that the same scheme can be used for identification of the prior even after the prior has been perturbed. Finally we summarize our findings in the conclusion of the chapter. Throughout the text we employ examples to improve the technical exposition of our results.

5.1 The model

The model considered in this chapter is a modification of the framework for routing games as introduced in Chapter 2. We recall from that chapter that a routing game is defined over a network, represented by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with a single origin-destination pair v_o, v_d, and the total amount of traffic, also called the demand, can be divided among the paths in $\mathcal{P}$, which is the set of all paths through the network from v_o to v_d. In this chapter we assume that the demand is constant, and for the sake of simplicity equal to one. That is, the feasible set is given by

$$\mathcal{F} := \Big\{ f \in \mathbb{R}^n_{\geq 0} \mid \sum_{p \in \mathcal{P}} f_p = 1 \Big\},$$

where $n = |\mathcal{P}|$. We recall that the edge-flow is then given by $f_{e_k} := \sum_{p \ni e_k} f_p$. To include uncertainty in our model, we assume that at any instant, the network can be in one of a finite number of states, drawn from the set of all states $\Theta := \{\theta_1, \cdots, \theta_m\}$. In any state $\theta_s \in \Theta$ each edge $e_k \in \mathcal{E}$ is associated with a cost function $C^{\theta_s}_{e_k} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $f_{e_k} \mapsto C^{\theta_s}_{e_k}(f_{e_k})$, which we assume to be known, continuous and strictly increasing on $\mathbb{R}_{\geq 0}$. As was the case in Chapter 2, the cost of traversing path p in state θ_s is simply the sum of costs in state θ_s of all edges contained in p:

$$C_p^{\theta_s}(f) = \sum_{e_k \in p} C_{e_k}^{\theta_s}(f_{e_k}). \tag{5.1}$$

We consider a Bayesian setting, where the users of the network are assumed to have a *prior* belief $q \in \Delta_1^m$ regarding the probability distribution of the state in which the network operates at any instant. That is, q_s is the probability with which the users believe the network will be in state θ_s, given that they have received no additional information. For $\varphi \in \mathbb{R}^m$, the *weighted cost* under φ of traversing a path p and an edge e_k are respectively given by

$$C_p^{\varphi}(f) := \sum_{s \in [m]} \varphi_s C_p^{\theta_s}(f), \quad C_{e_k}^{\varphi}(f) := \sum_{s \in [m]} \varphi_s C_{e_k}^{\theta_s}(f_{e_k}). \tag{5.2}$$

When $\varphi \in \Delta_1^m$, i.e., when φ is a probability distribution, we call these the *expected costs* under φ. For notational convenience, we define the following:

$$\begin{aligned} \mathcal{C} &:= \{C_{e_k}^{\theta_s}\}_{e_k \in \mathcal{E}, s \in [m]}, \\ C_p(f) &:= \left(C_p^{\theta_1}(f), \cdots, C_p^{\theta_m}(f)\right)^\top, \\ C^{\varphi}(f) &:= \left(C_1^{\varphi}(f), \cdots, C_n^{\varphi}(f)\right)^\top. \end{aligned}$$

Here, $\mathcal{C}$ is the set of all edge-cost functions, C_p is the vector of cost functions associated to path p per state, and C^{φ} is the vector of weighted costs under φ per path.

For a given probability distribution φ over the states Θ, we assume that the users aim to minimize their own expected cost of travelling, where the expectation is taken with respect to the distribution φ. To formalize which flows result from such rational decision-making of users, we define a modified version of the *Wardrop equilibrium*(WE) as follows:

Definition 5.1.1. *(φ-WE): Given a set of paths $\mathcal{P}$, states Θ, cost functions $\mathcal{C}$, and a probability distribution $\varphi \in \Delta_1^m$, a flow f^{φ} is said to be a φ-*based Wardrop equilibrium *(φ-WE) if $f^{\varphi} \in \mathcal{F}$ and for all $p \in \mathcal{P}$ such that $f_p^{\varphi} > 0$ we have*

$$C_p^{\varphi}(f^{\varphi}) \leq C_r^{\varphi}(f^{\varphi}) \quad \textit{for all } r \in \mathcal{P}. \tag{5.3}$$

The set of all φ-WE is denoted $\mathcal{W}^{\varphi}$.

Note that the set of φ-WE is simply the set of WE of a routing game without multiple states, where for all edges $e_k \in \mathcal{E}$ the edge-cost function is given by $C_{e_k}^{\varphi}$. The results in Chapter 2 therefore imply that the set $\mathcal{W}^{\varphi}$ is equal to the set of solutions of the *variational inequality*(VI) problem VI$(\mathcal{F}, C^{\varphi})$. We also recall from Chapter 2 that a

WE is not unique, but since we assume that the functions $C_{e_k}^{\varphi}$ are strictly increasing, we do have that the flow over the edges under WE is unique. That is, $\breve{f}^{\varphi}$ is a φ-WE if and only if

$$\breve{f}_{e_k}^{\varphi} = f_{e_k}^{\varphi} \text{ for all } e_k \in \mathcal{E}, f^{\varphi} \in \mathcal{W}^{\varphi}. \tag{5.4}$$

Throughout this chapter we use $f_{e_k}^{\varphi}$ to denote the edge-flow on edge e_k under φ-WE.

The last part of the model is a traffic information system (TIS), that observes the state θ_s of the network at any instant, and subsequently supplies information about this state to the drivers. The TIS has a set of signals $\mathcal{Z} := \{\zeta^1, \cdots, \zeta^z\}$ from which it chooses one to send to the users at any instant of the game. Before the traffic is routed, the TIS commits to a signalling scheme $\Phi : \Theta \mapsto \Delta_1^z$. Each state θ_s is mapped by Φ to a probability vector $\Phi(\theta_s) := \phi^{\theta_s} \in \Delta_1^z$. After observing state θ_s, the TIS randomly draws a signal from $\mathcal{Z}$ to send to the participants, where the probability of sending signal ζ^u is given by the u-th element of ϕ^{θ_s}. In our setting all participants receive the same signal, which is known as *public signalling*. Note that the signalling scheme Φ can be represented as a $z \times m$ column stochastic matrix; that is, $\Phi \in \mathsf{CS}(z, m)$, with the (u, s)-th entry, denoted ϕ_s^u, giving the probability of sending signal ζ^u after observing state θ_s. We will adhere to this matrix representation of Φ throughout this chapter.

After receiving a signal ζ^u, the users update their belief about the state of the network by forming a *posterior* $\widetilde{q}$ using Bayes' rule:

$$\widetilde{q}_s := \mathbb{P}[\theta_s|\zeta^u] = \frac{\mathbb{P}[\zeta^u|\theta_s]q_s}{\sum_{\ell \in [m]} \mathbb{P}[\zeta^u|\theta_\ell]q_\ell} = \frac{\phi_s^u q_s}{\sum_{\ell \in [m]} \phi_\ell^u q_\ell}, \tag{5.5}$$

for all $s \in [m]$, where $\mathbb{P}[\theta_s|\zeta^u]$ is the probability of the network being in state θ_s having received the signal ζ^u and $\mathbb{P}[\zeta^u|\theta_s]$ is the probability of sending signal ζ^u after observing state θ_s. The resulting flow is then assumed to be a $\widetilde{q}$-based Wardrop equilibrium. When no additional information regarding the state of the network is available to the users, the flow is assumed to depend on the prior q and is given by a q-WE denoted as f^q. Throughout this chapter we will use q to denote the prior, $\widetilde{q}^{\zeta^u}$ to denote the posterior with respect to the signal ζ^u, and use $\widetilde{q}$ when the signal is clear from the context. Associated sets of WE will be denoted as $\mathcal{W}^q$, $\mathcal{W}^{\zeta^u}$, and $\mathcal{W}^{\widetilde{q}}$, respectively. Similarly, given a distribution $\varphi \in \Delta_1^m$ we will use the notation $\mathcal{W}^{\varphi}$ for the set of φ-based WE, and $\mathcal{W}^{\theta_s}$ for a φ^{θ_s}-based WE, where the distribution φ^{θ_s} is defined by $\varphi_s^{\theta_s} = 1$. Note that when $q_s = 0$ for some $s \in [m]$, despite the TIS observing state θ_s, it is possible that $\widetilde{q}_\ell$ is ill-defined for some $\ell \in [m]$ as it may involve division by zero. To avoid this issue, we assume that $q_s > 0$ for all $s \in [m]$.

We finish this section with a motivating example showing how for a TIS that aims to design a signalling scheme to minimize social cost, a mismatch between the prior and the estimate of that prior made by that TIS can lead to an increase in social cost.

Example 5.1.2. *(Motivating example):* Consider a network with two nodes: the origin v_o and destination v_d, and two parallel paths from v_o to v_d as depicted in Figure 5.1.

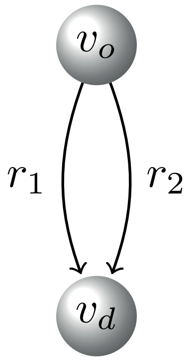


Figure 5.1: A parallel network with two roads.

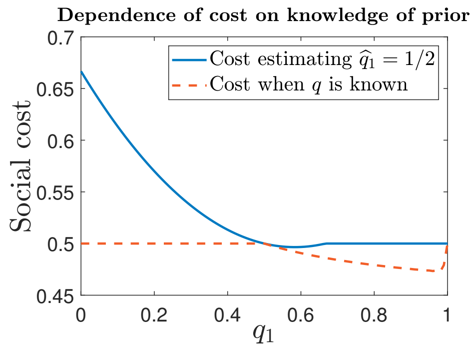


Figure 5.2: Cost for TIS with and without exact knowledge of q.

The network can be in two states, and the cost functions of the paths in these states are

$$C_1^{\theta_1}(f) = C_1^{\theta_2}(f) = 2f_1 + \frac{1}{2},$$
$$C_2^{\theta_1}(f) = 0, \text{ and } C_2^{\theta_2}(f) = 1.$$

The probability distribution of states θ_1 and θ_2 is given by $\varphi^{\text{true}} = (\varphi_1^{\text{true}}, \varphi_2^{\text{true}})$, where θ_1 occurs with probability $\varphi_1^{\text{true}} = 0.5$ and θ_2 occurs with probability $\varphi_2^{\text{true}} = 1 - \varphi_1^{\text{true}}$. The distribution φ^{true} is assumed to be known to the TIS. The goal of the TIS is to minimize the long-term average social cost, which is a function of the signalling scheme. For a general network, given the prior belief q, the state θ_s, and a message ζ^u, the incurred social cost is given by

$$J_q^{\text{stage}}(\zeta^u, \theta_s) := \sum_{p \in \mathcal{P}} \widetilde{f}_p^{\zeta^u} C_p^{\theta_s}(\widetilde{f}^{\zeta^u}), \tag{5.6}$$

where $\widetilde{f}^{\zeta^u}$ is a $\widetilde{q}^{\zeta^u}$-WE. Recall that for any two $\widetilde{q}^{\zeta^u}$-WE, say $\widetilde{f}^{\zeta^u}$ and $\check{f}^{\zeta^u}$, we have $\widetilde{f}_{e_k}^{\zeta^u} = \check{f}_{e_k}^{\zeta^u}$ for all $e_k \in \mathcal{E}$. From (5.1) and $f_{e_k} = \sum_{p \ni e_k} f_p$, we then conclude that (5.6) is independent of the choice of $\widetilde{q}^{\zeta^u}$-WE. The long-term average cost will be the sum of $J_q^{\text{stage}}(\zeta^u, \theta_s)$ over all possible combinations of signals ζ^u and states θ_s, weighted by the probability φ_s^{true} that θ_s occurs, and the probability ϕ_s^u of signal ζ^u being send when θ_s occurs. This will therefore depend on the number of signals that the scheme employs. However, in [54, Proposition 3] it is shown that a public signalling scheme $\Phi \in \mathsf{CS}(z, m)$ needs no more than m signals to achieve the optimum and therefore we set $z = m$. Summarizing this, the long-term average cost that the TIS aims to minimize is given by

$$J_q(\Phi) := \sum_{s \in [m]} \sum_{u \in [m]} \varphi_s^{\text{true}} \phi_s^u \sum_{p \in \mathcal{P}} \widetilde{f}_p^{\zeta^u} C_p^{\theta_s}(\widetilde{f}^{\zeta^u}).$$

When the TIS knows the prior belief q, it aims to find a scheme Φ that minimizes $J_q(\Phi)$. When the TIS does not know this prior belief, it assumes it to be the same as the probability distribution of states φ^{true}, and therefore employs a signalling scheme Φ that minimizes $J_{\varphi^{\text{true}}}(\Phi)$. Whenever $q \neq \varphi^{\text{true}}$, designing a signalling scheme using φ^{true} as an estimate of q can increase the social cost. This we show in Figure 5.2. The horizontal axis in the plot depicts the prior held by the users and since we only consider two states, it is completely specified by the first component q_1 of the two-dimensional vector q. The blue line shows the long-term average cost of the game when the TIS uses φ^{true} as an estimate of the prior q and employs a signalling scheme that minimizes $J_{\varphi^{\text{true}}}$. The dashed orange line shows the cost achieved when the TIS uses the exact knowledge of q and employs an optimal signalling scheme minimizing J_q. We see that the TIS with full knowledge performs better, with the difference becoming more pronounced as q moves further away from φ^{true}. •

5.2 Inferring the prior: General case

We have seen that to achieve optimal results, a TIS needs to know the prior that underlies the decisions of the population. Our goal is thus to find a way for the TIS to infer this prior based on the information available to it. With this goal in mind, we start by investigating the relation between the path-flows under φ-WE, edge-flows under φ-WE, and the distribution φ. To illustrate the intuition behind the ideas we will present on this subject, we first consider a simplified example where the TIS provides no information to the users and the resulting flow f^q is therefore a q-WE as players base their routing choices on the prior.[1] From Definition 5.1.1 we know that f^q satisfies (5.3) where φ is replaced by q. That is,

$$\begin{aligned} C_p^q(f^q) &= C_r^q(f^q), \quad \text{for all } p, r \in \mathcal{P} \text{ such that } f_p^q, f_r^q > 0, \\ C_p^q(f^q) &\leq C_r^q(f^q), \quad \text{for all } p, r \in \mathcal{P} \text{ such that } f_p^q > 0, f_r^q = 0. \end{aligned} \tag{5.7}$$

[1] The same situation can be achieved by using a signalling scheme which supplies no information, for instance by setting $\phi_s^u = \frac{1}{z}$ for all u, s.

Defining the matrix-valued map $C^{\mathtt{mat}} : \mathcal{F} \to \mathbb{R}_{\geq 0}^{n\times m}$ as

$$C^{\mathtt{mat}}(f) := \begin{pmatrix} C_1^{\theta_1}(f) & C_1^{\theta_2}(f) & \dots & C_1^{\theta_m}(f) \\ C_2^{\theta_1}(f) & C_2^{\theta_2}(f) & \dots & C_2^{\theta_m}(f) \\ \vdots & \vdots & & \vdots \\ C_n^{\theta_1}(f) & C_n^{\theta_2}(f) & \dots & C_n^{\theta_m}(f) \end{pmatrix},$$

we have $C^q(f^q) = C^{\mathtt{mat}}(f^q) \cdot q$ and so (5.7) can be rewritten as

$$\begin{aligned} \big(C_p^{\mathtt{mat}}(f^q) - C_r^{\mathtt{mat}}(f^q)\big)q &= 0, \quad \text{for all } p, r \text{ with } f_p^q, f_r^q > 0, \\ \big(C_p^{\mathtt{mat}}(f^q) - C_r^{\mathtt{mat}}(f^q)\big)q &\leq 0, \quad \text{for all } p, r \text{ with } f_p^q > 0, f_r^q = 0, \end{aligned} \tag{5.8}$$

where $C_p^{\mathtt{mat}}(f)$ denotes the p-th row of $C^{\mathtt{mat}}(f)$. Given a q-WE f^q, the above gives constraints on the possible values that the prior can take. We see that observing the WE flow f^q supplies us with constraints on the prior q, and in this way, observing equilibria can help us identify the prior. Most information can of course be obtained from the equality constraints, though it is also possible that a combination of equality and inequality constraints together result in additional equality constraints. We also note that in addition to the above we have $\sum_{s\in[m]} q_s = 1$, which is linearly independent from all equality constraints obtained from (5.8)[2]. We thus find a number of linearly independent equality constraints on q. However, since $q \in \mathbb{R}^m$ we need m such constraints to uniquely determine q, and there is no guarantee that observing a flow $f^q \in \mathcal{W}^q$ will give that number of independent constraints. If we do not obtain enough information by observing f^q, we can instead use a public signalling scheme Φ to induce different posteriors for the population. These posteriors will lead to different equilibrium flows resulting in equality constraints of the form (5.8), where q and f^q are replaced with $\widetilde{q}$ and $f^{\widetilde{q}}$, respectively. Using (5.5), these constraints on the posterior $\widetilde{q} = \widetilde{q}^{\zeta^u}$ can be rewritten into constraints on the prior q, by noting that

$$\big(C_p^{\mathtt{mat}}(f) - C_r^{\mathtt{mat}}(f)\big)\widetilde{q}^{\zeta^u} = \sum_{s\in[m]} \frac{\phi_s^u\big(C_p^{\theta_s}(f) - C_r^{\theta_s}(f)\big)q_s}{\sum_{\ell\in[m]} \phi_\ell^u q_\ell}.$$

Thus constraints on the prior q imposed by observing the equilibrium flow $\widetilde{f}^{\zeta^u}$ are of the form

$$\sum_{s\in[m]} \phi_s^u\big(C_p^{\theta_s}(\widetilde{f}^{\zeta^u}) - C_r^{\theta_s}(\widetilde{f}^{\zeta^u})\big)q_s = 0, \tag{5.9a}$$

[2] An intuitive way to see this is as follows. When f^q is fixed, for any q that satisfies the constraints in (5.7), cq will also satisfy these constraints for any $c \in \mathbb{R}_{\geq 0}$. This is clearly not the case for the constraint $\sum_{s\in[m]} q_s = 1$

$$\sum_{s\in[m]} \phi_s^u \big(C_p^{\theta_k}(\widetilde{f}^{\zeta^u}) - C_r^{\theta_s}(\widetilde{f}^{\zeta^u})\big) q_s \leq 0, \tag{5.9b}$$

where (5.9a) holds for all p, r with $\widetilde{f}_p^{\zeta^u}, \widetilde{f}_r^{\zeta^u} > 0$ and (5.9b) holds for all p, r such that $\widetilde{f}_p^{\zeta^u} > 0$ and $\widetilde{f}_r^{\zeta^u} = 0$. In the above conditions, the denominator has been dropped, since it is the same for each term in the summation, and assumed to be positive. When some of the constraints in (5.9) are linearly independent from those in (5.8), observing $\widetilde{f}^{\zeta^u}$ has thus supplied us with additional information on the prior. In this way, each of the individual signals induces a flow that potentially reveals more information about the prior to the TIS. To identify q completely, the set of all constraints obtained from observing the flows induced by each signal must contain at least m linearly independent constraints. For a given signalling scheme Φ we denote the set of all priors satisfying all obtained constraints from all signals as

$$\mathcal{Q}_\Phi = \{q \in \Delta_1^m \mid q \text{ satisfies (5.9) for all } u \in [z]\}.$$

Though the above seems to depend on which specific $\widetilde{q}^{\zeta^u}$-WE $\widetilde{f}^{\zeta^u}$ are observed we will show in upcoming results, (specifically Corollary 5.2.9) that this is not the case. We give the following definition:

Definition 5.2.1. *(q-identifying signalling scheme): Given a set of paths $\mathcal{P}$, states Θ, cost functions $\mathcal{C}$, and a prior $q \in \Delta_1^m$, a signalling scheme $\Phi \in \mathsf{CS}(s, m)$ is called q-identifying if $\mathcal{Q}_\Phi = \{q\}$.*

The goal of this chapter is to answer the question "How can we design Φ so as to ensure that it is q-identifying?" Before we are ready to address this issue, we must now first investigate the relations between the distribution φ, the associated φ-WE f^φ, and the related edge-flows $f_{e_k}^\varphi$.

5.2.1 Probability distribution and equilibrium

The results in upcoming sections build upon three lemmas presented here, which give insight into how the edge-flows under φ-WE, path-flows under φ-WE, and the distribution φ relate to each other. To ease the exposition of the first lemma, we introduce the following notation:

$$\mathcal{F}_e := \{v \in \mathbb{R}_{\geq 0}^{|\mathcal{E}|} \mid \exists f \in \mathcal{F} \text{ such that } v_k = f_{e_k} \quad \text{for all } k \in [|\mathcal{E}|]\}.$$

Note that since $\mathcal{F}$ is compact, so is $\mathcal{F}_e$. Our first lemma makes use of this set to show that the edge-flows under φ-WE change continuously with respect to φ.

Lemma 5.2.2. *(Continuity of φ-WE edge-flows): Let $\mathcal{P}$, Θ, and $\mathcal{C}$ be given. For every $\epsilon > 0$, there exists a $\delta > 0$ such that for any two distributions $\varphi, \xi \in \Delta_1^m$, we have*

$$\|\varphi - \xi\| < \delta \Rightarrow \left| f_{e_k}^{\varphi} - f_{e_k}^{\xi} \right| < \epsilon \quad \text{for all } e_k \in \mathcal{E}.$$

In other words, the edge-flows under φ-WE depend continuously on the distribution φ.

Proof. For $\varphi \in \Delta_1^m$, recall the notation of $C_{e_k}^{\varphi}$ from (5.2). Following [55], a flow vector $f^{\varphi} \in \mathcal{F}$ is a φ-WE if and only if it is a solution of the following optimization problem:

$$\min_{f \in \mathcal{F}} \sum_{e_k \in \mathcal{E}} \int_0^{f_{e_k}} C_{e_k}^{\varphi}(t) dt, \tag{5.10}$$

where for a path-flow f, the quantity f_{e_k} is the corresponding flow on edge e_k given by $f_{e_k} = \sum_{p \ni e_k} f_p$. Recall from [9] that while the φ-WE need not be unique, the edge-flows induced by them are. Thus, following (5.10), the edge-flows associated to φ-WE are given by the unique solution of the following problem:

$$\min_{f_e \in \mathcal{F}_e} \sum_{k \in [|\mathcal{E}|]} \int_0^{f_{e_k}} \sum_{s \in [m]} \varphi_s C_{e_k}^{\theta_s}(t) dt. \tag{5.11}$$

Consider the above optimization problem with φ as a parameter. Given φ, denote the optimal solution as $f_{\text{edge}}^{\varphi}$. Since the objective function of the above problem depends linearly on φ and the domain is compact and independent of φ, we deduce from [56, Proposition 4.4] that the map $\varphi \mapsto f_{\text{edge}}^{\varphi}$ is continuous. This concludes the proof. □

To ease the exposition of the next result, we define

$$\mathcal{R}_{\varphi}^{\text{use}} := \{p \in \mathcal{P} \mid \exists f^{\varphi} \in \mathcal{W}^{\varphi} \text{ such that } f_p^{\varphi} > 0\}.$$

That is, $\mathcal{R}_{\varphi}^{\text{use}}$ denotes the set of all paths p for which there exists a φ-WE such that a positive amount of flow is routed onto path p. We call these paths the *used* paths. The set of φ-WE then has the following useful properties:

Lemma 5.2.3. *(Characterizing used paths of φ-WE): Let $\mathcal{P}$, Θ, $\mathcal{C}$, and $\varphi \in \Delta_1^m$ be given. We have the following:*

1. *There exists an $f^{\varphi} \in \mathcal{W}^{\varphi}$ satisfying $f_p^{\varphi} > 0$ for all $p \in \mathcal{R}_{\varphi}^{\text{use}}$.*

2. *We have $p \in \mathcal{R}_{\varphi}^{\text{use}}$ if and only if $f_{e_k}^{\varphi} > 0$ for all $e_k \in p$.*

Proof. The first claim follows from the fact that the set $\mathcal{W}^\varphi$ is convex. This can be deduced from (5.4) and noting that if two path flows induce the same edge flow, then any convex combination of these flows will still induce that same edge flow. To see that the first claim then follows let the flow $f^{\varphi,r} \in \mathcal{W}^\varphi$ denote a WE satisfying $f_r^{\varphi,r} > 0$ for $r \in \mathcal{R}_\varphi^{\text{use}}$. Such a flow exists by the definition of $\mathcal{R}_\varphi^{\text{use}}$. Next select scalars $c_r > 0$ for all $r \in \mathcal{R}_\varphi^{\text{use}}$ such that $\sum_{r \in \mathcal{R}_\varphi^{\text{use}}} c_r = 1$. Using the selected WE flows and scalars, define $f^{\text{use}} := \sum_{r \in \mathcal{R}_\varphi^{\text{use}}} c_r f^{\varphi,r}$. Note that $f^{\text{use}} \in \mathcal{W}^\varphi$ as this set is convex. Finally, by definition of $\{f^{\varphi,r}, c_r\}$ and the fact that all WE flows are non-negative, we deduce that $f_r^{\text{use}} > 0$ for all $r \in \mathcal{R}_\varphi^{\text{use}}$. This establishes the first claim.

For the second claim, the "only if" part is easier to deduce. Let $p \in \mathcal{R}_\varphi^{\text{use}}$ and let $f^{\varphi,p} \in \mathcal{W}^\varphi$ satisfy $f_p^{\varphi,p} > 0$. Since $f_r^{\varphi,p} \geq 0$ for all $r \in \mathcal{P}$, and $f_{e_k} = \sum_{p \ni e_k} f_p$, it follows that $f_{e_k}^{\varphi,p} > 0$ for all $e_k \in p$. For the other direction we only provide a sketch of the arguments here. In it we make use of some properties of equilibrium flows which are straightforward, but cumbersome to establish formally, which is why, in the interest of space, we have chosen to omit a more formal proof. For a more detailed discussion on the subject we refer to [57].

For the sketch of the proof, first note that for a φ-WE, a total flow of unity enters and leaves the network at the origin and destination, respectively, while for all other vertices the flow satisfies mass-conservation constraints. That is, the total flow entering and leaving a vertex are equal. Second, it can also be shown that φ-WE does not contain any cycle with a positive amount of flow on all its edges. To see this, note that reducing the flow equally from all edges in such a cycle will preserve mass conservation and inflow and outflow constraints, while the value of (5.11) decreases. Thus, with the presence of a positive-flow cycle, the path-flow can not be a φ-WE. Lastly, consider any path p such that $f_{e_k}^\varphi > 0$ for all $e_k \in p$. Set $f_p^\varphi := \min_{e_k \in p} f_{e_k}^\varphi$ and then subtract f_p^φ of flow from all edges in p. The new flow will then still satisfy mass-conservation constraints, but the inflow and outflow at the origin and destination have both decreased by f_p^φ. Continue this procedure until all flow has been assigned and the result is a feasible flow f^φ which induces the same edge-flow as any φ-WE. Therefore, f^φ is a WE, and it satisfies $f_p^\varphi > 0$ for any desired $p \in \mathcal{R}_\varphi^{\text{use}}$ by construction, which concludes the proof. The procedure of assigning flow in this way is treated in more detail in [57, Theorem 2.1]. □

In our next result we show that for a given $f \in \mathcal{F}$, the set of all distributions φ such that $f \in \mathcal{W}^\varphi$ is compact and convex.

Lemma 5.2.4. *(Convexity of set of distributions inducing the same φ-WE): Let $\mathcal{P}$, Θ, $\mathcal{C}$, and $f \in \mathcal{F}$ be given. The set of distributions φ with $f \in \mathcal{W}^\varphi$ is compact and convex.*

Proof. For any distribution $\varphi \in \Delta_1^m$, we have $f \in \mathcal{W}^\varphi$ if and only if the constraints in (5.7) hold, where q and f^q are replaced with φ and f, respectively. Since f is fixed, the map $C^{\mathtt{mat}}(f)$ is also fixed and we see that (5.7) imposes a number of equality and non-strict inequality constraints on φ, all of which are affine. Therefore, the set of φ satisfying these constraints is convex and closed. Since distributions belong to a compact set Δ_1^m, the claim follows. □

We illustrate the implications of Lemma 5.2.4 using the following examples. For simplicity's sake, we have chosen examples such that the φ-WE are unique.

Example 5.2.5. *(Demonstration of Lemma 5.2.4):* Consider a 2-path, 2-state network, with cost functions given by

$$C_1^{\theta_1}(f) = 0.8f_2 + 0.7, \quad C_1^{\theta_2}(f) = 0.1f_2 + 0.2,$$
$$C_2^{\theta_1}(f) = 0.3f_1 + 0.2, \quad C_2^{\theta_2}(f) = 0.5f_1 + 0.5.$$

Figure 5.3 shows the relationship between the φ-WE and the distribution φ.

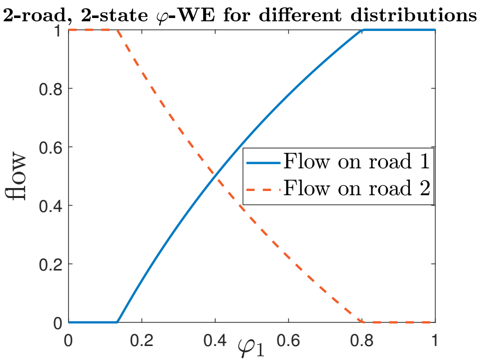


Figure 5.3: A graph showing the relationship between the φ-WE and the distribution φ for a 2-path, 2-state scenario.

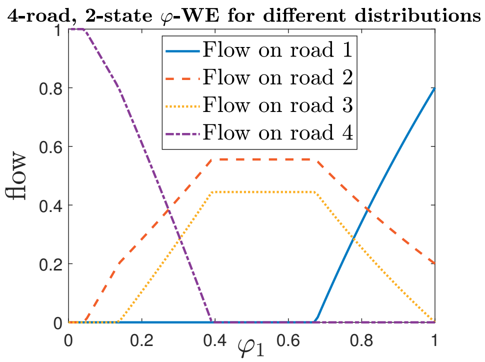


Figure 5.4: A graph showing the relationship between the φ-WE and the distribution φ for a 4-path, 2-state scenario.

Note that since $\varphi_2 = 1 - \varphi_1$ the distributions φ in the above figures are completely defined by the value φ_1. Figure 5.3 shows that the φ-WE remains constant in two convex regions, namely when $\varphi_1 \leq 0.133$ and when $\varphi_1 \geq 0.8$. In one of these cases we have $f^{\varphi} = f^{\theta_1} = (1, 0)^{\top}$ and in the other $f^{\varphi} = f^{\theta_2} = (0, 1)^{\top}$.

Next we consider a 4-path, 2-state network, with the following cost functions:

$$\begin{aligned} C_1^{\theta_1}(f) &= f_1 + 1, & C_1^{\theta_2}(f) &= 0.4f_1 + 4, \\ C_2^{\theta_1}(f) &= 0.5f_2 + 1.7, & C_2^{\theta_2}(f) &= 0.5f_2 + 1.7, \\ C_3^{\theta_1}(f) &= 0.4f_3 + 1.8, & C_3^{\theta_2}(f) &= 0.4f_3 + 1.8, \\ C_4^{\theta_1}(f) &= 0.4f_4 + 3.5, & C_4^{\theta_2}(f) &= 0.6f_4 + 1. \end{aligned} \tag{5.12}$$

Figure 5.4 shows the dependency between the distribution and the WE. We see that the situation has changed compared to the 2-path, 2-state case. Here we find a region in which the φ-WE remains constant while not equal to f^{θ_1} or f^{θ_2} or having all flows on one path. The sets of distributions in which the φ-WE remains constant are still convex, as claimed in 5.2.4

Although it is perhaps not directly apparent from Lemma 5.2.4, a consequence of that result is that for any distribution which is not in a convex set where the φ-WE remains constant, the associated φ-WE is unique to that distribution. When such a flow is observed, we can derive the unique distribution which induced it. If a φ-WE is observed that can be induced by multiple distributions, we can at best limit the distribution that induced the flow to a set. Intuitively regions of Δ_1^m where the φ-WE remains constant are less helpful in identifying q, and should be avoided when attempting to design a q-identifying signalling scheme. •

5.2.2 Existence of q-identifying signalling schemes

We are now ready to address the existence of signalling schemes which are q-identifying. Our strategy involves first showing existence for the simplified case where there are only two states; i.e. $\Theta = \{\theta_1, \theta_2\}$. Later we can use this result for the more general case $\Theta = \{\theta_1, \cdots, \theta_m\}$, $m \in \mathbb{N}$ by designing our signalling scheme Φ in such a way that the resulting posteriors only assign positive probability to exactly two states. A key element of designing such a scheme is the set of flows that provide information regarding the distribution that induced it. In particular, for the case $\Theta = \{\theta_1, \theta_2\}$, we define the set of *informative* flows $\mathcal{F}_{\text{inf}}$ as follows:

$$\mathcal{F}_{\text{inf}} := \{f \in \mathcal{F} \mid f \notin \mathcal{W}^{\theta_1} \cup \mathcal{W}^{\theta_2},\ \ f_p > 0\ \ \forall p \in \mathcal{R}^{\text{use}}_{\varphi^{\theta_1}}\}. \tag{5.13}$$

That is, $\mathcal{F}_{\text{inf}}$ is the set of all flows that are not in the set of φ^{θ_1}- or φ^{θ_2}-WE, but which do contain a positive amount of flow on all paths that have a positive amount of flow for some φ^{θ_1}-WE. The importance of this set lies in the fact that for the two-state case, observing a flow from this set allows us to uniquely identify which distribution induced that flow.

Remark 5.2.6. *(Sufficiency of $\mathcal{F}_{\text{inf}}$):* We note that it is not necessary for a flow f^{φ} to lie in $\mathcal{F}_{\text{inf}}$ in order to allow φ to be identified. Any flow f^{φ} that can only be induced by a unique distribution φ will, when observed, necessarily allow us to identify the distribution φ that induced it, while the set $\mathcal{F}_{\text{inf}}$ limits the attention to flows with a special relation to the flows in $\mathcal{W}^{\theta_1}$. The set $\mathcal{F}_{\text{inf}}$ is however of special import in the coming results because under mild assumptions, we can identify conditions that allow the flow induced by a signal to be contained in $\mathcal{F}_{\text{inf}}$. •

Before we discuss our results, we collect two useful properties of φ-WE here, both of which follow from the fact that a flow is a φ-WE if and only if it induces the same unique edge flow as all other φ-WE, as mentioned before in (5.4). The first property implies that for two distributions, the induced sets of WE overlap if and only if they are equal.

Lemma 5.2.7. *(Intersection of sets of WE induced by two distributions): Let $\mathcal{P}$, Θ, $\mathcal{C}$, and two distributions $\varphi, \xi \in \Delta_1^m$ be given. Then, $\mathcal{W}^\varphi \neq \mathcal{W}^\xi$ if and only if $\mathcal{W}^\varphi \cap \mathcal{W}^\xi = \emptyset$.*

The second property is that for two flows which are both WE induced by the same distribution, the sets of all distributions for which the first and the second flow is a WE, respectively, are equal.

Lemma 5.2.8. *(Equality of sets of distributions inducing two φ-WE): Let $\mathcal{P}$, Θ, and $\mathcal{C}$ be given. For $\xi \in \Delta_1^m$, if we have $f^\xi \in \mathcal{W}^\xi$ and $\check{f}^\xi \in \mathcal{W}^\xi$, then*

$$\{\varphi \in \Delta_1^m \mid f^\xi \in \mathcal{W}^\varphi\} = \{\varphi \in \Delta_1^m \mid \check{f}^\xi \in \mathcal{W}^\varphi\}.$$

A useful consequence of the above is that $\mathcal{Q}_\Phi$ is independent of which $\widetilde{q}^{\zeta^u}$-WE flow is observed for each signal ζ^u.

Corollary 5.2.9. *(Equal informativity of all q^{ζ^u}-WE.): Let $\mathcal{P}$, Θ, $\mathcal{C}$, a signalling scheme $\Phi \in \mathsf{CS}(z, m)$ and a signal ζ^u be given. For any $\widetilde{f}^{\zeta^u}$, $\check{f}^{\zeta^u} \in \mathcal{W}^{\zeta^u}$ the set of all priors $\varphi \in \Delta_1^m$ satisfying* (5.9) *is the same.*

Proof. The result follows by applying Lemma 5.2.8 to the routing game where the cost functions $C_p^{\theta_s}(\cdot)$ are replaced with $\phi_s^u C_p^{\theta_s}(\cdot)$. □

Our first result considers the two-state case, and shows that there exists a set of distributions which induce flows in $\mathcal{F}_{\inf}$.

Lemma 5.2.10. *(Distributions leading to $\mathcal{F}_{\inf}$): Let $\mathcal{P}$, Θ, and $\mathcal{C}$ be given, where $\Theta = \{\theta_1, \theta_2\}$ and $\mathcal{W}^{\theta_1} \neq \mathcal{W}^{\theta_2}$. Let $\mathcal{F}_{\inf}$ be as given in* (5.13)*. There exist distributions $\xi \neq \eta$ with $\xi_1 > \eta_1$ such that for any $\varphi_\mu := \mu\varphi^{\theta_1} + (1-\mu)\xi$, $\mu \in [0,1]$ we have $\mathcal{W}^{\varphi_\mu} = \mathcal{W}^{\theta_1}$, and for any $\varphi_\lambda := \lambda\xi + (1-\lambda)\eta$, $\lambda \in (0,1)$ there exists $f^{\varphi_\lambda} \in \mathcal{W}^{\varphi_\lambda}$ such that*

$$f^{\varphi_\lambda} \in \mathcal{F}_{\inf}. \tag{5.14}$$

Proof. First we aim to find the distribution $\xi := (\xi_1, 1-\xi_1)$. Pick any $f^{\theta_1} \in \mathcal{W}^{\theta_1}$. From Lemma 5.2.4, the set of distributions $\varphi \in \Delta_1^2$ with $f^{\theta_1} \in \mathcal{W}^\varphi$ is convex and compact. That is, there exists a $c \in [0,1]$ such that $f^{\theta_1} \notin \mathcal{W}^\varphi$ for all $\varphi \in \Delta_1^2$ with $\varphi_1 < c$ and $f^{\theta_1} \in \mathcal{W}^\varphi$ for all $\varphi \in \Delta_1^2$ with $\varphi_1 \geq c$. In addition, by Lemma 5.2.7, $f^{\theta_1} \in \mathcal{W}^\varphi$ for some φ if and only if $\mathcal{W}^\varphi = \mathcal{W}^{\theta_1}$. Combining these two facts and setting $\xi := (c, 1-c)$ yields that: (a) $\mathcal{W}^{\varphi_\mu} = \mathcal{W}^{\theta_1}$ for all $\varphi_\mu = \mu\varphi^{\theta_1} + (1-\mu)\xi$ and $\mu \in [0,1]$; and (b) for all $\varphi \in \Delta_1^2$ with $\varphi_1 < \xi_1$ and all $f^\varphi \in \mathcal{W}^\varphi$, we have $f^\varphi \notin \mathcal{W}^{\theta_1}$. The latter item (b) shows that $\xi_1 > 0$ which is essential for a η distribution with $\eta_1 < \xi_1$ to exist. To see that $\xi_1 > 0$, note that $\mathcal{W}^{\theta_1} \neq \mathcal{W}^{\theta_2}$ and by Lemma 5.2.7, $\mathcal{W}^{\theta_1} \cap \mathcal{W}^{\theta_2} = \emptyset$. This statement will contradict if $\xi_1 = 0$ as then, $\mathcal{W}^{\varphi_\mu} = \mathcal{W}^{\theta_2}$ for $\mu = 0$. The next step is to find the distribution $\eta := (\eta_1, 1-\eta_1)$. Let f^{ξ}_{edge} be the edge-flows associated with any ξ-WE. Pick any $p \in \mathcal{R}^{\text{use}}_\eta$ and by the second implication of Lemma 5.2.3, $(f^{\xi}_{\text{edge}})_k > 0$ for all $e_k \in p$. By continuity property of Lemma 5.2.2, there exist $\delta_\xi > 0$ such that $(f^{\eta}_{\text{edge}})_k > 0$ for edge-flows associated to any η-WE where $\|\xi - \eta\| < \delta_\xi$. This along with the second implication of Lemma 5.2.3 implies that $\mathcal{R}^{\text{use}}_\xi \subseteq \mathcal{R}^{\text{use}}_\eta$ for all η satisfying $\|\xi - \eta\| < \delta_\xi$. This along with the fact $\mathcal{W}^\xi = \mathcal{W}^{\theta_1}$ shown above, gives us

$$\mathcal{R}^{\text{use}}_{\varphi^{\theta_1}} \subseteq \mathcal{R}^{\text{use}}_\eta,$$

for all η satisfying $\|\xi - \eta\| < \delta_\xi$. From the first claim of Lemma 5.2.3, there exists $f^\eta \in \mathcal{W}^\eta$ such that $f^\eta_p > 0$ for all $p \in \mathcal{R}^{\text{use}}_{\varphi^{\theta_1}}$. Further, restricting our attention to η with $\eta_1 < \xi_1$, we also know that $f^\eta \notin \mathcal{W}^{\theta_1}$. To establish (5.14), we now show that $f^\eta \in \mathcal{F}_{\text{inf}}$. Given the above properties of f^η, all that remains to be shown to prove $f^\eta \in \mathcal{F}_{\text{inf}}$ is that setting δ_ξ small enough ensures $f^\eta \notin \mathcal{W}^{\theta_2}$. For this, note that since $\mathcal{W}^{\theta_1} \cap \mathcal{W}^{\theta_2} = \emptyset$, we have $f^{\theta_1}_{\text{edge}} \neq f^{\theta_2}_{\text{edge}}$. Consequently, by Lemma 5.2.2 and $\mathcal{W}^\xi = \mathcal{W}^{\theta_1}$, it follows that there exists $\delta_0 > 0$ such that $\|\xi - \eta\| < \delta_0$ gives $f^\eta \notin \mathcal{W}^{\theta_2}$. Thus, setting $\delta_\xi < \delta_0$ implies that for any $\eta \in \Delta_1^2$ with $\|\xi - \eta\| < \delta_\xi$ and $\eta_1 < \xi_1$ there exists $f^\eta \in \mathcal{W}^\eta$ such that $f^\eta \in \mathcal{F}_{\text{inf}}$. Fixing η_1 as the infimum over all values for which $\|\xi - \eta\| < \delta_\xi$ holds finishes the proof. □

Figure 5.4 can help us gain some intuition about the implications of Lemma 5.2.10. Under the given assumptions, the result divides the set Δ_1^2 of all distributions into three convex regions. The first region is compact, and for any distribution inside of it the induced flows are contained in $\mathcal{W}^{\theta_1}$. In Figure 5.4 we see that this region is the singleton set $\{(0,1)\}$. The second region is a convex and open set of distributions bordering the first region, for which the induced flows are in $\mathcal{F}_{\text{inf}}$. In Figure 5.4 this would be all distributions between $\varphi_1 = 1$ and the first point where the flow on path 1 becomes zero. Note that any flow in this region is induced by a unique distribution. The third region then contains all other distributions. Note that in this third region there are still flows that are uniquely associated with only one distribution.

The next result shows that if for a given distribution φ there exists a φ-WE f^φ such that $f^\varphi \in \mathcal{F}_{\text{inf}}$, then the constraints (5.8) for *any* φ-WE uniquely determine φ.

Lemma 5.2.11. *(Informativity of flows in $\mathcal{F}_{\text{inf}}$): Let $\mathcal{P}$, $\Theta = \{\theta_1, \theta_2\}$, $\mathcal{C}$, and $\varphi \in \Delta_1^2$ be given. If there exists $f^\varphi \in \mathcal{W}^\varphi$ satisfying $f^\varphi \in \mathcal{F}_{\text{inf}}$, then φ is the unique solution to* (5.8) *for any φ-WE $\breve{f}^\varphi \in \mathcal{W}^\varphi$.*[3]

Proof. Let $f^{\varphi,1} \in \mathcal{W}^\varphi$ be a flow such that $f_p^{\varphi,1} > 0$ for all $p \in \mathcal{R}_\varphi^{\text{use}}$, which exists by Lemma 5.2.3. By assumption, there exists a WE $f^{\varphi,2} \in \mathcal{W}^\varphi$ satisfying $f^{\varphi,2} \in \mathcal{F}_{\text{inf}}$. Let $\breve{f}^\varphi \in \mathcal{W}^\varphi$ be defined as $\breve{f}^\varphi := \lambda_1 f^{\varphi,1} + \lambda_2 f^{\varphi,2}$ for some $\lambda_1, \lambda_2 > 0$ with $\lambda_1 + \lambda_2 = 1$. Note that $\breve{f}_p^\varphi > 0$ for all $p \in \mathcal{R}_\varphi^{\text{use}} \cup \mathcal{R}_{\theta_1}^{\text{use}}$ and one can select λ_1 and λ_2 additionally to ensure $\breve{f}^\varphi \in \mathcal{F}_{\text{inf}}$. Picking such constants and noting the definition of $\mathcal{F}_{\text{inf}}$, we have $\breve{f}^\varphi \notin \mathcal{W}^{\theta_1}$, meaning that $\breve{f}^\varphi$ is not a φ^{θ_1}-WE. We will next show that φ is the unique solution to (5.8) where f^q is replaced with $\breve{f}^\varphi$ and q is treated as a variable to be solved for. Note that since $\breve{f}^\varphi$ is not a φ^{θ_1}-WE, there exist paths $p, r \in \mathcal{P}$ such that $\breve{f}_p^\varphi > 0$ and

$$C_p^{\theta_1}(\breve{f}^\varphi) > C_r^{\theta_1}(\breve{f}^\varphi). \tag{5.15}$$

Consider two cases: (a) $\breve{f}_r^\varphi > 0$ and (b) $\breve{f}_r^\varphi = 0$. For case (a), from (5.8), we obtain an equality constraint of the form

$$\begin{pmatrix} C_p^{\theta_1}(\breve{f}^\varphi) - C_r^{\theta_1}(\breve{f}^\varphi) & C_p^{\theta_2}(\breve{f}^\varphi) - C_r^{\theta_2}(\breve{f}^\varphi) \end{pmatrix} \varphi = 0.$$

Since $C_p^{\theta_1}(\breve{f}^\varphi) \neq C_r^{\theta_1}(\breve{f}^\varphi)$ this constraint along with $\varphi_1 + \varphi_2 = 1$ gives us two linearly independent equality constraints on φ. Since $\varphi \in \mathbb{R}^2$ this implies that φ is the only distribution that satisfies the constraints in (5.8). We next show that case (b), with $\breve{f}_r^\varphi = 0$, does not occur. To be precise, we claim that for $\breve{f}^\varphi \notin \mathcal{W}^{\theta_1}$, there exists at least one pair of paths $p, r \in \mathcal{P}$ satisfying (5.15) where both $\breve{f}_p > 0$ and $\breve{f}_r > 0$. To show this, we proceed with a contradiction argument. Assume there does not exist such a pair of paths. This implies two things: 1) $C_p^{\theta_1}(\breve{f}^\varphi) = C_r^{\theta_1}(\breve{f}^\varphi)$ for all $p, r \in \mathcal{P}$ such that $\breve{f}_p^\varphi > 0$ and $\breve{f}_r^\varphi > 0$; 2) if $C_p^{\theta_1}(\breve{f}^\varphi) > C_r^{\theta_1}(\breve{f}^\varphi)$ for some p such that $\breve{f}_p^\varphi > 0$ this implies $\breve{f}_r^\varphi = 0$. Now consider the graph $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ with $\mathcal{V}' = \mathcal{V}$ and $\mathcal{E}' \subseteq \mathcal{E}$ such that $e_k \in \mathcal{E}'$ if and only if $(\breve{f}_{\text{edge}}^\varphi)_k > 0$, where $\breve{f}_{\text{edge}}^\varphi$ is the vector of edge-flows associated to $\breve{f}^\varphi$. To all edges in $\mathcal{E}'$ associate the same state-dependent cost functions as in the original network, and consider the same set of states Θ. This defines a new routing game over the network $\mathcal{G}'$. Note that for any $p \in \mathcal{P}$ such that $\breve{f}_p^\varphi > 0$, we have by Lemma 5.2.3 that $(\breve{f}_{\text{edge}}^\varphi)_k > 0$ for all $e_k \in p$. Thus, when $\breve{f}_p^\varphi > 0$ and $e_k \in p$, then $e_k \in \mathcal{E}'$. Therefore, $\breve{f}_p^\varphi > 0$ implies $p \in \mathcal{P}'$. Thus, we can define a feasible flow for the modified game by setting $f_p'^\varphi := \breve{f}_p^\varphi$ for all p such that $\breve{f}_p^\varphi > 0$. Since cost functions over the used edges have not changed, if for two paths p and r we have $C_p^{\theta_1}(\breve{f}^\varphi) = C_r^{\theta_1}(\breve{f}^\varphi)$, then $C_p'^{\theta_1}(f'^\varphi) = C_r'^{\theta_1}(f'^\varphi)$. Now recall that if there was a path r in the original game such that $C_p^{\theta_1}(\breve{f}^\varphi) > C_r^{\theta_1}(\breve{f}^\varphi)$ for some p satisfying $\breve{f}_p^\varphi > 0$, then by 2) we have $\breve{f}_r^\varphi = 0$. Since $\breve{f}_p^\varphi > 0$ for all $p \in \mathcal{R}_\varphi^{\text{use}}$ by construction,

[3] Here we replace f^q in (5.8) with $\breve{f}^\varphi$ and treat q as a variable that can be solved for.

we deduce that $f_r = 0$ for all $f \in \mathcal{W}^{\varphi}$. Using the second implication of Lemma 5.2.3 we see that there exists some edge $e_k \in r$ such that $(\breve{f}^{\varphi}_{\text{edge}})_k = 0$. Therefore, the edge e_k has been removed in the modified game, and it follows that the path r is not present in the modified game. In conclusion, we have $C_p'^{\theta_1}(f'^{\varphi}) = C_r'^{\theta_1}(f'^{\varphi})$ for all $p, r \in \mathcal{P}'$ which implies that f'^{φ} is a φ^{θ_1}-WE for the modified game. Now consider any flow $f^{\theta_1} \in \mathcal{W}^{\theta_1}$. If $f_p^{\theta_1} > 0$, then $p \in \mathcal{R}^{\text{use}}_{\varphi^{\theta_1}}$, which implies $\breve{f}_p^{\varphi} > 0$. Repeating the above arguments then shows that $p \in \mathcal{P}'$. Thus, we can define a feasible flow for the modified game by setting $f_p''^{\theta_1} := f_p^{\theta_1}$. Similar to before we have that since f^{θ_1} is a φ^{θ_1}-WE of the original game this implies that f''^{θ_1} is a φ^{θ_1}-WE for the modified game. However, since $\breve{f}^{\varphi} \in \mathcal{F}_{\text{inf}}$, we have $\breve{f}^{\varphi} \notin \mathcal{W}^{\theta_1}$ which implies that $\breve{f}^{\varphi}_{\text{edge}} \neq f^{\theta_1}_{\text{edge}}$. This means we obtain two φ^{θ_1}-WE, namely f'^{φ} and f''^{θ_1} for the modified game with unequal edge-flows. This contradicts the uniqueness of edge-flow under φ^{θ_1}-WE. Thus we arrive at a contradiction. Therefore there do exist $p, r \in \mathcal{P}$ such that $\breve{f}_p^{\varphi} > 0$, $\breve{f}_r^{\varphi} > 0$, and (5.15) holds. Therefore, φ is uniquely determined by the constraints in (5.8). From Lemma 5.2.8, we have that for any $\breve{f}^{\varphi} \in \mathcal{W}^{\varphi}$ the set of priors satisfying the constraints imposed by (5.8) is the same, which then concludes the proof. □

Now we are ready to present the main result of this section. In it we make use of Lemma's 5.2.10 and 5.2.11 to design a signalling scheme for which all but one of the signals give an equality constraint on the prior, showing that there always exists a signalling scheme using m messages that is q-identifying.

Proposition 5.2.12. *(Existence of q-identifying signalling scheme): Let $\mathcal{P}$, Θ, $\mathcal{C}$, and q be given, and assume that $\mathcal{W}^{\theta_1} \neq \mathcal{W}^{\theta_2}$.*[4] *Then, there exists a signalling scheme $\Phi \in \mathsf{CS}(m, m)$ of m messages that is q-identifying.*

Proof. Our proof will be constructive. Recall the matrix notation of the signalling scheme, that is, $\Phi = (\phi_s^u)_{u,s\in[m]}$, where ϕ_s^u is the (u, s)-th entry of the matrix and denotes the probability of sending signal ζ^u under the state θ_s. We will proceed row-by-row starting from the second row of Φ.

Step 1: Constructing the second row: Set $\phi_s^2 = 0$ for all $s \in [m] \setminus \{1, 2\}$. Using (5.5) we obtain the posterior distribution under the message ζ^2 as

$$\begin{aligned} &\widetilde{q}_1^{\zeta^2} = \frac{\phi_1^2 q_1}{\phi_1^2 q_1 + \phi_2^2 q_2}, \quad \widetilde{q}_2^{\zeta^2} = \frac{\phi_2^2 q_2}{\phi_1^2 q_1 + \phi_2^2 q_2}, \\ &\widetilde{q}_s^{\zeta^2} = 0, \quad \text{for all } s \in [m] \setminus \{1, 2\}. \end{aligned} \tag{5.16}$$

Since q_1 and q_2 are non-zero by assumption, one can tune ϕ_1^2 and ϕ_2^2 to induce any posterior $\widetilde{q}^{\zeta^2}$ satisfying $0 < \widetilde{q}_1^{\zeta^2} < 1$ and $\widetilde{q}_1^{\zeta^2} = 1 - \widetilde{q}_2^{\zeta^2}$. By construction we then

[4]By relabelling the states we can see that this assumption is equivalent to assuming the existence of two states $\theta_k, \theta_\ell \in \Theta$ such that $\mathcal{W}^{\theta_k} \neq \mathcal{W}^{\theta_\ell}$.

have $\phi_1^2, \phi_2^2 > 0$. In the following, we will outline the procedure for tuning these parameters such that the flow induced by signal ζ^2 results in an equality constraint for the prior q.

Observe that when considering the signal ζ^2, we have simplified the situation by removing the influence from all but the first two states on the posterior (by setting $\phi_s^2 = 0$ for $s = 3, 4, \dots$). That is, we have effectively reduced the analysis to a two state case, as analysed in Lemma's 5.2.10 and 5.2.11. Consequently we can appeal to Lemma 5.2.10 to conclude that there exists a posterior $\widetilde{q}$, with $\widetilde{q}_1 \in (0,1)$, $\widetilde{q}_2 = 1 - \widetilde{q}_1$, and $\widetilde{q}_s = 0$ for all $s = [m] \setminus \{1,2\}$, such that there exists a $\widetilde{q}$-WE $f^{\widetilde{q}}$ satisfying $f^{\widetilde{q}} \in \mathcal{F}_{\text{inf}}$, where $\mathcal{F}_{\text{inf}}$ is given in (5.13). From Lemma 5.2.11 we know that if there exists a $\widetilde{q}$-WE that lies in $\mathcal{F}_{\text{inf}}$, then the constraints in (5.8) generated by any $\widetilde{q}$-WE allow for unique identification of $\widetilde{q}$.[5] Now pick ϕ_1^2 and ϕ_2^2 such that the posterior $\widetilde{q}$ with $f^{\widetilde{q}} \in \mathcal{F}_{\text{inf}}$ is induced under the signal ζ^2. Consequently, substituting $\widetilde{q} = \widetilde{q}^{\zeta^2}$ into (5.16) then gives the constraint

$$\frac{\phi_2^2}{\phi_1^2(1 - \widetilde{q}_1)} q_2 = q_1. \tag{5.17}$$

This constraint is well-posed and non-trivial since ϕ_2^2 and ϕ_1^2 are non-zero by design, and as noted $\widetilde{q}_1 = \widetilde{q}_1^{\zeta^2} < 1$. Thus, by tuning the values ϕ_1^2 and ϕ_2^2, we can find an equality constraint (5.17) on the prior.

Step 2: Constructing rows 3 through m: For row $s \notin \{1,2\}$ Lemma 5.2.7 and $\mathcal{W}^{\theta_1} \neq \mathcal{W}^{\theta_2}$ imply that we have either $\mathcal{W}^{\theta_s} = \mathcal{W}^{\theta_1}$, in which case $\mathcal{W}^{\theta_s} \neq \mathcal{W}^{\theta_2}$, or we have $\mathcal{W}^{\theta_s} \neq \mathcal{W}^{\theta_1}$. In other words, there exists a state θ_ℓ with $\ell < s$ such that $\mathcal{W}^{\theta_s} \neq \mathcal{W}^{\theta_\ell}$. By setting $\phi_{s'}^s = 0$ for all $s' \notin \{s, \ell\}$ we can, similar to before, induce any posterior $\widetilde{q} = \widetilde{q}^{\zeta^s}$ such that $\widetilde{q}_{s'} = 0$ for all $s' \notin \{s, \ell\}$ and so, $\widetilde{q}_s = 1 - \widetilde{q}_\ell$. We can then repeat the previous arguments to show that by tuning ϕ_s^s and ϕ_ℓ^s we can obtain a well-posed, non-trivial equality constraint on q of the form

$$\frac{\phi_s^s}{\phi_\ell^s(1 - \widetilde{q}_\ell)} q_s = q_\ell.$$

This equality constraint is necessarily linearly independent from the other equality constraints obtained in this manner. To see this note that the constraint generated by row s involves q_s, while the set of constraints generated by the rows $s' < s$ do not involve q_s by construction. Thus, in this manner we obtain $m-1$ linearly independent equality constraints q.

Step 3: Constructing the first row: Once we have constructed the rows 2 through m of Φ, we select the elements of the first row such that each column of Φ sums to one. This is always possible and a short procedure is given in Algorithm 1.

[5] That is, it allows us to determine $\widetilde{q}_1$, and $\widetilde{q}_2$. Since we already know that $\widetilde{q}_s = 0$ for all $s \notin \{1,2\}$ this fully identifies $\widetilde{q}$.

Finally, note that with the $m-1$ linearly independent equality constraints and the additional independent constraint $\mathbb{1}^\top q = 1$ derived from the condition that $q \in \Delta_1^m$ lies in the simplex, we obtain m constraints that uniquely identify q. $\square$

Remark 5.2.13. *(Drawbacks of signals limited to two states):* In the proof of Proposition 5.2.12 we make use of a specific kind of signalling scheme in which each signal except the first has a positive chance of being sent only in two states, and the first signal is used to ensure that the signalling scheme satisfies all the required constraints. Mathematically, the signalling scheme belongs to the set

$$\begin{aligned}\mathcal{S}_{\text{sig}} := \Big\{ &\Phi \in \mathsf{CS}(m,m) \mid \forall\, s \in [m] \setminus \{1\},\ \exists \ell < s \\ &\text{such that } \mathcal{W}^{\theta_s} \neq \mathcal{W}^{\theta_\ell}, \text{and } \phi^s_{s'} = 0 \Leftrightarrow s' \notin \{s, \ell\} \Big\}.\end{aligned} \tag{5.18}$$

Such a scheme is used because for each signal, as mentioned, the situation is effectively reduced to a two state case, allowing for simpler analysis. However, such a scheme is limited in that it can derive at most one equality constraint from a signal. If a signal can be sent in more than two states, more information may be gained. Analysis however becomes more difficult, since it is not clear if and how the result of Lemma 5.2.10 can be generalized to a case involving more than two states. •

5.2.3 Designing the signalling scheme

With the existence of a q-identifying signalling scheme guaranteed under mild conditions, the next step would be to give guidelines for how such a scheme can be designed. For this purpose we provide Algorithm 1, which using observations of the flow under various signals, updates a signalling scheme until it is q-identifying. The algorithm uses signalling schemes in the set (5.18), and requires the assumption of Proposition 5.2.12 that the sets $\mathcal{W}^{\theta_1}$ and $\mathcal{W}^{\theta_2}$ are not equal.

> *[Informal description of Algorithm 1]:* The procedure starts with an initial $\Phi(0)$ of the form (5.18), such that for each signal $s \in [m] \setminus \{1\}$ exactly two elements in the s-th row of $\Phi(0)$ are non-zero. One of these elements is ϕ^s_s and the other is denoted $\phi^s_{\ell(s)}$ (cf. Line 1). At each iteration N, and for each row $s \in \mathcal{I}$, we check whether the flow $\widetilde{f}^{\zeta^s}$ observed under the signalling scheme $\Phi(N)$ when sending signal ζ^s results in an equality constraint on q (cf. Line 8). If it does, then row s of $\Phi(N)$ is not updated in the for-loop and the ratio between ϕ^s_s and $\phi^k_{\ell(s)}$ remains the same for all subsequent iterations (cf. Lines 4 and 9). If not, then we consider two cases. In the first case, the flow $\widetilde{f}^{\zeta^s} \in \mathcal{W}^{\theta_s}$ and the values ϕ^s_s and $\phi^s_{\ell(s)}$ are updated so as to increase the ratio $\phi^s_{\ell(s)}/\phi^s_s$ in signalling scheme $\Phi(N+1)$.

In this way, the posterior induced by signal s in the next iteration will assign less probability to state θ_s. This increase in ratio is achieved in Lines 11 through 15. In the second case, $\widetilde{f}^{\zeta^s} \notin \mathcal{W}^{\theta_s}$ and we decrease the ratio $\phi^s_{\ell(s)}/\phi^s_s$ in Line 18. After modifying rows in this way, the signalling scheme $\Phi(N+1)$ is updated in Lines 21-24 so as to ensure that each column sums to unity while preserving the ratios $\phi^s_{\ell(s)}/\phi^s_s$.

The above procedure identifies the right signalling scheme, and can also determine the prior, since the obtained constraints define it uniquely. Next we establish the correctness of Algorithm 1.

Proposition 5.2.14. *(Convergence of Algorithm 1): Let $\mathcal{P}$, Θ, $\mathcal{C}$, and q be given, and assume that $\mathcal{W}^{\theta_1} \neq \mathcal{W}^{\theta_2}$. Then, Algorithm 1 terminates in a finite number of iterations N_f, and the resulting signalling scheme $\Phi(N_f)$ is q-identifying.*

Proof. For a signal $s \in [m] \setminus \{1\}$, we look at the properties of $\widetilde{f}^{\zeta^s}$, $\texttt{low}_s(N)$, $\texttt{up}_s(N)$, and $r_s(N) := \frac{\phi^s_{\ell(s)}}{\phi^s_s}$ as the algorithm iterates. We first show that $\texttt{low}_s(N) \leq \texttt{up}_s(N)$ for all N, which holds by definition for the initial iterate. We suppress the argument N in the following few statements for the sake of convenience. Observe that the signalling scheme maintains the same sparsity pattern, of the form (5.18), in all iterations. That is, $\phi^s_i = 0$ for all $i \notin \{\ell(s), s\}$ and all iterations. This effectively reduces the analysis to that of a two-state situation, meaning that the posterior under signal ζ^s, denoted $\widetilde{q} = \widetilde{q}^{\zeta^s}$, satisfies $\widetilde{q}_i = 0$ for all $i \notin \{\ell(s), s\}$ and any choice of $\phi^s_{\ell(s)}$, ϕ^s_s. From Lemma 5.2.10, there exist constants $a_s \in (0, 1]$ and $c_s \in (0, 1)$ with $a_s > c_s$ such that

$$\begin{aligned} \widetilde{q}_s \geq a_s &\Rightarrow \mathcal{W}^{\widetilde{q}} \cap \mathcal{W}^{\theta_s} = \emptyset, \\ a_s > \widetilde{q}_s > c_s &\Rightarrow \exists f^{\widetilde{q}} \in \mathcal{W}^{\widetilde{q}} \text{ such that } f^{\widetilde{q}} \in \mathcal{F}_{\text{inf}}.^6 \end{aligned} \tag{5.19}$$

From (5.5), we have

$$\widetilde{q}_s = \frac{\phi^s_s q_s}{\phi^s_s q_s + \phi^s_{\ell(s)} q_{\ell(s)}} = \frac{q_s}{q_s + \frac{\phi^s_{\ell(s)}}{\phi^s_s} q_{\ell(s)}}.$$

Note that the influence of Φ on $\widetilde{q}_s$ is completely determined by the ratio $r_s = \frac{\phi^s_{\ell(s)}}{\phi^s_s}$ and that $\widetilde{q}_s$ is monotonically decreasing in r_s with $\lim_{r_s \to \infty} \widetilde{q}_s = 0$ and $\lim_{r_s \to 0} \widetilde{q}_s = 1$. Thus, given (5.19), and the relationship between $\widetilde{q}_s$ and r_s, we deduce that there exist constants $b_s \geq 0$ and $d_s > b_s$ such that

$$\begin{aligned} r_s \leq b_s &\Rightarrow f^{\widetilde{q}} \in \mathcal{W}^{\theta_s}, \\ d_s > r_s > b_s &\Rightarrow f^{\widetilde{q}} \in \mathcal{F}_{\text{inf}}. \end{aligned} \tag{5.20}$$

[6] Here q and f^q in (5.8) are replaced with $\widetilde{q}$ and $f^{\widetilde{q}}$ respectively.

[7] Here, instead of considering all $f \in \mathcal{H}$ we only consider flows f for which there exists a $\varphi \in \{\xi \in \Delta^m_1 \mid \xi_{s'} = 0, \text{ for } s' \notin \{\ell(s), s\}\} \setminus \{\varphi^{\theta_s}\}_{s \in [m]}$ such that $f \in \mathcal{W}^{\varphi}$ in the definition of $\mathcal{F}_{\text{inf}}$.

Algorithm 1: Find q-identifying Signalling Scheme

Initialize: An index set $\mathcal{I} = \{2, 3, \cdots, m\}$, counter $N = 0$, lower and upper bounds $\texttt{low}_s(0) = 0$, $\texttt{up}_s(0) = \infty$ for all $s \in \mathcal{I}$, a signalling scheme $\Phi(0) \in \mathcal{S}_{\text{sig}}$ using (5.18)

For all $s \in \mathcal{I}$ set $\ell(s) \neq s$ such that $\phi^s_{\ell(s)}(0) \neq 0$
Compute f^{θ_s} for all $s \in [m]$ by solving $\text{VI}(\mathcal{H}, C^{\theta_s})$
while $\mathcal{I} \neq \emptyset$ **do**
    Set $\phi^j_i(N+1) \leftarrow \phi^j_i(N)$ for all $i \in [m]$ and $j \in [m] \setminus \mathcal{I}$
    for $s \in \mathcal{I}$ **do**
        Obtain $\widetilde{f}^{\zeta^s}$ under scheme $\Phi(N)$ and signal ζ^s
        Check if $\widetilde{q}^{\zeta^s}$ is uniquely determined by (5.8) [7]
        if $\widetilde{q}^{\zeta^s}$ is uniquely identified **then**
            Set $\mathcal{I} \leftarrow \mathcal{I} \setminus \{s\}$
        else if $\widetilde{f}^{\zeta^s} \in \mathcal{W}^{\theta_s}$ **then**
            Set $\texttt{low}_s(N+1) \leftarrow \frac{\phi^s_{\ell(s)}(N)}{\phi^s_s(N)}$, $\texttt{up}_s(N+1) \leftarrow \texttt{up}_s(N)$, and $\phi^s_s(N+1) \leftarrow \phi^s_s(N)$
            if $\texttt{up}_s(N+1) = \infty$ **then**
                Set $\phi^s_{\ell(s)}(N+1) \leftarrow 2\phi^s_{\ell(s)}(N)$
            else
                Set $\phi^s_{\ell(s)}(N+1) \leftarrow \frac{1}{2}\big(\texttt{low}_s(N+1) + \texttt{up}_s(N+1)\big)\phi^s_s(N)$
            end
        else
            Set $\texttt{up}_s(N+1) \leftarrow \frac{\phi^s_{\ell(s)}(N)}{\phi^s_s(N)}$, $\texttt{low}_s(N+1) \leftarrow \texttt{low}_s(N)$, and $\phi^s_s(N+1) \leftarrow \phi^s_s(N)$ $\phi^s_{\ell(s)}(N+1) \leftarrow \frac{1}{2}\big(\texttt{low}_s(N+1) + \texttt{up}_s(N+1)\big)\phi^s_s(N)$
        end
    end
    Set $a = \max_{i \in [m]} \sum_{j \in [m] \setminus \{1\}} \phi^j_i(N+1)$.
    Set $\Phi(N+1) \leftarrow \frac{1}{a}\Phi(N+1)$
    for $s \in [m]$ **do**
        Set $\phi^1_s(N+1) \leftarrow 1 - \sum_{j \in [m] \setminus \{1\}} \phi^j_s(N+1)$
    end
    Set $N \leftarrow N + 1$
end

With this in mind, we now analyse the evolution of $\texttt{low}_s$ and $\texttt{up}_s$. Note that $\texttt{low}_s$ is only changed in line 11 of the algorithm. Here we set $\texttt{low}_s(N) = r_s(N)$ whenever $f^{\widetilde{q}} \in \mathcal{W}^{\theta_s}$ and thus from (5.20), $\texttt{low}_s(N) \leq b_k$ for all N. Similarly, $\texttt{up}_s(N) = r_s(N)$ whenever $f^{\widetilde{q}} \notin \mathcal{W}^{\theta_s}$ and $\widetilde{q}^{\zeta^s}$ is not uniquely identified. As shown in the proof of Proposition 5.2.12, whenever $f^{\widetilde{q}} \in \mathcal{F}_{\text{inf}}$ we obtain an informative equality constraint. From (5.20) we then conclude that $\texttt{up}_s(N) \geq d_s$. We now have $\texttt{low}_s(N) < \texttt{up}_s(N)$ for all N. In fact, we have $(b_s, d_s) \subseteq \big(\texttt{low}_s(N), \texttt{up}_s(N)\big)$ for all N. We also note that since $d_s > b_s$ we have $d_s > 0$ and $b_s < \infty$.

Now we look at the evolution of $r_s(N)$. We will show that $r_s(N) \in (b_s, d_s)$ for some finite N and at that iteration, we obtain an informative equality constraint corresponding to signal s. This in turn proves the termination of the algorithm in a finite number of iterations. Consider three cases: (a) $d_s = \infty$; (b) $b_s = 0$ and $d_s < \infty$; and (c) otherwise. In case (a), at any N, we have either $r_s(N) \in (b_s, d_s)$ and we find an informative equality constraint, or $r_s(N) \leq b_s$, implying $f^{\widetilde{q}} = f^{\theta_s}$. In the latter case, $r_s(N+1) = 2r_s(N)$. Thus, there exists some $\bar{N}$ such that $r_s(\bar{N}) > b_s$, implying $r_s(\bar{N}) \in (b_s, d_s)$. Similarly, in case (b), we have either $r_s(N) \in (b_s, d_s)$ or $r_s(N) \geq d_s$. In the latter case, $r_s(N)$ is halved for the next iteration and so in a finite number of steps r_s reaches (b_s, d_s). In case (c), the arguments for case (a) and (b) can be repeated to show that there exists $\bar{N}$ such that $\texttt{low}_s(\bar{N}) > 0$ and $\texttt{up}_s(\bar{N}) < \infty$. Looking at the algorithm, we see that the quantity $\texttt{up}_s(N) - \texttt{low}_s(N)$ is halved in every subsequent iteration $N \geq \bar{N}$. Since $r_s(N)$ always belongs to the interval $\big(\texttt{low}_s(N), \texttt{up}_s(N)\big)$, it then reaches the set (b_s, d_s) in a finite number of iterations yielding an informative equality constraint. Following these facts, we conclude that an informative equality constraint is found in a finite number of iterations for each signal which completes the proof. □

Remark 5.2.15. *(Practical considerations of implementing Algorithm 1):* The purpose of Algorithm 1 is to demonstrate how insights from Proposition 5.2.12 can be applied. It gives a methodical approach for constructing a q-identifying signalling scheme. However, it has several drawbacks worth noting:

1) The TIS can only send one signal at any instance which is dependent on the observed state. Therefore, in practice, the TIS cannot send all signals in an ordered manner at each iteration of the algorithm and then update Φ. Instead, it would be best to update a row of Φ after each instance of a game when the used signal does not induce a useful equality constraint. We have presented the algorithm in its current form, rather than the practically implementable one, to simplify the exposition.

2) When additional information on the prior is available, such as a lower bound $q_s \geq \epsilon > 0$ which holds for all $s \in [m]$, it may be possible to determine in advance which signalling scheme will supply informative constraints on the prior. For instance

looking at Figure 5.3, we see that whenever $\varphi_1 \in (0.133, 0.8)$ the result is a WE belonging uniquely to the associated distribution. If we then have, for instance, $\varphi_1, \varphi_2 \geq 0.25$ it follows that for this example an uninformative scheme (with $\phi_s^u = 0.5$ for all $u, s \in [2]$) is q-identifying.

3) As mentioned in Remark 5.2.13, it may be beneficial to allow a signal to be sent in more than two states, in order to obtain multiple equality constraints from a single signal. This may significantly reduce the number of iterations required to identify the prior, especially in combination with the above mentioned possibility of using additional knowledge about the prior to determine a signalling scheme in advance that necessarily provides informative constraints.

4) Finally, we have only considered the question of identifying the prior. In practice, the social cost incurred during identification is also important. For instance, once a signal ζ^u has resulted in an equality constraint on q_ℓ and q_s, that specific signal is no longer required for identification and can be modified with the aim of minimizing the social cost. However, the comparison between the benefits of obtaining a better estimate of the prior and optimizing with respect to the current estimate is more involved and left for future work. •

Example 5.2.16. *(Application of Algorithm 1 in 4-path 2-state case):* To shed light on the conclusions of Proposition 5.2.12 and the workings of Algorithm 1, we revisit the 4-path, 2-state case in Example 5.2.5. Setting $q = (0.5, 0.5)^\top$, and using the initial signalling scheme

$$\Phi(0) = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}.$$

Using $\Phi(0)$ as a signalling scheme, (5.5) gives us $\widetilde{q}^{\zeta^2} = \widetilde{q} = (0.5, 0.5)^\top$. We can then use the functions $C_p^{\widetilde{q}}(f)$ and (5.3) to find that $f^{\widetilde{q}} = (0, \frac{5}{9}, \frac{4}{9}, 0)^\top$ is the flow observed after sending signal ζ^2. Even though two paths carry positive flow, the resulting constraint is trivial, since $C_2^{\theta_2}(\frac{5}{9}) - C_3^{\theta_2}(\frac{4}{9}) = 0$. In Figure 5.4 this can also be observed by noting that $\widetilde{q} = (0.5, 0.5)^\top$ is in a region of distributions where the flow remains constant. We do have $f^{\widetilde{q}} \neq f^{\theta_2}$ which means that we will update $\Phi(0)$ according to Line 18. Setting the values as prescribed there, we get $\phi_1^2(1) = 0.25$, $\phi_2^2(1) = 0.5$. In Lines 21-24 we then update the first row to ensure that all columns of $\Phi(1)$ sum to one, and thus we arrive at

$$\Phi(1) = \begin{pmatrix} 0.75 & 0.5 \\ 0.25 & 0.5 \end{pmatrix}.$$

Using the new signalling scheme we find $\widetilde{q}^{\zeta^2} = \widetilde{q} = (\frac{1}{3}, \frac{2}{3})^\top$, resulting in the equilibrium $f^{\widetilde{q}} = (0, \frac{32}{68}, \frac{23}{68}, \frac{13}{68})^\top$. Substituting $\phi_1^2(1)$, $\phi_2^2(1)$ and $f^{\widetilde{q}}$ into (5.9), where we set

$p = 2$, $r = 4$, we get

$$\begin{pmatrix} \frac{1}{4}(\frac{32}{68}\frac{1}{2} + 1.7 - \frac{2}{5}\frac{13}{68} - 3.5) \\ \frac{1}{2}(\frac{32}{68}\frac{1}{2} + 1.7 - \frac{3}{5}\frac{13}{68} - 1) \end{pmatrix}^{\top} q = 0.$$

' Solving this we find $q_1 = q_2$. Taken together with $q_1 + q_2 = 1$ this implies that $q = (0.5, 0.5)^{\top}$. Thus, the q-identifying scheme exists and is obtained in one iteration of the algorithm. •

5.3 Multiple priors and robust identification

We finish this chapter with a discussion about possible generalizations of our setup that can bring it closer to real-life implementation. First we discuss the case where the population does not have a common prior, but is instead heterogeneous, meaning that it is divided into groups, each adhering to their own prior, and later we show how the signalling schemes that we obtain have some robustness with respect to perturbations in the prior p.

5.3.1 Heterogeneous population

Consider the case where the population of users traversing the network are divided into K groups, each containing users that share a common prior. In particular, assume that $c^k \in (0, 1]$ is the fraction of users sharing the prior $q[k] \in \Delta_1^m$ and we have $\sum_{k=1}^{K} c^k = 1$. We assume that each group $k \in [K]$ uses the same set of available paths. Note that we considered $K = 1$ in the earlier sections. After a public signal ζ^u is received, each group k routes its fraction of the flow according to the $\widetilde{q}^{\zeta^u}[k]$-WE, where $\widetilde{q}^{\zeta^u}[k]$ is the posterior formed by group k under a signal ζ^u and some signalling scheme Φ. The aggregated flow observed by the TIS is

$$\widetilde{f}^{\zeta^u} := \sum_{k \in [K]} c^k \widetilde{f}^{u,k}, \tag{5.21}$$

where $\widetilde{f}^{u,k}$ is a $\widetilde{q}^{\zeta^u}[k]$-WE.

First, we note that for the case $K = 2$, where c^1, c^2 and $q[1]$ are known, then identification of $q[2]$ can be achieved by following Algorithm 1. This is so because for each signal we observe $\widetilde{f}^{\zeta^u}$ while we know $\widetilde{f}^{u,1}$. Thus, following (5.21), one gets $\widetilde{f}^{u,2} = \frac{\widetilde{f}^{\zeta^u} - c^1 \widetilde{f}^{u,1}}{c^2}$. Identification of $q[2]$ can then be done using Algorithm 1 by perceiving the second group as the only one being routed. Next examine the case where more than one prior is unknown. Here, even when the fractions c^1 and c^2 are known, it is not clear how to design a signalling scheme that can identify both priors. The reason being that now we have an additional m unknowns as compared to the case of single prior, while the amount of information that can be obtained from a signalling scheme does not grow.

Finally, consider the case where all priors $\{q[k]\}$ are known, but the fractions $\{c^k\}$ are not. Here, for a given signalling scheme $\Phi \in \mathsf{CS}(z, m)$, we define the following matrix:

$$M := \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \widetilde{f}^{1,1} & \widetilde{f}^{1,2} & \dots & \widetilde{f}^{1,K} \\ \widetilde{f}^{2,1} & \ddots & & \vdots \\ \vdots & & & \\ \widetilde{f}^{z,1} & \widetilde{f}^{z,2} & \dots & \widetilde{f}^{z,K} \end{pmatrix} \tag{5.22}$$

and present the following result.

Lemma 5.3.1. *(Identifying population size per prior): Let $\mathcal{P}$, Θ, $\mathcal{C}$ be given, together with pairs of fractions and priors $\{(c^k, q^k)\}_{k\in[K]}$, $K \in \mathbb{N}$ satisfying $c^k > 0$ for all $k \in [K]$ and $q^k \neq q^\ell$ for all $k \neq \ell$. A signalling scheme $\Phi \in \mathsf{CS}(z, m)$ allows us to uniquely identify the vector $c := (c^1, c^2, \cdots, c^K)^\top$ if and only if* $\operatorname{rank}(M) = K$.

Proof. We know that c must satisfy $\mathbb{1}^\top c = 1$, since the fractions sum up to the whole of the population. This, together with (5.21) and (5.22) implies that c must satisfy

$$Mc = \left(1,\ \widetilde{f}^{\zeta^1},\ \widetilde{f}^{\zeta^2},\ \cdots, \widetilde{f}^{\zeta^z}\right)^\top. \tag{5.23}$$

When $\operatorname{rank}(M) = K$, that is, M has full column rank, the above equation has a unique solution. If on the other hand $\operatorname{rank}(M) < K$, then the equality (5.23) still holds. However, in this case there also exists $\widetilde{c} \in \mathbb{R}^K$ such that $M\widetilde{c} = 0$ and $\widetilde{c} \neq 0$. Since $c > 0$, there exists $\epsilon > 0$ such that $c + \epsilon\widetilde{c} \geq 0$. We then have $M(c + \epsilon\widetilde{c}) = Mc$, which implies $c + \epsilon\widetilde{c}$ is in Δ_1^K and is a solution to (5.23). In other words, there exist multiple solutions to (5.23) in Δ_1^K. □

In general it is difficult to prescribe guidelines on how to design Φ in order to ensure that M has full row rank. However, when $z \geq K$ and flows $\{f^{\theta_s}\}_{s\in[k]}$ are linearly independent, one can design the signal ζ^k such that $\widetilde{q}^{\zeta^k}[\ell]$ is arbitrarily close to q^{θ_k}. In this way, the induced WE $\widetilde{f}^{u,\ell}$ will get arbitrarily close to f^{θ_k} for all ℓ. Since flows $\{f^{\theta_s}\}_{s\in[K]}$ are linearly independent, this will result in M having full column rank. Also note that when considering $K = 2$, all that is required is that there exist $k, \ell \in [K]$ and a $u \in [z]$ such that $\widetilde{f}^{u,k} \neq \widetilde{f}^{u,\ell}$.

5.3.2 Robustness of signalling schemes in identifying priors

One of the limitations of our results is that we consider the prior distribution that the population adheres to as fixed. However, we have the following robustness result on q-identifying signalling schemes with respect to perturbations in the prior, which shows that for a q-identifying signalling scheme Φ for which the obtained equality constraints are enough to identify q, there exists a neighbourhood of q such that for all priors $\breve{q}$ in this neighbourhood Φ is $\breve{q}$-identifying.

Lemma 5.3.2. *(Robustness of Φ for identifying prior): Let $\mathcal{P}$, Θ, $\mathcal{C}$, a prior q, and a q-identifying signalling scheme Φ be given. In addition, let $\mathcal{Q}_\Phi^=$ be defined as*

$$\mathcal{Q}_\Phi^= := \left\{ \varphi \in \mathbb{R}^m \;\middle|\; \begin{array}{c} \mathbb{1}^\top \varphi = 1, \\ \sum_{s\in[m]} \phi_s^u \big(C_p^{\theta_s}(\widetilde{f}^{\zeta^u}) - C_r^{\theta_s}(\widetilde{f}^{\zeta^u}) \big) \varphi_s = 0 \\ \forall u \in [z], \;\; p, r \in \mathcal{P} \text{ with } \widetilde{f}_p^{\zeta^u}, \widetilde{f}_r^{\zeta^u} > 0. \end{array} \right\}. \tag{5.24}$$

If $\mathcal{Q}_\Phi^= = \{q\}$, then there exists a $\delta > 0$ such that for all $\breve{q} \in \Delta_1^m$ with $\|q - \breve{q}\| < \delta$ the signalling scheme Φ is $\breve{q}$-identifying.

Proof. First, we note that as a consequence of Corollary 5.2.9, the set $\mathcal{Q}_\Phi$ is independent of which WE $\widetilde{f}^{\zeta^u} \in \mathcal{W}^{\zeta^u}$ are observed. Thus, to show that Φ is ξ-identifying for $\xi \in \Delta_1^m$, it is enough to show that there exists a set $\{\widetilde{f}^{\zeta^u}\}_{u\in[z]}$ of $\widetilde{\xi}^{\zeta^u}$-WE such that the obtained constraints identify ξ. Now consider $q \in \Delta_1^m$ and the signalling scheme Φ which by assumption is q-identifying. Since $\mathcal{Q}_\Phi^= = \{q\}$, it follows from (5.24) that there exist $m-1$ triplets $\{(p_i, r_i, u_i)\}_{i\in[m-1]}$ with $p_i, r_i \in \mathcal{P}$ and $u_i \in [m]$ such that $\widetilde{f}_{p_i}^{\zeta^{u_i}}, \widetilde{f}_{r_i}^{\zeta^{u_i}} > 0$ and the system of equations

$$Q(q)\varphi := \begin{pmatrix} \mathbb{1}^\top \\ \big(\alpha(q, p_1, r_1, u_1)\big)^\top \\ \vdots \\ \big(\alpha(q, p_{m-1}, r_{m-1}, u_{m-1})\big)^\top \end{pmatrix} \varphi = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{5.25}$$

has one solution $\varphi = q$. Here $\alpha(q, p_i, r_i, u_i) \in \mathbb{R}^m$ is given by

$$\alpha(q, p_i, r_i, u_i) := \begin{pmatrix} \phi_1^u \big(C_{p_i}^{\theta_1}(\widetilde{f}^{\zeta^{u_i}}) - C_{r_i}^{\theta_1}(\widetilde{f}^{\zeta^{u_i}}) \big) \\ \phi_2^u \big(C_{p_i}^{\theta_2}(\widetilde{f}^{\zeta^{u_i}}) - C_{r_i}^{\theta_2}(\widetilde{f}^{\zeta^{u_i}}) \big) \\ \vdots \\ \phi_m^u \big(C_{p_i}^{\theta_m}(\widetilde{f}^{\zeta^{u_i}}) - C_{r_i}^{\theta_m}(\widetilde{f}^{\zeta^{u_i}}) \big) \end{pmatrix},$$

where the dependence on q is via the dependence of $\widetilde{f}^{\zeta^{u_i}}$ on the posterior $\widetilde{q}^{\zeta^{u_i}}$. Note that from (5.25), the matrix $Q(q)$ has full rank, that is, $\text{rank}\big(Q(q)\big) = m$. From

Lemma 5.2.2, since $\widetilde{f}^{\zeta^{u_i}}_{p_i}, \widetilde{f}^{\zeta^{u_i}}_{r_i} > 0$, there exists a $\delta_f > 0$ such that $\|q - \breve{q}\| < \delta_f$ implies that for posteriors $\breve{q}^{\zeta^{u_i}}$ (based on the prior $\breve{q}$), there exist $\breve{q}^{\zeta^{u_i}}$-WE, denoted $\breve{f}^{\zeta^{u_i}}$, satisfying $\breve{f}^{\zeta^{u_i}}_{p_i}, \breve{f}^{\zeta^{u_i}}_{r_i} > 0$. That is, positive flow on p_i and r_i under a WE formed using signal ζ^{u_i} under scheme Φ and prior q implies that the same paths will have positive flow for some WE under the same signal ζ^{u_i} and scheme Φ but induced by a prior $\breve{q}$ that is close enough to q. This fact along with (5.1), Lemma 5.2.2, the continuity of functions $\{C_{e_k}(\cdot)\}$, and the fact that the flow over an edge is equal to the sum of the flow over the paths containing that edge implies that the entries of the matrix $Q(q)$ change continuously with respect to q. That is, there exists a $\delta_Q > 0$ such that $\|q - \breve{q}\| < \delta_Q$ implies $\text{rank}\big(Q(\breve{q})\big) = m$. [8] Thus, the linear system of equations $Q(\breve{q})\varphi = (1\ 0\ \cdots\ 0)^\top$ has a unique solution which is necessarily $\breve{q}$. That is, we have $\mathcal{Q}_\Phi = \{\breve{q}\}$. As we mentioned before, even though we use here that for specific $\breve{q}^{\zeta^{u_i}}$-WE we obtain $\breve{f}^{\zeta^{u_i}}_{p_i}, \breve{f}^{\zeta^{u_i}}_{r_i} > 0$, which supplies us with the required equality constraints, the set Q_Φ is independent of which specific $\breve{q}^{\zeta^{u_i}}$-WE is observed. The result follows. □

We note that the signalling schemes produced by Algorithm 1 are of the type considered in the above result. That is, Algorithm 1 produces signalling schemes for which the resulting equality constraints are enough to identify q. What is more, since for these schemes each signal (except the first) supplies one independent equality constraint relating two elements q_k and q_ℓ of the prior, each signal can be analysed separately to find the region of priors for which it is guaranteed to still supply an equality constraint. For example, let Φ be a signalling scheme for a given instance of the game such that under the signal ζ^2 we have $\widetilde{q}_s = 0$ for all $s \geq 3$ and let the remaining dependency of the WE on the posterior $\widetilde{q}$ be given in Figure 5.3, where $\widetilde{q}_1 = \varphi_1$. If $\widetilde{q}_1 \in (0.133, 0.8)$ the resulting flow gives us an equality constraint on q. Additionally, for any perturbation $\breve{q}$ of q, the signalling scheme Φ will still provide an equality constraint on $\breve{q}$ as long as the induced change in posterior does not take it outside of the set $(0.133, 0.8)$. Furthermore, once q has been identified, the scheme can be modified so as to ensure $\widetilde{q}_1$ is in the centre of the interval $(0.133, 0.8)$ thereby increasing the robustness of this signal for identification purposes.

5.4 Conclusions

In this chapter we have studied the problem of identifying the unknown prior belief that a population holds on the state of the network in a routing game, on the

[8] To see this, take a square, non-singular submatrix of $Q(q)$ and note that the determinant depends continuously on the coefficients of $Q(q)$. It follows that for small enough perturbations, the determinant of the submatrix does not become 0, and thus $Q(q)$ retains full column rank.

basis of observations of Wardrop equilibria. First we provided an analysis of how observations of WE induce constraints on the underlying prior, and how a signalling scheme can be used to obtain such constraints. We have shown that under very mild conditions there always exists a signalling scheme with as many signals as there are states, that allows for full identification of the prior. Using the insights gained from proving this existence we have also provided a way in which such a signalling scheme can be designed by iteratively updating an initial scheme based on observed WE. We finished the chapter with a discussion on the challenges and potential of generalizing presented results to a scenario in which not all participants adhere to the same prior, and by highlighting how the signalling schemes designed by the given procedure are robust in that they can still identify the prior after a small perturbation.

Chapter 6

CVaR-Based variational inequalities

A useful tool in the field of game theory is the *variational inequality*(VI) problem [16]. For example, the set of *Nash equilibria* of a game, or the *Wardrop equilibria* of a routing game are under mild conditions given by the set of solutions of a VI, which is often relatively simple to obtain. The specifics of the to be solved VI depend on the specifics of the game; in particular, the feasible set of the considered game and the involved cost functions. Of course in real-life, these cost functions which the participants of the game wish to optimize can be subject to uncertainty, and faced with such randomness, the risk-preferences of the participants are an important factor in the decision making process. Instead of optimizing a certain cost, or optimizing the expected value of an uncertain cost, participants can choose to optimize a particular risk-measure of the uncertain costs. Consequently finding the equilibria of such a game involves solving a VI defined by risk-measures of uncertain costs. Motivated by this setup, we consider in this chapter VIs defined by the *conditional value-at-risk*(CVaR) of random costs, and develop some *stochastic approximation*(SA) schemes for solving these VIs.

Literature review

The most popular way of incorporating uncertainty in VIs is to formulate a *stochastic variational inequality* (SVI) problem, see e.g. [58] and references therein. Here, the map associated with the VI is the expectation of a random function. SA methods for solving SVI are well studied [58, 59]. A key feature of such schemes is the availability of an unbiased estimator of the map using any number of samples of the uncertainty. This leads to strong convergence guarantees under a mild set of assumptions. However, the empirical estimator of CVaR, while being consistent, is biased [60]. Therefore, depending on the required level of precision, more samples are required to estimate the CVaR. This biasedness poses challenges in the convergence analysis of SA schemes. For a general discussion on risk-based VIs, including CVaR, and their potential applications, see [61]. In [62], a sample average approximation method for estimating the solution of CVaR-based VI was discussed. Our work also broadly relates to [63] and [64] where sample-based methods are used for optimizing the CVaR and other risk measures, respectively. The convergence analysis of our iterative methods consists of approximating the asymptotic behaviour of iterates with a trajectory of a continuous-time dynamical system and studying their stability. See [65] and [66] for a comprehensive account of such analysis.

Organization

After introducing some preliminaries, the starting point of this chapter is the definition of the CVaR-based variational inequality (VI), where the map defining the VI consists of components that are the CVaR of random functions. We motivate this setup with two examples of non-cooperative games for which CVaR-based VIs are relevant. We then introduce and analyse three different *stochastic approximation*(SA) algorithms for solving these VIs. The first scheme we term the *projected method*. This iterative method consists of moving along the empirical estimate of the map defining the VI and projecting each iterate onto the feasibility set. We show that under strict monotonicity, the projected algorithm asymptotically converges to any arbitrary neighbourhood of the solution of the VI, where the size of the neighbourhood is determined by the number of samples used to form the empirical estimate in each iteration. The second scheme, which we call the *subspace-constrained method*, overcomes the computational burden of calculating projections onto the feasibility set by dealing with equality and inequalities differently. In particular, the proximity to satisfying inequality constraints is ensured using penalty functions and iterates are constrained to lie on the subspace generated by linear equality constraints by pre-multiplying the iteration step by an appropriate matrix. We establish that under strict monotonicity, the algorithm converges asymptotically to any neighbourhood of the solution of the VI. In the third scheme, which we call the *multiplier-driven* method, projections are discarded altogether by introducing a multiplier for the inequality constraints. Satisfaction of equality constraints is guaranteed in the same way as for the subspace-constrained method by using matrix pre-multiplication. The iterates are shown to converge asymptotically under strict monotonicity to any neighbourhood of the solution of the VI. With the convergence of all three methods established, we also supply a result that directly relates the accuracy of the convergence to number of samples required to achieve that level of precision. Finally, we demonstrate the behaviour of the algorithms using a network routing example.

6.1 Preliminaries

In this section we take the time to introduce some definitions and preliminary results concerning the subjects of variational inequalities, conditional value-at-risk and projected dynamical systems which are required for the analysis in the rest of this chapter.

Variational inequalities and KKT points

We recall the definition of a variational inequality problem, as given in Chapter 2:

Definition 6.1.1. *(Variational inequalities (VIs)): Given a map $C : \mathbb{R}^n \to \mathbb{R}^n$ and a set $\mathcal{F} \subset \mathbb{R}^n$, the associated variational inequality problem, denoted $\mathrm{VI}(\mathcal{F}, C)$ is to find $f^* \in \mathcal{F}$ such that the following holds:*

$$C(f^*)^\top (f - f^*) \geq 0, \text{ for all } f \in \mathcal{F}. \tag{6.1}$$

The set of solutions $f^* \in \mathcal{F}$ satisfying the above property is denoted as $\mathrm{SOL}(\mathcal{F}, C)$. •

A useful concept in the context of variational inequalities is monotonicity of functions.

Definition 6.1.2. *(Monotone functions): The map $C : \mathbb{R}^n \to \mathbb{R}^n$ is called monotone if for all $f, \check{f} \in \mathbb{R}^n$ we have*

$$\big(C(f) - C(\check{f})\big)^\top (f - \check{f}) \geq 0.$$

When the inequality holds strictly for all $f \neq \check{f}$, the map C is called s strictly monotone. *If instead we have*

$$\big(C(f) - C(\check{f})\big)^\top (f - \check{f}) \geq c\|f - \check{f}\|^2,$$

for all $f, \check{f} \in \mathbb{R}^n$, then C is called strongly monotone *with constant $c > 0$.*

Important for this chapter is that, under the assumption that the map C is strictly or strongly monotone and the feasible set $\mathcal{F}$ is compact and convex, $\mathrm{VI}(\mathcal{F}, C)$ has a unique solution [16, Corollary 2.2.5, Theorem 2.2.3].

Lemma 6.1.3. *(Unique solutions to $\mathrm{VI}(\mathcal{F}, C)$): Let $C : \mathbb{R}^n \to \mathbb{R}^n$ be a strictly or strongly monotone, continuous function, and let $\mathcal{F}$ be a compact, convex set. In that case $\mathrm{VI}(\mathcal{F}, C)$ has a unique solution.*

Under some additional assumptions on the constraints determining $\mathcal{F}$, there exists another useful characterization of $\mathrm{SOL}(\mathcal{F}, C)$, namely as the set as the set of Karush-Kuhn-Tucker (KKT) points.

Lemma 6.1.4. *(KKT points of $\mathrm{VI}(\mathcal{F}, C)$): Let*

$$\mathcal{F} := \{f \in \mathbb{R}^n \mid Af = b,\ \ q^i(f) \leq 0,\ \ \forall i \in [s]\},$$

where $A \in \mathbb{R}^{l \times n}$, $b \in \mathbb{R}^l$ *and* $l \in \mathbb{N}$, *and the functions* $q^i : \mathbb{R}^n \to \mathbb{R}$, $i \in [s]$, $s \in \mathbb{N}$ *are convex and continuously differentiable. For* $q(f) := (q^1(f), \dots, q^s(f))^\top \in \mathbb{R}^s$, *let* $Dq(f) \in \mathbb{R}^{s \times n}$ *be the Jacobian at* f. *For any* $f^* \in \mathbb{R}^n$, *if there exists a multiplier* $(\lambda^*, \mu^*) \in \mathbb{R}^s \times \mathbb{R}^l$ *satisfying*

$$\begin{aligned} C(f^*) + \big(Dq(f^*)\big)^\top \lambda^* + A^\top \mu^* = 0, \\ Af^* = b, \quad q(f^*) \leq 0, \quad \lambda^* \geq 0, \quad \lambda^{*\top} q(f^*) = 0, \end{aligned} \tag{6.2}$$

then we have $f^* \in \mathrm{SOL}(\mathcal{F}, C)$. *Such a point* (f^*, λ^*, μ^*) *is referred to as a KKT point of the* $\mathrm{VI}(\mathcal{F}, C)$. *Conversely, for* $f^* \in \mathrm{SOL}(\mathcal{F}, C)$, *let* $\mathcal{I}_{f^*} = \{i \in [s] \mid q^i(f^*) = 0\}$. *If the vectors* $\{\nabla q^i(f^*)\}_{i \in \mathcal{I}_{f^*}}$ *and the row vectors* $\{A_j\}_{j \in [l]}$ *are linearly independent, or in other words, the LICQ holds at* f^*, *then there exists a* (λ^*, μ^*) *satisfying* (6.2).

The above result is well known in the context of convex optimization. The extension to the VI setting can be deduced in a straightforward manner, for instance from [67, Proposition 3.46], [68, Theorem 12.1], and noting that if $f^* \in \mathrm{SOL}(\mathcal{F}, C)$, then it is also a minimizer of the function $\check{f} \mapsto \check{f}^\top C(f^*)$ subject to $\check{f} \in \mathcal{F}$.

Conditional Value-at-Risk

The following is the definition of the *Conditional Value-at-Risk* (CVaR) of a real-valued random variable Z defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

Definition 6.1.5. *Let* Z *be a real-valued random variable, defined on the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *The* Conditional Value-at-Risk *at level* $\alpha \in (0, 1]$ *is given by*

$$\mathrm{CVaR}_\alpha[Z] := \inf_{\eta \in \mathbb{R}} \big\{\eta + \alpha^{-1} \mathbb{E}[Z - \eta]^+\big\}.$$

The value α in the above definition is a parameter that characterizes risk-averseness. Using lower values of α puts more weight on extreme scenarios where Z is high, such that minimizing $\mathrm{CVaR}_\alpha[Z]$ will result in conservative, less risky strategies. Thus lower values of α correspond to less willingness to take risks.

Note that calculating $\mathrm{CVaR}_\alpha[Z]$ requires knowledge of the distribution of Z, due to the presence of an expected value. When the distribution of Z is not directly available to us, we will instead have to resort to empirical estimation.Given N i.i.d samples $\{\widehat{Z}_j\}_{j \in [N]}$ of Z, one can approximate $\mathrm{CVaR}_\alpha[Z]$ using the following empirical estimate:

$$\widehat{\mathrm{CVaR}}^N_\alpha[Z] = \inf_{\eta \in \mathbb{R}} \big\{\eta + (N\alpha)^{-1} \textstyle\sum_{j=1}^N [\widehat{Z}_j - \eta]^+\big\}. \tag{6.3}$$

This estimator is biased, but consistent [69, Page 300].

Lemma 6.1.6. *(Consistency of* $\widehat{\mathrm{CVaR}}_\alpha^N$*): For a real-valued random variable* Z *we have*

$$\widehat{\mathrm{CVaR}}_\alpha^N[Z] \to \mathrm{CVaR}_\alpha[Z] \quad \textit{almost surely as } N \to \infty.$$

Projected dynamical systems

Part of the analysis in this chapter relies on approximating the asymptotic behaviour of iterates produced by our suggested schemes with the trajectories of continuous-time dynamical systems. For this we require the definition of a *projected dynamical system*.

Definition 6.1.7. *For a given* $C : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^n$ *and a closed set* $\mathcal{F} \subseteq \mathbb{R}^n$*, the associated projected dynamical system is given by*

$$\dot{f}(t) = \Pi_{\mathcal{T}_{\mathcal{F}}(f(t))}\big(C(f, t)\big).$$

In the above $\mathcal{T}_{\mathcal{F}}(f)$ is the tangent cone of $\mathcal{F}$ at f (see Section 1.4). We say that a map $\bar{f} : [0, \infty) \to \mathcal{F}$ with $\bar{f}(0) \in \mathcal{F}$ is a solution of the above system when $\bar{f}(\cdot)$ is absolutely continuous and $\dot{\bar{f}}(t) = \Pi_{\mathcal{T}_{\mathcal{F}}(\bar{f}(t))}\Big(F\big(\bar{f}(t), t\big)\Big)$ for almost all $t \in [0, \infty)$. Note that $\bar{f}(t) \in \mathcal{F}$ for all t. Throughout this chapter we use the terms solution and trajectory interchangeably.

6.2 Problem statement and motivating examples

The objective of this chapter is to provide stochastic approximation(SA) algorithms to solve the variational inequality problem $\mathrm{VI}(\mathcal{F}, C)$, where the map C is a vector of conditional values-of-risk of a given set of uncertain cost functions. Specifically consider a set of functions $c_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $i \in [n]$, $(f, \xi) \mapsto c_i(f, \xi)$, where ξ represents a random variable with distribution $\mathbb{P}$. For a fixed f, $c_i(f, \xi)$ is therefore a real-valued random variable. Define the map $C_i : \mathbb{R}^n \to \mathbb{R}$ as the CVaR of c_i at level $\alpha \in (0, 1]$:

$$C_i(f) := \mathrm{CVaR}_\alpha\big[c_i(f, \xi)\big], \quad \text{for all } i \in [n]. \tag{6.4}$$

For notational convenience, let $C : \mathbb{R}^n \to \mathbb{R}^n$ be the element-wise concatenation of the maps $\{C_i\}_{i \in [n]}$, respectively. Let $\mathcal{F} \subseteq \mathbb{R}^n$ be a non-empty closed set of the form

$$\mathcal{F} := \{f \in \mathbb{R}^n \mid Af = b, \;\; q^i(f) \le 0, \;\; \forall i \in [s]\}, \tag{6.5}$$

where $A \in \mathbb{R}^{l \times n}$, $b \in \mathbb{R}^l$ and $l \in \mathbb{N}$, and the functions $q^i : \mathbb{R}^n \to \mathbb{R}$, $i \in [s]$, $s \in \mathbb{N}$ are convex and continuously differentiable. The VIs whose solutions we aim to approximate are then of the form $\mathrm{VI}(\mathcal{F}, C)$. Before we move on to the analysis, we discuss two motivating examples for our setup.

6.2.1 CVaR-based routing games

For our first example, we consider a routing game, as introduced in Chapter 2, with some minor modifications. In particular, we generalize the setup in Chapter 2 by considering a *set* of origin-destination(OD) pairs, instead of a single OD pair. Therefore, consider a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = [N]$ is the set of vertices, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges and let $\mathcal{O} \subseteq \mathcal{V} \times \mathcal{V}$ be the set of origin-destination (OD) pairs. An OD-pair w is given by an ordered pair (v_o^w, v_d^w), where $v_o^w, v_d^w \in \mathcal{V}$ are called the origin and the destination of w, respectively. The set of all paths in $\mathcal{G}$ from the origin to the destination of w is denoted $\mathcal{P}_w$. The set of all paths relevant to the game is given by $\mathcal{P} = \cup_{w \in \mathcal{O}} \mathcal{P}_w$, and $n = |\mathcal{P}|$.

Given a demand D_w for each OD pair in $\mathcal{O}$ the feasible set of the routing game is then given by

$$\mathcal{F} = \{f \in \mathbb{R}^n \mid \sum_{p \in \mathcal{P}_w} f_p = D_w \text{ for all } w \in \mathcal{O}, \text{ and } f_p \geq 0 \text{ for all } p \in \mathcal{P}\}.$$

To each of the paths $p \in \mathcal{P}$ a cost function $c_p : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, (f, \xi) \mapsto c_p(f, \xi)$ is associated, which depends on the flow f, as well as on the uncertainty $\xi \in \mathbb{R}^m$. faced with this uncertainty, we assume that agents are risk-averse, in that they aim to minimize the conditional value at risk of their travel time. That is, they aim to choose the path $p \in \mathcal{P}_w$ that minimizes $\mathrm{CVaR}_\alpha \left[c_p(f, \xi)\right]$. The result is a CVaR-based routing game [62], to which we assign the following notion of equilibrium: the flow $f^* \in \mathcal{H}$ is said to be a CVaR-based Wardrop equilibrium (CWE) of the CVaR-based routing game if, for all $w \in \mathcal{O}$ and all $p, p' \in \mathcal{P}_w$ such that $f_p^* > 0$, we have

$$\mathrm{CVaR}_\alpha \left[c_p(f^*, \xi)\right] \leq \mathrm{CVaR}_\alpha \left[c_{p'}(f^*, \xi)\right].$$

Assuming that the functions c_p are continuous, the set of CWE is then equal to the set of solutions of $\mathrm{VI}(\mathcal{F}, C)$, where $C : \mathbb{R}^n \to \mathbb{R}^n$ takes the form (6.4).

6.2.2 CVaR-based Nash equilibrium

A more general example of our setup would be in finding the Nash equilibrium of a non-cooperative game [16, section 1.4.2]. Let there be N players with individual cost functions $c_i : \mathbb{R}^{nN} \to \mathbb{R}$, $f \mapsto c_i(f)$ and possible strategy sets $\mathcal{F}_i \subseteq \mathbb{R}^n$. Here $f \in \mathbb{R}^{nN}$ denotes the vector containing the strategies of all players, where $f_i \in \mathbb{R}^n$ is the strategy of player i. We assume without loss of generality that the strategy sets of each player are of the same dimension n, and we use the shorthand notation $c_i(f) = c_i(f_i, f_{-i})$, to denote the cost function of agent i, where f_{-i} is the vector containing the strategies of all players except i. Each player i aims to minimize its cost c_i by choosing its own strategy optimally. That is, for any fixed $\breve{f}_{-i}$ they solve

$$\begin{aligned} \text{minimize} \quad & c_i(f_i, \breve{f}_{-i}), \\ \text{subject to} \quad & f_i \in \mathcal{F}_i. \end{aligned}$$

A Nash equilibrium of such a game is a solution vector f^* such that none of the players can reduce their costs by changing their strategy. Under the assumption that the sets $\mathcal{F}_i$ are convex and closed, and the functions $f_i \mapsto c_i(f_i, \breve{f}_{-i})$ are convex and continuously differentiable for any $\breve{f}_{-i}$, a joint strategy vector f^* is a Nash equilibrium if and only if it is a solution to $\mathrm{VI}(\mathcal{F}, C)$, where $C(f) := (\nabla_{f_i} c_i(f))_{i=1}^N$ is the concatenation of the gradients of c_i functions, and $\mathcal{F} = \prod_{i=1}^n \mathcal{F}_i$. Consider the functions c_i of the form

$$c_i(x) := \mathrm{CVaR}_\alpha[g_i(f_i, f_{-i})h(\xi) + \breve{g}_i(g_i, g_{-i})],$$

where g_i, h and $\breve{g}_i$ are real-valued, ξ models the uncertainty, and $g_i(f_i, f_{-i}) \geq 0$ for all f. Then, $\mathrm{VI}(\mathcal{F}, C)$ is a CVaR-based variational inequality. Specifically, in this case, since CVaR is positive-homogeneous and shift-invariant [69, Chapter 6], we have

$$c_i(f) = \mathrm{CVaR}_\alpha[h(\xi)]g_i(f_i, f_{-i}) + \breve{g}_i(f_i, f_{-i}).$$

As a consequence, we get

$$\nabla c_i(f) = \mathrm{CVaR}_\alpha[h(\xi)]\nabla_{f_i} g_i(f_i, f_{-i}) + \nabla_{f_i}\breve{g}_i(f_i, f_{-i}).$$

Under the assumption that $\nabla_{f_i} g_i$ is non-negative for all $f \in \mathcal{F}$, we get

$$\nabla c_i(f) = \mathrm{CVaR}_\alpha[h(\xi)\nabla_{f_i} g_i(f_i, f_{-i}) + \nabla_{f_i}\breve{g}_i(f_i, f_{-i})].$$

where CVaR is understood component-wise. Thus, C can be written as the concatenation of CVaR of various functions and finding the Nash equilibrium of this game is equivalent to solving $\mathrm{VI}(\mathcal{F}, C)$, which fits into our presented framework.

6.3 Algorithms for solving $\mathrm{VI}(\mathcal{H}, F)$

In this section, we introduce stochastic approximation(SA) algorithms for solving CVaR-based VIs, along with their convergence analysis. All introduced schemes approximate C with the estimator given in (6.3). Given N independently and identically distributed samples $\left\{(\widehat{c_i(f,\xi)})_j\right\}_{j=1}^N$ of the random variable $c_i(f, \xi)$, let

$$\widehat{C}_i^N(f) := \inf_{t \in \mathbb{R}} \left\{ t + (N\alpha)^{-1} \sum_{j=1}^N \left[(\widehat{c_i(f,\xi)})_j - t \right]^+ \right\}$$

stand for the estimator of $C_i(f)$. Analogously, the estimator of $C(f)$ formed using the element-wise concatenation of $\widehat{C}_i^N(f)$, $i \in [n]$, is denoted by $\widehat{C}^N(f)$. We assume that the N samples of each cost function are a result of the same set of N events, that is, the distribution of $\widehat{C}^N(f)$ depends on $\mathbb{P}^N$. We start our analysis with an algorithm that employs projection of the iterates onto the feasible set $\mathcal{F}$.

6.3.1 Projected algorithm

For a given sequence of step-sizes $\{\gamma^k\}_{k=0}^{\infty}$, with $\gamma^k > 0$ for all k, a sequence $\{N_k\}_{k=0}^{\infty} \subset \mathbb{N}$, and an initial vector $f^0 \in \mathcal{F}$, the first algorithm under consideration, which we will refer to as the *projected algorithm*, is given by

$$f^{k+1} = \Pi_{\mathcal{F}}\big(f^k - \gamma^k \widehat{C}^{N_k}(f^k)\big), \tag{6.6}$$

where $\Pi_{\mathcal{F}}$ is the projection operator (see Section 1.4) and f^k is the k-th iterate produced by the algorithm. The above algorithm is inspired by the SA schemes for solving a stochastic VI problem, see [58] for details on other SA schemes. The key difference from the setup in [58] is the fact that there the map C is the expected value of a random variable for which an unbiased estimator $\widehat{C}$ is available. In our case the employed estimator is biased posing limitations on the sample requirements for convergence of the algorithms. We can write the projected algorithm (6.6) equivalently as

$$f^{k+1} = \Pi_{\mathcal{F}}\Big(f^k - \gamma^k\big(C(f^k) + \widehat{\beta}^{N_k}\big)\Big), \tag{6.7}$$

where $\widehat{\beta}^{N_k} := \widehat{C}^{N_k}(f^k) - C(f^k)$ is the error introduced by the estimation. For this and the upcoming algorithms, common assumptions on the sequence $\{\gamma^k\}$ are

$$\textstyle\sum_{k=0}^{\infty} \gamma^k = \infty, \qquad \sum_{k=0}^{\infty} (\gamma^k)^2 < \infty. \tag{6.8}$$

Our first result gives sufficient conditions for convergence of (6.7) to any neighbourhood of the solution f^* of VI$(\mathcal{F}, C)$.

Proposition 6.3.1. *(Convergence of the projected algorithm* (6.6)*): Let C as defined in* (6.4) *be a strictly monotone, continuous function, and let $\mathcal{F}$ be a compact convex set of the form* (6.5)*. For the algorithm* (6.7)*, assume that the sequence $\{\gamma^k\}$ satisfies* (6.8) *and the sequence $\{N_k\}$ is such that $\{\widehat{\beta}^{N_k}\}$ is bounded with probability one. Then, for any $\epsilon > 0$ there exists $N_\epsilon \in \mathbb{N}$ such that $N_k \geq N_\epsilon$ for all k implies, with probability one,*

$$\lim_{k\to\infty} \|f^k - f^*\| \leq \epsilon.$$

Proof. To ease the exposition of this proof, we split the error as $\widehat{\beta}^{N_k} = e^{N_k} + \widehat{\varepsilon}^{N_k}$, where $e^{N_k} = \mathbb{E}[\widehat{\beta}^{N_k}]$. Note that we then have $\mathbb{E}[\widehat{\varepsilon}^{N_k}] = 0$, and by the boundedness assumption, there exists a constant $B_e > 0$ such that $\|e^{N_k}\| \leq B_e$ for all k. The first step of the proof is to show that the sequence $\{f^k\}$ converges to a trajectory of the following continuous-time projected dynamical system:

$$\dot{\bar{f}}(t) = \Pi_{\mathcal{T}_{\mathcal{F}}(\bar{f}(t))}\Big(- C\big(\bar{f}(t)\big) - e(t)\Big), \quad \bar{f}(0) \in \mathcal{F}. \tag{6.9}$$

Here $e(\cdot)$ is a uniformly bounded measurable map satisfying $\|e(t)\| \leq B_e$ for all t (see Section 6.1 for further details on how solutions to projected dynamical systems are defined). For the sake of rigour, we note that the existence of a trajectory of (6.9) starting from any point in $\mathcal{F}$ is guaranteed by [70, Lemma A.1]. To make precise the convergence of the sequence generated by (6.7) to a trajectory of (6.9), we say that $\{f^k\}$ converges to a trajectory $\bar{f}(\cdot)$ of (6.9) if

$$\lim_{i\to\infty} \sup_{j\geq i} \Big\| f^j - \bar{f}\Big(\sum_{k=i}^{j-1} \gamma^k\Big)\Big\| = 0. \tag{6.10}$$

That is, the discrete-time trajectory formed by the linear interpolation of the iterates $\{f^k\}$ approaches the continuous time trajectory $t \mapsto \bar{f}(t)$. The proof of the existence of a map $\bar{f}(\cdot)$ satisfying (6.10) is similar to that of [65, Theorem 5.3.1], with the only change being the existence of an error term $e(t)$ in dynamics (6.9) which is absent in the cited reference. The inclusion of the error term is facilitated by reasoning presented in the proof of [65, Theorem 5.2.2]. We avoid repeating these arguments here in the interest of space.

Convergence of the sequence $\{f^k\}$ can now be analysed by studying the asymptotic stability of (6.9). To this end, we consider the candidate Lyapunov function

$$V(\bar{f}) = \frac{1}{2}\|\bar{f} - f^*\|^2,$$

where f^* is the unique solution of VI($\mathcal{F}, C$). We first look at the case $e(\cdot) \equiv 0$. For notational convenience, define the right-hand side of (6.9) in such a case by the map $X_{e\equiv 0} : \mathbb{R}^n \to \mathbb{R}^n$. The Lie derivative of V along $X_{e\equiv 0}$ is then given by

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) = (\bar{f} - f^*)^\top \Pi_{\mathcal{T}_{\mathcal{F}}(\bar{f})}\big(- C(\bar{f})\big). \tag{6.11}$$

We want to show that the right-hand side of the above equation is negative for all $\bar{f} \neq f^*$. We first note that by Moreau's decomposition theorem [71, Theorem 3.2.5], for any $v \in \mathbb{R}^n$ and $\bar{f} \in \mathcal{F}$, we have $\Pi_{\mathcal{T}_{\mathcal{F}}(\bar{f})}(v) = v - \Pi_{\mathcal{N}_{\mathcal{F}}(\bar{f})}(v)$, where $\mathcal{N}_{\mathcal{F}}(\bar{f})$ is the normal cone to $\mathcal{F}$ at $\bar{f}$. Using the above relation in (6.11) gives

$$\begin{aligned}\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) &= - (\bar{f} - f^*)^\top C(\bar{f}) + (f^* - \bar{f})^\top \Pi_{\mathcal{N}_{\mathcal{F}}(\bar{f})}\big(- C(\bar{f})\big) \\ &\leq - (\bar{f} - f^*)^\top C(\bar{f}),\end{aligned} \tag{6.12}$$

where the inequality is due to the definition of the normal cone (see Section 1.4) and $f^* \in \mathcal{F}$. Due to strict monotonicity of C, we have $(\bar{f} - f^*)^\top C(\bar{f}) > (\bar{f} - f^*)^\top C(f^*)$ whenever $\bar{f} \neq f^*$. Since $f^* \in \mathrm{SOL}(\mathcal{F}, C)$ we also know that $(\bar{f} - f^*)^\top C(f^*) \geq 0$ for all $\bar{f} \in \mathcal{F}$. Combining these two facts implies that the function $W(\bar{f}) := (\bar{f} - f^*)^\top C(\bar{f})$ satisfies $W(\bar{f}) > 0$ whenever $\bar{f} \neq f^*$. Using this in the inequality (6.12) yields

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) \leq -W(\bar{f}) < 0 \tag{6.13}$$

whenever $\bar{f} \neq f^*$. Now let $\overline{\mathcal{F}}_\epsilon := \{f \in \mathcal{F} \mid \|f - f^*\| \geq \epsilon\}$. Since $\mathcal{F}$ is compact, $\overline{\mathcal{F}}_\epsilon$ is compact. Since W is continuous, there exists a $\delta > 0$ such that $W(\bar{f}) \geq \delta$ for all $\bar{f} \in \overline{\mathcal{F}}_\epsilon$. Therefore we get, from (6.13),

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) \leq -\delta, \quad \text{for all } \bar{f} \in \overline{\mathcal{F}}_\epsilon. \tag{6.14}$$

Next, we drop the assumption that $e(\cdot) \equiv 0$ and use the map $X : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^n$ to denote the right-hand side of (6.9). Consider any trajectory $t \mapsto \bar{f}(t)$ of (6.9). Since the map is absolutely continuous and V is differentiable, we have for almost all $t \geq 0$ and for $\bar{f}(t) \in \overline{\mathcal{F}}_\epsilon$,

$$\frac{dV}{dt}(t) = \nabla V(\bar{f}(t))^\top X(\bar{f}(t), t) \leq -\delta - (\bar{f}(t) - f^*)^\top e(t),$$

where for obtaining the above inequality we have first used Moreau's decomposition as before to get rid of the projection operator in X and then employed (6.14). Next we bound the error term in the above inequality. Since $\mathcal{F}$ is compact and $f^* \in \mathcal{F}$, there exists $B_f > 0$ such that $\|\bar{f} - f^*\| \leq B_f$ for all $\bar{f} \in \mathcal{F}$. In addition $\|e(t)\| \leq B_e$ for all t, where B_e is the bound satisfying $\|e^{N_k}\| \leq B_e$. Since the empirical estimate of the CVaR is consistent, we know that B_e can be made arbitrarily small by selecting N_k to be appropriately large for all k. That is, there exists $N_\epsilon \in \mathbb{N}$ such that when $N_k > N_\epsilon$ we have $\|e^{N_k}\| < \frac{\delta}{B_f}$. Consequently, if $N_k > N_\epsilon$ for all k, then $\|e(t)\| < \frac{\delta}{B_f}$ for all t. By selecting such a sample size at each iteration and thus bounding the error term, we obtain

$$\begin{aligned}\frac{dV}{dt}(t) &\leq -\delta - (\bar{f}(t) - f^*)^\top e(t) \\ &\leq -\delta + \|\bar{f}(t) - f^*\|\|e(t)\| < -\delta + B_f \frac{\delta}{B_f} \leq 0,\end{aligned}$$

which holds for almost all t whenever $\bar{f}(t) \in \overline{\mathcal{F}}_\epsilon$. That is, the trajectory converges to the set $\{f \in \mathcal{F} \mid \|f - f^*\| \leq \epsilon\}$ as $t \to \infty$. This concludes the proof. □

In the above result, the restriction $N_k \geq N_\epsilon$ does not need to hold for all k. The result also holds if there exists a $K \in \mathbb{N}$ such that $N_k \geq N_\epsilon$ for all $k \geq K$. Regarding boundedness of $\{\widehat{\beta}^{N_k}\}$ we note that it is ensured if for example each C_i is bounded over the set $\mathcal{F} \times \Xi$, where Ξ is the support of ξ.

Despite the convergence property established in Proposition 6.3.1, the algorithm in (6.7) suffers from some disadvantages. Most notably, the algorithm requires computing projections onto the set $\mathcal{F}$ at each iteration, which can be computationally expensive. To address these issues we propose two algorithms that achieve similar convergence to any neighbourhood of the solution of the $\mathrm{VI}(\mathcal{F}, C)$. The first requires projection onto inequality constraints only and the second does not involve any projection on the primal iterates and instead ensures feasibility using dual variables. As in Proposition 6.3.1, we will impose continuity and monotonicity assumptions on F in the upcoming results. We provide the following general result on the continuity and monotonicity properties of C.

Lemma 6.3.2. *(Sufficient conditions for monotonicity and continuity of C): The following hold:*

- *If for any $\epsilon > 0$ there exist a $\delta > 0$ such that $\|c_i(f, \xi) - c_i(\check{f}, \xi)\| \le \epsilon$ holds for all $i \in [n]$ whenever $\|f - \check{f}\| \le \delta$, then C is continuous.*
- *Let $g_i : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^m \to \mathbb{R}$ satisfy $c_i(f, \xi) \equiv g_i(f) + h_i(\xi)$, for all $i \in [n]$. Let $g(f) := \big(g_1(f), \dots, g_n(f)\big)$. Then, C is monotone (resp. strictly) if g is monotone (resp. strictly monotone).*

Proof. Continuity follows by arguments similar to the proof of [62, Lemma IV.8]. For the second part, note that CVaR satisfies $\mathrm{CVaR}_\alpha\big[C_i(f, \xi)\big] = g_i(f) + \mathrm{CVaR}_\alpha\big[h_i(\xi)\big]$, for all f and $i \in [n]$ [69, Page 261]. The proof then follows from the fact that $C(f) - C(f^*) = g(f) - g(f^*)$. □

In the above result, the continuity condition, that may be difficult to check in practice, holds if ξ has a compact support and for any fixed ξ, the functions C_i are continuous with respect to h.

6.3.2 Subspace-constrained algorithm

Too address the computational intensiveness of the projected algorithm, we take a closer look at the form of $\mathcal{F}$ given in (6.5) and design an algorithm that handles inequality and equality constraints independently. To this end, we us the notation $\mathcal{F}_{\mathrm{aff}} := \{f \in \mathbb{R}^n \mid Af = b\}$, and $\mathcal{F}_{\mathrm{ineq}} := \{f \in \mathbb{R}^n \mid q^i(f) \le 0, \ \forall i \in [s]\}$ for the sets of points satisfying the equality and inequality constraints, respectively. We then have $\mathcal{F} = \mathcal{F}_{\mathrm{aff}} \cap \mathcal{F}_{\mathrm{ineq}}$. It turns out that, using matrix operation, we can ensure that the iterates of our algorithm always remain in $\mathcal{F}_{\mathrm{aff}}$. The method works as follows. Let

$\{a_1, \cdots, a_l\}$ be the row vectors of A, and let $\{u_1, \cdots, u_n\}$ be an orthonormal basis for $\mathbb{R}^n$ such that the first $M \in \mathbb{N}$ vectors $\{u_1, \cdots, u_M\}$ form a basis for the span of vectors $\{a_1, \cdots, a_l\}$. Then, for the subspace $\mathcal{S} = \{g \in \mathbb{R}^n \mid Ag = 0\}$, we have

$$\Pi_{\mathcal{S}}(v) = \Big(I - \textstyle\sum_{i=1}^{M} u_i u_i^\top\Big)v, \qquad \text{for any } v \in \mathbb{R}^n.$$

This well known expression follows from [72, Theorem 7.10] combined with the fact $\Pi_{\mathcal{S}}(v) = v - \Pi_{\mathcal{S}^\perp}(v)$, where $\mathcal{S}^\perp$ is the set of vectors orthogonal to the subspace $\mathcal{S}$. Thus, the projection onto $\mathcal{S}$ is achieved by pre-multiplying with the matrix

$$L := I - \sum_{i=1}^{M} u_i u_i^T. \tag{6.15}$$

Consequently, for any vector z of the form $z = Lv$, $v \in \mathbb{R}^n$ we have $Az = \mathbb{0}$. To construct L one can find the orthonormal basis vectors $\{u_i\}_{i\in[l]}$ for the span of $\{a_j\}_{j\in[l]}$ and $\mathbb{R}^n$ by using the Gram-Schmidt orthogonalization process [72, Section 6.4]. Alternatively, if A has full row rank one can use $L := I - A^\top(AA^\top)^{-1}A$, see e.g., [73]. We use this projection operator to define our next method called the *subspace-constrained algorithm*:

$$f^{k+1} = f^k - \gamma^k L\Big(C(f^k) + d\big(f^k - \Pi_{\mathcal{F}_{\text{ineq}}}(f^k)\big) + \widehat{\beta}^{N_k}\Big), \tag{6.16}$$

where the initial iterate $f^0 \in \mathcal{F}_{\text{aff}}$. In the above, $d > 0$ is a parameter to be specified later in the convergence result, the error sequence $\{\widehat{\beta}^{N_k}\}$ is as defined in (6.7), and $L \in \mathbb{R}^{n\times n}$ is as defined in (6.15).

Due to the presence of L in the above algorithm, the direction in which the iterate moves in each iteration is projected onto the subspace $\mathcal{S}$. Hence, $f^k \in \mathcal{F}_{\text{aff}}$ for all k. We formally establish this in the below result. Furthermore, convergence to a neighbourhood of the set $\mathcal{F}_{\text{ineq}}$ is achieved through the term $f^k - \Pi_{\mathcal{F}_{\text{ineq}}}(f^k)$. That is, the higher the value of the design parameter c, the closer the limit of $\{f^k\}$ is to $\mathcal{F}_{\text{ineq}}$. Together, these mechanisms ensure that we keep iterates close to $\mathcal{F}$ and ultimately drive them to a neighbourhood of f^*.

Proposition 6.3.3. *(Convergence of subspace-constrained algorithm* (6.16)*): Let C as defined in* (6.4) *be a strictly monotone, continuous function, and let $\mathcal{F}$ be a compact convex set of the form* (6.5). *For the algorithm* (6.16), *assume that the step-sizes sequence $\{\gamma^k\}$ satisfies* (6.8) *and that the sequence $\{N^k\}$ is such that there exists $B_{\text{traj}} \in \mathbb{R}$ satisfying $\|f^k\| \le B_{\text{traj}}$ and $\{\widehat{\beta}^{N_k}\}$ is bounded with probability one. Then, for any $\epsilon > 0$, there exist $d_\epsilon(B_{\text{traj}}) > 0$ and $N_\epsilon(B_{\text{traj}}) \in \mathbb{N}$ such that $d \ge d_\epsilon(B_{\text{traj}})$ and $N_k \ge N_\epsilon(B_{\text{traj}})$ for all k imply that the iterates of* (6.16) *satisfy, with probability one,*

$$\lim_{k\to\infty} \|f^k - f^*\| \le \epsilon.$$

Proof. First we show that $f^k \in \mathcal{F}_{\text{aff}}$ for all k. To see this, recall that $ALv = 0$ for any $v \in \mathbb{R}^n$. Using this in (6.16) implies $Af^{k+1} = Af^k$ for all k. Consequently, for all k, we have $Af^k = Af^0 = b$ and therefore $f^k \in \mathcal{F}_{\text{aff}}$.

Analogous to the proof of Proposition 6.3.1, it can be established that $\{f^k\}$ converges with probability one, in the sense of (6.10), to a trajectory of the following dynamics

$$\dot{\bar{f}}(t) = -L\Big(C\big(\bar{f}(t)\big) + c\Big(\bar{f}(t) - \Pi_{\mathcal{F}_{\text{ineq}}}\big(\bar{f}(t)\big)\Big) - e(t)\Big), \tag{6.17}$$

with the initial state $\bar{f}(0) \in \mathcal{F}_{\text{aff}}$. Here, $e(\cdot)$ is a uniformly bounded measurable map satisfying $\|e(t)\| \leq B$ for all t. We will use the above fact to establish convergence of the sequence $\{f^k\}$ by analysing the asymptotic stability of (6.17). Note that $A\dot{\bar{f}}(t) = 0$ for all t, and therefore a trajectory $\bar{f}(\cdot)$ of (6.17) satisfies $\bar{f}(t) \in \mathcal{F}_{\text{aff}}$ for all $t \geq 0$ as $\bar{f}(0) \in \mathcal{F}_{\text{aff}}$. Now consider the Lyapunov candidate

$$V(\bar{f}) = \frac{1}{2}\|\bar{f} - f^*\|^2,$$

where f^* is the unique solution of VI($\mathcal{F}, C$), that follows from strict monotonicity. As was the case for the previous result, we will first analyse the evolution of V along (6.17) when $e \equiv 0$. Therefore, we define the notation $X_{e\equiv 0} : \mathbb{R}^n \to \mathbb{R}^n$ to represent the right-hand side of (6.17) with $e \equiv 0$. The Lie derivative of V along $X_{e\equiv 0}$ is

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) = -(\bar{f} - f^*)^\top L\Big(C(\bar{f}) + d\big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big)\Big). \tag{6.18}$$

Since $\bar{f}, f^* \in \mathcal{F}_{\text{aff}}$, we have $A(\bar{f} - f^*) = 0$ and so $(\bar{f} - f^*) \in \mathcal{S}$. Consequently, for any vector $v \in \mathbb{R}^n$, we have

$$\begin{aligned}(\bar{f} - f^*)^\top v &= (\bar{f} - f^*)^\top \big(\Pi_{\mathcal{S}}(v) + \Pi_{\mathcal{S}^\perp}(v)\big) \\ &= (\bar{f} - f^*)^\top \Pi_{\mathcal{S}}(v) = (\bar{f} - f^*)^\top Lv.\end{aligned}$$

Using the above equality in (6.18) gives

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) = -(\bar{f} - f^*)^\top \Big(C(\bar{f}) + d\big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big)\Big). \tag{6.19}$$

We first upper bound the second term on the right-hand side of the above equality. We have

$$\begin{aligned}&-d(\bar{f} - f^*)^\top \big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big), \\ &= -d\Big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f}) + \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f}) - f^*\Big)^\top \big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big), \\ &= -d\|\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\|^2 + d\big(f^* - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big)^\top \big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big) \leq 0,\end{aligned} \tag{6.20}$$

where for the inequality we have used $\big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big)^\top \big(f^* - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big) \leq 0$ for any $\bar{f} \in \mathbb{R}^n$ (see [74, Thm. 3.1.1]). Note that the inequality (6.20) is strict whenever $\bar{f} \neq \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})$. We now turn our attention towards the first term in (6.19). Due to strict monotonicity of C and the fact that $f^* \in \text{SOL}(\mathcal{F}, C)$, we obtain

$$-(\bar{f} - f^*)^\top C(\bar{f}) < -(\bar{f} - f^*)^\top C(f^*) \leq 0 \tag{6.21}$$

whenever $\bar{f} \in \mathcal{F}$ and $\bar{f} \neq f^*$. The above inequality along with (6.20) shows $\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) \leq 0$ for any $\bar{f} \in \mathcal{F}$. However, recalling the approach in the proof of Proposition 6.3.1, what we require in order to establish convergence is the existence of $\delta > 0$ such that

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) \leq -\delta \text{ for all } \bar{f} \in \overline{\mathcal{F}}_\epsilon, \tag{6.22}$$

where $\overline{\mathcal{F}}_\epsilon := \{f \in \mathcal{F}_{\text{aff}} \mid \|f - f^*\| \geq \epsilon\}$. We obtain this bound below. Note that the strict inequality (6.21) along with continuity of C imply that for any $f \in \mathcal{F} \setminus \{f^*\}$, there exists $\varepsilon_f > 0$ such that

$$-(\check{f} - f^*)^\top C(\check{f}) < 0 \text{ for all } \check{f} \in \mathcal{C}_{\varepsilon_f}(f), \tag{6.23}$$

where we recall that $\mathcal{C}_{\varepsilon_f}(f)$ is the open ε_f-ball centred at f. Now let $\mathcal{F}_\epsilon := \mathcal{F} \setminus \mathcal{C}_\epsilon(f^*)$. Since $\mathcal{F}$ is compact, so is $\mathcal{F}_\epsilon$. Using this property and (6.23), we deduce that there exists $\varepsilon_0 > 0$ such that for every $f \in \mathcal{F}_\epsilon$ we have

$$-(\check{f} - f^*)^\top C(\check{f}) < 0 \text{ for all } \check{f} \in \mathcal{C}_{\varepsilon_0}(f). \tag{6.24}$$

Next define

$$\Delta_{\varepsilon_0} := \{\bar{f} \in \mathcal{F}_{\text{aff}} \setminus \mathcal{C}_\epsilon(f^*) \mid \bar{f} \notin \mathcal{C}_{\varepsilon_0}(\mathcal{F}_\epsilon) \text{ and } \|\bar{f}\| \leq B_{\text{traj}}\}.$$

Here, $\mathcal{C}_{\varepsilon_0}(\mathcal{F}_\epsilon)$ is the open ε_0-ball of the set $\mathcal{F}_\epsilon$ and $B_{\text{traj}} > 0$ is used as an upper bound on any trajectory $\bar{f}(\cdot)$ of (6.17). Note that Δ_{ε_0} is compact. Therefore, there exists $B_C > 0$ satisfying

$$-\,(\bar{f} - f^*)^\top C(\bar{f}) \leq B_C \text{ for all } \bar{f} \in \Delta_{\varepsilon_0}. \tag{6.25}$$

Furthermore, by definition, if $\bar{f} \in \Delta_{\varepsilon_0}$, then $\bar{f} \notin \mathcal{F}$ and $\bar{f} \in \mathcal{F}_{\text{aff}}$. Thus, $\bar{f} \in \Delta_{\varepsilon_0}$ implies $\bar{f} \notin \mathcal{F}_{\text{ineq}}$. That is, for such a point, the inequality (6.20) holds strictly. This along with compactness of Δ_{ε_0} implies that there exists $B_\Pi > 0$ such that

$$-\,(\bar{f} - f^*)^\top \big(\bar{f} - \Pi_{\mathcal{F}_{\text{ineq}}}(\bar{f})\big) \leq -B_\Pi \text{ for all } \bar{f} \in \Delta_{\varepsilon_0}. \tag{6.26}$$

Using (6.25) and (6.26) in (6.20) and setting $d > \frac{B_C}{B_\Pi}$ yields

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) < 0 \text{ for all } \bar{f} \in \Delta_{\varepsilon_0}. \tag{6.27}$$

Now consider $\bar{f}$ satisfying $\bar{f} \notin \Delta_{\epsilon_0} \cup \mathcal{C}_\epsilon(f^*)$ and $\|\bar{f}\| \leq B_{\text{traj}}$. Note that such a point belongs to $\mathcal{F}_{\text{aff}} \cap \mathcal{C}_{\varepsilon_0}(\mathcal{F}_\epsilon) \cap \mathcal{C}_{B_{\text{traj}}}(0)$. Thus, by (6.24), we have $-(\bar{f} - f^*)^\top C(\bar{f}) < 0$ for such a point. This fact combined with (6.27) leads us to the conclusion that

$$\nabla V(\bar{f})^\top X_{e\equiv 0}(\bar{f}) < 0 \text{ for all } \bar{f} \in \overline{\mathcal{F}}_\epsilon.$$

Since the left-hand side of the above equation is a continuous function and $\overline{\mathcal{F}}_\epsilon$ is compact, we deduce that (6.22) holds. The rest of the proof is then analogous to the corresponding section of the proof in Proposition 6.3.1. □

Remark 6.3.4. *(Practical considerations of* (6.16)*):* In Proposition 6.3.3, for small values of ϵ, one would require a large value of d to ensure convergence. This may result in large oscillations of f^k when γ^k remains large. Such behaviour can be prevented by either starting with small values of γ^k or increasing d along iterations, until it reaches a predetermined size. The result is then still valid but the convergence can only be guaranteed once d reaches the required size.

We note that the required assumption of boundedness of $\{f^k\}$ can be ensured by constraining the iterates in $\{f^k\}$ to lie in a hyper-rectangle containing $\mathcal{F}$ (cf. [65, Page 40]). However, on the boundary of the hyper-rectangle, one would have to make use of steps of the form (6.7) to ensure that the iterates remain in the feasible set. •

6.3.3 Multiplier-driven algorithm

Although the projection required for the subspace-constrained algorithm is less involved than that of the projection algorithm, it does involve projection onto $\mathcal{F}_{\text{ineq}}$, which can still be computationally burdensome. Our next algorithm overcomes this limitation. We assume $\mathcal{F}$ to be of the form (6.5) and introduce a multiplier variable $\lambda \in \mathbb{R}^s_{\geq 0}$ that enforces satisfaction of the inequality constraint as the algorithm progresses. In order to simplify the coming equations we introduce the notation $H(f, \lambda) := C(f) + Dq(f)^\top \lambda$, where $Dq(f)$ is the Jacobian of q at f. The *multiplier-driven algorithm* is now given as

$$\begin{aligned} f^{k+1} &= f^k - \gamma^k L\big(H(f^k, \lambda^k) + \widehat{\beta}^{N_k}\big), \\ \lambda^{k+1} &= \big[\lambda^k + \gamma^k q(f^k)\big]^+. \end{aligned} \tag{6.28}$$

Here L is as defined in (6.15). Also recall that $\widehat{\beta}^{N_k}$ is the error due to empirical estimation of F. The next result establishes the convergence properties of (6.28) to a KKT point of the VI (see Section 6.1 for definitions) and thus to a solution of the VI.

Proposition 6.3.5. *(Convergence of the multiplier-driven algorithm* (6.28)*): Let C, as defined in* (6.4)*, be a strictly monotone, continuous function, and let $\mathcal{F}$ be a compact convex* *set of the form* (6.5)*, where functions* q^i*,* $i \in [s]$*, are affine. Assume that the LICQ holds for* $f^* \in \mathrm{SOL}(\mathcal{F}, C)$*, and let* (f^*, λ^*, μ^*) *be an associated KKT point. For algorithm* (6.28)*, assume that the step-size sequence* $\{\gamma^k\}$ *satisfies* (6.8) *and let* $\{N_k\}$ *be such that* $\{\widehat{\beta}^{N_k}\}$*,* $\{f^k\}$*, and* $\{\lambda^k\}$ *are bounded with probability one. Then, for any* $\epsilon > 0$*, there exists an* $N_\epsilon \in \mathbb{N}$ *such that if* $N_k \geq N_\epsilon$ *for all* k*, then, with probability one,*

$$\lim_{k\to\infty} \|f^k - f^*\| \leq \epsilon.$$

Proof. Analogous to the proof of Proposition 6.3.1, the first step establishes convergence with probability one of the sequence $\{(f^k, \lambda^k)\}$, in the sense of (6.10), to a trajectory $(\bar{f}(\cdot), \bar{\lambda}(\cdot))$ of a continuous time dynamical system. To express the system we define the following notation: For $q, \lambda \in \mathbb{R}$, the operator $[q]_\lambda^+$ equals q if $\lambda > 0$ and it equals $\max\{0, q\}$ if $\lambda \leq 0$, and for the vectors $q, \lambda \in \mathbb{R}^n$, the i-th element of the vector $[q]_\lambda^+$ is given by $[q_i]_{\lambda_i}^+$. The sequence $\{(f^k, \lambda^k)\}$ then converges to the trajectory given by

$$\dot{\bar{f}}(t) = -L\Big(H\big(\bar{f}(t), \bar{\lambda}(t)\big) + e(t)\Big), \tag{6.29a}$$

$$\dot{\bar{\lambda}}(t) = \Big[q\big(\bar{f}(t)\big)\Big]_{\bar{\lambda}(t)}^+, \tag{6.29b}$$

with initial condition $\bar{f}(0) \in \mathbb{R}^n$ and $\bar{\lambda}(0) \in \mathbb{R}^l_{\geq 0}$. Note that as a consequence of (6.29b), $\bar{\lambda}$ is contained in the non-negative orthant along any trajectory of the system. The map $\bar{e}(\cdot)$ is uniformly bounded and so, as before, we have $\|e(t)\| \leq B_e$ for all t. The proof of convergence of the iterates to a continuous trajectory is similar to that of [65, Theorem 5.2.2] and is not repeated here in the interest of space. Note that, as was the case for Proposition 6.3.3, multiplication with the matrix L ensure that $f^k, \bar{f}(t) \in \mathcal{F}_{\text{aff}}$ for all k and $t \geq 0$. Next, we analyse the convergence of (6.29). We will occasionally use $\bar{x}$ as shorthand for $(\bar{f}, \bar{\lambda})$. Define the candidate Lyapunov function

$$V(\bar{f}, \bar{\lambda}) := \frac{1}{2}\big(\|\bar{f} - f^*\|^2 + \|\bar{\lambda} - \lambda^*\|^2\big), \tag{6.30}$$

where f^* is the unique solution of $\mathrm{VI}(\mathcal{F}, C)$ and there exist $\mu^* \in \mathbb{R}^l$ such that (f^*, λ^*, μ^*) is an associated KKT point.We analyse the evolution of (6.30) for the case $e \equiv 0$. Denoting the right-hand side of (6.29) for this case by $X_{e\equiv 0}$, the Lie derivative of V along (6.29) is

$$\nabla V(\bar{x})^\top X_{e\equiv 0}(\bar{x}) = -(\bar{f} - f^*)^\top H(\bar{f}, \bar{\lambda}) + (\bar{\lambda} - \lambda^*)^\top\big(q(\bar{f}) + [q(\bar{f})]_{\bar{\lambda}}^+ - q(\bar{f})\big). \tag{6.31}$$

Here we have dropped the matrix L from the term $(\bar{f} - f^*)^\top LH(\bar{f}, \bar{\lambda})$, which is justified by the same argument used for deriving (6.19). Note $\big([q(\bar{f})]_{\bar{\lambda}}^+\big)_i = \big(q(\bar{f})\big)_i$ if $\bar{\lambda}_i > 0$ for any i. Also, if $\bar{\lambda}_i = 0$, then $\bar{\lambda}_i - \lambda_i^* \leq 0$. Consequently we find that

$(\bar{\lambda}-\lambda^*)^\top\big([q(\bar{f})]^+_{\bar{\lambda}}-q(\bar{f})\big)\leq 0$. Since q is affine, we have $Dq(\bar{f})=Dq(f^*)$ for all $\bar{f}\in\mathbb{R}^n$. Combined with strict monotonicity this gives, for $\bar{f}\neq f^*$,

$$\begin{aligned}0&<(\bar{f}-f^*)^\top\big(H(\bar{f},\bar{\lambda})-H(f^*,\bar{\lambda})\big)\\&=(\bar{f}-f^*)^\top\big(H(\bar{f},\bar{\lambda})-H(f^*,\lambda^*)+Dq(f^*)^\top\lambda^*-Dq(f^*)^\top\bar{\lambda}\big).\end{aligned}\tag{6.32}$$

From (6.2) we have $-H(f^*,\lambda^*)=A^\top\mu^*$. Since we have $\bar{f},f^*\in\mathcal{F}_{\text{aff}}$ it follows that $-(\bar{f}-f^*)^\top H(f^*,\lambda^*)=0$. Then, using the assumption that q is affine, (6.32) gives us $-(\bar{f}-f^*)^\top H(\bar{f},\bar{\lambda})<(\lambda^*-\bar{\lambda})^\top\big(q(\bar{f})-q(f^*)\big)$ Combining these derivations, and writing $W(\bar{f})$ for the right-hand side of (6.31) we get that for $\bar{f}\neq f^*$, $W(\bar{f})<(\bar{\lambda}-\lambda^*)^\top q(f^*)$. From (6.2) we have $\lambda^{*\top}q(f^*)=0$ and $\bar{\lambda}^\top q(f^*)\leq 0$, which then implies $\nabla V(\bar{f},\bar{\lambda})X_{e\equiv 0}(\bar{f},\bar{\lambda})<0$ for almost all t with $\bar{f}(t)\neq f^*$. The rest of the proof is analogous to the corresponding section of the proof of Proposition 6.3.1. □

Remark 6.3.6. *(Implementation aspects of Proposition 6.3.5):* In Proposition 6.3.5 we require boundedness of $\{f^k\}$, $\{\lambda^k\}$. When upper bounds on $\|\lambda^*\|$ are known beforehand, projection onto hyper-rectangles can ensure boundedness of $\{\lambda^k\}$, while the result remains valid, (cf. [65, Page 40, Theorem 5.2.2]). For boundedness of $\{f^k\}$, see Remark 6.3.4. •

6.4 Estimation error, sample sizes and accuracy

In all of the provided algorithms, the convergence depends on the bias of the estimator, given by $b_{N_k}:=\mathbb{E}[\widehat{\beta}^{N_k}]$. When F is assumed to be strongly monotone, we can give an explicit bound on $\|b_{N_k}\|$ sufficient for ensuring convergence.

Corollary 6.4.1. *Assume that C is strongly monotone with constant c, and define*

$$b_{N_k}:=\mathbb{E}[\widehat{\beta}^{N_k}]$$

In addition let the conditions of Proposition 6.3.1 (respectively Proposition 6.3.3 and 6.3.5) hold, and let $\{f^k\}$ be generated by (6.7) *(respectively* (6.16) *and* (6.28)*). Then*

$$\|b_{N_k}\|<\epsilon c$$

implies $\lim_{k\to\infty}\|f^k-f^*\|\leq\epsilon$ with probability one.

Proof. In the proof of Proposition 6.3.1 we derive that the derivative of the Lyapunov function satisfies

$$\frac{dV}{dt}(t)\leq-\delta-\big(\bar{f}(t)-f^*\big)^\top e(t).$$

When F is strongly monotone we can replace $-\delta$ with $-c\|\bar{f}(t) - f^*\|^2$. Using the Cauchy-Schwarz inequality we then obtain

$$\begin{aligned}\frac{dV}{dt}(t) &= -c\|\bar{f}(t) - f^*\|^2 - \big(\bar{f}(t) - f^*\big)^\top e(t),\\ &\leq -c\|\bar{f}(t) - f^*\|^2 + \|\bar{f}(t) - f^*\|\|e(t)\|\\ &= -c\|\bar{f}(t) - f^*\|^2 + \|\bar{f}(t) - f^*\|\|b^{N_k}\|\\ &= \|\bar{f}(t) - f^*\|\big(\|b^{N_k}\| - c\|\bar{f}(t) - f^*\|\big).\end{aligned}$$

It is clear that this derivative is negative whenever $\|\bar{f}(t) - f^*\| > \frac{\|b^{N_k}\|}{c}$. This show that the derivative is negative when $\|b^{N_k}\| < \epsilon c$ and $\|\bar{f}(t) - f^*\| \geq \epsilon$, which shows the result for the case considering the projected algorithm. For the other two algorithms the proof is identical. □

The next step is to translate the bound on $\|b_{N_k}\|$ into a bound on N_k. In order to do so, we impose the condition that the random variables $c_i(f, u)$ are not supported outside the range $[z_1, z_2]$ for all f and i. We can then make use of the following result.

Lemma 6.4.2. *(Relation between estimation error and sample size): Let C be given by* (6.4)*, where $c_i(f, u) \in [z_1, z_2]$, $z_2 \geq z_1$, for all f, u and i. For $b^{N_k} = \mathbb{E}\big[C(f^k) - \widehat{C}^{N_k}(f^k)\big]$ we then have*

$$\|b^{N_k}\| \leq \frac{3}{2}\sqrt{\frac{5n\pi}{N_k\alpha}}(z_2 - z_1).$$

Proof. Let Z be any random variable supported on an interval $[a, b]$, and let $\widehat{N}$ be the number of samples taken to obtain $\widehat{\mathrm{CVaR}}_\alpha[Z]$. In [69, Chapter 6] it is shown that we have $\mathbb{E}\big[\widehat{\mathrm{CVaR}}_\alpha[Z]\big] \leq \mathrm{CVaR}_\alpha[Z]$. Therefore

$$\begin{aligned}\|\mathbb{E}\big[\,\mathrm{CVaR}_\alpha[Z] - \widehat{\mathrm{CVaR}}_\alpha[Z]\big]\| &= \mathbb{E}\big[\,\mathrm{CVaR}_\alpha[Z] - \widehat{\mathrm{CVaR}}_\alpha[Z]\big]\\ &\leq \mathbb{E}\big[\,\mathrm{CVaR}_\alpha[Z] - \widehat{\mathrm{CVaR}}_\alpha[Z]\big]_+.\end{aligned}$$

Here $[x]_+ := \max(0, x)$. From [75, Theorem 3.1], we have the concentration bound

$$\mathbb{P}\big[\,\mathrm{CVaR}_\alpha[Z] - \widehat{\mathrm{CVaR}}_\alpha[Z] \geq z\big] \leq 3e^{-\frac{1}{5}\alpha\left(\frac{z}{z_2 - z_1}\right)^2 \widehat{N}}.$$

It follows that

$$\begin{aligned}&\mathbb{E}\big[\,\mathrm{CVaR}_\alpha[Z] - \widehat{\mathrm{CVaR}}_\alpha[Z]\big]_+\\ &= \int_0^\infty \mathbb{P}\big[\,\mathrm{CVaR}_\alpha[Z] - \widehat{\mathrm{CVaR}}_\alpha[Z]\big] \geq z]dz\\ &\leq \frac{3}{2}\sqrt{\frac{5\pi}{\widehat{N}\alpha}}(b - a).\end{aligned}$$

Setting $Z = C_i(f, u)$, $a = z_1$, $b = z_2$ and $\widehat{N} = N_k$, and taking into account that $C(f^k)$, $\widehat{C}^{N_k}(n^k)$ and b^{N_k} are n-dimensional vectors instead of scalars, the above gives

$$\|b^{N_k}\| \leq \frac{3}{2}\sqrt{\frac{5n\pi}{N_k\alpha}}(z_2 - z_1).$$

□

Using the obtained relation between estimation error and sample size, we can now give a lower bound on N_k that ensures convergence to $\mathcal{N}_\epsilon(f^*)$.

Corollary 6.4.3. *(Sample size bounds under strong monotonicity): Assume that C is strongly monotone with constant c. In addition let the conditions of Proposition 6.3.1 (respectively Proposition 6.3.3 and 6.3.5) hold, and let $\{f^k\}$ be generated by* (6.7) *(respectively* (6.16) *and* (6.28)*). For $\epsilon > 0$, if*

$$N_k > \frac{45n\pi(z_2 - z_1)^2}{4\alpha\epsilon c_F} \text{ for all } k \in \mathbb{N},$$

then $\lim_{k\to\infty}\|f^k - f^*\| \leq \epsilon$ with probability one.

6.5 Simulations

Here we demonstrate an application of the presented stochastic approximation algorithms for finding the solutions of a CVaR-based variational inequality. The example is an instance of a CVaR-based routing game (see Section 6.2.1) based on the Sioux Falls network [76]. The network consists of 24 nodes and 76 edges. To each of the edges, we associate an affine cost function given by $c_e(f_e, u_e) = t_e(1 + u_e\frac{100}{a_e}f_e)$, where f_e is the flow over edge e, and t_e and a_e are the free-flow travel time and capacity of edge e, respectively, as obtained from [76]. The uncertainty u_e has the uniform distribution over the interval $[0, 0.5]$ for all edges connected to the vertices 10, 16, or 17. For the rest of the edges, u_e is set to zero. This defines the cost functions for all edges, and consequently defines the costs of all paths through the network as well. We consider three origin destination(OD) pairs $\mathcal{O} = \big\{(1, 19), (13, 8), (12, 18)\big\}$, and for each of these paths we select the ten paths that have the smallest free-flow travel time associated with them. The set of these 30 paths we denote as $\mathcal{P}$. The demands for each OD-pair are given by $d_{1,19} = 300$, $d_{13,8} = 600$, $d_{12,18} = 200$. We aim to find a CVaR-based Wardrop equilibrium which is equivalent to finding a solution of the VI problem defined by a map $C(f) := Af + b + \mathrm{CVaR}_\alpha[\xi]$, and a feasible set

$$\mathcal{F} = \{f \in \mathbb{R}^{30} \mid f \geq 0, \sum_{i=1}^{10} f_i = 300, \sum_{i=11}^{20} f_i = 600, \sum_{i=21}^{30} f_i = 200\}$$

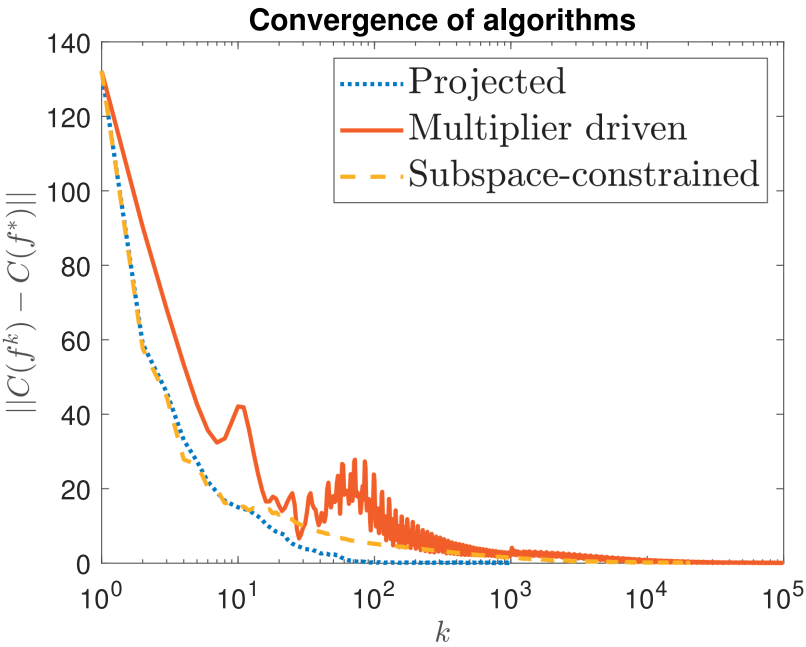


Figure 6.1: Plot illustrating the convergence of the algorithms for the routing example explained in Section 6.5. The initial condition for all algorithms is set as f_0 defined as $(f_0)_i = 30$ for $i \in \{1, \ldots, 10\}$, $(f_0)_i = 60$ for $i \in \{11, \ldots, 20\}$ and $(f_0)_i = 20$ for $i \in \{21, \ldots, 30\}$.

. Here, $f, b \in \mathbb{R}^{30}$, $A \in \mathbb{R}^{30\times 30}$, and $\alpha = 0.05$. The exact values of A and b and the distribution of ξ are constructed using the cost functions and the network structure, see [77, Section 6] for details.

In Figure 6.1, we see the evolution of the error for each of the different algorithms. The stepsize sequence for the projected, subspace-constrained, and multiplier-driven algorithms are $\gamma^k = \frac{100}{100+k}$, $\gamma^k = \frac{200}{200+k}$, and $\gamma^k = \min(\frac{100}{100+k}, \frac{1}{2})$, respectively. In addition, for the subspace-constrained algorithm we initially let c depend on k, to prevent unstable behaviour. We used $c = \min(\frac{1}{\gamma^k}, 200)$. For the multiplier driven algorithm, for similar reasons, we used a modified step-size sequence for updating the multipliers λ given by $\gamma^k_\lambda = 2\gamma^k$ for $k < 1000$ and $\gamma^k_\lambda = 0.5\gamma^k$, otherwise. The figure shows that all algorithms converge to a neighbourhood of the solution of the variational inequality, albeit requiring a different number of iterations. Specifically, the number of iterations taken by the projected algorithm to converge is two orders of magnitude less than that of the subspace-constrained and multiplier-driven algorithms. The quality of convergence is summarized in Table 6.1, where we can see both the accuracy of the achieved convergence as well as the effect of increasing the

Samples per iteration	25	50	100
Projected	0.3875	0.2062	0.1157
Subspace constrained	0.3780	0.2015	0.1332
Multiplier-driven	0.3889	0.1987	0.1064

Table 6.1: Table illustrating the performance of algorithms in the regime where iterates have converged to a significant level of accuracy. For the row related to the projected method, each number denotes the average error $\|F(h^k) - F(h^*)\|$ over iterates k after the error has become less than the value 0.6, 0.3 and 0.15, respectively, using 25, 50 and 100 samples in each iteration respectively. That is, using 25 samples in each iteration, the average error after the error hits a value 0.6 is 0.3875. The number of total iterations for the projected method is 1000. Similar average errors are denoted for subspace-constrained and multiplier-driven methods but the total number of iterates used are 50000 and 100000, respectively.

sample sizes. It is important to note that the errors shown in Fig. 6.1 and Table 6.1 are in terms of the deviation in the value of the map $\|C(f^k) - C(f^*)\|$, rather than the deviation in the solution $\|f^k - f^*\|$. This is because the solution f^* is not unique for the formulated VI. However, for any two solutions $f^*, \breve{f}^* \in \mathrm{SOL}(\mathcal{F}, C)$ we do have $C(\breve{f}^*) = C(f^*)$.

6.6 Conclusions

In this chapter we have studied variational inequalities involving a map that contains the conditional value at risk of uncertain costs. We have proposed three different algorithms for approximating the solution of these CVaR based VIs, and shown that the estimates produced by these algorithms converge asymptotically to the set of solutions of the VI, where any desired level of accuracy can be achieved by appropriately selecting the sample size used in estimating the CVaR of the unknown cost. We have also supplied an explicit relation between the achieved accuracy and the sample size. We have then compared the performance of these algorithms when employed to find the set of equilibria of a routing game example.

Chapter 7

Conclusions

In this thesis we studied some of the difficulties encountered when aiming to design and regulate routing games. In this conclusion we recall our main contributions and discuss some of the potential directions of future research.

7.1 Contributions

In Chapter 3 we investigate how the Wardrop equilibrium(WE) of a routing game is influenced by the total amount of demand that needs to be routed over the network. For the specific case of networks with affine cost functions over the edges we give a rigorous and complete characterization of how the set of WE evolves as the demand increases. Making use of the fact that the non-negative real line is divided into a finite number of intervals on which the evolution of the set of WE remains constant, we show how these results can be used to obtain an explicit expression for the WE in the final interval, as well as the lower endpoint of this final interval, without having to compute any WE for lower values of demand. In Chapter 4 we build upon the result of Chapter 3 to analyse the relationship between the total demand and the presence of Braess's paradox(BP). We have given several sufficient conditions for the presence of BP, which are computationally feasible to check. Some of these conditions relate in an interesting manner to the results from Chapter 3 on how information about WE in the final interval can be obtained. As a result some of these conditions supply very efficient methods for detecting sets of paths in the network that are 'suspect', meaning that they either cause BP at some level of demand, or can be removed from the routing game without changing the cost under WE at any level of demand. We also give a necessary and sufficient condition for the presence of BP, which is difficult to check in full, but still provides an efficient method to obtain upper-bounds on the achievable cost under WE for all demands, such that whenever the WE-cost exceeds such an upper bound, BP is known to occur. Finally we show how the obtained results grant a different perspective on BP. Instead of the paradox being a purely negative phenomenon, it turns out it is more a matter of balance. Any decrease in efficiency caused by a set of paths at some demand is accompanied by an increase in efficiency at another level of demand. Whether removing a set of paths from the network is beneficial or not, thus requires investigation of the effects at different levels of demand, as well as a choice of a measure that quantifies the value of a set of paths to the game.

In Chapter 5 we have studied routing games subject to uncertainty, in which participants of the routing game make use of a traffic information system(TIS) that supplies information about the current state of the network. The aim of a TIS is to minimize the average cost of the participants, and it can influence the game by selectively supplying information. We have studied how the potential for optimization by the TIS depends on the prior belief of the participants about the current state of the network. We have shown that under mild conditions, a TIS is able to use its information provision strategy to identify this prior belief based on observed equilibria, which then allows the TIS to achieve optimal performance by updating its information provision strategy. We also provided a design method for an information provision strategy that will allow the TIS to fully identify the prior belief.

In Chapter 6 we have explored the subject of variational inequalities(VIs) in which the involved mappings are the Conditional-Value-at-Risk(CVaR) of an uncertain cost, a framework that, among other things, can be used to model routing games with risk-averse participants. We have adapted three different algorithms for solving VIs to mka them suitable for solving CVaR-based VIs using stochastic approximation, where samples are used to estimate the CVaR of the uncertain costs. We have shown that each of these algorithms asymptotically converges to a neighbourhood of the solution, where the size of this neighbourhood can be made arbitrarily small by making the sample size sufficiently large. We have also given an explicit upper bound on the number of samples required to achieve a given level of accuracy, and have compared the performance of the three methods using an example of a routing game.

7.2 Future work

There are several potentially interesting research directions in which the material of this thesis can be extended. Starting with the results of Chapter 3, the most obvious step is to drop the assumption that the cost functions are affine, and consider more general (continuous, non-decreasing) cost functions. This is especially true since many of the functions that are important to modelling of traffic networks are not affine [78], and it would be useful to further enable the analysis of routing games involving these types of functions. For the specific case of strictly increasing cost functions such an extension of our results concerning the characterization of the evolution of the set of WE can already be derived from results in [12], and using the techniques presented in this thesis, an extension to non-decreasing functions seems plausibly obtainable. We note however that when the assumption of affine cost functions is dropped, there no longer necessarily exists a finite set of intervals on which the evolution of the WE is constant, and results concerning the final of these interval are therefore unlikely to be extended in this direction. Along similar lines an extension of the results of Chapter 4 to a case considering more general cost functions is also desirable. A number of the sufficient conditions for the presence of BP may have parallels in this context, though it is in this case perhaps more feasible to consider constraining the amount of demand a (set of) path(s) can facilitate instead of complete removal of a path.

Another interesting question is to what extent the results of Chapters 3 and 4, and specifically the necessary and sufficient condition for the presence of BP, can be used to derive subsets of paths that give restrictive or optimal upper bounds on the WE-cost, such the exceeding these bounds reveals the presence of BP. We note that this closely relates to finding the optimal set of *edges* in a network for minimizing the cost under WE, which has been proven to be an NP-hard problem [30]. Therefore this is expected to be a very difficult question to answer, but there is some intuition in the result on computing all WE in the final interval that is worth exploring. For instance, a simplistic idea would be to simply remove one or more of the paths that is used in the final interval, and has a relatively high contribution to the free flow cost β^M derived from (3.22). Any result on why this does or does not work could provide useful insights.

For the material presented in Chapter 5, one of the major shortcomings is the assumption that all participants share the same prior. However, we have already shown that the potential for public signalling schemes to address learning multiple unknown priors is limited. For future research we are thus interested in the question of whether different types of information provision strategies, such as private signalling schemes, can potentially address this problem.

Appendix A

Example of a routing game not representable by a graph

Consider a routing game defined over the network in Figure A.1, where $C_{e_k}(f_{e_k}) = f_{e_k}$ for all $k \in [6]$. We have

$$\begin{aligned} \mathcal{P} &= \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8\} \\ &:= \big\{(e_1, e_2, e_3), (e_1, e_2, e_6), (e_1, e_5, e_3), (e_1, e_5, e_6), \\ &\qquad (e_4, e_2, e_3), (e_4, e_2, e_6), (e_4, e_5, e_3), (e_4, e_5, e_6)\big\}, \end{aligned}$$

with a path-cost function given by $C(f) = Af$ where

$$A = \begin{pmatrix} 3 & 2 & 2 & 1 & 2 & 1 & 1 & 0 \\ 2 & 3 & 1 & 2 & 1 & 2 & 0 & 1 \\ 2 & 1 & 3 & 2 & 1 & 0 & 2 & 1 \\ 1 & 2 & 2 & 3 & 0 & 1 & 1 & 2 \\ 2 & 1 & 1 & 0 & 3 & 2 & 2 & 1 \\ 1 & 2 & 0 & 1 & 2 & 3 & 1 & 2 \\ 1 & 0 & 2 & 1 & 2 & 1 & 3 & 2 \\ 0 & 1 & 1 & 2 & 1 & 2 & 2 & 3 \end{pmatrix}.$$

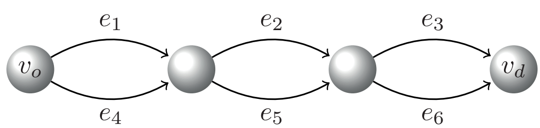


Figure A.1: Example for constructing a routing game not representable by a graph.

Now consider the modified game over only the first seven paths. That is, a routing game with a path-cost function given by $\breve{C}(\breve{f}) = \breve{A}\breve{f}$ where

$$\breve{A} = \begin{pmatrix} 3 & 2 & 2 & 1 & 2 & 1 & 1 \\ 2 & 3 & 1 & 2 & 1 & 2 & 0 \\ 2 & 1 & 3 & 2 & 1 & 0 & 2 \\ 1 & 2 & 2 & 3 & 0 & 1 & 1 \\ 2 & 1 & 1 & 0 & 3 & 2 & 2 \\ 1 & 2 & 0 & 1 & 2 & 3 & 1 \\ 1 & 0 & 2 & 1 & 2 & 1 & 3 \end{pmatrix}.$$

Our goal is now to show that there is no graph $\mathcal{G}$ that represents this routing game. We argue by contradiction, so let $\breve{\mathcal{G}}$ be a graph where $\breve{v}_o$ and $\breve{v}_d$ are the origin en destination in this graph and let $\breve{\mathcal{C}}$ be an associated set of cost functions over the edges such that the the resulting path-cost function is given by $\breve{C}(\breve{f}) = \breve{A}\breve{f}$. Note that this implies that there are only seven paths from origin to destination in this graph $\breve{\mathcal{G}}$. From the structure of A we can deduce the following:

$$\sum_{e_k \in p_1} \breve{\alpha}_{e_k} = 3, \qquad \sum_{e_k \in p_1 \cap p_2} \breve{\alpha}_{e_k} = 2$$
$$\sum_{e_k \in p_1 \cap p_7} \breve{\alpha}_{e_k} = 1, \qquad \sum_{e_k \in p_2 \cap p_7} \breve{\alpha}_{e_k} = 0.$$

From this we can deduce there exist sets of edges, suggestively denoted E_3 and $E_{1,2}$, such that

$$E_{1,2} \subseteq p_1 \cap p_2, \ \ E_{1,2} \cap p_7 = \emptyset, \sum_{e_k \in E_{1,2}} \breve{\alpha}_{e_k} = 2,$$
$$E_3 \subseteq p_1 \cap p_7, \ \ E_3 \cap p_2 = \emptyset, \quad \sum_{e_k \in E_3} \breve{\alpha}_{e_k} = 1,$$

and

$$E_{1,2} \cup E_3 = \{e_k \in p_1 \mid \breve{\alpha}_{e_k} > 0\}.$$

Similarly, we can show that there exist sets of edges, denoted $E_{1,3}$ and E_2 such that

$$E_{1,3} \subseteq p_1 \cap p_3, \ \ E_{1,3} \cap p_6 = \emptyset, \sum_{e_k \in E_{1,3}} \breve{\alpha}_{e_k} = 2,$$
$$E_2 \subseteq p_1 \cap p_6, \ \ E_2 \cap p_3 = \emptyset, \quad \sum_{e_k \in E_2} \breve{\alpha}_{e_k} = 1,$$

and

$$E_{1,3} \cup E_2 = \{e_k \in p_1 \mid \breve{\alpha}_{e_k} > 0\}.$$

Combining these observations with the fact that $\sum_{e_k \in p_2 \cap p_3} \breve{\alpha}_{e_k} = 1$, it follows that in fact there exist sets of edges E_1, E_2 and E_3 such that

$$\begin{aligned}
&E_1 \subseteq p_1 \cap p_2 \cap p_3, \quad E_1 \cap (p_6 \cup p_7) = \emptyset, \sum_{e_k \in E_1} \breve{\alpha}_{e_k} = 1,\\
&E_2 \subseteq p_1 \cap p_2 \cap p_6, \quad E_2 \cap (p_3 \cup p_7) = \emptyset, \sum_{e_k \in E_2} \breve{\alpha}_{e_k} = 1,\\
&E_3 \subseteq p_1 \cap p_3 \cap p_7, \quad E_3 \cap (p_2 \cup p_3) = \emptyset, \sum_{e_k \in E_3} \breve{\alpha}_{e_k} = 1.
\end{aligned}$$

and

$$E_1 \cup E_2 \cup E_3 = \{e_k \in p_1 \mid \breve{\alpha}_{e_k} > 0\}.$$

Repeating this line of reasoning for p_4 and p_5 we find

$$\begin{aligned}
&E_1 \subseteq p_1 \cap p_2 \cap p_3 \cap p_4, \quad E_1 \cap (p_5 \cup p_6 \cup p_7) = \emptyset,\\
&E_2 \subseteq p_1 \cap p_2 \cap p_5 \cap p_6, \quad E_2 \cap (p_3 \cup p_4 \cup p_7) = \emptyset,\\
&E_3 \subseteq p_1 \cap p_3 \cap p_5 \cap p_7, \quad E_3 \cap (p_2 \cup p_3 \cup p_4) = \emptyset.
\end{aligned}$$

Further analysing the structure imposed by $\breve{A}$ using the same type of arguments, we find that there must also exist sets of edges E_4, E_5, and E_6 such that

$$\begin{aligned}
&E_4 \subseteq p_5 \cap p_6 \cap p_7, \quad E_4 \cap (p_1 \cup p_2 \cup p_3 \cup p_4) = \emptyset,\\
&E_5 \subseteq p_3 \cap p_4 \cap p_7, \quad E_5 \cap (p_1 \cup p_2 \cup p_5 \cup p_6) = \emptyset,\\
&E_6 \subseteq p_2 \cap p_4 \cap p_6, \quad E_6 \cap (p_1 \cup p_3 \cup p_5 \cup p_7) = \emptyset,
\end{aligned}$$

and

$$\sum_{e_k \in E_4} \breve{\alpha}_{e_k} = \sum_{e_k \in E_5} \breve{\alpha}_{e_k} = \sum_{e_k \in E_6} \breve{\alpha}_{e_k} = 1.$$

It follows that none of the sets in $\{E_k\}_{k \in [6]}$ overlap. Therefore, there exist at least six edges in $\breve{\mathcal{G}}$, which satisfy

$$\begin{aligned}
&\breve{e}_1 \in p_1 \cap p_2 \cap p_3 \cap p_4, && \breve{e}_1 \notin p_5 \cup p_6 \cup p_7,\\
&\breve{e}_2 \in p_1 \cap p_2 \cap p_5 \cap p_6, && \breve{e}_2 \notin p_3 \cup p_4 \cup p_7,\\
&\breve{e}_3 \in p_1 \cap p_3 \cap p_5 \cap p_7, && \breve{e}_3 \notin p_2 \cup p_3 \cup p_4,\\
&\breve{e}_4 \in p_5 \cap p_6 \cap p_7, && \breve{e}_4 \notin p_1 \cup p_2 \cup p_3 \cup p_4,\\
&\breve{e}_5 \in p_3 \cap p_4 \cap p_7, && \breve{e}_5 \notin p_1 \cup p_2 \cup p_5 \cup p_6,\\
&\breve{e}_6 \in p_2 \cap p_4 \cap p_6, && \breve{e}_6 \notin p_1 \cup p_3 \cup p_5 \cup p_7.
\end{aligned}$$

For the next part, we assume that the edges $\breve{e}_1$, $\breve{e}_2$ and $\breve{e}_3$ appear in that order in the path p_1. That is, when traversing p_1, we first encounter $\breve{e}_1$, then $\breve{e}_2$ and finally $\breve{e}_3$.

The upcoming arguments can be repeated for any other ordering to arrive at similar contradictions.

Note that given this ordering, we know that there exists a path from $\breve{v}_1^{\text{out}}$ to $\breve{v}_2^{\text{in}}$, and a path from $\breve{v}_2^{\text{out}}$ to $\breve{v}_3^{\text{in}}$. In addition, since p_5 traverses $\breve{e}_2$ but not $\breve{e}_1$ there must exist a path from $\breve{v}_o$ to $\breve{v}_2^{\text{in}}$ that doesn't traverse $\breve{e}_1$. Thus there exists a path that does not contain $\breve{e}_1$ but does contain $\breve{e}_2$ and $\breve{e}_3$. The only option is that this is p_5, and that it goes through the edges $\breve{e}_4$, $\breve{e}_2$, $\breve{e}_3$ in that order. We see that there exists a path from $\breve{v}_o$ to $\breve{v}_4^{\text{in}}$ and from $\breve{v}_4^{\text{out}}$ to $\breve{v}_2^{\text{in}}$. Next note that since p_7 traverses $\breve{e}_3$ but not $\breve{e}_1$ and $\breve{e}_2$ there must exists a path from $\breve{v}_o$ to $\breve{v}_3^{\text{in}}$, and by similar arguments as before we find that $\breve{e}_4$ and $\breve{e}_5$ are part of this path. If this goes from $\breve{v}_o$ to $\breve{v}_5^{\text{in}}$ and then from $\breve{v}_5^{\text{out}}$ to $\breve{v}_4^{\text{in}}$ then there exists a path containing $\breve{e}_5$, $\breve{e}_4$, $\breve{e}_2$, which is a contradiction. Thus this path goes instead from $\breve{v}_o$ to $\breve{v}_4^{\text{in}}$ and then from $\breve{v}_4^{\text{out}}$ to $\breve{v}_5^{\text{in}}$. Thus there exists a path from $\breve{v}_4^{\text{out}}$ to $\breve{v}_5^{\text{in}}$. With the same reasoning considering p_4 instead of p_7, we can see that there must exist a path from $\breve{v}_5^{\text{out}}$ to $\breve{v}_6^{\text{in}}$, and a path from $\breve{v}_6^{\text{out}}$ to $\breve{v}_d$. Taking all this together, there exists a path containing $\breve{e}_4$, $\breve{e}_5$ and $\breve{e}_6$, which is once again a contradiction. We arrive at the conclusion that our premise is false, and that the given graph, origin and destination and cost functions do not induce a routing game for which $\breve{C}(\breve{f}) = \breve{A}\breve{f}$, and therefore there exists no combination of a graph, a single origin and destination pair and cost functions that does.

Appendix B

Proof of Corollary 3.1.1

Corollary. *(Piecewise constant evolution of active and used sets): Let $\mathcal{P}$ and $\mathcal{C} \subset \mathcal{K}$ be given. There exists a finite set of points $\mathcal{D} := (D_0, D_1, \cdots, D_M, D_{M+1}) \subset \mathbb{R}_{\geq 0} \cup \{+\infty\}$ with $D_0 = 0$, $D_{M+1} = \infty$ and $D_j > D_{j-1}$ for all $j \in [M+1]$, and corresponding sets of subsets of $\mathcal{P}$ denoted $\{\mathcal{J}_0^{\text{act}}, \mathcal{J}_1^{\text{act}}, \cdots, \mathcal{J}_M^{\text{act}}\}$ and $\{\mathcal{J}_0^{\text{use}}, \mathcal{J}_1^{\text{use}}, \cdots, \mathcal{J}_M^{\text{use}}\}$, such that, for all $i \in [M]_0$ and $D \in (D_i, D_{i+1})$, we have*

$$\mathcal{R}_D^{\text{act}} = \mathcal{J}_i^{\text{act}}, \quad \mathcal{R}_D^{\text{use}} = \mathcal{J}_i^{\text{use}}.$$

Furthermore, $\mathcal{J}_i^{\text{act}} \neq \mathcal{J}_j^{\text{act}}$ and $\mathcal{J}_i^{\text{use}} \neq \mathcal{J}_j^{\text{use}}$ for all $i \neq j$.

Proof. The claim about active sets $\mathcal{R}_D^{\text{act}}$ is shown in [9, Section 4]. That is, there exist points $\mathcal{D} := (D_0, D_1, \cdots, D_M, D_{M+1})$ with $D_0 = 0$, $D_{M+1} = \infty$ and $D_j > D_{j-1}$ for all $j \in [M+1]$ along with sets $\{\mathcal{J}_0^{\text{act}}, \mathcal{J}_1^{\text{act}}, \cdots, \mathcal{J}_M^{\text{act}}\}$ such that $\mathcal{R}_D^{\text{act}} = \mathcal{J}_i^{\text{act}}$ for all $i \in [M]$ and $D \in (D_i, D_{i+1})$. Using this result and the defined points in $\mathcal{D}$ we will show the existence of sets $\{\mathcal{J}_0^{\text{use}}, \mathcal{J}_1^{\text{use}}, \cdots, \mathcal{J}_M^{\text{use}}\}$ such that $\mathcal{R}_D^{\text{use}} = \mathcal{J}_i^{\text{use}}$ for all $D \in (D_i, D_{i+1})$ and all i. Pick some $D_i, D_{i+1} \in \mathcal{D}$. Let $D \in (D_i, D_{i+1})$ and $p \in \mathcal{R}_D^{\text{use}}$. Consider any other demand $D' \in (D_i, D_{i+1})$ with $D' \neq D$. Associated to D and D', select T such that $T \in (D_i, D_{i+1})$ and D' can be written as a convex combination of T and D. That is, $D' = \mu D + (1-\mu)T$ for some $\mu \in (0,1)$. Since $p \in \mathcal{R}_D^{\text{use}}$, there exists f^D such that $f_p^D > 0$. Furthermore, $f_p^T \geq 0$. Since $\mathcal{R}_D^{\text{act}} = \mathcal{R}_T^{\text{act}}$, from Lemma 3.1.2-2, there exists $f^{D'} \in \mathcal{W}_{D'}$ such that $f^{D'} = \mu f^D + (1-\mu) f^T$ and so $f_p^{D'} > 0$. Thus, $p \in \mathcal{R}_{D'}^{\text{use}}$. Since D and D' were selected arbitrarily, we conclude that the used set remains the same in the interval (D_i, D_{i+1}).

Now, for $D_i, D_{i+1}, D_j, D_{j+1} \in \mathcal{D}$ with $i \neq j$ let D^-, D^+ satisfy $D^- \in (D_i, D_{i+1})$ and $D^+ \in (D_j, D_{j+1})$. Note that therefore $\mathcal{R}_{D^-}^{\text{act}} \neq \mathcal{R}_{D^+}^{\text{act}}$. For the sake of contradiction assume that $\mathcal{R}_{D^-}^{\text{use}} = \mathcal{R}_{D^+}^{\text{use}}$. For any $p \in \mathcal{R}_{D^-}^{\text{use}}$ we know that $C_p(f^{D^-}) \leq C_r(f^{D^-})$ for all $r \in \mathcal{P}$ and $f^{D^-} \in \mathcal{W}_{D^-}$. Similarly $C_p(f^{D^+}) \leq C_r(f^{D^+})$ for all $r \in \mathcal{P}$ and $f^{D^+} \in \mathcal{W}_{D^+}$. Since $C(\cdot)$ is affine(see (3.2)) it follows that for $f^T = \text{coco}_\mu(f^{D^-}, f^{D^+})$ with $f^{D^-} \in \mathcal{W}_{D^-}$, $f^{D^+} \in \mathcal{W}_{D^+}$, $\mu \in [0,1]$ and $T = \text{coco}_\mu(D^-, D^+)$ we have that $C_p(f^T) \leq C_r(f^T)$ for all $p \in \mathcal{R}_{D^-}^{\text{use}}$. Also note that $f_p^T > 0$ implies $f_p^{D^-} > 0$ or $f_p^{D^+} > 0$, which in turn gives us $p \in \mathcal{R}_{D^-}^{\text{use}}$. In other words f^T is a WE. Since $\mathcal{R}_{D^-}^{\text{act}} \neq \mathcal{R}_{D^+}^{\text{act}}$ we can without loss of generality pick $r \in \mathcal{P}$ such that $r \in \mathcal{R}_{D^-}^{\text{act}}$ and

$r \notin \mathcal{R}^{\mathrm{act}}_{D^+}$. In other words, the cost of path p is minimal at the demand D^-, but not minimal at demand D^+. Using (3.2) and the fact that f^T is a WE it follows that the cost of r is not minimal at demand $T = \mathrm{coco}_\mu(D^-, D^+)$. Since $D^- \in (D_i, D_{i+1})$ we can set μ such that $T \in (D_i, D_{i+1})$. This would then imply that $\mathcal{R}^{\mathrm{act}}_T \neq \mathcal{R}^{\mathrm{act}}_{D^-}$, which contradicts the already established results concerning the active set. Therefore our premise is false, which shows that $\mathcal{R}^{\mathrm{use}}_{D^-} \neq \mathcal{R}^{\mathrm{use}}_{D^+}$, finishing the proof. $\square$

Appendix C

WE and WE-costs of routing games in Example 4.3.6

Consider the original routing game in Example 4.3.6. The path-cost function is given by

$$C(f) = Af + b = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 2 & 3 \\ 1 & 1 & 2 & 3 & 3 \\ 1 & 1 & 3 & 3 & 4 \end{pmatrix} f + \begin{pmatrix} 7 \\ 7 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

The WE for this game are given by

$$f^D = \begin{cases} \begin{pmatrix} 0 & 0 & 0 & 0 & D \end{pmatrix}^\top & D \in [0,1], \\ \begin{pmatrix} 0 & 0 & D-1 & D-1 & 2-D \end{pmatrix}^\top & D \in [1,2], \\ \begin{pmatrix} 0 & 0 & \frac{D}{2} & \frac{D}{2} & 0 \end{pmatrix}^\top & D \in [2,4], \\ \begin{pmatrix} \frac{3D-12}{4} & \frac{3D-12}{4} & \frac{12-D}{4} & \frac{12-D}{4} & 0 \end{pmatrix}^\top & D \in [4,8], \\ \begin{pmatrix} D-5 & D-5 & 9-D & 9-D & D-8 \end{pmatrix}^\top & D \in [8,9], \\ \begin{pmatrix} \frac{3D-7}{5} & \frac{3D-7}{5} & 0 & 0 & \frac{14-D}{5} \end{pmatrix}^\top & D \in [8,14], \\ \begin{pmatrix} \frac{D}{2} & \frac{D}{2} & 0 & 0 & 0 \end{pmatrix}^\top & D \in [14,\infty). \end{cases}$$

For the modified game with path p_5 removed the WE are given by

$$\widetilde{f}^D = \begin{cases} \begin{pmatrix} 0 & 0 & \frac{D}{2} & \frac{D}{2} \end{pmatrix}^\top & \text{for } D \in [0,4], \\ \begin{pmatrix} \frac{3D-12}{4} & \frac{3D-12}{4} & \frac{12-D}{4} & \frac{12-D}{4} \end{pmatrix}^\top & \text{for } D \in [4,12], \\ \begin{pmatrix} \frac{D}{2} & \frac{D}{2} & 0 & 0 \end{pmatrix}^\top & \text{for } D \in [12,\infty). \end{cases}$$

For the last case, with paths p_3, p_4 and p_5 the WE satisfy

$$\breve{f}^D = \begin{pmatrix} \frac{D}{2} & \frac{D}{2} \end{pmatrix}^\top,$$

for $D \geq 0$. The associated WE-costs are

$$\lambda^{\mathrm{WE}}(D) = \begin{cases} 4D & \text{for } D \in \quad [0,1], \\ 2D+2 & \text{for } D \in \quad [1,2], \\ \frac{5D}{2}+1 & \text{for } D \in \quad [2,4], \\ \frac{D}{4}+10 & \text{for } D \in \quad [4,8], \\ 12 & \text{for } D \in \quad [8,9], \\ \frac{2D}{5}+8\frac{2}{5} & \text{for } D \in \ [9,14], \\ \frac{D}{2}+7 & \text{for } D \in [14,\infty), \end{cases}$$

$$\widetilde{\lambda}^{\mathrm{WE}}(D) = \begin{cases} \frac{5D}{2}+1 & \text{for } D \in \quad [0,4], \\ \frac{D}{4}+10 & \text{for } D \in \ [4,12], \\ \frac{D}{2}+7 & \text{for } D \in [12,\infty), \end{cases}$$

$$\breve{\lambda}^{\mathrm{WE}}(D) = \frac{D}{2}+7 \quad \text{for } D \geq 0.$$

Appendix D

WE of routing game in Example 4.4.1

Consider the routing game in Example 4.4.1, defined by the path-cost function

$$C(f) = Af + b = \begin{pmatrix} 3 & 0 & 2 & 0 \\ 0 & 4 & 2 & 1 \\ 2 & 2 & 4 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix} f + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 5 \end{pmatrix}.$$

The WE for this game are given by

$$f^D = \begin{cases} \begin{pmatrix} 0 & 0 & D & 0 \end{pmatrix}^\top & \text{for } D \in [0, \frac{1}{2}], \\ \begin{pmatrix} \frac{8D-4}{12} & \frac{6D-3}{12} & \frac{2D-7}{12} & 0 \end{pmatrix}^\top & \text{for } D \in [\frac{1}{2}, \frac{7}{2}], \\ \begin{pmatrix} \frac{4D}{7} & \frac{3D}{7} & 0 & 0 \end{pmatrix}^\top & \text{for } D \in [\frac{7}{2}, \frac{35}{9}], \\ \begin{pmatrix} \frac{7D+15}{19} & \frac{3D+20}{19} & 0 & \frac{9D-35}{19} \end{pmatrix}^\top & \text{for } D \in [\frac{35}{9}, 6], \\ \begin{pmatrix} \frac{10D+27}{29} & \frac{4D+34}{29} & \frac{D-6}{29} & \frac{14D-55}{29} \end{pmatrix}^\top & \text{for } D \in [6, \infty). \end{cases}$$

Bibliography

[1] D. Braess, “Über ein paradoxon aus der verkehrsplanung,” *Unternehmensforschung*, vol. 12, pp. 258–268, 1968.

[2] D. Braess, A. Nagurney, and T. Wakolbinger, “On a paradox of traffic planning,” *Transportation science*, vol. 39, no. 4, pp. 446–450, 2005.

[3] A. Nagurney, D. Parkes, and P. Daniele, “The internet, evolutionary variational inequalities, and the time-dependent braess paradox,” *Computational Management Science*, vol. 4, pp. 355–375, 2007.

[4] J. E. Cohen and P. Horowitz, “Paradoxical behaviour of mechanical and electrical networks,” *Nature*, vol. 352, no. 6337, pp. 699–701, 1991.

[5] T. Roughgarden, “Routing games,” *Algorithmic game theory*, vol. 18, pp. 459–484, 2007.

[6] M. J. Smith, “The existence, uniqueness and stability of traffic equilibria,” *Transportation Research Part B: Methodological*, vol. 13, no. 4, pp. 295–304, 1979.

[7] J. R. Correa and N. E. Stier-Moses, “Wardrop equilibria,” *Encyclopedia of Operations Research and Management Science. Wiley*, 2011.

[8] T. Roughgarden and E. Tardos, “How bad is selfish routing?” *Journal of the ACM*, vol. 49, no. 2, pp. 236–259, 2002.

[9] R. Cominetti, V. Dose, and M. Scarsini, “The price of anarchy in routing games as a function of the demand,” *Mathematical Programming*, pp. 1–28, 2021.

[10] M. A. Hall, “Properties of the equilibrium state in transportation networks,” *Transportation Science*, vol. 12, no. 3, pp. 208–216, 1978.

[11] C. Fisk, “More paradoxes in the equilibrium assignment problem,” *Transportation Research Part B: Methodological*, vol. 13, no. 4, pp. 305–309, 1979.

[12] M. Patriksson, “Sensitivity analysis of traffic equilibria,” *Transportation Science*, vol. 38, no. 3, pp. 258–281, 2004.

[13] J.-P. Aubin and H. Frankowska, *Set-valued analysis*. Springer Science & Business Media, 2009.

[14] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer Science & Business Media, 2009, vol. 317.

[15] S. M. Robinson, "Solution continuity in monotone affine variational inequalities," *SIAM Journal on Optimization*, vol. 18, no. 3, pp. 1046–1060, 2007.

[16] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems, volume I*. Springer Science & Business Media, 2003.

[17] D. Braess, "https://homepage.ruhr-uni-bochum.de/dietrich.braess/," 2023.

[18] J. D. Murchland, "Braess's paradox of traffic flow," *Transportation Research*, vol. 4, no. 4, pp. 391–394, 1970.

[19] N. Stewart, "Equilibrium vs system-optimal flow: some examples," *Transportation Research Part A: General*, vol. 14, no. 2, pp. 81–84, 1980.

[20] M. Frank, "The braess paradox," *Mathematical Programming*, vol. 20, no. 1, pp. 283–302, 1981.

[21] R. Steinberg and W. I. Zangwill, "The prevalence of braess' paradox," *Transportation Science*, vol. 17, no. 3, pp. 301–318, 1983.

[22] S. Dafermos and A. Nagurney, "On some traffic equilibrium theory paradoxes," *Transportation Research Part B: Methodological*, vol. 18, no. 2, pp. 101–110, 1984.

[23] E. I. Pas and S. L. Principio, "Braess' paradox: Some new insights," *Transportation Research Part B: Methodological*, vol. 31, no. 3, pp. 265–276, 1997.

[24] I. Milchtaich, "Network topology and the efficiency of equilibrium," *Games and Economic Behavior*, vol. 57, no. 2, pp. 321–346, 2006.

[25] G. Valiant and T. Roughgarden, "Braess's paradox in large random graphs," in *Proceedings of the 7th ACM conference on Electronic commerce*, 2006, pp. 296–305.

[26] L. Baker, "Removing roads and traffic lights speeds urban travel," *Scientific American*, pp. 20–21, 2009.

[27] J. N. Hagstrom and R. A. Abrams, "Characterizing braess's paradox for traffic networks," in *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585)*. IEEE, 2001, pp. 836–841.

[28] C. Ma, Q. Cai, S. Alam, B. Sridhar, and V. N. Duong, "Airway network management using braess's paradox," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 565–579, 2019.

[29] S. A. Bagloee, A. Ceder, M. Tavana, and C. Bozic, "A heuristic methodology to tackle the braess paradox detecting problem tailored for real road networks," *Transportmetrica A: Transport Science*, vol. 10, no. 5, pp. 437–456, 2014.

[30] T. Roughgarden, "On the severity of braess's paradox: Designing networks for selfish users is hard," *Journal of Computer and System Sciences*, vol. 72, no. 5, pp. 922–953, 2006.

[31] V. Zverovich and E. Avineri, "Braess' paradox in a generalised traffic network," *Journal of Advanced Transportation*, vol. 49, no. 1, pp. 114–138, 2015.

[32] J. Macfarlane, "Your navigation app is making traffic unmanageable," *IEEE Spectrum*, pp. 22–27, 2019.

[33] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Informational Braess' paradox: The effect of information on traffic congestion," *Operations Research*, vol. 66, no. 4, pp. 893–917, 2018. [Online]. Available: https://doi.org/10.1287/opre.2017.1712

[34] E. Kamenica and M. Gentzkow, "Bayesian persuasion," *American Economic Review*, vol. 101, no. 6, pp. 2590–2615, 2011.

[35] S. Das, E. Kamenica, and R. Mirka, "Reducing congestion through information design," in *allerton*, 2017, pp. 1279–1284.

[36] S. Vasserman, M. Feldman, and A. Hassidim, "Implementing the wisdom of Waze," in *International Joint Conference on Artificial Intelligence*, 2015, pp. 660–666.

[37] O. Massicot and C. Langbort, "On the comparative performance of information provision policies in network routing games," in *Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 1434–1438.

[38] H. Tavafoghi and D. Teneketzis, "Strategic information provision in routing games," https://hamidtavaf.github.io/infodesign_routing.pdf, 2019.

[39] Y. Zhu and K. Savla, "On the stability of optimal Bayesian persuasion strategy under a mistrust dynamics in routing games," in *allerton*, 2018, pp. 92–99.

[40] ——, "A semidefinite approach to information design in non-atomic routing games," *arXiv preprint arXiv:2005.03000*, 2020.

[41] J. Liu, "On traffic disruptions: event detection from visual data and bayesian congestion games," Ph.D. dissertation, Massachusetts Institute of Technology, 2019.

[42] M. Wu, S. Amin, and A. E. Ozdaglar, "Value of information in Bayesian routing games," *Operations Research*, vol. 69, no. 1, pp. 148–163, 2021.

[43] H. Tavafoghi, A. Shetty, K. Poolla, and P. Varaiya, "Strategic information platforms in transportation networks," in *allerton*. IEEE, 2019, pp. 816–823.

[44] G. Bianchin and F. Pasqualetti, "Routing apps may deteriorate stability in traffic networks: Oscillating congestions and robust information design," *arXiv preprint arXiv:2003.10018*, 2020.

[45] R. Lindsey, T. Daniel, E. Gisches, and A. Rapoport, "Pre-trip information and route-choice decisions with stochastic travel conditions: Theory," *Transportation Research Part B: Methodological*, vol. 67, pp. 187–207, 2014.

[46] K. Lam, W. Krichene, and A. Bayen, "On learning how players learn: estimation of learning dynamics in the routing game," in *ACM/IEEE International Conference on Cyber-Physical Systems*, 2016, pp. 1–10.

[47] J. Thai and A. Bayen, "Learnability of edge cost functions in routing games," in *cdc*, 2017, pp. 6422–6429.

[48] L. J. Ratliff and T. Fiez, "Adaptive incentive design," *IEEE Transactions on Automatic Control*, vol. 66, no. 8, pp. 3871–3878, 2021.

[49] S. Krichene, W. Krichene, R. Dong, and A. Bayen, "Convergence of heterogeneous distributed learning in stochastic routing games," in *allerton*, 2015, pp. 480–487.

[50] E. Meigs, F. Parise, and A. Ozdaglar, "Learning dynamics in stochastic routing games," in *allerton*, 2017, pp. 259–266.

[51] M. Wu and S. Amin, "Learning an unknown network state in routing games," *IFAC-PapersOnLine*, vol. 52, no. 20, pp. 345–350, 2019.

[52] P. N. Brown and J. R. Marden, "Studies on robust social influence mechanisms: Incentives for efficient network routing in uncertain settings," *IEEE Control Systems Magazine*, vol. 37, no. 1, pp. 98–115, 2017.

[53] B. L. Ferguson, P. N. Brown, and J. R. Marden, "Avoiding unintended consequences: How incentives aid information provisioning in bayesian congestion games," *arXiv preprint arXiv:2204.06046*, 2022.

[54] O. Massicot and C. Langbort, "Public signals and persuasion for road network congestion games under vagaries," *IFAC-PapersOnLine*, vol. 51, no. 34, pp. 124–130, 2019, iFAC Conference on Cyber-Physical and Human Systems 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896319300564

[55] M. J. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. Rand Corporation, 1955.

[56] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.

[57] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, "Network flows," Ph.D. dissertation, Cambridge, Mass.: Alfred P. Sloan School of Management, Massachusetts ..., 1988.

[58] U. V. Shanbhag, "Stochastic variational inequality problems: Applications, analysis, and algorithms," *INFORMS TutORials in Operations Research*, pp. 71–107, 2013.

[59] Y. Chen, G. Lan, and Y. Ouyang, "Accelerated schemes for a class of variational inequalities," *Mathematical Programming*, vol. 165, p. 113–149, 2014.

[60] R. Russo and N. Shyamalkumar, "Bounds for the bias of the empirical CTE," *Insurance: Mathematics and Economics*, vol. 47, pp. 352–357, 2010.

[61] U. Ravat, "On the analysis of stochastic optimization and variational inequality problems," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2014.

[62] A. Cherukuri, "Sample average approximation of CVaR-based wardrop equilibrium in routing under uncertain costs," in *IEEE Conference on Decision and Control*, 2019, pp. 3164–3169.

[63] A. Tamar, Y. Chow, M. Ghavamzadeh, and S. Mannor, "Sequential decision making with coherent risk," *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 332–3338, 2017.

[64] C. Jie, L. Prashanth, M. Fu, S. Marcus, and C. Szepesvári, "Stochastic optimization in a cumulative prospect theory framework," *IEEE Transactions on Automatic Control*, vol. 42, no. 4, pp. 2867–2882, 2018.

[65] H. J. Kushner and D. S. Clark, *Stochastic approximation methods for constrained and unconstrained systems*, ser. Appl. Math. Sci. New York, NY: Springer, 1978.

[66] V. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Hindustan Book Agency, 2008.

[67] D. Steck, "Lagrange multiplier methods for contraint optimization and variational problems in banach spaces," Ph.D. dissertation, Julius Maximilians Universität Würzburg, 2018.

[68] U. Faigle, W. Kern, and G. Still, *Algorithmic Principles of Mathematical Programming*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2010, vol. 24.

[69] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on stochastic programming*. Philadelphia, PA: SIAM, 2014.

[70] M. Shakarami, A. Cherukuri, and N. Monshizadeh, "Steering the aggregative behavior of noncooperative agents: a nudge framework," *Automatica*, vol. 136, p. 110003, 2022.

[71] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, ser. Grundlehren Text Editions. New York: Springer, 1993.

[72] D. C. Lay, S. R. Lay, and J. J. McDonald, *Linear algebra and its applications, fifth edition*. Pearson, 2016.

[73] C. D. Meyer, *Matrix analysis and applied linear algebra*. Siam, 2000, vol. 71.

[74] J. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer-Verlag Berlin Heidelberg New York, 2001.

[75] Y. Wang and F. Gao, "Deviation inequalities for an estimator of the conditional value-at-risk," *Operations Research Letters*, vol. 38, pp. 236–239, 2010.

[76] Transportation Networks for Research Core Team., "Transportation networks for research," https://github.com/bstabler/TransportationNetworks. Accessed on 12.08.2020.

[77] A. Cherukuri, "Sample average approximation of conditional value-at-risk based variational inequalities," *Optimization Letters*, pp. 1–26, 2023.

[78] L. F. Huntsinger and N. M. Rouphail, "Bottleneck and queuing analysis: Calibrating volume–delay functions of travel demand models," *Transportation research record*, vol. 2255, no. 1, pp. 117–124, 2011.

Summary

We start this thesis with an analysis of how the set of Wardrop equilibria(WE) of a routing game with affine cost functions is influenced by the total demand that needs to be routed over the network. We give a rigorous analysis of the effects of changing demand on various aspects of the game, such as the set of paths that carry flow and the cost under WE, culminating in a full characterization of the set of directions in which the WE changes as the demand increases. We also use these results to show how one can directly compute the WE flow for high levels of demand. Next we turn our attention to the subject of Braess's paradox(BP), and use the results for varying demand to give different ways in which the evolution of the set of WE can reveal the presence of BP. We also translate these results into a necessary and sufficient condition that allows us to construct affine functions which serve as upper bounds on the WE-cost for all levels of demand, such that exceeding one of these bounds implies the presence of BP. We then discuss how our results reveal that BP is not a purely negative phenomenon. We show that any inefficiency caused by a set of paths at some demand is related to an increase in efficiency at some other level of demand, thus care should be taken with removing paths from a network, even when it is known that they cause BP.

Then we turn our attention to routing games involving uncertain costs, in which the participants rely on a central traffic information system(TIS) for information about the current state of the network. The TIS tries to leverage its position to minimize the average travel time of all agents. We show how the performance of the TISs strategies depends on the knowledge it has about the prior belief that the population has about the state of the network, and give mild conditions under which the TIS can use observations of the WE of the game to identify this prior. We also provide a way in which the TIS can design its strategy to learn this prior.

Finally we take a look at the concept of variational inequalities(VIs), an important tool in the study of routing games, and the challenges of computing the solutions of these VIs when the involved mappings are given by the Conditional-Value-at-Risk(CVaR) of an uncertain cost. We study three different stochastic approximation algorithms for solving these types of problems, and show that they converge to a neighbourhood of the solution, where the size of the neighbourhood can be tuned by controlling the number of samples taken at each iteration. We also give an upper bound on the required sample size for achieving a given precision.

Samenvatting

We beginnen met een analyse van de invloed die verandering in de totale vraag die over een netwerk moet worden getransporteerd heeft op de set van Wardrop equilibria(WE) van een navigatiespel. We geven een grondige analyse van het effect van een veranderende vraag op verschillende aspecten van het spel, zoals de set aan paden waar verkeer aan wordt toegewezen en de transportkosten die worden ervaren in een WE, en geven uiteindelijk een volledige karakterisatie van de richtingen waarin de set van WE zich beweegt als de vraag toeneemt. Ook gebruiken we deze resultaten om een direct methode te ontwikkelen voor het berekenen van WE als de vraag hoog is. Vervolgens richten we ons op het fenomeen "Braess's paradox"(BP), en gebruiken de resultaten voor variërende vraag om verschillende manieren te presenteren waarop de evolutie van de set van WE de aanwezigheid van BP kan onthullen. We vertalen deze resultaten ook naar een noodzakelijke en afdoende voorwaarde voor de aanwezigheid van BP, die ons toestaat lineare functies te construeren die we gebruiken als bovengrens voor de kosten gegeven een WE. Dat betekent dat wanneer één van deze grenzen wordt overschreden we zeker weten dat BP aanwezig is. Vervolgens bespreken we hoe onze resultaten een nieuw perspectief geven op BP. In plaats van BP te zien als een puur negatief fenomeen laten we zien dat het eerder een kwestie is van balans. Elk verlies aan efficiënte veroorzaakt door een set aan paden op een niveau van de vraag gaat noodzakelijk gepaard met een verhoogde efficiëntie op een ander niveau van de vraag. Voorzichtigheid is dus geboden als men paden uit een netwerk wil verwijderen, zelfs als men weet dat ze BP veroorzaken.

Vervolgens richten we ons op navigatiespellen met onzekere kosten, waarin de deelnemers vertrouwen op een centraal verkeersinformatiesysteem(TIS) die informatie vrijgeeft over de huidige staat van het netwerk. Het TIS probeert zijn positie te gebruiken om de gemiddelde reistijd te minimaliseren. We laten zien hoe de prestaties van de strategie van het TIS zijn afhangen van de kennis die het heeft over de verwachting van de deelnemers over de huidige staat van het netwerk. We geven ook milde voorwaarden waaronder het TIS observaties van de equilibria kan gebruiken om kennis te verkrijgen van deze verwachting en laten zien hoe de TIS een strategie kan ontwerpen om deze verwachting volledig vast te stellen.

Afsluitend kijken we naar variatiestellingen(VIs) die een belangrijk rol spelen in het bestuderen van navigatiespellen en de uitdagingen die gepaard gaan met het berekenen van oplossing voor deze VIs wanneer de relevante functies de Conditionele Waarde op Risico van een onzekere kostenfunctie. We onderzoeken drie verschillende stochastische benaderingsmethodes voor het berekenen van oplossing van dit type VIs en laten zien dat de algoritmes convergeren naar de nabijheid van de oplossing, waar de uiteindelijke acuraatheid van de convergentie kan worden gereguleerd door het aantal gebruikte waarnemingen vast te stellen. We geven ook een explicitie bovengrens aan het vereiste aantal waarnemingen dat nodig is om een bepaalde precisie te garanderen.