The Boundedly Rational User Equilibrium: A Parametric Optimization Approach

Oskar Eikenbroek

October 14, 2016

Title The Boundedly Rational User Equilibrium:
A Parametric Optimization Approach

Master Thesis

Date October 14, 2016

Author Oskar A.L. Eikenbroek
Civil Engineering & Management, Applied Mathematics
University of Twente

oskar.eikenbroek@gmail.com

Committee Prof. dr. ir. E.C. van Berkum
Centre for Transport Studies

Dr. W. Kern
Discrete Mathematics and Mathematical Programming

Dr. G.J. Still
Discrete Mathematics and Mathematical Programming

Prof. dr. J.L. Hurink
Discrete Mathematics and Mathematical Programming

Dr. J.C.W. van Ommeren
Stochastics Operations Research

M.A. van Essen MSc.
Centre for Transport Studies

Abstract

This study proposes a static traffic assignment model in which travelers are boundedly rational in route choice: users only switch routes when a new path significantly improves travel time. We introduce a bilevel optimization problem which finds the best-performing boundedly rational user equilibrium flow with respect to travel time. We use a parametric optimization approach to study continuity of the feasible set, optimal value function and optimal link flow solution with respect to perturbations in the indifference band. We propose two algorithms that approach the best-performing flow distribution and compare our results with approaches in literature.

Summary

Ex-ante evaluation of measures in transportation systems requires accurate prediction of the choice behavior of travelers. Traffic assignment models often presume perfect rationality in route choice decision making: travelers are assumed to be selfish, fully informed and can perfectly assess the consequence of choosing an alternative. In fact, any change in the system leads to a new user equilibrium in which no traveler can unilaterally change routes and improve his travel time. However, empirical studies suggest that this equilibrium does not arise in practice. Consequently, real-world application of measures founded under perfect rationality conditions may show undesirable results.

We develop an assignment model which adopts a more realistic behavioral perspective on decision making. We adopt the notion of bounded rationality from Mahmassani and Chang (1987) in which users switch routes when the new path significantly improves travel time. Under bounded rationality, the response to an intervention of an authority is subject to uncertainty. In contrast to the user equilibrium, a range of possible responses exists and it is unknown which boundedly rational user equilibrium (BRUE) is most likely to be realized in practice.

In the Network Design Problem we need to make assumptions on the response of travelers but this may lead to adverse effects. For instance, a pricing scheme, based on the assumption that the best-performing BRUE with respect to travel time will be realized, may lead to inferior performance in the worst-performing BRUE. We therefore consider the extremes of possible network performances under bounded rationality: an indication what a policy could achieve under uncertainty.

The mathematical problem to find the best-performing BRUE flow is a program with complementarity constraints. Current approaches in literature require complex and computationally expensive algorithms to find solutions and guarantee neither local nor global optimality. So, these approaches are due to their complexity less appropriate for application on large networks. Our approach assesses performance in large networks and maintains limited computation times.

The best-performing BRUE assignment reduces to the user equilibrium and system-optimal assignment under appropriate choices of the parameter. We apply a parametric optimization approach since both the user equilibrium and system-optimal problem are

easy to solve. In other words, we study the behavior of the optimization problem under perturbations in the indifference band.

Whereas the best-performing BRUE problem is a difficult program, we offer a new approach that replaces the difficult constraints by easier ones. We develop an equivalent bilevel optimization problem in which the leader of the problem chooses its decision variables so that the BRUE in the lower-level problem has minimum travel time (Best-case BRUE). Although the lower-level program is a convex optimization problem, the bilevel problem itself is difficult since it does not satisfy a regularity condition.

The parametric analysis shows that the feasible set and optimal value function of the lower-level problem behave continuously with respect to perturbations in the indifference band. However, the upper-level problem does not have these favorable properties and a small perturbation in the indifference band may lead to a substantial difference in flow on an arbitrary link. Intuitively, the indifference band should be a perfectly-calibrated parameter.

We offer a regularized approach, since the unfavorable properties of the Best-case BRUE lead to difficulties in calculating the flow. This approach perturbs the objective function of the lower-level problem. The corresponding problem approximates the Best-case BRUE flow distribution and the feasible set and optimal value function of this problem are continuous with respect to perturbations in the indifference band. Moreover, this reformulation allows us to construct a descent method to approximate a Best-case BRUE flow distribution. This approach solely uses convex and linear programs and any optimization tool can thus be used. Furthermore, we offer a relaxation approach that solves a sequence of nonlinear programs to find a Best-case BRUE flow distribution.

A test network assesses performance of both algorithms. The relaxation approach outperforms the descent method and the current approach in literature since we find a BRUE flow which performs better (i.e. less total travel time). Moreover, our approach only requires a few seconds to find a solution. The numerical experiments show that steering the network towards the Best-case BRUE distribution is beneficial. The Best-case BRUE is a traffic state in which all travelers take an acceptable route while it minimizes the total travel time in the system.

Samenvatting

Ex-ante evaluatie van maatregelen in verkeerssystemen vraagt om accurate voorspellingen van het keuzegedrag van reizigers. Huidige verkeersdistributie-modellen nemen aan dat gebruikers perfect rationeel zijn: reizigers zijn zelfzuchtig, volledig geïnformeerd en in staat de consequenties van elk alternatief perfect in te schatten. Onder deze aannames leidt elke verandering in een network tot een gebruikersevenwicht waarin geen reziger zijn reistijd kan verbeteren door unilateraal van route te veranderen. Empirische studies laten echter zien dat een gebruikersenwicht in niet in de praktijk voorkomt. Als gevolg hiervan zijn maatregelen, ontworpen onder de aannames van perfecte rationaliteit, minder effectief in de praktijk.

Wij ontwikkelen een verkeersdistributie-model met een realistisch perspectief op routekeuze gedrag van reizigers. We gebruiken de notie van beperkte rationaliteit van Mahmassani and Chang (1987) waarbij wordt aangenomen dat gebruikers alleen van route veranderen indien er een route beschikbaar is met significant minder reistijd. Onder de aannames van dit gedrag wordt de reactie op een interventie onzeker. In tegenstelling tot het gebruikersevenwicht, zorgt elke interventie voor een verzameling van mogelijke reacties en het is onduidelijk welke *boundedly rational user equilibrium* (BRUE) verkeersstroom in werkelijkheid wordt gerealiseerd.

In het Netwerk-Ontwerp-Probleem moeten we echter aannames doen over de *lower-level* reactie. De aanname kan in de praktijk leiden tot negatieve effecten. Een tolschema, optimaal onder de aannames van perfecte rationaliteit, kan bijvoorbeeld tot een hogere totale reistijd leiden. We stellen dan ook voor om naar de uitersten van alle mogelijke BRUE verkeersdistributies te kijken. Dit geeft, ondanks de onzekerheid, een indicatie wat een interventie kan bereiken.

Het mathematisch probleem dat de best-presterende BRUE stroom vindt, is een mathematisch programma met complementariteitscondities. Huidige benaderingen in literatuur gebruiken complexe en rekenkundig intensieve algoritmen om een oplossing te vinden. Deze algoritmen kunnen niet garanderen dat een lokaal dan wel globaal minimium wordt gevonden. De complexiteit van de methodes zorgt ervoor dat ze minder geschikt zijn voor toepassing op grotere netwerken. Onze benadering verschilt omdat we de prestatie in grote netwerken bekijken terwijl we de rekentijd gelimiteerd houden.

De scriptie laat zien dat het vaak beschouwde gebruikersevenwicht en systeem optimum speciale gevallen van de BRUE verkeersdistributie zijn. Aangezien de mathematische programma's van het gebruikersevenwicht en het systeem optimum gemakkelijk op te lossen zijn, gebruiken we de technieken van parametrische optimalisatie in het analyseren van de best-presterende BRUE verkeersstroom.

Omdat het vinden van de best-presterende BRUE distributie correspondeert met een moeilijk optimalisatieprobleem, vervangen we de moeilijke condities door een makkelijker te bestuderen probleem. In het resulterende bilevel probleem kiest de leider in het *upper-level* zijn variabelen zodanig dat in het lower-level het BRUE wordt gerealiseerd die voor de minste totale reistijd zorgt. Ondanks dat het lower-level programma een convex optimalisatieprobleem is, is het bilevel probleem een programma dat niet aan een regulariteitsconditie voldoet.

De parametrische analyse toont dat het toegelaten gebied en de optimale waarde functie van het lower-level probleem zich continu gedragen onder verstoringen in de parameter. Echter, het upper-level probleem kent deze eigenschappen niet. Een kleine verstoring in de *onverschilligheidsband* kan leiden tot een substantieel verschil in de vervoersstroom op een wegvak. Intuïtief houdt dit in dat de onverschilligheidsband parameter perfect gecalibreerd moet worden.

De genoemde eigenschappen leiden tot problemen in het berekenen van BRUE vervoersstromen. Wij introduceren daarom een geregulariseerd problem. Dit probleem kent een kleine verstoring toe aan de doelfunctioe van het lower-level probleem. Daarmee benaderen wat het mathematische optimalisatieprobleem van de best-presterende BRUE distributie. Het toegelaten gebied, optimale waarde function en oplossingsset van het geïntroduceerde programma gedragen zich wel continu onder verstoringen in de parameter. Deze aanpak geeft de mogelijkheid om een *descent* methode toe te passen. De methode gebruikt louter convexe en lineaire optimalisatieproblemen om de best-presterende BRUE vervoersstroom te benaderen. Willekeurige optimalisatiesoftware kan dus worden gebruikt om de vervoersstrom te benaderen. Daarnaast introduceren we ook nog een relaxatie methode die het best-presterende BRUE benadert door een reeks niet-lineaire programma's op te lossen.

We testen de methodes op een numeriek netwerk. Uit deze experimenten blijkt dat de relaxatie methode beter presteert dan de descent methode en de methode uit de literatuur. Ons algortime komt tot een BRUE verkeersdistributie met minder totale reistijd. De numerieke experimenten laten ook zien dat de best-presterende BRUE distributie een geambieerde verkeersstroom is: een vervoersstroom die zowel voor iedere reiziger acceptabel is als goed presteert qua totale reistijd.

Preface

I would like to acknowledge those who made this thesis possible.

My deepest gratitude goes to Georg Still. Georg, thank you for the interesting discussions, your valuable contributions and all pleasant meetings. I could not have wished for a more inspiring supervisor.

I would like to thank my supervisors Walter Kern and Eric van Berkum. Walter, I admire your eye for detail. Eric, I am grateful for your support and guidance throughout my research.

To my family and to Mani, thank you for everything.

Oskar Eikenbroek
Enschede, October 2016.

Contents

Chapter 1

Introduction

Transport authorities and traffic engineers encounter challenges to limit delays in transportation networks since these delays lead to economic wealth losses and may affect the quality of life. A wide variety of policy measures have been taken to limit congestion. Expanding infrastructure, by adding lanes or road construction, seems to be the most straightforward and popular measure. Nevertheless, road capacity expansion might lead to inferior performance of the system as a whole as illustrated by Braess et al. (2005). This *Braess' paradox* holds in practice (e.g. in Manhattan, as was shown by Kolata (1990)) and perfectly illustrates the need for policy instrument appraisal: authorities need to assess whether transport measures achieve the desired outcomes (van Wee et al., 2013).

The *Network Design Problem* (NDP) determines network settings so that the transportation system behaves optimally. It evaluates a set of measures with respect to global performance and takes into account the expected behavioral response of individuals to interventions (Brands, 2015). The NDP is frequently modeled as a *bilevel optimization problem* (Farahani et al., 2013). The upper level represents the behavior of a government or road authority, while the lower level represents the behavior of travelers. The authority has a certain objective to achieve (e.g. minimizing total delay) and has for this purpose the opportunity to impose measures on the network. Implementation of a policy induces a response of travelers which is expressed in a network state (distribution of traffic flow over the network) and from that the system objective is derived (Brands and van Berkum, 2014).

To assess impact in the NDP, the prediction of travelers' choice behavior should be accurate. Assuming route choice as users' only degree of freedom, traffic assignment models often adopt Wardrop's first principle (Wardrop, 1952) to predict the network state. This economic view on decision making assumes users to be perfectly rational: agents are selfish (i.e. choose the alternative which maximizes their individual utility), fully informed (the route choice set is known beforehand) and can perfectly assess the consequence of choosing a particular alternative (Simon, 1997; Vreeswijk et al., 2013a). Hence, "agents act as if they perform exhaustive searches over all possible decisions and then pick the best" (Conlisk, 1996, p. 675). Under these assumptions, any change in network settings leads to a new *Wardrop* or *user equilibrium*: a network state in which no traveler (marginally) benefits by unilaterally switching routes.

Wardrop's first principle has major computational benefits in the static traffic assignment. The corresponding optimization problem, defined by Beckmann et al. (1956), is convex and under mild assumptions the existence and uniqueness of the equilibrium link flow in the traffic network is guaranteed. Any upper-level measure leads to a unique behavioral response and uniquely determines the performance of the system. The bilevel problem implicitly assumes that the realized flows in practice are well approximated by the Wardrop equilibrium flows of the traffic assignment since we choose those network settings which turn out to be optimal in the NDP (Ban et al., 2009).

However, empirical studies show that the economic assumptions of Wardrop are debatable. Consequently, a user equilibrium does not necessarily arise in practice. For instance, only 34% (Zhu and Levinson, 2010) to 75% (Thomas and Tutert, 2008) of all travelers follow the shortest time path. In addition, network changes turn out to be *irreversible*: while initial network settings are re-obtained by revoking a change, the initial traffic state is not (Guo and Liu, 2011). These findings illustrate that the evaluation of network settings under Wardrop's condition is naive. Real-world application of policy interventions designed under Wardrop's condition are not necessarily effective and may show undesirable results. Hence, we need a more realistic and thus behavioral perspective on route choice decision making in traffic assignment models.

Bounded Rationality

We incorporate a realistic view on decision making in the static traffic assignment and adopt the notion of bounded rationality. Bounded rationality presumes decision makers to be limited in knowledge and computational capacity since a decision situation is complex (Simon, 1955, 1997). In the context of route choice decision making, boundedly rational travelers make *suboptimal* choices due to limited awareness or the existence of thresholds (Chorus and Timmermans, 2009). In other words, travelers choose a fast - but not necessarily the fastest - route. This notion of bounded rationality in route choice was introduced by Mahmassani and Chang (1987) and explained 90% of the trips in the study of Zhu and Levinson (2010).

A more realistic behavioral view in assignment models increases complexity compared to the economic user equilibrium. Mahmassani and Chang (1987) pointed out that bounded rationality in the static traffic assignment may lead to a non-unique link flow pattern. Consequently, any measure in the upper level of the NDP leads to a set of lower-level solutions and each lower-level solution refers to a different behavioral response of travelers

(Ban et al., 2009). The non-unique response to an intervention induces uncertainty among upper-level performances and adds a new dimension to the bilevel optimization problem (Figure 1.0.1). Whereas it is not known which lower-level flow is most likely to be realized in practice (*cf.* Ban et al. (2009)), it is unclear which upper-level measure optimizes the system objective.

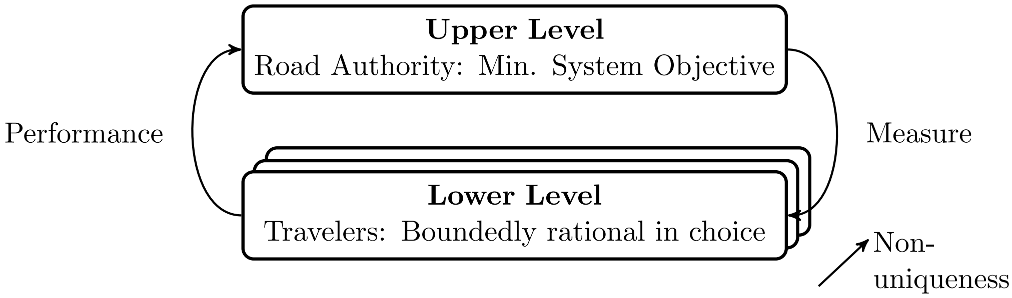


Figure 1.0.1: Network Design Problem as a bilevel problem. Adapted from Brands (2015).

Since authorities generally evaluate a (large) set of alternative measures and scenarios on large network instances, we need an efficient approach to find traffic flows as result of boundedly rational route choice behavior, as also stressed by Sun et al. (2015). Therefore we have to cope with the uncertainty in the lower-level response. In literature, it is suggested to apply the policy measure which shows (1) best worst-case performance with respect to the global objective, (2) optimal best-case performance, or (3) best performance assuming a probability distribution over all possible solutions (Ban et al., 2009; Di and Liu, 2016). Similar as in Lou et al. (2010), we consider the extremes of possible network performances under bounded rationality: an indication what a policy could achieve under uncertainty.

The currently available analyses are solely useful for assignment problems which are small in terms of network size and number of routes. Moreover, the mathematical properties of *boundedly rational user equilibrium* (BRUE) flows - traffic flows as result of bounded rationality in route choice decision making - are rather unexplored. Lou et al. (2010) showed that a BRUE assignment problem leads to a mathematical program with complementarity constraints. They formulated the more restrictive link-based BRUE so that an earlier developed algorithm (see Lawphongpanich and Yin (2010)) could be applied in finding best- and worst-case link-based BRUE flows with respect to the total travel time. Di et al. (2013) (and successive studies in Di et al. (2014a, 2016)) used complementarity conditions in deriving all possible boundedly rational path flows using a two-stage approach; first they identified a set of possible used paths and, second, they used this route choice set in distributing traffic. For each stage, a complementarity constraint program is formulated which is difficult to solve numerically and different local minimizers may co-exist. The attached complex and computationally expensive algorithms to approach local minima suggest that the mentioned attempts are less appropriate for larger networks.

Based on our analysis, we make the following observations:

- The NDP requires accurate predictions of travel behavior to determine the optimal network settings. The conventional model presumes perfect rationality in route choice decision making and lacks empirical validity. Consequently, real-world application of measures founded under perfect rationality conditions may show undesirable results. Empirical studies suggest that integration of bounded rationality in the traffic assignment problem increases validity of the NDP;

- Adoption of bounded rationality in the lower level of the NDP adds a new dimension to the design problem. We evaluate the extremes of the possible network performances under bounded rationality to assess performance of a measure. However, the corresponding optimization program is a complementarity constraint problem which is non-convex and there are limited approaches available to handle this difficult program;

- Current approaches in literature require complex and computationally expensive algorithms to find the BRUE flow distributions. So, these methods are less appropriate to apply in the NDP. Moreover, since authorities evaluate a set of possible scenarios, we require an algorithm that approaches BRUE flows in limited computation time.

Research Objective

The behavioral response of travelers mainly determines the effect of a policy measure on network performance. However, current traffic assignment models make naive economic assumptions on users' decision making processes. Assignment models should adopt a realistic view on decision making to better predict the expected outcomes of measures in the NDP. The notion of bounded rationality in route choice has the potential to predict the behavioral response of users.

This leads to the following research objective:

> *Development and analysis of a static traffic assignment model that incorporates bounded rationality: travelers only switch routes when a path significantly improves travel time.*

We adopt the notion of bounded rationality from Mahmassani and Chang (1987): users solely switch routes when an alternative path improves travel time by more than a certain threshold (*indifference band*). Although this notion was earlier integrated in the static traffic assignment by Di et al. (2013) and Lou et al. (2010), the problem is, from a mathematical point of view, still underexplored.

Our research contributes to literature since we offer an in-depth mathematical analysis of the structure (topology) of the best and worst-performing BRUE flow with respect to travel time (Best-/Worst-case BRUE). To support our study, we add an intuitive analysis of this problem under affine linear latencies. We identify the main difficulties of the adoption of bounded rational choice behavior in the NDP by numerous examples. In addition, we overcome the current complex algorithms. Indeed, we propose two algorithms that both approach the Best-case BRUE flow distribution in limited time. These algorithms allow incorporation of boundedly rational choice behavior in the NDP. We theoretically and numerically compare the proposed assignment model to the approaches of Di et al. (2013) and Lou et al. (2010).

Our theoretical analysis is necessary if we want to integrate boundedly rational behavior in the static traffic assignment for large network instances. Moreover, the analysis supports future development of algorithms that efficiently find bounded rational link flow solutions.

Approach

This research takes a sensitivity or continuity analysis point of view. This type of analysis evaluates whether for instance the optimal value function and link flows change smoothly as the parameter changes. We provide a sufficient condition for the continuity of the feasible set and optimal value function of the best-case boundedly rational traffic assignment problem. We reformulate the problem as a bilevel optimization problem. In particular, we study continuity of the feasible set, optimal value function and optimal link flow solution set with respect to perturbations in the indifference band parameter for the Best-case BRUE. We approach the indifference band as a parameter since the problem of finding BRUE flows reduces for certain values of the indifference band to an easy to solve problem.

This study relates to the recent paper of Di et al. (2016). Di et al. (2016) studied the continuity of the feasible set and optimal value function of the second-best toll pricing problem in a linear latency context. However, Di et al. (2013, 2016) assumed that the used route choice set is known by sequentially solving a complementarity constraint program beforehand. This assumption has serious consequences for large network instances and, therefore, we refrain from making such an assumption in our study. Han et al. (2015) investigated mathematical properties of the bounded rational dynamic assignment. Although the results presented can be particularly applied to the static assignment, the study mainly focused on boundedly rational flow distributions in the selection of departure times and routes. Our study, however, aims to find the BRUE flow which minimizes the total travel time in a static context.

We assume that users possess full information and, consequently, any significant improvement in transit time induces a path switch. Although we emphasize that a bounded rational route choice results from a dynamic, day-to-day, process (e.g. a repetitive choice which becomes habitual), we neither investigate these processes (see e.g. Guo and Liu (2011)) nor make any assumptions on the cognitive processes which led to a possibly suboptimal choice (see e.g. Vreeswijk et al. (2013b)). In other words, we focus on the outcomes of boundedly rational route choice behavior rather than the decision making process itself.

Outline

The thesis is organized as follows. We discuss and illustrate different approaches to model bounded rationality in non-cooperative games in Chapter 2. Chapter 3 treats the formal problem description of the static traffic assignment including Wardrop's principles. Chapter 4 formally introduces the boundedly rational user equilibrium and includes examples to illustrate the main difficulties for the NDP. Chapter 5 treats the problem of finding the Best-case BRUE flow distribution with respect to total travel time. We discuss convexity and regularity conditions of this problem and introduce branches of convex problems to reduce the Best-case BRUE to an easier to study problem assuming affine linear latencies. Chapter 6 discusses continuity of the feasible set, optimal value function and optimal solution set of these branches. We generalize these results to the Best-case BRUE problem for linear latencies and more general latency functions in Chapter 7 and Chapter 8, respectively. In Chapter 9 we propose a regularized approach in order to approximate the Best-case BRUE problem and we give a descent approach to this flow distribution. We perform numerical experiments to show performance of the descent approach and to discuss implications for policy evaluation (Chapter 10). Lastly, Chapter 11 draws the main conclusions and discusses topics for further research.

Chapter 2

Modeling Bounded Rationality

The economic assumptions of Wardrop state that users are able to maximize their utility function (i.e. minimize their own travel time). Users sequentially consider all possible routes and the consequences of choosing the alternative. Mathematically, we model this behavior as follows. Let Ω be the set of route alternatives for an individual and let s be the disutility function corresponding to the paths, $s : \Omega \to \mathbb{R}$. Abstractly, the problem for each user to find alternative which maximizes utility is then:

$$\min_{x\in\Omega} s(x). \tag{2.0.1}$$

In other words, find point $x^\star \in \Omega$ so that $s(x)$ has its minimum at $s(x^\star)$:

$$x^\star = \arg\min_{x\in\Omega} s(x).$$

The assumption of bounded rationality in decision making relaxes the economic principles of Wardrop. Basically, there exist two perspectives to model bounded rationality in route choice. The first conceptualization assumes users to have limited awareness, while the second conceptualization says that the existence of indifference bands or thresholds makes that users consciously violate Wardrop's first principle (Chorus and Timmermans, 2009).

As mentioned, the first perspective assumes users to subconsciously violate Wardrop's principle due to limited awareness: travelers do not choose the shortest path since they either do not know it is fastest or do not perceive it as such (Chorus and Timmermans, 2009; Di et al., 2013). In modeling, this perspective maintains the utility maximization behavior of users but either (i) imposes additional constraints on the set of alternatives or (ii) perturbs the characteristics of the alternatives (Hato and Asakura, 2000; Mallard, 2012).

The latter approach introduces *noise* to account for the perception error (Sheffi, 1985; Zhao and Huang, 2016). This approach introduces a new disutility function $\tilde{s} : \Omega \to \mathbb{R}$

but the set of alternatives Ω remains. The optimization problem (2.0.1) becomes

$$\min_{x\in\Omega} \tilde{s}(x).$$

Hato and Asakura (2000) imposed additional constraints on the choice set and endowed every street in a transportation network with a probability that it is observed. A chosen route only consists of links which are perceived by the user (the *mental map*). Zhang and Yang (2015) assigned a probability to the paths in the network since users tend to choose familiar routes. The latter studies define $\tilde{\Omega} \subset \Omega$ as the observed set of alternatives and every agent seeks for point $\tilde{x} \in \tilde{\Omega}$ so that

$$\tilde{x} = \arg\min_{x\in\tilde{\Omega}} s(x).$$

The second conceptualization assumes agents to consciously violate Wardrop's first principle because of the existence of indifference bands or thresholds. Users are then *satisficers* (Simon, 1997), solely inclined to change their itinerary when a significant travel time improvement can be achieved (Mahmassani and Chang, 1987). Travelers may possess full information on (all) the alternatives but are not necessarily utility maximizers. In this approach, users try to find point $\bar{x} \in \Omega$ so that

$$\min_{x\in\Omega} s(x) + \epsilon \geq s(\bar{x}) \tag{2.0.2}$$

holds. Here, $\epsilon > 0$ represents the indifference band or threshold. This conceptualization was introduced by Mahmassani and Chang (1987) and later formalized by Di et al. (2013) and Lou et al. (2010). In these static traffic assignment studies, travelers are assumed to make a suboptimal choice, i.e. the chosen route is said to be *acceptable* as long as the corresponding disutility is at most ϵ worse than the alternative having minimum disutility (Lou et al., 2010). Zhao and Huang (2016) assumed the *aspiration level* $\min_{x\in\Omega} s(x) + \epsilon$ to be fixed: users consider any alternative which meets this level. In a day-to-day setting, condition (2.0.2) assumes users to repeat their route choice as long as an alternative cannot improve travel time more than ϵ (Wu et al., 2013).

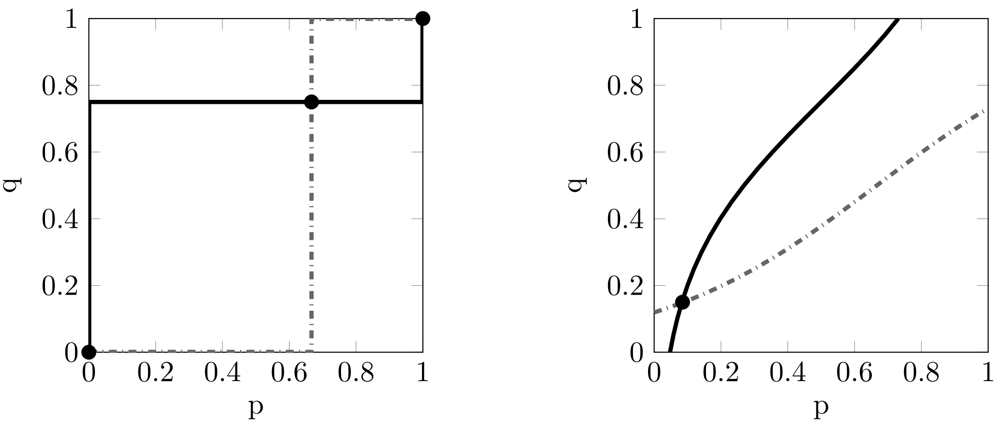
2.1 Bounded Rationality in a 2×2-game

We introduced three approaches on modeling bounded rationality. This section illustrates the consequences of the integration of these approaches in a simple (finite player) non-cooperative game. The game we use was described in Peters (2008). In this two-player game, Player 1 has the opportunity to play strategies T and B, while Player 2 can choose from either L or R. Each strategy combination has a corresponding payoff:

$$\begin{array}{cc} & \textit{Player 2} \\ & \begin{array}{cc} L & \;\; R \end{array} \\ \textit{Player 1} \quad \begin{array}{c} T \\ B \end{array} & \begin{pmatrix} 2,2 & 0,1 \\ 1,1 & 3,3 \end{pmatrix} \end{array}$$

Here, strategy (T, R) leads to payoff-vector $(0, 1)$: a payoff of 0 for Player 1 and a payoff of 1 for Player 2. Each player chooses to play a pure strategy or picks each strategy with a certain probability. Central in this game and in non-cooperative games in general is the notion of *best reply*. A selfish player chooses the strategy which maximizes utility, given the knowledge about his available strategies and the strategies chosen by the others (Peters, 2008). We restrict ourselves for the sake of clarity to this easy problem and use a graphical solution method to find all *equilibria* under different notions of bounded rationality. In this context, an equilibrium refers to a state in which no player has the incentive to unilaterally change strategies, i.e. there is no *force* that tries to get the game in another state.

This section limits itself to some basic examples and relies on the recent elaborate literature study from Di and Liu (2016).



(a) Perfectly Rational Equlibrium. **(b)** Limited awareness equilibrium.

Figure 2.1.1: Equilibria in 2×2-game.

Perfectly Rational Equilibrium

A perfect rationality setting assumes players to be cognitively unhindered in making a decision. For this game, this means that each player observes what the other player chooses and picks a corresponding strategy to maximize his own payoff. Suppose Player 2 plays strategy L, then Player 1 picks T since this leads to a utility of 2. Choosing B, on the other hand, leads to a payoff of 1. We have two pure equilibria in this game: (T, L) and (B, R).

More generally, assume Player 2 plays $(q, 1-q)$, i.e. he chooses L with probability q and R with probability $1-q$. The best reply $\beta_1(q, 1-q)$ of Player 1 is to play strategy T as long as

$$2q > q + 3(1-q), \qquad \text{or} \qquad q > \frac{3}{4},$$

holds. In other words, Player 1 will use strategy T if $q > \frac{3}{4}$. In case $q < \frac{3}{4}$ Player 1 uses B as best reply. The player is *indifferent* among strategies T and B if $q = \frac{3}{4}$. Henceforth, any strategy combination $(p, 1-p)$, where $0 \leq p \leq 1$, of T and B is the best reply. That is (Peters, 2008):

$$\beta_1(q, 1-q) = \begin{cases} \{(1,0)\} & \text{if } \frac{3}{4} < q \leq 1; \\ \{(p, 1-p) \mid 0 \leq p \leq 1\} & \text{if } q = \frac{3}{4}; \\ \{(0,1)\} & \text{if } 0 \leq q < \frac{3}{4}. \end{cases}$$

Analogously, we find $\beta_2(p, 1-p)$ as the best reply of Player 2 against $(p, 1-p)$:

$$\beta_2(p, 1-p) = \begin{cases} \{(1,0)\} & \text{if } \frac{2}{3} < p \leq 1; \\ \{(q, 1-q) \mid 0 \leq q \leq 1\} & \text{if } p = \frac{2}{3}; \\ \{(0,1)\} & \text{if } 0 \leq p < \frac{2}{3}. \end{cases}$$

Figure 2.1.1a draws both curves: the solid line is the best reply curve of Player 1 and the dotted line is the best reply curve of Player 2. We find points of intersection. These three points are all Perfectly Rational Equilibria: $((0,1),(0,1)), ((\frac{2}{3},\frac{1}{3}),(\frac{3}{4},\frac{1}{4}))$ and $((1,0),(1,0))$. At such an equilibrium no player can unilaterally change strategies and improve utility. As we will see later, this definition is similar to the definition of a *Wardrop equilibrium* or *user equilibrium* in the static traffic assignment.

Limited Awareness

The conceptualization of limited awareness assumes players to subconsciously violate the perfect rationality condition since travelers either do not know all strategies or the utility function is perturbed. We start with the latter approach and introduce noise to account for the perception error of a player. Let us assume that Player 2 uses strategy $(q, 1-q)$ and that Player 1 has a certain probability P_T of playing strategy T. The probability P_T equals the probability that the utility U_T of strategy T is greater than or equal to utility U_B corresponding to playing B:

$$P_T = Pr(U_T \geq U_B).$$

Following Ben-Akiva and Lerman (1985), the utility is divided into a deterministic and random part:

$$P_T = Pr(2q + \eta_T \geq q + 3(1-q) + \eta_B).$$

We assume that the random variables η_T, η_B are independently and identically distributed and are Gumbel-distributed with scale parameter μ. The probability that Player 1 will play T in response to strategy $(q, 1-q)$ is then (we omit the calculus, see Ben-Akiva and Lerman (1985)):

$$P_T = \frac{e^{\mu(2q)}}{e^{\mu(2q)} + e^{\mu(q+3(1-q))}}.$$

This results in the best reply sets for both players:

$$\beta_1(q, 1-q) = \ \{(p, 1-p) \mid p = \tfrac{e^{\mu(2q)}}{e^{\mu(2q)} + e^{\mu(q+3(1-q))}}\} \quad \text{for } 0 \leq q \leq 1,$$

$$\beta_2(p, 1-p) = \ \{(q, 1-q) \mid q = \tfrac{e^{\mu(2p+1-p)}}{e^{\mu(2p+1-p)} + e^{\mu(p+3(1-p))}}\} \quad \text{for } 0 \leq p \leq 1.$$

Again, Figure 2.1.1b plots the curves and we find the point of intersection (assuming $\mu = 1$) numerically. We find a single equilibrium, known as the *quantal response equilibrium* (Di and Liu, 2016) or, in a traffic assignment setting, as the stochastic user equilibrium (Sheffi, 1985). Here, no traveler thinks that a switch in routes improves the travel time. Intuitively, μ is the rationality parameter: if $\mu = 0$ users are perfectly rational while $\mu \to \infty$ assumes users to be completely indifferent (i.e. the probability P_T is independent of utilities U_T and U_B).

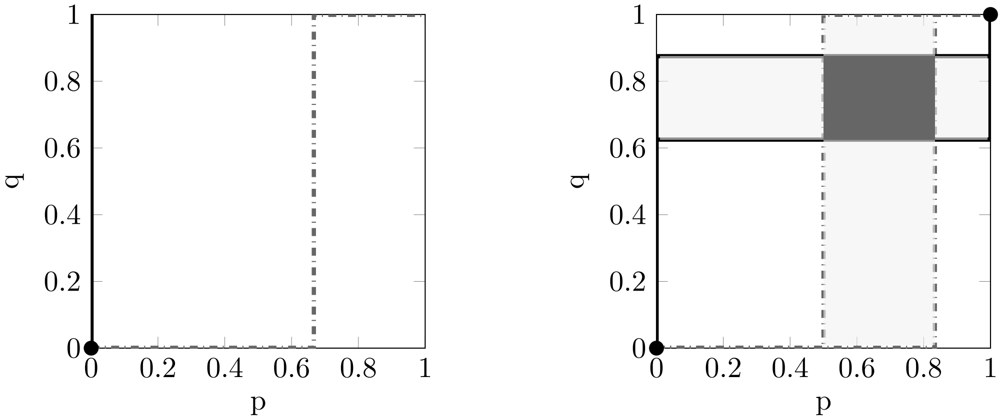
Another approach to model limited awareness in a game is by imposing additional constraints. For instance, using the approach of Hato and Asakura (2000), we assign a probability for *observing* strategy T. Player 1 is solely aware that strategy T can be played in, say, $\frac{1}{4}$ of the cases. In the other cases, he can only play B. In $\frac{1}{4}$ of the cases a perfectly rational equilibrium arises since we assume that players maintain utility maximization in decision making. With probability $\frac{1}{4}$, the best reply function $\beta_1(q, 1-q)$ becomes:

$$\beta_1(q, 1-q) = \{(0, 1) \mid \text{if } 0 \leq q \leq 1\}.$$

Let us assume that the set of best replies for Player 2 remains. Figure 2.1.2a plots both curves and we find the equilibrium at $((0, 1), (0, 1))$. Observe that with probability $\frac{3}{4}$ the best reply function is as in Figure 2.1.1a.

ϵ-Equilibrium

The ϵ-equilibrium assumes players to play a satisfactory rather than optimal strategy. Still, individuals are assumed to have perfect knowledge but have an $\epsilon > 0$ that denotes the indifference band: one will only change strategies if another strategy improves utility



(a) Limited awareness equlibrium.

(b) ϵ-Equilibrium.

Figure 2.1.2: Equilibria in 2×2-game assuming bounded rationality.

by at least ϵ. Recall the earlier example. Assume Player 2 uses strategy $(q, 1-q)$. Player 1 solely plays strategy T if strategy T gives at least ϵ higher utility than B:

$$2q > q + 3(1-q) + \epsilon, \qquad \text{or} \qquad q > \frac{3+\epsilon}{4}.$$

Player 1 plays $(1,0)$ as reply if $q > \frac{3+\epsilon}{4}$. Similarly, the player solely uses B if $q < \frac{3-\epsilon}{4}$ and is indifferent between T and B (any strategy will do) if $\frac{3-\epsilon}{4} \leq q \leq \frac{3+\epsilon}{4}$. We use the set of *sufficing* replies to find all ϵ-Equilibria:

$$\beta_1(q, 1-q) = \begin{cases} \{(1,0)\} & \text{if } \frac{3+\epsilon}{4} < q \leq 1; \\ \{(p, 1-p) \mid 0 \leq p \leq 1\} & \text{if } \frac{3-\epsilon}{4} \leq q \leq \frac{3+\epsilon}{4}; \\ \{(0,1)\} & \text{if } 0 \leq q < \frac{3-\epsilon}{4}. \end{cases}$$

In a similar way, we find $\beta_2(p, 1-p)$ as the reply of Player 2 against $(p, 1-p)$:

$$\beta_2(p, 1-p) = \begin{cases} \{(1,0)\} & \text{if } \frac{2+\epsilon}{3} < p \leq 1; \\ \{(q, 1-q) \mid 0 \leq q \leq 1\} & \text{if } \frac{2-\epsilon}{3} \leq p \leq \frac{2+\epsilon}{3}; \\ \{(0,1)\} & \text{if } 0 \leq p < \frac{2-\epsilon}{3}. \end{cases}$$

This notion is similar to the notion of the BRUE used by Mahmassani and Chang (1987) in the static traffic assignment. In the traffic assignment, individuals try to find a satisfactory route rather than the optimal one. In this setting we find a set of possible solutions assuming $\epsilon = \frac{1}{2}$ (Figure 2.1.2b): $((0,1),(0,1)), ((1,0),(1,0))$ and $((p, 1-p), (q, 1-q) \mid 3 - \epsilon \leq 4q \leq 3 + \epsilon, 3 - \epsilon \leq 6p \leq 3 + \epsilon)$. Observe that the perfectly rational equilibrium is also an ϵ-equilibrium.

2.2 Bounded Rationality in Route Choice

Common practice is to model route choice behavior by a perfectly rational equilibrium. As mentioned, this equilibrium lacks empirical validity but is easy to calculate. We compare the outcome of this equilibrium with the limited awareness equilibrium in Figure 2.1.1a and Figure 2.1.1b. We see that the analysis of a single scenario under different behavioral assumptions may lead to completely different outcomes in terms of link flows and travel time. Consequently, the severity of congestion on certain links (i.e. bottlenecks) is possibly different and we may derive other optimal settings from the NDP.

In the remainder of this study we adopt the notion of the ϵ-equilibrium, also called BRUE in our context. Major benefit of this notion is that we study policy implications under realistic behavioral assumptions and do not need to make any assumptions on the cognitive processes of traffic users. By this we mean that we model the outcome of users' decision making by deriving the indifference band ϵ from observed choice behavior without the necessity to study the perceptions of travelers. The other discussed approaches all require a set of rules to derive the choice strategy.

A large body of empirical studies confirms that chosen routes deviate from the fastest one. Zhu and Levinson (2010) found that only 35% of the trips were made via the shortest time path, while 90% of all trips could be explained by an indifference band of 5 minutes. A small GPS study from Jan et al. (2000) revealed an average indifference band of 0.6 minutes with respect to shortest travel time path. The fastest route was only traveled by 31 of the 124 trips. The analysis of Ciscal-Terry et al. (2016), recording 89 drivers over a period of 17 months, indicated that the observed average travel time was on average 4 minutes more than the shortest time path, including a bias due to setup and finishing operations of the driver. Although the width of the indifference band is not of interest here, the mentioned studies empirically support the incorporation of the BRUE in the traffic assignment.

From a mathematical perspective, Figure 2.1.2b illustrates that the boundedly rational user equilibrium comes with more complexity: for $\epsilon > 0$, there exist an infinite number of equilibria.

2.3 Synthesis

In this chapter, we showed that different assumptions on behavior in decision making may have serious implications for the system as a whole. In other words, the strategy of the players, representing route choice in our setting, varies under different behavioral assumptions. In the next chapter we formally introduce the boundedly rational user equilibrium, which comes, compared to the perfectly rational equilibrium, with more complexity. For instance, there exist an infinite number of equilibra for a single scenario.

Chapter 3

Formal Problem Description

3.1 Notation & Traffic Assignment Formulation

We study the static traffic assignment. Given is a directed traffic network $G = (V, E)$, with V being the set of nodes and E is the set of directed edges (road, links, or arcs) $e = (i, j)$, where $i, j \in V$. The network includes a set of *origin-destination pairs* (OD pairs) $\mathcal{K} \subseteq V \times V$, with static demand $d_k > 0, k \in \mathcal{K}$. We refer to each OD pair $k \in \mathcal{K}$ as *commodity* k and each OD pair is connected by a set of simple directed paths, denoted by $\mathcal{P}_k$. The set $\mathcal{P}$ of all paths in the network is the union of the path sets per commodity: $\mathcal{P} = \cup_{k \in \mathcal{K}} \mathcal{P}_k$.

A feasible *traffic flow* or *flow* for given demand $d \in \mathbb{R}_+^{|\mathcal{K}|}$ (we denote by $|.|$ the cardinality of a set) is a pair of vectors $(f, x) \in \mathbb{R}^{|P|} \times \mathbb{R}^{|E|} = (f_p, p \in \mathcal{P}; x_e, e \in E)$ so that

$$x = \Gamma f, \qquad \Lambda f = d, \qquad \text{and} \qquad f \geq 0. \tag{3.1.1}$$

Here, the matrix $\Gamma \in \mathbb{R}^{|E| \times |\mathcal{P}|}$ denotes the *link-path incidence matrix* in which $\Gamma_{ep} = 1$ if edge e is in route p and $\Gamma_{ep} = 0$ otherwise. $\Lambda \in \mathbb{R}^{|\mathcal{K}| \times |\mathcal{P}|}$ is the *OD-path incidence matrix*; $\Lambda_{kp} = 1$ if $p \in \mathcal{P}_k$ and $\Lambda_{kp} = 0$ otherwise. The conditions in (3.1.1) say that (i) the flow on a link equals the sum of all path flows that pass through this link (i.e. (f, x) is consistent), (ii) the sum of flows on all paths for a commodity meet the demand, and (iii) path flows (and thus link flows) are non-negative.

Each link $e \in E$ in the network has a flow dependent *travel time* or *cost* $l_e(x)$. The travel time $c(f)$ $\left(c : \mathbb{R}_+^{|\mathcal{P}|} \to \mathbb{R}_+^{|\mathcal{P}|}\right)$ along a route is *additive*. In other words, the cost of a route $c_p(f), p \in \mathcal{P}$, is the sum of travel costs on all edges in that path:

$$c_p(f) = \sum_{e \in p} l_e(x) \qquad \text{or} \qquad c(f) = \Gamma^T l(x) = \Gamma^T l(\Gamma f)$$

in vector-form.

We assume that the travel time of a trip is the only determinant in route choice. Furthermore, in choosing routes, travelers are assumed to value travel time homogeneously. Literature refers to these settings as the *fixed demand static traffic assignment* or *nonatomic routing game.*

Assumption 1. *Let us assume throughout the study that the travel time function $l(x)$ is separable (i.e. $l_e(x) = l_e(x_e)$), continuous, convex and strictly monotone: $l_e(x_e) < l_e(\bar{x}_e)$ provided $x_e < \bar{x}_e$.*

We emphasize that Assumption 1 is not necessarily a very strong one and applies to general travel cost functions, including the well-known BPR-function (Bureau of Public Roads) or a - more general - polynomial with positive coefficients.

3.2 Wardrop's Principles

Wardrop (1952) formulated two alternative criteria to determine the distribution of flow over a traffic network. The first principle was formulated as: *"the journey times on all the routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route"* (Wardrop, 1952, p. 345). Wardrop's first principle assumes travelers to be perfectly rational in making route choice decisions: users maximize own utility by considering and evaluating all possible alternatives. The resulting traffic flow pattern (f^n, x^n) (the superscript n refers to John F. Nash, who introduced the perfectly rational equilibrium in non-cooperative games) under the assumptions of this behavior in a static environment is a traffic state in which no traveler can unilaterally change routes to decrease its own travel time.

Definition 3.2.1 (PRUE). *Consider a traffic flow (f^n, x^n) as in (3.1.1) with corresponding cost vector $c(f^n)$. Flow (f^n, x^n) is said to be a Perfectly Rational User Equilibrium (PRUE) if for all $k \in \mathcal{K}$ the following condition holds for all $p, q \in \mathcal{P}_k$:*

$$f_p^n > 0 \Rightarrow \begin{cases} c_p(f^n) = c_q(f^n) & \text{if } f_q^n > 0; \\ c_p(f^n) \leq c_q(f^n) & \text{if } f_q^n = 0. \end{cases} \tag{3.2.1}$$

In particular, at PRUE as in (3.2.1) all *flow-carrying paths* (i.e. $p \in \mathcal{P} : f_p^n > 0$) for a commodity $k \in \mathcal{K}$ experience equal (in fact, minimum) cost $\min_{q \in \mathcal{P}_k} c_q(f^n)$.

Beckmann et al. (1956) formulated an optimization problem to a PRUE flow distribution:

$$\min_{(f,x)} z(x) \qquad \text{s.t.} \qquad (f,x) \text{ as in (3.1.1)}, \tag{Q_n}$$

where

$$z(x) := \sum_{e \in E} \int_0^{x_e} l_e(\tau) d\tau.$$

(f^n, x^n) solves (Q_n) under Assumption 1 if and only if this flow satisfies the *Karush-Kuhn-Tucker (KKT) conditions* corresponding to (Q_n) (note that the feasible set is defined by linear (in)equalities and the objective function $z(x)$ is strictly convex). The system of KKT conditions with corresponding multipliers (μ, ϕ, λ) for this problem reads, assuming $l(x)$ to be continuously differentiable, as follows:

$$\begin{aligned} \nabla_x s(x) - \mu &= 0 \\ \Gamma^T \mu - \Lambda^T \lambda + \phi &= 0 \\ f^T \phi &= 0 \\ (\Gamma f - x)^T \mu &= 0 \\ (d - \Lambda f)^T \lambda &= 0 \end{aligned} \qquad \begin{aligned} \phi &\geq 0 \\ f &\geq 0 \\ d &= \Lambda f \\ \Gamma f &= x \end{aligned} \tag{3.2.2}$$

Since $\nabla_x s(x) = l(x)$, the system (3.2.2) corresponding to (Q_n) reveals that flow-carrying paths share (minimum) travel cost as in (3.2.1). Indeed, simple analysis shows that λ_k represents the minimum cost to travel for commodity $k \in \mathcal{K}$. We obtain, where the travel time on the minimum cost path for commodity $k \in \mathcal{K}$ is denoted by λ_k^n, that traffic flow (f^n, x^n) as in (3.1.1) is a PRUE flow if and only if

$$f_p^n(c_p(x^n) - \lambda_k^n) = 0, \qquad \text{and} \qquad c_p(x^n) \geq \lambda_k^n \tag{3.2.3}$$

is satisfied for all $p \in \mathcal{P}_k, k \in \mathcal{K}$. The system defined by (3.1.1) and (3.2.3) is the *nonlinear complementarity problem* (NCP) for finding the PRUE flow distribution.

Note that since the objective function $z(x)$ in problem (Q_n) is strictly convex on the feasible set defined by the (in)equalities in (3.1.1), a PRUE (f^n, x^n) exists and is unique with respect to x^n (and λ^n).

Remark 3.2.1 (Non-unique route flows). *As widely discussed in literature, there may exist a set of route flows f corresponding to a single link flow distribution x (see Lu and Nie (2010)). Let us regard the set $\mathcal{H}(x)$ of route flows for known link flow x of pair (f, x) (and fixed d) as:*

$$\mathcal{H}(x) = \{f \in \mathbb{R}^{|\mathcal{P}|} \mid \Lambda f = d, \Gamma f = x, f \geq 0\}.$$

Solely under the condition that the matrix $\begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix}$ has full column rank there is only a single route flow solution f corresponding to link flow x. With exception from parallel-link networks, we cannot expect that this condition is satisfied in general.

The second principle of Wardrop is known as the *system-optimal traffic assignment.* Qualitatively, a traffic flow distribution is a system optimum if *"the average journey time is a minimum"* (Wardrop, 1952, p. 345). The system optimum is a desired traffic state for an authority, but will not occur in practice: this state requires full cooperation among travelers or coordination by an authority in all route choice (Yang and Huang, 2005).

Formally, a flow (f^s, x^s), superscript s referring to system optimum, as in (3.1.1) is said to be a system-optimal traffic flow if the flow minimizes total travel time. In other words, (f^s, x^s) solves the following optimization problem:

$$\min_{(f,x)} s(x) \qquad \text{s.t.} \qquad (f,x) \text{ as in (3.1.1)}, \tag{Q_s}$$

where

$$s(x) = \sum_{e \in E} x_e l_e(x_e). \tag{3.2.4}$$

Under Assumption 1, the objective function $s(x)$ - representing total travel time - is (strictly) convex. Then, (f^s, x^s) solves (Q_s) if and only if (f^s, x^s) satisfies the KKT conditions of (Q_s) with corresponding Lagrange multiplier vectors. These conditions are similar to (3.2.2) with the only difference that $\nabla_x s(x) = l(x) + x^T \nabla_x l(x)$. This reveals that the resulting traffic flow is a PRUE flow using marginal cost $\frac{\partial}{\partial x_e}(x_e l_e(x_e))$ as latency function for each link $e \in E$. Equivalent to the PRUE condition in (3.2.1), a traffic flow (f^s, x^s) as in (3.1.1) is a system optimum if and only if

$$f_p^s > 0 \Rightarrow \begin{cases} \tilde{c}_p(x^s) = \tilde{c}_q(x^s) & \text{if } f_q^s > 0; \\ \tilde{c}_p(x^s) \leq \tilde{c}_q(x^s) & \text{if } f_q^s = 0, \end{cases} \tag{3.2.5}$$

is satisfied for all $p \in \mathcal{P}_k, k \in \mathcal{K}$, where

$$\tilde{c}_p(x) := \sum_{e \in p} \left(l_e(x_e) + x_e l'_e(x_e) \right),$$

and $l'_e(x_e) = \frac{\partial l_e(x)}{\partial x_e}$.

Again, the objective function $s(x)$ of (Q_s) is strictly convex on the feasible set defined by the (in)equalities in (3.1.1) and the system-optimal traffic flow distribution (f^s, x^s) exists and is unique with respect to x^s.

3.3 Synthesis

We introduced the PRUE and system-optimal traffic assignment. Both flow distributions are unique with respect to the link flow x under Assumption 1. However, both assignments are unrealistic because the first assumes users to be perfectly rational and the latter requires full cooperation or coordination. In the next chapter we introduce the assignment under the assumptions of boundedly rational individuals.

Chapter 4

Boundedly Rational User Equilibrium

4.1 Formal Definition

The assumptions of perfect rationality in route choice decision making are, from a behavioral perspective, naive. The boundedly rational equilibrium condition in (4.1.1) states that travelers are satisficers so that unilaterally switching routes cannot lead to a travel time improvement of more than an individually-tailored and situation-specific tolerance threshold: the indifference band (Mahmassani and Chang, 1987; Vreeswijk et al., 2013a). Although this behavior might be in practice the result of limited awareness, our modeling perspective assumes travelers to be fully informed.

Definition 4.1.1 (BRUE). *Given the indifference band $\epsilon \in \mathbb{R}_+^{|\mathcal{K}|}$, a traffic flow (f, x) as in (3.1.1) with corresponding path costs $c(f)$ is called a Boundedly Rational User Equilibrium (BRUE) if for all $k \in \mathcal{K}$ the following condition is satisfied for all $p \in \mathcal{P}_k$:*

$$f_p > 0 \Rightarrow c_p(f) \leq \min_{q \in \mathcal{P}_k} c_q(f) + \epsilon_k. \tag{4.1.1}$$

BRUE condition (4.1.1) was first articulated by Mahmassani and Chang (1987), and was, among others, formalized by Di et al. (2013) and Lou et al. (2010). Condition (4.1.1) formulates a range of allowable travel times for a user by introducing indifference band or *tolerance vector* $\epsilon \in \mathbb{R}_+^{|\mathcal{K}|}$. This is clearly in contrast to PRUE condition (3.2.1) that only contains a single value of allowed travel times for each OD pair: $\min_{q \in \mathcal{P}_k} c_q(f), k \in \mathcal{K}$. Consequently, the BRUE flow distribution (i.e. satisfies (4.1.1) is a traffic state in which the travel time of any flow-carrying path is within the formulated range. A BRUE flow may possess the property that all minimum cost paths carry no flow (assuming $\epsilon > 0$). In the remainder, we call a path $p \in \mathcal{P}_k, k \in \mathcal{K}$, *acceptable* if $c_p(f) \leq \min_{q \in \mathcal{P}_k} c_q(f) + \epsilon_k$ is satisfied for this path.

We assume that the indifference band $\epsilon_k, k \in \mathcal{K}$, in (4.1.1) is exogenous, i.e. the range of acceptable travel times is independent of traffic state (f, x). Moreover, we assume that ϵ_k does not differ among users of the same commodity $k \in \mathcal{K}$. These assumptions might not reflect real-world situations. For example, some travelers might be oblivious in their route choice with respect to congestion (they choose their route with respect to shortest distance) while others, more frequent users, choose routes based on previous experienced delays. This behavior can easily be incorporated by introducing heterogeneous users (adding a superscript on user type dependent variables and parameters) or by approaching the indifference band as a random variable. An endogenous tolerance vector $\epsilon(f)$, i.e. ϵ is a function of f, as in Han et al. (2015), might be more appropriate in a dynamic setting.

Remark 4.1.1 (Multiplicative indifference band). *It remains an open question whether it is preferable to incorporate an additive (as in condition (4.1.1)) or multiplicative indifference band with respect to the minimum cost path (see Christodoulou et al. (2011), Roughgarden and Tardos (2002)):*

$$f_p > 0 \Rightarrow c_p(f) \leq \min_{q \in \mathcal{P}_k} c_q(f)(1 + \epsilon_k). \tag{4.1.2}$$

Empirical studies do not (yet) answer this question. From a mathematical analysis point of view the condition in (4.1.1) has major advantages. We will highlight results (see Remark 6.3.1 and Remark 6.4.1) in the remainder of this study in which the analysis does not hold under assumptions of a multiplicative indifference band as in (4.1.2).

4.2 Illustrative Examples

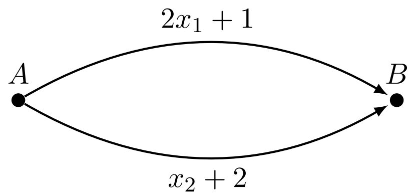
In previous sections, we discussed the differences in modeling route choice decision making based on perfectly rational and boundedly rational choice behavior. In this section, illustrative examples show the implications using BRUE rather than PRUE in the NDP.

Given a network instance, the PRUE link flow distribution x^n is uniquely determined. So the corresponding system's objective value $s(x^n)$ is uniquely determined as well. Let us show that under BRUE condition (4.1.1) uncertainty among performances with respect to $s(x)$ and x arises.

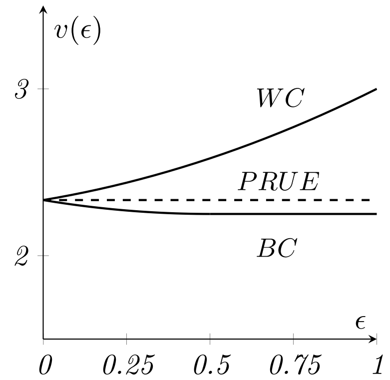
Example 4.2.1 (Uncertainty among performances). *Consider the two-link network in Figure 4.2.1a. Both paths* $(1, 2)$ *connect commodity* (A, B) *and are endowed with a strictly monotone linear latency function. We send demand* $d = 1$ *from node* A *to* B*.*

The PRUE link flow distribution x^n *(see condition (3.2.1)) is uniquely determined:*

$$x^n = \left(\frac{2}{3}, \frac{1}{3}\right).$$



(a) Example Network.



(b) Uncertainty among performances.

Figure 4.2.1: Network and performance results for Example 4.2.1. *(BC = Best-case, WC = Worst-case.)*

Given $\epsilon \geq 0$ we have a set of link flows satisfying (4.1.1). This set of feasible flows is denoted by $\mathcal{F}(\epsilon)$ and given by

$$\mathcal{F}(\epsilon) = \left\{ \ (x_1, (1 - x_1)) \ \middle| \ \tfrac{2-\epsilon}{3} \leq x_1 \leq \tfrac{2+\epsilon}{3}, \ 0 \leq x_1 \leq 1 \ \right\}.$$

As discussed, non-uniqueness of BRUE flow distributions may lead to difficulties in determining performance of a certain measure. Let us consider the best and worst-performing BRUE with respect to total travel time for a given $\epsilon \geq 0$. Hence, for any $\epsilon \geq 0$ we find the flows that satisfy (4.1.1) and minimizes and maximizes $s(x)$ respectively.

Figure 4.2.1b plots the total travel time $v(\epsilon)$ in both cases for each $\epsilon \in [0, 1]$ in . We show that performance may differ significantly. Further, we make the observation that the Best- and Worst-case BRUE provide a lower and upper bound respectively for the total travel time under Wardrop's first principle (i.e. $s(x^n)$). Note that for $\epsilon = 0$ both the Best- and Worst-case BRUE reduces to the PRUE flow x^n.

We interpret Example 4.2.1 as follows. Compared to the PRUE, the BRUE flow distribution is not necessarily unique. In the NDP, any upper-level policy instrument leads to a set of BRUE flows among which performance may differ substantially. If the lower level of the NDP adapts the PRUE condition (3.2.1), a naive measure - optimal under this condition - might be implemented. Indeed, in the next example we show that policy interventions designed under Wardrop's first principle might lead to an inferior performing Worst-case BRUE. Basically, Example 4.2.2 illustrates that if we use bounded rationality in the static traffic assignment, then we should base or tolling scheme on the BRUE assignment as well.

Example 4.2.2 (Optimal toll setting). *A typical application of the NDP is the derivation of optimal toll prices for (certain) links in the network (see Yang and Huang (2005)*

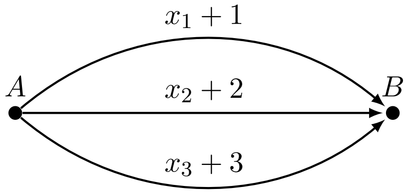
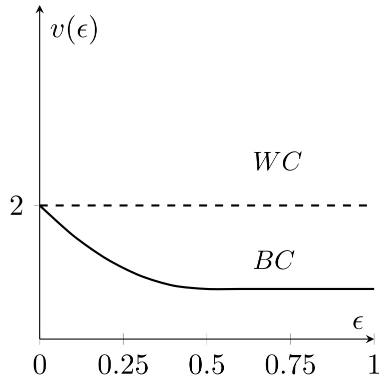
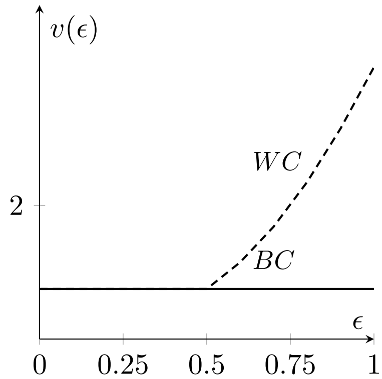


Figure 4.2.2: Network in Example 4.2.2.

for an overview). Assume we have the possibility to set (time-independent) tolls τ on all edges in the network.



(a) Best- and Worst-case performance with no tolls.



(b) Performance with tolls.

Figure 4.2.3: Performance results for Example 4.2.2. *(BC = Best-case, WC = Worst-case.)*

We look at the simple parallel-link network (three paths connect commodity (A, B)) (Figure 4.2.2). Based on Beckmann's minimization problem (Q_n) and the system optimum (Q_s) we arrive at the PRUE and system-optimal flow distribution x^n, x^s, respectively:

$$x^n = (1, 0, 0)\,, \qquad x^s = \left(\frac{3}{4}, \frac{1}{4}, 0\right).$$

We derive the first-best pricing strategy: a toll pricing strategy that allows all links to be tolled with any toll. Recall that we have for a system-optimal flow (f^s, x^s) the equilibrium KKT condition

$$\sum_{e \in p} l_e(x_e^s) + x_e^s l'_e(x_e^s) \geq \lambda_k^s, \qquad \textit{for all } p \in \mathcal{P}_k, k \in \mathcal{K}.$$

Note that such a flow distribution will not occur under perfect rationality conditions since individuals then only consider $l_e(x_e)$ and not $l'_e(x_e)x_e$ (i.e. the impact of their choice on others). The first-best tolling strategy easily follows from previous equation if the disutility function incorporates $l'_e(x_e)x_e$. Assume Wardrop's equilibrium will be realized. An optimal *toll vector $\tau \geq 0$ that satisfies the following conditions leads to a system optimum under perfectly rational decision making (Yang and Huang, 2005):*

$$l_e(x_e^s) + \tau_e \geq \min_{e \in E}(l_e(x_e^s) + x_e^s), \qquad \sum_{e \in E}(l_e(x_e^s) + \tau_e)x_e^s = d \min_{e \in E}(l_e(x_e^s) + x_e^s).$$

So, a first-best tolling scheme would be to impose $\tau_e = x_e^s l_e'(x_e^s) = x_e^s$ for all $e \in E$. Figure 4.2.4 depicts the link cost functions including the tolls.

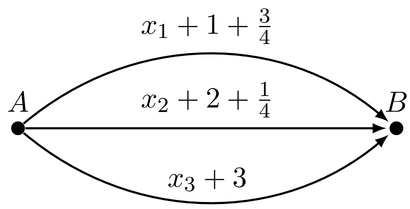


Figure 4.2.4: Latency functions with tolls for Example 4.2.2 .

Suppose in practice a BRUE flow, $\epsilon \geq 0$, is realized. We show again the Best- and Worst-case performing BRUE (with and without the induced toll) and compare global performance in Figure 4.2.3a and Figure 4.2.3b.

The results in Figure 4.2.3b show that the first-best tolling scheme always induces a system optimum if we assume that the Best-case BRUE is realized. This follows from the observation in Example 4.2.1: the Best-case BRUE performs at least as good as the PRUE and since the PRUE reduces to a system optimum, the Best-case BRUE reduces to this global optimum as well.

However, for certain values of $\epsilon \geq 0$, edge 3 becomes acceptable (i.e. travel time on this edge is within ϵ from the shortest). This was not taken into account in the first-best tolling scheme since $x_3^s = x_3^n = 0$. In the Worst-case BRUE, path 3 might be utilized and therefore worsens overall performance of the network.

Example 4.2.2 supports our earlier claim that the toll scheme should be adapted to its behavioral assumptions. It might be suggested that an authority induces a toll after observing flows in a transport network: the naive PRUE flow distribution will not arise in practice. However, a PRUE flow is also a BRUE traffic flow and thus the PRUE distribution might be observed after all. Example 4.2.2 shows that the Worst-case BRUE performance with the tolls might be worse compared to the Worst-case BRUE without tolls. We suggest that a pricing scheme which minimizes Worst-case BRUE performance is more appropriate in this setting.

Another example of a worse performing BRUE flow after imposing tolls was presented in Lou et al. (2010). However, there, Assumption 1 was violated and a constant latency function was used on one of the paths.

In practice, not only link flow x is of importance, but route flow f as well. For instance, relieving a bottleneck by giving pre-trip route advice requires that the link flow is disentangled by commodities. As widely discussed, there may exist several route flows f that correspond to a single link flow x (Lu and Nie, 2010). An arbitrary route flow f corresponding to x (e.g. as by-product of a convex combination method in the PRUE) may lead to several difficulties. First, in a PRUE, it may cause discontinuity in f with respect to perturbations; a small change in the demand leads, from a behavioral perspective, to completely different route choice behavior (Lu and Nie, 2010). Second, several route flow distributions corresponding to the same x are incomparable (Borchers et al., 2015). Again, let us regard the set $\mathcal{H}(x)$ of route flows for known link flow distribution x of BRUE solution (f, x, λ) (and fixed d) as a point-to-set mapping:

$$\mathcal{H}(x) = \left\{ f \in \mathbb{R}^{|\mathcal{P}|} \mid \Lambda f = d, \Gamma f = x, f \geq 0 \right\}.$$

It is desirable to have a signle realistic route flow and therefore we impose additional conditions on f. Conventional methods in literature use a two-stage approach to derive PRUE route flow f^n out of x^n (e.g. by maximizing entropy). That is: first calculate link flow x and, independently, use x to find f. We indicate that additional difficulties arise under bounded rationality compared to perfect rationality. If we choose $f \in \mathcal{H}(x)$ arbitrarily this may not only lead to discontinuous and incomparable route flows in a PRUE setting, it may also generate route flows which are infeasible under BRUE condition (4.1.1) (see Example 4.2.3 below).

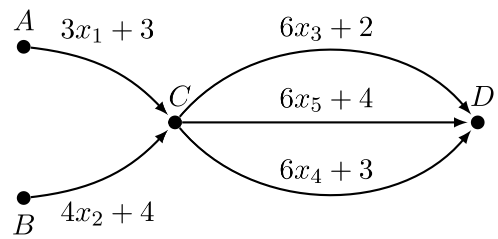


Figure 4.2.5: Considered network in Example 4.2.3.

Example 4.2.3 (Determining f out of x). *Consider the network in Figure 4.2.5. Assume there are two commodities (A, D) and (B, D), both with demand 1. We have six path flows: f_{13}, f_{14}, f_{15}, f_{23}, f_{24}, f_{25}. Given link flow x, route flow f is not uniquely determined since*

$$\mathit{rank} \begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix} = 4 < 6.$$

Suppose $\epsilon_{(A,D)} = 2$ and $\epsilon_{(B,D)} = 1$. A BRUE link flow is:

$$x = (x_1, x_2, x_3, x_4, x_5) = \left(1, 1, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right),$$

and the corresponding path cost vector $c(x)$:

$$c(x) = (c_{13}(x), c_{14}(x), c_{15}(x), c_{23}(x), c_{24}(x), c_{25}(x)) = (12, 13, 14, 14, 15, 16).$$

Choose two route flows $f^1, f^2 \in \mathcal{H}(x)$:

$$f^1 = (f^1_{13}, f^1_{14}, f^1_{15}, f^1_{23}, f^1_{24}, f^1_{25}) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right),$$

$$f^2 = \left(0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{1}{3}, 0\right).$$

In case f^1 is the resulting route flow, the flow is infeasible since $c_{25}(x) - c_{23}(x) = 2 > \epsilon_{(B,D)} = 1$. However, if we choose f^2, $f^2_{25} = 0$ and all paths that carry flow are in the range of acceptable travel times:

$$\tilde{f}_{25}\left(c_{25}(x) - c_{23}(x) - \epsilon_{(B,D)}\right) = 0.$$

We interpret Example 4.2.3 as follows. The choice of route flows $f \in \mathcal{H}(x)$ for a BRUE flow x influences the *experienced travel time*. Given $f^1, f^2 \in \mathcal{H}(x)$ and both correspond to BRUE link flow distribution x (as in the example). The example shows that the maximum travel time among all flow-carrying paths differs for f^1 and f^2. Indeed, for f^1 the maximum travel time for commodity (B, D) is 16 while the maximum travel time under f^2 is 15.

All in all, given BRUE flow x, not all $f \in \mathcal{H}(x)$ are feasible BRUE route flows. Moreover, if we want to give route advice to relieve a bottleneck, the choice of f corresponding to x influences the travel time for travelers.

From a mathematical perspective, this conclusion is trivial since the problem adds additional conditions compared to the problem to find the PRUE flow. Later we will see that the non-uniqueness of route flows plays an essential role in the continuity of the feasible set and optimal value function of the Best-case BRUE flow distribution.

Remark 4.2.1. *We note that the issue of infeasible route flows $f \in \mathcal{H}(x)$ does not occur in a PRUE and system-optimal setting. Nonetheless, in a system optimum (and any other flow not in PRUE), the choice of route flow f corresponding to x influences the experienced travel time. In a system-optimal traffic assignment, travelers are assumed to be memoryless: the choice of a bypass at a node is independent of the travel time up to that node. In other words: agents have an unbounded range of acceptable travel times.*

4.3 BRUE in Other Settings

BRUE flow distributions also appear in other settings. The first analogy concerns system-optimal route guidance in traffic networks with user constraints, the second setting describes a more general matrix game.

Route Guidance with User Constraints

Opposed to individual performance, road authorities are in general concerned with the performance of the transport system as a whole. It is well-studied that individual and system interests in this context may interfere: an individual optimum may lead to a degradation of the transport system as a whole (Roughgarden, 2005). The system-optimal traffic flow distribution (f^s, x^s) is known to require cooperation among travelers or complete coordination of the authority in all route choice (Yang and Huang, 2005). A possible measure to achieve this system-optimal traffic assignment is by means of route guidance systems.

Current navigation systems, however, provide descriptive or advisory information and guide travelers in a self-interest way based on shortest instantaneous travel times. As mentioned, this may lead to an inferior performance of the system. A traffic routing strategy that aims for a system optimum diverts some individuals to significantly longer routes to minimize global performance (Jahn et al., 2005). As a result, these drivers may not comply with such advices and the optimal state will not be achieved, i.e. in a system-optimal assignment unfairness or inequity among individuals arises.

For our purpose, let us define unfairness $\eta_k(f)$ for traffic flow (f, x) as the maximum difference in travel time between a used route and the cost the fastest alternative for commodity $k \in \mathcal{K}$:

$$\eta_k(f) = \max\left(\ c_p(f) - \min_{q \in \mathcal{P}_k} c_q(f) \ \middle| \ f_p > 0, p \in \mathcal{P}_k \ \right). \tag{4.3.1}$$

We show in the next proposition that a system optimum suffers under unfairness issues, earlier presented for a single commodity in Correa et al. (2007). In particular, if we consider the BPR-function, in a system-optimal route guidance some travelers end up traveling 5 times as long as others. Note that $\eta_k(f^n) = 0$ for all $k \in \mathcal{K}$ where (f^n, x^n) satisfies (3.2.1)

Proposition 4.3.1. *(Correa et al., 2007). Suppose $\delta \in \mathbb{N}$ is the maximum degree of the polynomial of the link cost function $l(x)$ used in the network. Given that (f^s, x^s) solves (Q_s) then*

$$\eta_k(f^s) \leq (\delta + 1) \min_{q \in \mathcal{P}_k} \sum_{e \in q} l_e({x_e}^s)$$

for all $k \in \mathcal{K}$.

Proof. The proof is similar to the proof in Correa et al. (2007). It is shown by Roughgarden (2002) that under Assumption 1, $l_e(x_e)+x_e l_e'(x_e) \le (\delta+1)l_e(x_e)$ for all $e \in E$ and for all $x_e \ge 0$. For instance, for monomials (i.e. $l_e(x_e) = {x_e}^{\delta}$), $l_e(x_e) + x_e l_e'(x_e) = (\delta+1){x_e}^{\delta}$. Hence, we can upper bound the marginal cost of a path by a constant factor of the latency function. A lower bound on the marginal cost for a link is straightforward under Assumption 1: $l_e(x_e) + x_e l_e'(x_e) \ge l_e(x_e)$.

Let λ_k^s denote the minimum marginal cost for commodity $k \in \mathcal{K}$ for system-optimal traffic flow (f^s, x^s). Recall that all flow-carrying paths experience these marginal cost (3.2.5). Pick $p \in \mathcal{P}_k, k \in \mathcal{K}$ so that $f_p > 0$ and take $q \in \mathcal{P}_k$ as a minimum latency path for commodity $k \in \mathcal{K}$, we arrive at:

$$\lambda_k^s = \sum_{e\in p} l_e(x_e^s) + {x_e}^s l_e'({x_e}^s) \ge \sum_{e\in p} l_e({x_e}^s) \ge \sum_{e\in q} l_e({x_e}^s),$$

and

$$\lambda_k^s \le \sum_{e\in q} l_e(x_e^s) + {x_e}^s l_e'({x_e}^s) \le \sum_{e\in q} (\delta+1) l_e({x_e}^s).$$

Hence, it turns out that the travel time is bounded by a constant factor from the travel time on the shortest path:

$$\sum_{e\in p} l_e({x_e}^s) \le (\delta+1) \min_{q\in\mathcal{P}_k} \sum_{e\in q} l_e({x_e}^s).$$

The claim of the proposition follows directly. $\square$

van Essen et al. (2016) pointed out that users are willing to use alternatives with higher travel times as long as the difference with respect to the fastest path is found to be acceptable. Therefore, Jahn et al. (2005) proposed a system-optimal routing strategy incorporating unfairness among individuals with respect to an apriori determined *normal length*. Unfairness with respect to this normal length - independent of flow - was chosen for computational purposes. We suggest that the distribution of travel time among users is a more appropriate indicator to measure unfairness.

Let us assume that the travel time differences as defined in (4.3.1) solely determines the acceptability of a route guidance system. The problem of finding a system-optimal traffic route guidance system with user constraints (i.e. every traveler uses an acceptable route) becomes:

$$\min_{(f,x)} s(x) \qquad \text{s.t.} \qquad \begin{array}{rl} (f,x) & \text{as in (3.1.1);} \\ \eta_k(f) \le \epsilon_k & \text{for all } k \in \mathcal{K}. \end{array} \tag{4.3.2}$$

In this problem, $\epsilon_k \ge 0$ is the upper bound in travel time difference for commodity $k \in \mathcal{K}$. Obviously, problem (4.3.2) is equivalent to the problem that finds the BRUE flow distribution as in (4.1.1) which minimizes $s(x)$ (i.e. Best-case BRUE).

ϵ-Nash Equilibria

Game Theory often uses Nash equilibria to model the behavior of individuals in non-cooperative games (*cf.* Section 2.1). Here, the player's strategy is the best reply given the other players strategy (Peters, 2008). Then, an equilibrium is a solution of a game in which no player can unilaterally change strategies and improve his or her utility (note the equivalence with PRUE condition (3.2.1)). An ϵ-(approximate) Nash equilibrium is a strategy profile in which no player can improve payoff by more than $\epsilon \geq 0$ (Nisan et al., 2007). Let us formally define this solution concept using the notation of Nisan et al. (2007).

Adopting the definition of Nash equilibrium of Nisan et al. (2007), strategy vector $s \in S$ is said to be an *ϵ-Nash equilibrium*, $\epsilon \geq 0$, if for all players $i \in \{1, \ldots, n\}$ and each alternative strategy $s'_i \in S_i$, we have that

$$u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i}) - \epsilon. \tag{4.3.3}$$

Here, s_i denotes the strategy played by i and s_{-i} the strategies played by all other players. $u_i(s_i, s_{-i})$ is the utility of player i. Trivially, condition (4.3.3) reduces to the definition of a conventional Nash equilibrium when $\epsilon = 0$.

ϵ-Nash equilibria gained attention in behavioral game theory (incorporating experimental economics into strategic decision making) to efficiently approximate *true* Nash equilibria (Nisan et al., 2007) and in (non-)atomic routing games (Christodoulou et al., 2011). In non-atomic routing games, condition (4.3.3) reduces to (4.1.1). An interesting open problem in the latter field is upper bounding the Price of Anarchy for flows (f, x) that satisfy condition (4.1.1).

4.4 Synthesis

We introduced the BRUE in a static traffic assignment. The BRUE is an equilibrium in which no traveler can substantially improve his or her travel time by changing routes. In addition, we gave basic examples that showed the main difficulties of boundedly rational choice behavior in the NDP. We indicated that uncertainty among performances arises, a naive toll setting in the network may lead to worse performance of the system, and additional conditions should be imposed on the set of route flows.

Chapter 5

Best-Case BRUE

Authorities concern themselves with the impact of measures on transportation systems. Lou et al. (2010) highlighted that a BRUE flow distribution is not necessarily unique and causes uncertainty among performances (see also Section 4.2). So it is important to assess performance of flow distributions with best and worst-case performance with respect to system's objective (Best/Worst-case BRUE). Following our discussion, the Best/Worst-case BRUE flow is a solution of the program (5.0.1).

$$\min / \max_{(f,x)} s(x) \qquad \text{s.t.} \qquad (f, x) \text{ satisfies (4.1.1)} \tag{5.0.1}$$

In our case, the system cost function $s(x)$ in (5.0.1) will be the total travel time (3.2.4). Nevertheless, any other objective function can be used as long as it is strictly convex with respect to link flow x. The upcoming sections particularly focus on the Best-case BRUE problem.

We emphasize that the formulation in (5.0.1) does not necessarily find the BRUE flow distribution which is most likely to be realized in practice. Some BRUE flow distributions may be unrealistic in real-world applications. Since we assume that an authority has no control over which BRUE will be realized for a given tolerance vector, we only consider the extreme cases (best, respectively worst, system performance).

5.1 Problem Definition

We formulate the boundedly rational user equilibrium model, qualitatively and quantitatively defined in previous chapters, as a mathematical optimization problem. We discuss basic topological properties of the problem. The next sections extend the discussion and consider continuity of the feasible set, optimal value function and optimal solution set with respect to perturbations in the indifference band parameter. Before we continue, let us introduce the topological definitions.

- We denote the n-dimensional Euclidean space and its non-negative orthant by $\mathbb{R}^n$ and $\mathbb{R}^n_+$ respectively;
- The *Euclidean norm* of a vector $x \in \mathbb{R}^n$ is defined by $\|x\| := \left(\sum_{i=1:n}|x_i|^2\right)^{\frac{1}{2}}$.
- The *Frobenius norm* of a matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ is defined as $\|A\| = \|(a_{ij})\|_2$;
- A *multifunction* or *point-to-set mapping* F of $X \subseteq \mathbb{R}^n$ into $Y \subseteq \mathbb{R}^m$ assigns to each $x \in X$ a (possibly empty) subset $F(x)$ of Y;
- The *domain* of a multifunction F consists of all $x \in X$ for which $F(x)$ is nonempty: $\operatorname{dom}(F) := \{x \in X \mid F(x) \neq \emptyset\}$;
- $U_r(A)$ is defined as a *neighborhood* of radius $r \in \mathbb{R}_+$ around set $A : U_r(A) := \{x \in X \mid \inf_{y\in A}\|x - y\| < r\}$.

The optimization problem to find the Best-case BRUE flow distribution(s) (5.0.1) is equivalently formulated as follows. We minimize total travel time while the constraints assure that the difference in travel costs between users of the same commodity $k \in \mathcal{K}$ does not exceed indifference band $\epsilon_k \geq 0$. More precisely, the Best-case BRUE is a solution of

$$\min_{(f,x,\lambda)} s(x) \qquad \text{s.t.} \qquad (f,x,\lambda) \in \mathcal{F}(\epsilon), \tag{$Q(\epsilon)$}$$

where the feasible set $\mathcal{F}(\epsilon)$ is given by

$$\mathcal{F}(\epsilon) = \left\{ (f,x,\lambda) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} \times \mathbb{R}^{|\mathcal{K}|} \;\middle|\; \begin{array}{l} (f,x) \in \mathcal{F}_0 \\ (f,x,\lambda) \in \mathcal{F}_l \\ (f,x,\lambda) \in \mathcal{F}_u(\epsilon) \end{array} \right\}.$$

Here, $\mathcal{F}_0$ assures that the resulting traffic flow (f,x) is feasible (i.e. (f,x) is as in (3.1.1)). Due to the linear (in)equalities, $\mathcal{F}_0$ is a convex set and is formally defined to be

$$\mathcal{F}_0 = \{(f,x) \mid x = \Gamma f, \Lambda f = d, f \geq 0\}.$$

$\mathcal{F}_l$ and $\mathcal{F}_u(\epsilon)$ arise due to the bounded rationality condition (4.1.1). The set $\mathcal{F}_l$ is defined as

$$\mathcal{F}_l = \{(f,x,\lambda) \mid -c_p(x) + \lambda_k \leq 0 \text{ for all } p \in \mathcal{P}_k, k \in \mathcal{K}\},$$

and $\mathcal{F}_u(\epsilon)$, $\epsilon \geq 0$, is defined to be

$$\mathcal{F}_u(\epsilon) = \{(f,x,\lambda) \mid f_p > 0 \Rightarrow c_p(x) - \lambda_k \leq \epsilon_k \text{ for all } p \in \mathcal{P}_k, k \in \mathcal{K}\}.$$

The term λ_k is (a lower bound on) the minimum travel cost for commodity $k \in \mathcal{K}$. The set $\mathcal{F}_l$ assures that the travel time on a path is at least the minimum travel cost for that commodity, while $\mathcal{F}_u(\epsilon)$ defines the travel time of a flow-carrying path to be within the range $[\lambda_k, \lambda_k + \epsilon_k]$. Obviously, any flow $(f, x) \in \mathcal{F}(\epsilon)$ satisfies the BRUE condition (4.1.1). Conversely, any (f, x) that satisfies (4.1.1) is contained in $\mathcal{F}(\epsilon)$ by taking $\lambda_k = \min_{p \in \mathcal{P}_k} c_p(x)$ for all $k \in \mathcal{K}$. Hence, we will refer to any point $(f, x) \in \mathcal{F}(\epsilon)$ with corresponding λ as a BRUE (flow distribution).

Consider $\epsilon \in \mathbb{R}_+^{|\mathcal{K}|}$ as a parameter vector. Our problem $(Q(\epsilon))$ is then a *parametric optimization problem.* The objective function $s : \mathbb{R}^{|E|} \to \mathbb{R}$ of $(Q(\epsilon))$ is defined in (3.2.4), is independent of the parameter ϵ and is strictly convex with respect to link flows x.

The parametric problem $(Q(\epsilon))$ assesses performance of BRUE flow distributions with best-case performance (with respect to the system cost). That is, find BRUE flow distribution which minimizes travel time among all possible BRUE flows for a given indifference band ϵ. Figure 5.1.1 abstractly illustrates to process to find the Best-case BRUE: we seek for the flow distribution $(\tilde{f}, \tilde{x})$ closest to (the unique with respect to x) (f^s, x^s) with respect to $s(x)$.

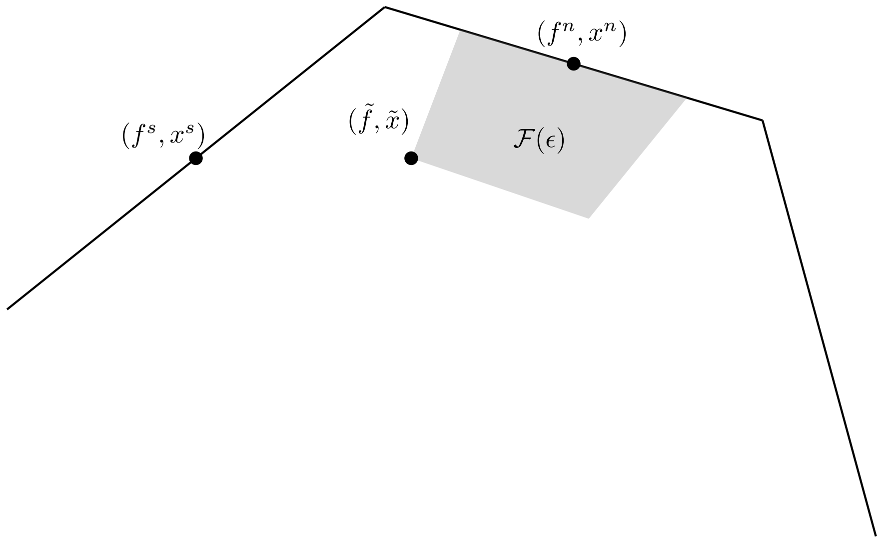


Figure 5.1.1: Abstract illustration of the feasible BRUE set $\mathcal{F}(\epsilon)$. $(Q(\epsilon))$ finds $(\tilde{f}, \tilde{x})$ closest to (f^s, x^s) with respect to $s(x)$.

Unfortunately, although as such illustrated in Figure 5.1.1, we cannot expect $(Q(\epsilon))$ to be a convex optimization problem for any $\epsilon \in \operatorname{dom}(\mathcal{F})$. Even worse, the next section indicates that certain regularity conditions do not hold in general.

5.2 Properties of $\mathcal{F}(\epsilon)$

Given tolerance vector $\epsilon \geq 0$, the set $\mathcal{F}_0 \cap \mathcal{F}_l \cap \mathcal{F}_u(\epsilon) = \mathcal{F}(\epsilon)$ can be defined equivalently by complementarity:

$$\mathcal{F}(\epsilon) = \left\{ (f, x, \lambda) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} \times \mathbb{R}^{|\mathcal{K}|} \;\middle|\; \begin{array}{r} (f, x) \in \mathcal{F}_0 \\ f_p(c_p(x) - \lambda_k - \epsilon_k) \leq 0 \text{ for all } p \in \mathcal{P}_k, k \in \mathcal{K} \\ -c_p(x) + \lambda_k \leq 0 \text{ for all } p \in \mathcal{P}_k, k \in \mathcal{K} \end{array} \right\}. \tag{5.2.1}$$

Equivalence follows directly for a traffic flow (f, x, λ) satisfying (5.2.1) since

$$f_p > 0 \Rightarrow c_p(x) \leq \lambda_k + \epsilon_k, \qquad \text{and} \qquad f_p \geq 0 \Rightarrow c_p(x) \geq \lambda_k$$

holds for all $p \in \mathcal{P}_k, k \in \mathcal{K}$. $\mathcal{F}(\epsilon)$ as in (5.2.1) requires no explicit information on which paths carry flow (which is required in the system of inequalities defining $\mathcal{F}_u(\epsilon)$).

Lou et al. (2010) illustrated by a simple example that the feasible set $\mathcal{F}(\epsilon)$ for the problem $(Q(\epsilon))$ is not necessarily convex. We further extend this observation by means of an illustration.

For simplicity, we reduce the problem to one dimension in the reals. Consider, for a fixed $\epsilon \geq 0$, the following system of inequalities:

$$\theta(\upsilon - \epsilon) \leq 0, \qquad \theta \geq 0, \qquad \upsilon \geq 0. \tag{5.2.2}$$

Note the similarity to the system in (5.2.1). Figure 5.2.1 depicts the feasible set that corresponds to the system of inequalities (5.2.2). Clearly, taking (θ^1, υ^1) where $\theta^1 = 0$, $\upsilon^1 > \epsilon$ and (θ^2, υ^2) so that $\theta^2(\upsilon^2 - \epsilon) < 0$, we cannot draw a line between these two points so that all points on this line remain feasible.

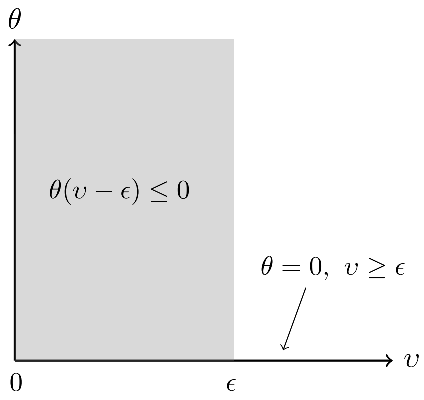


Figure 5.2.1: Illustration of the feasible set corresponding to the system of inequalities (5.2.2) for arbitrary $\epsilon > 0$.

Due to the complementarity constraint in (5.2.1), the problem ($Q(\epsilon)$) turns into a (parametric) *mathematical program with complementarity constraints.* These programs are difficult to solve numerically and different local minimizers may co-exist. The difficulties arise since these programs lack regularity conditions. In other words, at a minimum point the KKT condition does not necessarily hold and standard optimization software cannot be used.

We show that a feasible point $(f, x, \lambda) \in \mathcal{F}(\epsilon)$ not necessarily satisfies the *Mangasarian-Fromovitz constraint qualification* (MFCQ). As a consequence, we cannot express the gradient of the objective function as a linear combination of the gradients of the constraint functions.

For notational purposes, we consider the system defined in (5.2.2). We find the following gradients with respect to (θ, υ), i.e. $\nabla = \nabla_{(\theta,\upsilon)}$, of the constraint functions given $\epsilon \geq 0$:

$$\nabla(\theta(\upsilon - \epsilon)) = \begin{bmatrix} \upsilon - \epsilon \\ \theta \end{bmatrix}, \qquad \nabla(-\theta) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \qquad \nabla(-\upsilon) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

In case the constraint $\theta(\upsilon - \epsilon) \leq 0$ is active (i.e. $\theta(\upsilon - \epsilon) = 0$), it follows that $\theta = 0$ and/or $\upsilon = \epsilon$. Assume the former and $\upsilon - \epsilon > 0$. We cannot find an $\eta \in \mathbb{R}^2$ so that for the active gradients

$$\begin{bmatrix} \upsilon - \epsilon \\ 0 \end{bmatrix}^T \eta < 0, \qquad \text{and} \qquad \begin{bmatrix} -1 \\ 0 \end{bmatrix}^T \eta < 0$$

holds. MFCQ does not hold: the active inequality constraints are linear dependent under the mentioned assumptions.

Numerous studies in literature discuss the difficulties as result of the violation of the MFCQ and tried to overcome these issues by using a *disjunctive* or *piecewise* approach (see Luo et al. (1996)). Our problem does not fulfill the strict definition of complementarity constraints defined in literature. Nonetheless, our problem ($Q(\epsilon)$) suffers under the same difficulties, even for fixed $\epsilon \in \operatorname{dom}(\mathcal{F})$. Later in this study, we proceed by using some techniques of the disjunctive approach to analyze the topological structure of ($Q(\epsilon)$).

5.3 Simple Cases of ($Q(\epsilon)$)

We regard ϵ as a parameter vector whereas ($Q(\epsilon)$) reduces in certain cases of ϵ to an easier to solve convex optimization problem. In particular, problem ($Q(\epsilon)$) reduces to the (Q_n) problem in case $\epsilon = 0$ and to (Q_s) when the tolerance vector ϵ is sufficiently large. So, we use the characteristics of these optimization problems to analyze ($Q(\epsilon)$) for an arbitrary $\epsilon \geq 0$.

Recall the complementarity definition of $\mathcal{F}(\epsilon)$ of (5.2.1) and assume $\epsilon = 0$, we obtain:

$$\mathcal{F}(0) = \left\{ (f, x, \lambda) \;\middle|\; \begin{array}{c} (f, x) \in \mathcal{F}_0 \\ f_p(c_p(x) - \lambda_k) \leq 0 \text{ for all } p \in \mathcal{P}_k, k \in \mathcal{K} \\ -c_p(x) + \lambda_k \leq 0 \text{ for all } p \in \mathcal{P}_k, k \in \mathcal{K} \end{array} \right\}.$$

Since $f_p \geq 0$ and $-c_p(x) + \lambda_k \leq 0$ holds for all $p \in \mathcal{P}_k, k \in \mathcal{K}$, the first set of defining inequalities of $\mathcal{F}(0)$ becomes $f_p(c_p(x) - \lambda_k) = 0$. Then, $\mathcal{F}(0)$ is nothing else than complementarity condition (3.2.3) for the PRUE. In other words, in case $\epsilon = 0$, BRUE reduces to a PRUE, which was also noted by Lou et al. (2010). Furthermore, $\mathcal{F}(0)$ is a singleton with respect to link flow x under Assumption 1 (we made this observation in Example 4.2.1).

When the tolerance vector ϵ becomes large, it turns out that problem $(Q(\epsilon))$ reduces to the system-optimal traffic assignment. The traffic flow $(f^s, x^s) \in \mathcal{F}_0$ is the (unique with respect to x) flow which minimizes total travel time, i.e. a solution of (Q_s). Straightforward analysis shows that if $(f^s, x^s) \in \mathcal{F}_0$ is a system-optimal traffic flow and if we define

$$\epsilon_k^{so} := \max_{p \in \mathcal{P}_k : f_p > 0} c_p(x^s) - \lambda_k^s, \tag{5.3.1}$$

with $\lambda_k^s = \min_{q \in \mathcal{P}_k} c_q(f^s)$, the feasibility constraints

$$f_p^s\,(c_p(x^s) - \epsilon_k^{so} - \lambda_k^s) \leq 0, \qquad \text{and} \qquad -c_p(x^s) + \lambda_k^s \leq 0,$$

of $(Q(\epsilon^{so}))$ are satisfied for all $p \in \mathcal{P}_k, k \in \mathcal{K}$. So, any $\epsilon \geq \epsilon^{so}$ in the Best-case BRUE problem $(Q(\epsilon))$ leads to a system-optimal traffic flow (f^s, x^s), i.e. (f^s, x^s) satisfies (3.2.5).

We showed in this section that the BRUE problem $(Q(\epsilon))$ reduces to a convex optimization problem for particular cases of ϵ. Known methods (e.g. Frank-Wolfe method) can easily solve these problems. Due to the lack of regularity conditions of $(Q(\epsilon))$ for $\epsilon \in (0, \epsilon^{so})$ we cannot apply these methods to this problem in these cases. However, we use the characteristics of (Q_n) and (Q_s) to analyze BRUE problem $(Q(\epsilon))$.

5.4 Existence and Uniqueness of the BRUE

For optimization problems, it is essential to know whether a mathematical program has a solution and, if so, whether it is unique or not. We discussed that under some choices of ϵ the complementarity constraint problem $(Q(\epsilon))$ turns into a convex problem which has under mild assumptions (i.e. Assumption 1) a unique corresponding link flow distribution x.

We show that for any $\epsilon \geq 0$, the feasible set $\mathcal{F}(\epsilon)$ is nonempty. Notice that the next relation holds for $\epsilon, \in \mathrm{dom}\,(\mathcal{F})$ and $\tilde{\epsilon} \geq \epsilon$:

$$(f, x, \lambda) \in \mathcal{F}(\epsilon) \Rightarrow (f, x, \lambda) \in \mathcal{F}(\tilde{\epsilon}).$$

A path, which transit time is within the acceptable range bounded by indifference band ϵ, is certainly within the bounds defined by $\tilde{\epsilon} \geq \epsilon$, assuming $\epsilon \in \mathrm{dom}\,(\mathcal{F})$. Indeed (Di et al., 2013),

$$c_p(f) \leq \lambda_k + \epsilon_k \leq \lambda_k + \tilde{\epsilon}_k, \qquad \text{for all } p \in \mathcal{P}_k, k \in \mathcal{K}.$$

$\mathcal{F}(\epsilon)$ is nonempty for $\epsilon \geq 0$ is then proven if $\mathcal{F}(0)$ is nonempty. We discussed that $\mathcal{F}(0) \neq \emptyset$ as $(f^n, x^n, \lambda^n) \in \mathcal{F}(0)$ (where (f^n, x^n, λ^n) is the solution of (Q_n)). Consequently,

$$\mathrm{dom}\,(\mathcal{F}) = \{\epsilon \mid \epsilon \geq 0\}. \tag{5.4.1}$$

The feasible set $\mathcal{F}(\epsilon)$ is in general not a singleton with respect to link flow x for any $0 \neq \epsilon \in \mathrm{dom}\,(\mathcal{F})$. We return to the network of Example 4.2.1. We showed that by increasing indifference band $\epsilon > 0$, more feasible BRUE link flows become available (see Figure 5.4.1). Han et al. (2015) investigated conditions under which non-emptiness of $\mathcal{F}(\epsilon)$ can be guaranteed in a dynamic BRUE setting.

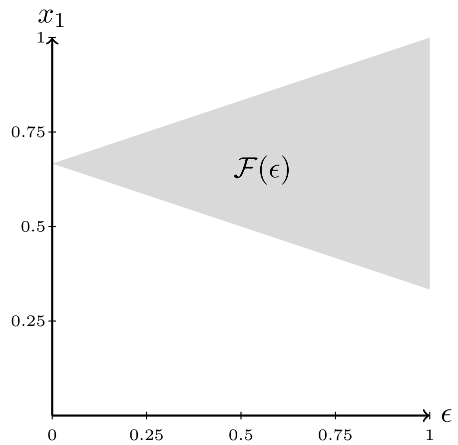


Figure 5.4.1: Example of feasible link flow set $\mathcal{F}(\epsilon)$ for a parallel two-link network.

Remark 5.4.1 (Cyclic flows). *The KKT conditions corresponding to problem (Q_s) and (Q_n) reveal that utilized routes in te mentioned problems are acyclic. Indeed, consider the KKT system (3.2.2) corresponding to (Q_n), we obtain for used route $p \in \mathcal{P}_k, k \in \mathcal{K}$:*

$$f_p^n > 0 \Rightarrow \sum_{e \in p} l_e(x_e^n) = \lambda_k^n.$$

Suppose now that flow-carrying path p contains cycle c. Then, we can find a route $q = p \setminus c$, which has strict less travel cost. This contradicts the fact that in a PRUE flow *distribution all routes have at least cost λ_k^n. However, in particular in a Worst-case BRUE, cyclic routes carrying flow may exist. From a behavioral perspective cyclic route flows are unrealistic and therefore we assume in our case that these flows do not arise.*

5.5 A Branch Approach

We showed that a minimizer of the complementarity constraint problem ($Q(\epsilon)$) does not necessarily satisfy the standard KKT conditions as the constraints of the program violate regularity conditions (e.g. MFCQ). This section proposes to decompose the problem into *branches* of optimization problems to overcome the difficulties of ($Q(\epsilon)$). These branches are easier to study, and benefit the analysis on the the topological structure of ($Q(\epsilon)$).

Choose a subset of paths $P \subseteq \mathcal{P}$, then define a branch of ($Q(\epsilon)$) for $\epsilon \geq 0$ by

$$\min_{(f,x,\lambda)} s(x) \qquad \text{s.t.} \qquad (f,x,\lambda) \in \mathcal{F}_P(\epsilon), \tag{$Q_P(\epsilon)$}$$

where the feasible set $\mathcal{F}_P(\epsilon)$ is a point-to-set mapping from the indifference band ϵ to a set of traffic flows (f,x) and minimum travel cost λ, defined by

$$\mathcal{F}_P(\epsilon) = \left\{ (f,x,\lambda) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} \times \mathbb{R}^{|\mathcal{K}|} \;\middle|\; \begin{array}{l} (f,x) \in \mathcal{F}_0(P) \\ (f,x,\lambda) \in \mathcal{F}_l \\ (f,x,\lambda) \in \mathcal{F}_u(\epsilon,P) \end{array} \right\}.$$

The flow feasibility set $\mathcal{F}_0(P)$ forces (i) the flow on an edge to be the sum of flows on paths $P \subseteq \mathcal{P}$ which pass through this link, (ii) the flow on paths $P \subseteq \mathcal{P}$ to meet demand vector d, and (iii) flows to be non-negative. We emphasize that $\mathcal{F}_0(P)$ is a convex set:

$$\mathcal{F}_0(P) = \left\{ (f,x) \;\middle|\; \begin{array}{ll} f_p = 0 & \text{for all } p \notin P \\ \multicolumn{2}{l}{\Gamma_{\cdot,P} f - x = 0, \Lambda_{\cdot,P} f = d, f_P \geq 0} \end{array} \right\}.$$

(For a given matrix $A \in \mathbb{R}^{m \times n}$ and $L \subseteq \{1,\ldots,n\}$, $A_{\cdot,L}$ is the submatrix with columns of matrix A whose indices are in L).

Recall that $\mathcal{F}_l$ defines λ_k to be (a lower bound on) the minimum travel cost for OD pair $k \in \mathcal{K}$. Note that the constraint $c_p(f) \geq \lambda_k$ is imposed on all paths (including those not in P). $\mathcal{F}_u(\epsilon,P)$ assures the travel costs for paths in P to be within band ϵ from the path with minimum cost, i.e.

$$\mathcal{F}_u(\epsilon,P) = \{(f,x,\lambda) \mid c_p(x) - \lambda_k - \epsilon_k \leq 0, p \in P_k, k \in \mathcal{K}\}.$$

Here, $P_k := P \cap \mathcal{P}_k$, $k \in \mathcal{K}$. ($Q_P(\epsilon)$) is a standard (parametric) optimization problem and overcomes the complementarity constraints of ($Q(\epsilon)$) by choosing a subset $P \subseteq \mathcal{P}$

beforehand. The travel costs of these paths (i.e. $p \in P_k$, $k \in \mathcal{K}$) are forced to be within the indifference band ϵ_k from the minimum travel cost λ_k. The remaining paths (i.e. $p \in \mathcal{P} \setminus P$) carry no flow.

Given $P \subseteq \mathcal{P}$, we can expect the constraints in $\mathcal{F}_P(\epsilon)$ to satisfy the regularity conditions for any $\epsilon \in \mathrm{dom}\,(\mathcal{F}_P)$. However, without any additional information, for any $\epsilon \geq 0$, $2^{|\mathcal{P}|}$ possible choices for P exist. We say that $(Q(\epsilon))$ suffers under a *combinatorial curse.*

In Lemma 5.5.1 we stipulate relations between problems $Q_P(\epsilon))$ and $(Q(\epsilon))$. We use statements of the lemma in the upcoming analysis.

Lemma 5.5.1. *The following relations hold between* $\mathcal{F}_P(\epsilon)$ *and* $\mathcal{F}(\epsilon)$*:*

(i) Any feasible flow for $\mathcal{F}_P(\epsilon)$ *is also feasible for* $\mathcal{F}(\epsilon)$*. That is, given* $P \subseteq \mathcal{P}$ *and* $\epsilon \in$ *dom*$\,(\mathcal{F}_P)$*:*
$$(f, x, \lambda) \in \mathcal{F}_P(\epsilon) \Rightarrow (f, x, \lambda) \in \mathcal{F}(\epsilon);$$

(ii) For all $\epsilon \geq 0$ *and* $(f, x, \lambda) \in \mathcal{F}(\epsilon)$ *there exists a subset of paths* $P^{\epsilon,x} \subseteq \mathcal{P}$ *so that* $(f, x, \lambda) \in \mathcal{F}_{P^{\epsilon,x}}(\epsilon)$*;*

(iii) Assume $\epsilon = 0$*, there exists a subset of paths* $P \subseteq \mathcal{P}$ *so that* $(Q_P(\epsilon))$ *reduces to* (Q_n)*;*

(iv) There exists a tolerance vector ϵ *and subset of paths* $P \subseteq \mathcal{P}$ *so that* $(Q_P(\epsilon))$ *reduces to* (Q_s)*;*

(v) (Di et al., 2013). Assume $0 \leq \epsilon \leq \tilde{\epsilon}$*, then:*
$$\mathcal{F}(\epsilon) = \bigcup_{P \subseteq \mathcal{P} : \mathcal{F}_P(\epsilon) \neq \emptyset} \mathcal{F}_P(\epsilon) \subseteq \bigcup_{P \subseteq \mathcal{P} : \mathcal{F}_P(\tilde{\epsilon}) \neq \emptyset} \mathcal{F}_P(\tilde{\epsilon}) = \mathcal{F}(\tilde{\epsilon}).$$

Proof. We prove the statements of the lemma one by one.

(i) Assume $P \subseteq \mathcal{P}$ and $\epsilon \in \mathrm{dom}\,(\mathcal{F}_P)$. Let $(f, x, \lambda) \in \mathcal{F}_P(\epsilon)$ be given. $(f, x) \in \mathcal{F}_0(P) \Rightarrow (f, x) \in \mathcal{F}_0$ by definition. As we have that for all paths not in P the route flow is null, i.e. $p \notin P \Rightarrow f_p = 0$, we get $f_p(c_p(x) - \lambda_k - \epsilon_k) = 0$ for all these paths. For the routes included in P, we have $p \in P : f_p \geq 0$ it holds that $c_p(x) - \lambda_k - \epsilon_k \leq 0$. Hence, $f_p(c_p(x) - \lambda_k - \epsilon_k) \leq 0$. The claim follows.

(ii) Assume a non-negative indifference band $\epsilon \geq 0$ and let $(f, x, \lambda) \in \mathcal{F}(\epsilon)$. Choose for all commodities $k \in \mathcal{K}$ those paths whose travel times are within the tolerance band ϵ_k from the minimum cost λ_k. That is:
$$P_k^{\epsilon,x} = \{p \in \mathcal{P}_k \mid c_p(x) - \lambda_k - \epsilon_k \leq 0\}, \qquad \text{for all } k \in \mathcal{K}.$$
For each $p \notin P_k^{\epsilon,x}$, $k \in \mathcal{K}$, we have then $f_p = 0$ as these paths are too expensive. Hence, $(f, x, \lambda) \in \mathcal{F}_{P^{\epsilon,x}}(\epsilon)$.

(iii) We discussed earlier that in case $\epsilon = 0$ $(Q(\epsilon))$ reduces to Beckmann's formulation of a PRUE (Q_n). (Q_n) and $(Q_P(\epsilon))$ are equivalent by choosing $\epsilon = 0$ and, for each $k \in \mathcal{K}$,

$$P_k = P_k^n := \{p \in \mathcal{P}_k \mid c_p(x^n) - \lambda_k^n = 0\}. \tag{5.5.1}$$

Here (f^n, x^n), with corresponding KKT multiplier λ^n, is the (unique with respect to link flows x) solution of (Q_n) under Assumption 1.

P is not necessarily uniquely determined. In other words, different choices of P may lead to the same result.

(iv) Denote by (f^s, x^s) with corresponding multiplier vector λ^s the system optimum, i.e. the solution of problem (Q_s). Set, for all $k \in \mathcal{K}$, ϵ_k^{so} as defined in (5.3.1) and choose

$$P_k = \{p \in \mathcal{P}_k \mid c_p(x^s) - \lambda_k^s - \epsilon_k^{so} \le 0\}.$$

In this case $(Q_P(\epsilon^{so}))$ reduces to (Q_s).

(v) Following our discussion, the feasible set $\mathcal{F}(\epsilon)$, $\epsilon \ge 0$, is the union of all possible choices of subsets of paths: $\mathcal{F}(\epsilon) = \cup_{P \subseteq \mathcal{P}} \mathcal{F}(\epsilon)$. Furthermore, let P be fixed and assume $0 \le \epsilon < \tilde{\epsilon}$. We obtain that

$$(f, x, \lambda) \in \mathcal{F}_P(\epsilon) \Rightarrow (f, x, \lambda) \in \mathcal{F}_P(\tilde{\epsilon}),$$

and, thus,

$$\mathcal{F}(\epsilon) \subseteq \mathcal{F}(\tilde{\epsilon}).$$

Which is the claim.

□

As discussed earlier, $\mathcal{F}(\epsilon) \neq \emptyset$ for all $\epsilon \ge 0$. Lemma 5.5.1 (ii) states that for all $\epsilon \ge 0$ there exists a subset of paths $P^\epsilon \subseteq \mathcal{P}$ (i.e. depending on ϵ) so that $\mathcal{F}_P(\epsilon)$ is nonempty. However, for an arbitrary $\epsilon \ge 0$ and arbitrary choice of $P \subseteq \mathcal{P}$ we cannot expect the set $\mathcal{F}_P(\epsilon)$ to be nonempty. Nonetheless, choosing $P_k = P_k^n$ for all $k \in \mathcal{K}$ as defined in (5.5.1), $\mathcal{F}_{P^n}(\epsilon)$ is nonempty for all $\epsilon \ge 0$. More precisely, given $P \subseteq \mathcal{P}$, $\epsilon \in \operatorname{dom}(\mathcal{F}_P) \Rightarrow \tilde{\epsilon} \in \operatorname{dom}(\mathcal{F}_P)$ for all $\tilde{\epsilon} \ge \epsilon$.

The discussed relations between the feasible sets of the defined optimization problems $(Q(\epsilon))$ and $(Q_P(\epsilon))$ relate the optimal value functions of these problems. The optimal value function corresponds to the total travel time of the best-performing BRUE flow distribution. The optimal value functions for the mentioned problems are thus defined as follows:

$$v(\epsilon) := \min_{x \in \mathcal{F}(\epsilon)} s(x), \qquad \text{and} \qquad v_P(\epsilon) := \min_{x \in \mathcal{F}_P(\epsilon)} s(x), \tag{5.5.2}$$

for $P \subseteq \mathcal{P}$. Recall from Lemma 5.5.1 that for given $\epsilon \geq 0$:

$$\bigcup_{P \subseteq \mathcal{P}} \mathcal{F}_P(\epsilon) = \mathcal{F}(\epsilon). \tag{5.5.3}$$

Consequently, for a given $\epsilon \geq 0$ and $P \subseteq \mathcal{P}$ so that $\mathcal{F}_P(\epsilon) \neq \emptyset$ the following relation holds:

$$v(\epsilon) \leq v_P(\epsilon).$$

From Lemma 5.5.1 (ii), for all $\epsilon \geq 0$, there exists a subset of paths $P^\epsilon \subseteq \mathcal{P}$ so that

$$v(\epsilon) = v_{P^\epsilon}(\epsilon)$$

holds.

Given $P \subseteq \mathcal{P}$, we can check the nonlinear programming optimality conditions corresponding to problem $(Q_P(\epsilon))$. Under affine linear latency assumptions and Assumption 1, $(Q_P(\epsilon))$ becomes a convex mathematical program for each $P \subseteq \mathcal{P}$. So, the solutions of these problems are unique with respect to link flow x. Nonetheless, without any additional information on the choice of the set $P \subseteq \mathcal{P}$ the decomposition by branches is not able to overcome the combinatorial cure of the complementarity constraint problem (i.e. $2^{|\mathcal{P}|}$ choices for P exist).

5.6 Synthesis

The Best-case BRUE problem, introduced in this chapter, is a mathematical program with complementarity constraints. The feasible set of this program is not convex and a regularity condition does not necessarily hold. However, we showed that we can decompose the problem into branches of easier problems. We indicated that the user equilibrium is a special case of the BRUE, since the BRUE reduces to the user equilibrium when the indifference band is 0. Moreover, the Best-case BRUE problem reduces to the problem to find the system-optimal flow distribution if the indifference band is sufficiently large. Since the user equilibrium and system-optimal traffic assignment are easy to solve problems, we use a parametric optimization approach: we study the behavior of the optimization problem under perturbations in the indifference band. In the next chapter we study the behavior of a branch problem under perturbations in the tolerance vector.

Chapter 6

Behavior of $(Q_P(\epsilon))$

This chapter discusses the behavior of the feasible set $\mathcal{F}_P(\epsilon)$, optimal value function $v_P(\epsilon)$ and optimal solution set $\mathcal{S}_P(\epsilon)$ corresponding to problem $(Q_P(\epsilon))$ with respect to perturbations in parameter ϵ for a given $P \subseteq \mathcal{P}$. Let us introduce an equivalent formulation of $(Q_P(\epsilon))$:

$$\min_x s(x) \qquad \text{s.t.} \qquad x \in \mathcal{F}_P^x(\epsilon).$$

Where, for each $\epsilon \geq 0$, $\mathcal{F}_P^x(\epsilon)$ is the projection of $\mathcal{F}_P(\epsilon)$ onto the x-space:

$$\mathcal{F}_P^x(\epsilon) = \left\{ x \in \mathbb{R}^{|E|} \mid \exists (f, \lambda) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|\mathcal{K}|} \text{ s.t. } (f, x, \lambda) \in \mathcal{F}_0(P) \cap \mathcal{F}_l \cap \mathcal{F}_u(\epsilon, P) \right\}.$$

We introduce the necessary definitions of the mentioned functions for given $P \subseteq \mathcal{P}$. The optimal value function is defined as

$$v_P(\epsilon) := \min_{x \in \mathcal{F}_P^x(\epsilon)} s(x),$$

and the projection of the optimal solution set onto the x-space is defined by

$$\mathcal{S}_P^x(\epsilon) := \arg\min_{x \in \mathcal{F}_P^x(\epsilon)} s(x). \tag{6.0.1}$$

In our study, a (multi-)function is said to be continuous if a small change in the parameter (input) leads to a small change in the output. For instance, the analysis evaluates whether link flows change smoothly as the parameter ϵ changes in problem $(Q_P(\epsilon))$. Such a function is also called *stable* in other studies (Lu and Nie, 2010; Smith, 1979). Throughout this section, we will assume that the link latency functions in the network are affine linear. Under the following Assumption 2, the set $\mathcal{F}_P^x(\epsilon)$ is convex for given $P \subseteq \mathcal{P}$ and $\epsilon \geq 0$.

Assumption 2 (Affine linear latencies). *Let us assume that the link cost functions are affine linear and strictly monotone, i.e.:* $l(x) = Ax + b$ *where* $A = diag(a_e, e \in E)$,

$a_e > 0$, for all $e \in E$. The route cost vector is then

$$c(f) = \Gamma^T l(\Gamma f) = \Gamma^T A \Gamma f + \Gamma^T b = \tilde{A} f + \tilde{b},$$

where $\mathbb{R}_+^{|\mathcal{P}| \times |\mathcal{P}|} \ni \tilde{A} := \Gamma^T A \Gamma$ and $\mathbb{R}_+^{|\mathcal{P}|} \ni \tilde{b} := \Gamma^T b$.

Remark 6.0.1. *Since A is positive definite, $\tilde{A}$ is positive semidefinite. If Γ has full column rank, then $\tilde{A}$ is positive definite.*

Remark 6.0.2. *Under Assumption 2, $\mathcal{F}_P^x(\epsilon)$ is a closed polyhedron for all $\epsilon \in dom(\mathcal{F}_P)$.*

6.1 Related Work

Continuity analysis of parametric problems in the (fixed demand) static traffic assignment has mostly been focusing on local sensitivity. That is: analyzing solutions in a neighborhood around a parameter to account for changes in the parameter. This type of analysis is particularly useful in anticipating on the effects on link flows by small changes in the demand d and cost function (Patriksson, 2004).

Most studies in literature on this topic (e.g. Yang (1997)) are based on the seminal study of Tobin and Friesz (1988) and follow-up study of Cho et al. (2000). The cited studies used the Implicit Function Theorem to analyze local sensitivity. Let us introduce a simple parametric program to discuss the limitations of this approach.

For given ϵ, find local minimizer $x = x(\epsilon)$ of

$$\min_{x \in \mathbb{R}^n} s(x) \qquad \text{s.t.} \qquad x \in \mathcal{F}(\epsilon) := \{x \mid g_j(x, \epsilon) \leq 0, j \in J\}. \tag{$P(\epsilon)$}$$

Here, we assume that $s(x)$ and $g_j(x, \epsilon)$ are twice continuously differentiable. We are interested in finding local minimizers of $(P(\epsilon))$ where ϵ is close to $\bar{\epsilon}$. We look at the following system with Lagrange multiplier vector $\mu \geq 0$:

$$F(x, \epsilon, \mu) := \begin{bmatrix} \nabla_x s(x) + \sum_{j \in J_0(\bar{x}, \bar{\epsilon})} \mu_j \nabla_x g_j(x, \epsilon) \\ g_{j \in J_0(\bar{x}, \bar{\epsilon})}(x, \epsilon) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$J_0(\bar{x}, \bar{\epsilon})$ is the active index set of $g_j(x, \epsilon), j \in J$, for given $\bar{\epsilon}$ and feasible $\bar{x} \in \mathcal{F}(\bar{\epsilon})$. We turned our system $(P(\epsilon))$ into a system of equalities $F(x, \epsilon, \mu) = 0$. The Implicit Function Theorem states that if the KKT condition corresponding to $(P(\epsilon))$ is satisfied with multiplier vector $\bar{\mu}$, such that (i) the linear independence constraint qualification (LICQ) holds at $\bar{x}$ with respect to $\mathcal{F}(\bar{\epsilon})$, (ii) strict complementarity holds (i.e. $\bar{\mu}_j > 0$, for all $j \in J_0(\bar{x}, \bar{\epsilon})$), and (iii) $\bar{x}$ is a first or second order minimizer, then there exists a neighborhood $U_\delta(\bar{\epsilon})$ of $\bar{\epsilon}$ and continuously differentiable function $x(\epsilon)$, $\mu(\epsilon)$ defined on this neighborhood then $x(\epsilon)$ is a first or second order minimizer of $(P(\epsilon))$ for any $\epsilon \in U_\delta(\bar{\epsilon})$. The derivatives of $x(\epsilon)$,

$\mu(\epsilon)$ are calculated by

$$\begin{bmatrix} \nabla_\epsilon x(\epsilon) \\ \nabla_\epsilon \mu(\epsilon) \end{bmatrix} = - \left[\nabla_{(x,\mu)} F(x(\epsilon), \epsilon, \mu(\epsilon)) \right]^{-1} \nabla_\epsilon F(x(\epsilon), \epsilon, \mu(\epsilon)).$$

In the context of the static traffic assignment, the conditions of the Implicit Function Theorem are as follows. First, the problem should be reduced to a problem where only routes with positive link flows are included (Chung et al., 2014; Patriksson, 2004), assuming strict complementarity. Second, since $\nabla_{(x,\mu)} F(x(\epsilon), \epsilon, \mu(\epsilon))$ should be invertible, a path flow solution $f(\bar{\epsilon})$ corresponding to $x(\bar{\epsilon})$ should be chosen so that this path flow is non-degenerate (Chung et al., 2014).

Application of the Implicit Function Theorem is particularly useful in the numerical evaluation of small perturbations. However, the results only hold for the neighborhood $U_\delta(\bar{\epsilon})$ of $\bar{\epsilon}$. Although the assumptions of the Implicit Function Theorem were relaxed in e.g. Patriksson (2004), the provided results only hold locally. Moreover, since LICQ does not hold for problem $(Q(\epsilon))$ we need to study continuity on a *global* and more abstract level so that we can generalize the continuity results of $(Q_P(\epsilon))$ to problem $(Q(\epsilon))$.

Global continuity was studied in the fixed demand static traffic assignment setting by Lu and Nie (2010). We take a similar approach in this chapter. Although the setting in this chapter is different (we study perturbations with respect to indifference vector ϵ, Lu and Nie (2010) considered perturbations in the demand and cost vector), some results can easily be translated into the setting of Lu and Nie (2010). Continuity for problem $(Q(\epsilon))$ was partially studied by Di et al. (2016).

6.2 Topology

We distinguish upper and lower semicontinuity. An upper semicontinuous multifunction assures that a set does not explode after a small perturbation in the parameter. Similarly, lower semicontinuity assures that the set does not implode after perturbing ϵ.

Definition 6.2.1 (Upper semicontinuity). *A multifunction F is said to be upper semicontinuous at a point $\bar{\epsilon} \in dom(F)$, if for each $\tau > 0$ there exists a $\delta > 0$ so that the following condition holds:*

$$F(\epsilon) \subseteq U_\tau(F(\bar{\epsilon})), \qquad \textit{for all } \epsilon \in U_\delta(\bar{\epsilon}). \tag{6.2.1}$$

Definition 6.2.2 (Lower semicontinuity). *A multifunction F is said to be lower semicontinuous at $\bar{\epsilon} \in dom(F)$, if for each $\tau > 0$ there exists a $\delta > 0$ such that*

$$F(\bar{\epsilon}) \subseteq U_\tau(F(\epsilon)), \qquad \textit{for all } \epsilon \in U_\delta(\bar{\epsilon}) \tag{6.2.2}$$

holds.

Remark 6.2.1. *The definitions of upper and lower semicontinuity in (6.2.1) and (6.2.2) are due to Hausdorff. Rather than using upper (lower) semicontinuity, in the upcoming proofs we frequently use the definition of outer (inner) semicontinuity (see Still (2006)). Since our setting considers compact spaces, the definitions are equivalent.*

In case a multifunction is both lower and upper semicontinuous at $\bar{\epsilon}$ it is said to be *continuous* at $\bar{\epsilon}$. A multifunction F, continuous for all $\epsilon \in \mathrm{dom}\,(F)$, is said to be continuous relative (to its domain). The upcoming definition of Lipschitz continuity is even stronger and implies continuity and requires closedness of the point-to-set mapping.

Definition 6.2.3 (Closed mapping). *A multifunction F is said to be closed at $\bar{\epsilon} \in dom\,(F)$, if for any sequences ϵ^n, x^n, $n \in \mathbb{N}$, with $\epsilon^n \to \bar{\epsilon}$, $x^n \in F(\epsilon^n)$, $x^n \to \bar{x}$ implies $\bar{x} \in F(\bar{\epsilon})$.*

Definition 6.2.4 (Lipschitz continuity). *The multifunction F is said to be Lipschitz continuous at $D \subseteq dom\,(F)$, if $F(\epsilon)$ is closed for every $\epsilon \in D$ and there exists $\kappa \geq 0$ such that*

$$F(\bar{\epsilon}) \subseteq U_{\kappa\|\bar{\epsilon}-\epsilon\|}\left(F(\epsilon)\right), \qquad \text{for all } \epsilon, \bar{\epsilon} \in D. \tag{6.2.3}$$

For completeness, we give the definitions of upper and lower semicontinuity of a single-valued function.

Definition 6.2.5 (Upper semicontinuous single-valued function). *A single-valued function v is said to be upper semicontinuous at $\bar{\epsilon} \in dom\,(v)$, if for each $\tau > 0$, exists $\delta > 0$ so that*

$$v(\epsilon) \leq v(\bar{\epsilon}) + \tau, \qquad \text{for all } \|\epsilon - \bar{\epsilon}\| < \delta.$$

Definition 6.2.6 (Lower semicontinuous single-valued function). *A single-valued function v is said to be lower semicontinuous at $\bar{\epsilon} \in dom\,(v)$, if for each $\tau > 0$, exists $\delta > 0$ so that*

$$v(\bar{\epsilon}) \leq v(\epsilon) + \tau, \qquad \text{for all } \|\epsilon - \bar{\epsilon}\| < \delta.$$

Under Assumption 2 and given $\epsilon \geq 0$, $\mathcal{F}^x_P(\epsilon)$, $P \subseteq \mathcal{P}$, is a closed polyhedron (Remark 6.0.2). Then, $\mathcal{F}(\epsilon)$ is the union of a finite number of polyhedra but not necessarily a polyhedron itself. In order to construct $\mathcal{F}^x(\epsilon)$ for given $\epsilon \geq 0$ out of a set of $\mathcal{F}^x_P(\epsilon)$, $P \subseteq \mathcal{P}$, it is required to know for which $\epsilon \geq 0$, $\mathcal{F}^x_P(\epsilon)$ is nonempty. In other words, we are interested in $\mathrm{dom}\,(\mathcal{F}^x_P)$. The next lemma shows that $\mathrm{dom}\,(\mathcal{F}^x_P)$ is closed for all $P \subseteq \mathcal{P}$.

Lemma 6.2.1 ($\mathbf{dom}\,(\mathcal{F}^x_P)$ closed). *$dom\,(\mathcal{F}^x_P)$ is a (closed) polyhedron for all $P \subseteq \mathcal{P}$.*

Proof. For any (fixed) $P \subseteq \mathcal{P}$ and $\epsilon \geq 0$ the feasible set $\mathcal{F}^x_P(\epsilon)$ is defined by linear inequalities and equations (see Remark 6.0.2). There exist vector $c \in \mathbb{R}^m$ and matrices

$\mathcal{A} \in \mathbb{R}^{m\times|\mathcal{K}|}, \mathcal{B} \in \mathbb{R}^{m\times|\mathcal{K}|}$ so that for all ϵ the feasible set is described by

$$\mathcal{F}_P^x(\epsilon) = \left\{x \in \mathbb{R}^{|E|} \mid \mathcal{A}x + \mathcal{B}\epsilon \leq c\right\}. \tag{6.2.4}$$

Consider H, the space where $x = 0$, i.e.:

$$H = \left\{(x, \epsilon) \in \mathbb{R}^{|E|} \times \mathbb{R}^{|\mathcal{K}|} \mid x = 0\right\}.$$

We project $\mathcal{F}_P^x$ onto H. $\mathrm{proj}_H(\mathcal{F}_P)$ is then the domain of $\mathcal{F}_P^x$:

$$\mathrm{proj}_H(\mathcal{F}_P^x) = \{(0, \epsilon) \mid \text{exists } x \text{ s.t. } x \in \mathcal{F}_P^x(\epsilon)\}.$$

By a well-known result (e.g. Theorem 2.5 in Faigle et al. (2013)) $\mathrm{proj}_H(\mathcal{F}_P^x)$ is also the solution set of the system $\tilde{\mathcal{B}}\epsilon \leq \tilde{c}$, $\tilde{\mathcal{B}} \in \mathbb{R}^{m\times|\mathcal{K}|}$, $\tilde{c} \in \mathbb{R}^m$, which is a closed polyhedron. □

6.3 Behavior of $\mathcal{F}_P(\epsilon)$

Lemma 6.3.1 shows that the feasible set $\mathcal{F}_P^x(\epsilon)$, $P \subseteq \mathcal{P}$ is compact. We emphasize that $\mathcal{F}_0(P)$ is bounded since $0 \leq x_e \leq \sum_{k\in\mathcal{K}} d_k$ for all $e \in E$ and $0 \leq f_p \leq d_k$ for all $p \in \mathcal{P}_k, k \in \mathcal{K}$.

Lemma 6.3.1 ($\mathcal{F}_P^x(\epsilon)$ compact). *Given $P \subseteq \mathcal{P}$, the set $\mathcal{F}_P^x(\epsilon)$ is a closed mapping on dom$(\mathcal{F}_P^x)$ and $\mathcal{F}_P^x(\epsilon)$ is a compact set for any $\epsilon \in$ dom$(\mathcal{F}_P^x)$.*

Proof. Let $P \subseteq \mathcal{P}$ arbitrary and $\epsilon^0 \in \mathrm{dom}(\mathcal{F}_P^x)$. Since $c_p(x) - \lambda_k - \epsilon_k$ is a continuous function with respect to (x, λ, ϵ), we get that for a sequence $\epsilon^l \to \epsilon^0$ and $(x^l, \lambda^l) \in \mathcal{F}_P^x(\epsilon^l)$, $(x^l, \lambda^l) \to (x^0, \lambda^0)$:

$$c_p(x^0) - \lambda_k^0 - \epsilon_k^0 = \lim_{l\to\infty} (c_p(x^l) - \lambda_k^l - \epsilon_k^l) \leq 0,$$

holds for all $p \in \mathcal{P}_k, k \in \mathcal{K}$. A similar relation holds for the other defining (in)equalities. Hence, $x^0 \in \mathcal{F}_P^x(\epsilon^0)$ and $\mathcal{F}_P^x(\epsilon^0)$ is a closed mapping at ϵ^0. Since ϵ^0 is arbitrarily chosen, the lemma holds.

Furthermore, since $\mathcal{F}_u(P, \epsilon)$ and $\mathcal{F}_l$ are closed and $\mathcal{F}_0(P)$ is compact, the intersection of these sets, defining $\mathcal{F}_P(\epsilon)$, is compact. It follows that $\mathcal{F}_P^x(\epsilon)$ is compact. □

We are ready to state the first continuity result of feasible set $\mathcal{F}_P^x(\epsilon)$, $P \subseteq \mathcal{P}$. Therefore we utilize the next proposition, stated and proven in Mangasarian and Shiau (1987).

Proposition 6.3.1. *(Mangasarian and Shiau, 1987). Let the multifunction $F(\epsilon)$ be defined by $F(\epsilon) := \{x \mid Ax \leq \epsilon\}$. Then there exists $K > 0$ such that for any $\epsilon^0, \epsilon^1 \in$ dom(F)*

it holds that for each $x^0 \in F(\epsilon^0)$ there is an $x^1 \in F(\epsilon^1)$ whereby

$$\|x^1 - x^0\| \le K\|\epsilon^1 - \epsilon^0\|$$

holds.

Under Assumption 2, the set $\mathcal{F}_P^x(\epsilon)$ is a polyhedron for all $\epsilon \in \operatorname{dom}(\mathcal{F}_P)$ (Lemma 6.3.1). Combining this fact with Proposition 6.3.1 gives the next corollary in which we show that the feasible set $\mathcal{F}_P^x(\epsilon)$ is Lipschitz continuous with respect to perturbations in ϵ. Intuitively, the feasible set does not ex- or implode after a perturbation in ϵ.

Corollary 6.3.1 ($\mathcal{F}_P^x(\epsilon)$ Lipschitz continuous). *For any $P \subseteq \mathcal{P}$ exists $\mathcal{M} > 0$ so that for all $\epsilon^0, \epsilon^1 \in dom(\mathcal{F}_P)$ and for all $x(\epsilon^0) \in \mathcal{F}_P^x(\epsilon^0)$ there exists an $x(\epsilon^1) \in \mathcal{F}_P^x(\epsilon^1)$ with the property:*

$$\|x(\epsilon^1) - x(\epsilon^0)\| \le \mathcal{M}\|\epsilon^1 - \epsilon^0\|.$$

Hence, $\mathcal{F}_P^x(\epsilon)$ is (Lipschitz) continuous relative to $dom(\mathcal{F}_P)$.

Proof. By Proposition 6.3.1 we find that under the assumptions of the corollary the following holds (see (6.2.4)):

$$\begin{aligned}\|x(\epsilon^1) - x(\epsilon^0)\| &\le K\|\mathcal{B}\epsilon^1 - B\epsilon^0\|\\ &\le K\|\mathcal{B}\|\|\epsilon^1 - \epsilon^0\|\\ &= \mathcal{M}\|\epsilon^1 - \epsilon^0\|.\end{aligned}$$

The second inequality follows from the matrix norm property $\|\mathcal{B}y\| \le \|\mathcal{B}\|\|y\|$ for any vector y. □

Remark 6.3.1. *The previous results and the results in the remainder of this chapter depend on the continuity and convexity of inequality constraints $c_p(x) - \lambda_k - \epsilon_k \le 0$ and $-c_p(x) + \lambda_k \le 0$, $p \in P_k, k \in \mathcal{K}$ with respect to x, λ and ϵ. We lose convexity of the feasible set if we apply a multiplicative indifference band rather than the additive one (Remark 4.1.1). Moreover, the results of this chapter do not necessarily hold for higher polynomial degree latency functions. Then, $\mathcal{F}_P(\epsilon)$ is not defined by linear (in)equalities.*

6.4 Behavior of $v_P(\epsilon)$

We discuss continuity of the optimal value function $v_P(\epsilon)$, $P \subseteq \mathcal{P}$, as defined in (5.5.2) with respect to perturbations in the indifference vector ϵ. Recall that we assume the link cost functions to be affine linear (Assumption 2).

The next proposition is the result of the non-decreasing feasible set for increasing ϵ, by Lemma 5.5.1 (v).

Proposition 6.4.1 (**$v_P(\epsilon)$ monotonically non-increasing**). *Suppose $\epsilon^0, \epsilon^1 \in dom(\mathcal{F}_P)$, then $v_P(\epsilon^0) \geq v_P(\epsilon^1)$, $P \subseteq \mathcal{P}$.*

The statement of Proposition 6.4.1 is straightforward: an increase in tolerance vector ϵ widens the range of acceptable travel times: $\mathcal{F}_P(\epsilon^0) \subseteq \mathcal{F}_P(\epsilon^1)$, provided $\epsilon^1 \geq \epsilon^0$. Improvement of the performance of the system requires cooperative behavior among individuals, i.e. some people have to suffer in favor of others (Jahn et al., 2005). An increase in the indifference band gives users the opportunity to act in favor of the system.

The following inequality follows directly from Proposition 6.4.1:

$$v(\epsilon) \leq v(0), \qquad \text{for all } \epsilon \geq 0.$$

Hence, the PRUE provides an upper bound on the performance with respect to total travel time of the Best-case BRUE. In other words, the Best-case BRUE performs as least as good as the PRUE (we made this observation in Example 4.2.1).

It turns out that the optimal value function $v_P(\epsilon)$ is convex with respect to ϵ for given $P \subseteq \mathcal{P}$. This result is formalized in Lemma 6.4.1.

Lemma 6.4.1 (**$v_P(\epsilon)$ convex**). *The optimal value function $v_P(\epsilon)$, $P \subseteq \mathcal{P}$, is convex relative to $dom(\mathcal{F}_P)$. That is, for $\epsilon^\beta = \beta\epsilon^0 + (1-\beta)\epsilon^1$,*

$$v_P(\epsilon^\beta) \leq \beta v_P(\epsilon^0) + (1-\beta) v_P(\epsilon^1),$$

where $\epsilon^0 \neq \epsilon^1 \in dom(\mathcal{F}_P)$ and $\beta \in [0,1]$.

Proof. Let the assumptions of the Lemma hold. Assume $x^0 \in \mathcal{F}^x_P(\epsilon^0)$ and $x^1 \in \mathcal{F}^x_P(\epsilon^1)$. x^0 (x^1 respectively) is the optimal solution corresponding to $v_P(\epsilon^0)$ ($v_P(\epsilon^1)$ resp.). Define $x^\beta := \beta x^0 + (1-\beta)x^1$ (and $\lambda^\beta, \epsilon^\beta$ similarly). By Assumption 2, we claim that $x^\beta \in \mathcal{F}^x_P(\epsilon^\beta)$. Indeed, using $l(x) = Ax + b$,

$$\begin{aligned}
\Gamma^T(Ax^\beta + b) - \Lambda^T\lambda^\beta - \Lambda^T\epsilon^\beta &= \Gamma^T(Ax(\beta x^0 + (1-\beta)x^1) + b) \\
&\quad - \Lambda^T(\beta\lambda^0 + (1-\beta)\lambda^1) - \Lambda^T(\beta\epsilon^0 + (1-\beta)\epsilon^1) \\
&= \beta\Gamma^T(Ax^0 + b) + (1-\beta)\Gamma^T(Ax^1 + b) \\
&\quad - \beta\Lambda^T\lambda^0 - (1-\beta)\Lambda^T\lambda^1 - \beta\Lambda\epsilon^0 - (1-\beta)\Lambda\epsilon^1 \\
&= \beta\left(\Gamma^T(Ax^0 + b) - \Lambda^T\lambda^0 - \Lambda^T\epsilon^0\right) \\
&\quad + (1-\beta)\left(\Gamma^T(Ax^1 + b) - \Lambda^T\lambda^1 - \Lambda^T\epsilon^1\right).
\end{aligned}$$

Obviously, a similar analysis holds for the other (in)equalities that define $\mathcal{F}^x_P(\epsilon)$. The claim of the lemma follows by convexity of the objective function $s(x)$ with respect to link flows x:

$$\begin{aligned} v_P(\epsilon^\beta) \leq s(x^\beta) &= s(\beta x^0 + (1-\beta)x^1) \\ &\leq \beta s(x^0) + (1-\beta)s(x^1) \\ &= \beta v_P(\epsilon^0) + (1-\beta)v_P(\epsilon^1). \end{aligned}$$

We have proven that the optimal value function is convex on dom $(\mathcal{F}_P)$. □

Remark 6.4.1. *The result of Lemma 6.4.1 does not hold when the indifference band is assumed to be multiplicative (see (4.1.2)). In particular, the feasible set is then not convex with respect to (x, ϵ).*

Theorem 6.4.1 ($v_P(\epsilon)$ continuous). *$v_P(\epsilon)$, $P \subseteq \mathcal{P}$, is Lipschitz continuous with respect to dom$(\mathcal{F}_P)$. That is, there exists $D > 0$ so that for all $\epsilon^0, \epsilon^1 \in$ dom$(\mathcal{F}_P)$:*

$$\|v_P(\epsilon^0) - v_P(\epsilon^1)\| \leq D\|\epsilon^0 - \epsilon^1\|.$$

Proof. Take any $\epsilon^0, \epsilon^1 \in$ dom $(\mathcal{F}_P)$. Let $x^0 \in \mathcal{F}_P^x(\epsilon^0)$, $x^1 \in \mathcal{F}_P^x(\epsilon^1)$ be the optimal solutions corresponding to these indifference bands respectively, in other words: $s(x^0) = v_P(\epsilon^0)$ and $s(x^1) = v_P(\epsilon^1)$. By Corollary 6.3.1, exists $\mathcal{M} > 0$ and $\tilde{x}^0 \in \mathcal{F}_P^x(\epsilon^0)$ and $\tilde{x}^1 \in \mathcal{F}_P^x(\epsilon^1)$ so that:

$$\|\tilde{x}^0 - x^1\| \leq \mathcal{M}\|\epsilon^0 - \epsilon^1\|, \qquad \text{and} \qquad \|\tilde{x}^1 - x^0\| \leq \mathcal{M}\|\epsilon^0 - \epsilon^1\|.$$

We find

$$v_P(\epsilon^0) - v_P(\epsilon^1) \leq s(\tilde{x}^0) - s(x^1) \leq K\|\tilde{x}^0 - x^1\| \leq K\mathcal{M}\|\epsilon^0 - \epsilon^1\|$$

by Lipschitz continuity of $s(x)$ relative to compact set $\mathcal{F}_0$ (since $s(x)$ is continuously differentiable, see Rockafellar and Wets (2009)). Similarly,

$$v_P(\epsilon^1) - v_P(\epsilon^0) \leq s(\tilde{x}^1) - s(x^0) \leq K\|\tilde{x}^0 - x^1\| \leq K\mathcal{M}\|\epsilon^0 - \epsilon^1\|.$$

The statement of the theorem follows by defining $D = K\mathcal{M}$. □

We observe that the optimal value function $v(\epsilon)$ for $\epsilon = 0$ is equal to the total travel time in the PRUE. Hence, $v(0) = v_{P^n}(0)$ where P^n consists of all the paths with minimum cost in the PRUE. As ϵ gradually increases, $v_{P^n}(\epsilon)$ changes (Lipschitz) continuously. Moreover, $v_{P^n}(\epsilon)$ is close to $v(0)$ when ϵ is close to 0. In the next section, we strengthen the result of Theorem 6.4.1 and show that $v_P(\epsilon)$ is *piecewise quadratic* relative to dom $(\mathcal{F}_P)$.

We conclude this section with a simple illustration. We indicated that for a given subset of paths P, $v_P(\epsilon)$ is globally Lipschitz continuous, decreasing, convex and piecewise quadratic. Figure 6.4.1 illustrates this behavior for a one dimensional ϵ.

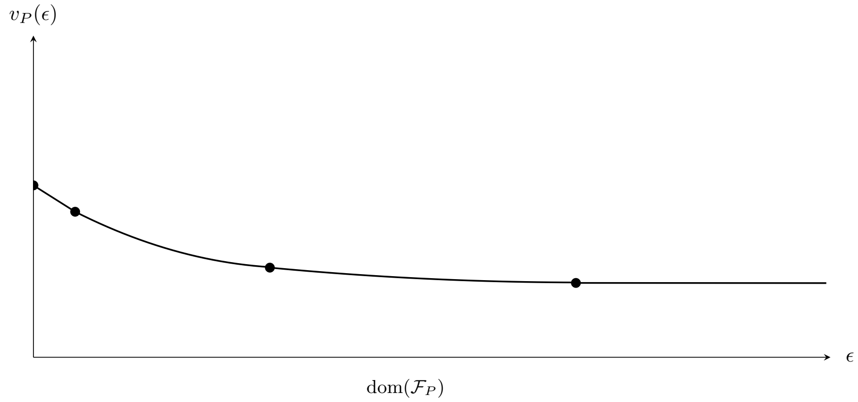


Figure 6.4.1: Illustration of piecewise quadratic, monotonically non-increasing behavior of value function $v_P(\epsilon)$, $P \subseteq \mathcal{P}$, relative to $\mathrm{dom}\,(\mathcal{F}_P)$. *The dots indicate the borders between pieces in which $v_P(\epsilon)$ behaves quadratic.*

6.5 Behavior of $\mathcal{S}_P(\epsilon)$

Assumption 2 assures that the optimal link flow solution is unique for $(Q_P(\epsilon))$, $P \subseteq \mathcal{P}$, for given $\epsilon \in \mathrm{dom}\,(\mathcal{F}_P)$. Indeed, $(Q_P(\epsilon))$ is a convex optimization problem with a strictly convex objective function. Thus, the optimal solution set for this problem with respect to link flows is a singleton, i.e. $\mathcal{S}_P^x(\epsilon) = \{x_P(\epsilon)\}$. Note that corresponding route flows f to x, and corresponding (lower bound) minimum path cost vector λ do not influence the value function $s(x)$. We can prove locally $\frac{1}{2}$-Hölder continuity of $x_P(\epsilon)$ relative to its domain by utilizing the second-order sufficient optimality conditions over a polyhedron (see Still and Streng (1996)). However, since the objective function is quadratic, we prove Lipschitz-continuity of the optimal solution $x_P(\epsilon)$. We start by a theorem that proves continuity of the optimal link flow solution.

Theorem 6.5.1 ($x_P(\epsilon)$ continuous). *$x_P(\epsilon)$ is a continuous function relative to dom $(\mathcal{F}_P)$.*

Proof. First we show that $\mathcal{S}_P^x(\epsilon)$ is a closed mapping at each $\epsilon \in \mathrm{dom}\,(\mathcal{F}_P)$, using the approach of Bank et al. (1983). We equivalently define $\mathcal{S}_P^x(\epsilon)$ as the intersection of two sets:

$$\mathcal{S}_P^x(\epsilon) = \mathcal{F}_P^x(\epsilon) \cap \{x \mid s(x) \leq v_P(\epsilon)\}.$$

Since the intersection of two closed mappings is closed, we solely have to prove that the mapping defined by $\{x \mid s(x) \leq v_P(\epsilon)\}$ is a closed mapping.

Take $\epsilon^0 \in \mathrm{dom}\,(\mathcal{F}_P)$ arbitrary. Let the sequences $\epsilon^n \to \epsilon^0$ be given and take x^l s.t. $s(x^l) \leq v_P(\epsilon^n)$ for all $l \in \mathbb{N}$. Then, $x^l \to x^0$ implies that $s(x^0) \leq v_P(\epsilon^0)$. Indeed,

$$s(x^l) \leq v_P(\epsilon^l) \qquad \text{implies} \qquad s(x^0) \leq v_P(\epsilon^0),$$

by continuity of $s(x)$ and $v_P(\epsilon)$.

We show continuity of $x_P(\epsilon)$. Take $\epsilon^0 \in \mathrm{dom}\,(\mathcal{F}_P)$, $x^0 \in \mathcal{S}_P^x(\epsilon^0)$, $\epsilon^l \to \epsilon^0$ and $x^l \in \mathcal{S}_P^x(\epsilon^n)$. We have to show that $x^l \to x^0$. Assume to the contrary that there exists a subsequence $x^{l_k} \not\to x^0$. By compactness there exists a subsequence $x^{l_{k_j}}$ of x^{l_k} which converges, i.e. $x^{l_{k_j}} \to \bar{x} \in \mathcal{S}_P^x(\epsilon^0)$ by closedness of $\mathcal{S}_P^x(\epsilon)$ at any $\epsilon \in \mathrm{dom}\,(\mathcal{F}_P)$. Since $\mathcal{S}_P^x(\epsilon^0)$ is a singleton, we get $\bar{x} = x^0$, leading to a contradiction. □

In the previous section we proved that $v_{P^n}(\epsilon)$ is close to $v(0)$ as long as ϵ is close to 0. The result of Theorem 6.5.1 shows that the corresponding link flow is close as well.

To prove that $x_P(\epsilon)$ is Lipschitz continuous we consider quadratic (convex) minimization problem $(\tilde{Q}_P(\epsilon))$, given $P \subseteq \mathcal{P}$, as equivalent of $(Q_P(\epsilon))$. We project $\mathcal{F}_P(\epsilon)$ onto $\mathbb{R}^{|E|}$ and obtain $\mathcal{F}_P^x(\epsilon)$, defined by linear (in)equalities. For the purpose of our argument we define $\mathcal{F}_P^x(\epsilon) := \mathcal{A}x \leq \epsilon$.

$$\min_x x^T A x + b^T x \qquad \text{s.t.} \qquad \mathcal{A}x \leq \epsilon. \tag{$\tilde{Q}_P(\epsilon)$}$$

Here, $\mathcal{A} \in \mathbb{R}^{m \times k}$. A is a diagonal matrix and is positive definite (Assumption 2). So, for fixed $P \subseteq \mathcal{P}$ and $\epsilon \in \mathrm{dom}\,(\mathcal{F}_P)$ we have a unique solution $\bar{x} = x_P(\epsilon)$ of quadratic problem $(Q_P(\epsilon))$.

Given $P \subseteq \mathcal{P}$, for each $\epsilon \in \mathrm{dom}\,(\mathcal{F}_P)$ the associated Lagrange dual problem of $(\tilde{Q}_P(\epsilon))$ is given by

$$\begin{aligned} 2Ax + b + \mathcal{A}^T\mu &= 0, \\ \mathcal{A}x &\leq \epsilon, \\ (\mathcal{A}x - \epsilon)^T\mu &= 0, \\ \mu &\geq 0, \end{aligned} \tag{6.5.1}$$

with corresponding multiplier $\mu \in \mathbb{R}^m$. By the mentioned assumptions, x solves $(\tilde{Q}_P(\epsilon))$ if and only if (x, μ) solves the system in (6.5.1). We utilize the Lagrange dual system of (6.5.1) to prove that the optimal solution $x_P(\epsilon)$ of problem $(Q_P(\epsilon))$ is a piecewise affine function with respect to its domain. We use a comparable approach as in Bank et al. (1983).

Lemma 6.5.1 ($x_P(\epsilon)$ piecewise affine). *$x_P(\epsilon)$ is a piecewise affine function with respect to dom$\,(\mathcal{F}_P)$.*

Proof. Let us define, given $\epsilon \geq 0$, the index set $\mathcal{I}_x(\epsilon) \subseteq \{1, \ldots, n\}$ with respect to a feasible x of problem $(\tilde{Q}_P(\epsilon))$ *($\mathcal{A}_{j,\cdot}$ is the jth row of matrix $\mathcal{A}$)*:

$$\mathcal{I}_x(\epsilon) = \{j \mid \mathcal{A}_{j,\cdot}x = \epsilon_j\}$$

By positive definiteness of matrix A, $\bar{x}$ is the unique solution of $(\tilde{Q}_P(\epsilon))$. So, we can uniquely define the characterizing index set $\mathcal{I} = \mathcal{I}_{\bar{x}}(\epsilon)$ corresponding to this optimal solution. Hence, $\bar{x}$ is a solution of the following system of equations:

$$\begin{aligned} 2Ax + \mathcal{A}^T\mu &= -b; \\ \mathcal{A}_{j,\cdot}x &= \epsilon_j, \qquad j \in \mathcal{I}; \\ \mu_j &= 0, \qquad j \notin \mathcal{I}. \end{aligned} \tag{6.5.2}$$

Define the associated matrix $M_{\mathcal{I}}$ to (6.5.2) by

$$M_{\mathcal{I}} = \begin{bmatrix} 2A & \mathcal{A}^T \\ A_{j,\cdot} & 0 \\ 0 & I_{m\setminus|\mathcal{I}|} \end{bmatrix},$$

where I_m denotes an identity matrix of size $m \times m$. $M_{\mathcal{I}}$ is not necessarily non-singular and thus we need to define $M_{\mathcal{I}}^+$ to be the *pseudoinverse* of $M_{\mathcal{I}}$ (Bank et al., 1983, p. 135). Then,

$$\begin{bmatrix} x^+ \\ \mu^+ \end{bmatrix} = M_{\mathcal{I}}^+ \begin{bmatrix} -b \\ \epsilon \\ 0 \end{bmatrix}$$

is the corresponding solution (x, μ) of system (6.5.2) with smallest Euclidean norm. Now, for any $\epsilon \in \operatorname{dom}(\mathcal{F}_P)$, solution $x_P(\epsilon)$ of $(\tilde{Q}_P(\epsilon))$ is given by $(\mathcal{I} = \mathcal{I}_{x_P(\epsilon)}(\epsilon))$

$$x_P(\epsilon) = -M_{\mathcal{I}}^{+,1}b + M_{\mathcal{I}}^{+,2}\epsilon, \tag{6.5.3}$$

obviously linear in tolerance vector ϵ. Here,

$$M_{\mathcal{I}}^+ = \begin{bmatrix} M_{\mathcal{I}}^{+,1} & M_{\mathcal{I}}^{+,2} & M_{\mathcal{I}}^{+,3} \\ M_{\mathcal{I}}^{+,4} & M_{\mathcal{I}}^{+,5} & M_{\mathcal{I}}^{+,6} \end{bmatrix}.$$

Since $\mathcal{I} \subseteq \{1, \ldots, n\}$ only a finite number of active index sets $\mathcal{I}$ can occur. By continuity of optimal solution $x_P(\epsilon)$, we have that $x_P(\epsilon)$ is a piecewise affine function with respect to ϵ. □

We have proven that the optimal solution $x_P(\epsilon)$ behaves continuously relative to $\operatorname{dom}(\mathcal{F}_P)$ and can be partitioned so that on each partition it is linear in tolerance vector ϵ. The continuous and piecewise linear behavior of this function makes that $x_P(\epsilon)$ is globally Lipschitz continuous.

Theorem 6.5.2 ($x_P(\epsilon)$ Lipschitz continuous). *Given $P \subseteq \mathcal{P}$, there exists $W > 0$ so* *that for all* $\epsilon^0, \epsilon^1 \in$ *dom* $(\mathcal{F}_P)$,

$$\|x_P(\epsilon^0) - x_P(\epsilon^1)\| \leq W\|\epsilon^0 - \epsilon^1\|$$

holds. Here, $x_P(\epsilon) = \arg\min_{x \in \mathcal{F}^x_P(\epsilon)} v_P(\epsilon)$.

Proof. We claim that for a given index set $\mathcal{I} \subseteq \{1, \ldots, m\}$ the solution set of KKT system (6.5.2) is convex. We define the *stability set* $\Omega(\mathcal{I})$ as the domain corresponding to this index $\mathcal{I}$:

$$\Omega(\mathcal{I}) = \left\{\epsilon \in \operatorname{dom}(\mathcal{F}_P) \mid \mathcal{I}_{x_P(\epsilon)}(\epsilon) = \mathcal{I}\right\}.$$

Clearly, $\Omega(\mathcal{I})$ for all possible $\mathcal{I}$ gives a partitioning of dom $(\mathcal{F}_P)$.

Regard $\Omega(\mathcal{I})$ as the projection of the system

$$\begin{aligned} 2Ax + b + \mathcal{A}^T\mu &= 0; \\ \mathcal{A}x &\leq \epsilon; \\ \mathcal{A}_{j,\cdot}x &= \epsilon_j, \qquad j \in \mathcal{I}; \\ \mu_j &= 0, \qquad j \notin \mathcal{I}; \end{aligned}$$

onto the ϵ-space. Since the projection of a closed polyhedron is a closed polyhedron, $\Omega(\mathcal{I})$ is a closed polyhedron.

Segment $[\epsilon^0, \epsilon^1] = \{(1-\beta)\epsilon^0 + \beta\epsilon^1 | 0 \leq \beta \leq 1\}$ can be covered by finitely many $\Omega(\mathcal{I})$. In other words: there exist finitely many points $0 = \beta_0 < \beta_1 < \cdots < \beta_n = 1$ so that

$$[0, 1] = [\beta_0, \beta_1] \cup [\beta_1, \beta_2] \cup \cdots \cup [\beta_{n-1}, \beta_n].$$

Define $\epsilon^\beta = ((1-\beta)\epsilon^0 + \beta\epsilon^1)$, $\beta \in [0, 1]$. For $\beta \in [\beta_j, \beta_{j+1}]$, $j = 0, \ldots, n-1$, with corresponding $\mathcal{I}_j$, $x_P(\epsilon)$ is given by the linear system (6.5.3) with $\mathcal{I} = \mathcal{I}_j$. Then:

$$\begin{aligned} x_P(\epsilon^{\beta_j}) - x_P(\epsilon^{\beta_{j+1}}) &= M^{+,2}_{\mathcal{I}_j}(\epsilon^{\beta_j} - \epsilon^{\beta_{j+1}}), \\ \|x_P(\epsilon^{\beta_j}) - x_P(\epsilon^{\beta_{j+1}})\| &\leq \|M^{+,2}_{\mathcal{I}_j}\|\|\epsilon^{\beta_j} - \epsilon^{\beta_{j+1}}\| \leq W\|\epsilon^{\beta_j} - \epsilon^{\beta_{j+1}}\| \end{aligned}$$

for all $j = 0, \ldots, n-1$ with

$$W := \max_j \|M^{+,2}_{\mathcal{I}_j}\|.$$

$M^+_{\mathcal{I}_j}$ refers to the pseudoinverse of the matrix associated to the Langrage system (6.5.2) for $\beta \in [\beta_j, \beta_j + 1]$. It follows by continuity of $x_P(\epsilon)$ relative to its domain (Lemma 6.5.1) that there exists $W > 0$ so that for all $\epsilon^0, \epsilon^1 \in \mathrm{dom}\,(\mathcal{F}_P)$ the following holds:

$$\begin{aligned} \|x_P(\epsilon^0) - x_P(\epsilon^1)\| &\leq \sum_{j=0}^{n-1} \|x_P(\epsilon^{\beta_j}) - x_P(\epsilon^{\beta_{j+1}})\| \\ &\leq W \sum_{j=0}^{n-1} \|\epsilon^{\beta_j} - \epsilon^{\beta_{j+1}}\| \\ &= W\|\epsilon^0 - \epsilon^1\|. \end{aligned} \tag{6.5.4}$$

So, $x_P(\epsilon)$ is Lipschitz continuous relative to its domain. □

Since $x_P(\epsilon)$ is piecewise affine with respect to the domain, $v_P(\epsilon)$ is a piecewise quadratic function. The corollary follows directly from the fact that $s(x)$ is a quadratic function given Assumption 2.

Corollary 6.5.1 ($v_P(\epsilon)$ piecewise quadratic function). *Optimal value function $v_P(\epsilon)$, $P \subseteq \mathcal{P}$, is a (Lipschitz) continuous piecewise quadratic function with respect to the domain dom $(\mathcal{F}_P)$.*

6.6 Example

In the upcoming example we show the piecewise quadratic behavior of the optimal value function $v_P(\epsilon)$ and the piecewise linearity of $x_P(\epsilon)$. The example is an adapted Braess' network and was earlier studied by Di et al. (2013, 2016) and Lou et al. (2010).

Consider the network in Figure 6.6.1 with corresponding route set $\mathcal{P}$ that consists of four paths. Namely:

$$a = \{1, 6\}, \qquad b = \{1, 5, 4\}, \qquad c = \{2, 3, 4\}, \qquad d = \{2, 6\},$$

that all connect OD pair (A, D). We send total demand of $d_{(A,D)} = 6$.

We solve $(Q(\epsilon))$ for $\epsilon = 0$, which is equivalent to solving convex problem (Q_n) (see Section 5.3). We find

$$x^n = (x^n_1, x^n_2, x^n_3, x^n_4, x^n_5, x^n_6) = (4, 2, 0, 4, 2, 2).$$

Since in this particular case route flow f is uniquely determined out of link flow distribution x, we have:

$$f^n = (f^n_a, f^n_b, f^n_c, f^n_d) = (2, 2, 0, 2).$$

Solving for the system optimum gives:

$$f^s = (2.846, 0, 0.308, 2.846).$$

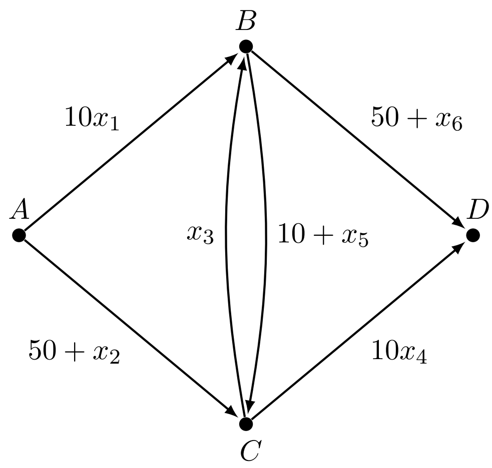


Figure 6.6.1: Extended Braess' network (Lou et al., 2010).

Therefore, let us choose $P = P^n = \{a, b, d\}$ and $\mathcal{P} = \{a, b, c, d\}$. Solving the program in (6.6.1) for P and $\mathcal{P}$ respectively, gives us $\operatorname{dom}(\mathcal{F}_P) = [0, \infty)$ and $\operatorname{dom}(\mathcal{F}_{\mathcal{P}}) = [6.5, \infty)$, respectively.

$$\min_{(\epsilon)} \epsilon \qquad \text{s.t.} \qquad \exists (f, x, \lambda) \in \mathcal{F}_P(\epsilon) \tag{6.6.1}$$

We recall that $\mathcal{F}(\epsilon) = \cup_{P \subseteq \mathcal{P}} \mathcal{F}_P(\epsilon)$ and thus:

$$\mathcal{F}(\epsilon) = \begin{cases} \mathcal{F}_P(\epsilon) & \text{for } \epsilon \in [0, 6.5); \\ \mathcal{F}_P(\epsilon) \cup \mathcal{F}_{\mathcal{P}}(\epsilon) & \text{for } \epsilon \in [6.5, \infty). \end{cases}$$

We consider the branches of problems $(Q_P(\epsilon))$ and $(Q_{\mathcal{P}}(\epsilon))$ separately. We illustrate our approach for the subset of paths P, but the analysis can easily be repeated for problem $(Q_{\mathcal{P}}(\epsilon))$. $(Q_P(\epsilon))$ is a convex quadratic minimization problem. Define the index sets $I = (I_f,\, I_l,\, I_u)$ for $(f, x, \lambda) \in \mathcal{F}_P(\epsilon)$, given $P \subseteq \mathcal{P}$, as follows:

$$\begin{aligned} I_f &:= \{p \in P \mid f_p = 0\}; \\ I_l &:= \{p \in P \mid c_p(f) = \lambda_k\}; \\ I_u &:= \{p \in P \mid c_p(f) = \lambda_k + \epsilon_k\}. \end{aligned}$$

Although numerous possibilities for $I = (I_f, I_l, I_u)$ exist, we can decompose $\operatorname{dom}(\mathcal{F}_P)$ by defining stability sets $\Omega(I)$ and using only four index sets:

$$\Omega(I^0) = \{0\}, \qquad \Omega(I^1) = [0, 13], \qquad \Omega(I^2) = \{13\}, \qquad \Omega(I^3) = [13, \infty).$$

We define the following index sets:

	I_f	I_l	I_u
I^0	$\emptyset$	$\{a,b,d\}$	$\{a,b,d\}$
I^1	$\emptyset$	$\{b\}$	$\{a,d\}$
I^2	$\{b\}$	$\{b\}$	$\{a,d\}$
I^3	$\{b\}$	$\{b\}$	$\emptyset$

Simple calculus leads to the optimal route flow $f_P(\epsilon)$ corresponding to link flow solution $x_P(\epsilon)$ (we used f rather than x for convenience):

$$f_P(\epsilon) = \begin{cases} (2,2,0,2) & \text{if } \epsilon \in \Omega(I^0); \\ (2+\frac{1}{13}\epsilon, 2-\frac{2}{13}\epsilon, 0, 2+\frac{1}{13}\epsilon) & \text{if } \epsilon \in \Omega(I^1); \\ (3,0,0,3) & \text{if } \epsilon \in \Omega(I^2); \\ (3,0,0,3) & \text{if } \epsilon \in \Omega(I^3). \end{cases}$$

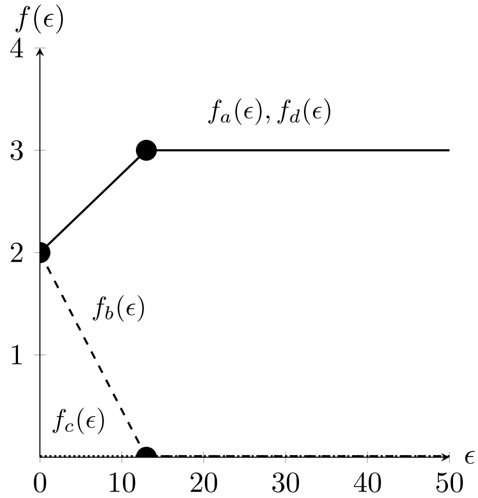
Further,

$$v_P(\epsilon) = \begin{cases} 552 & \text{if } \epsilon \in \Omega(I^0); \\ 552 + \frac{2}{13}(\epsilon^2 - 40\epsilon) & \text{if } \epsilon \in \Omega(I^1); \\ 498 & \text{if } \epsilon \in \Omega(I^2); \\ 498 & \text{if } \epsilon \in \Omega(I^3). \end{cases}$$

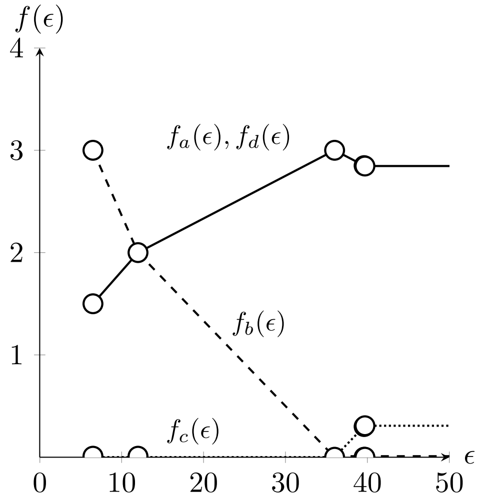
Clearly $f_P(\epsilon)$ (and thus $x_P(\epsilon)$) is piecewise affine and $v_P(\epsilon)$ is a piecewise quadratic function. We apply the same method to find $v_{\mathcal{P}}(\epsilon)$ and $x_{\mathcal{P}}(\epsilon)$ on $\mathrm{dom}\,(\mathcal{F}_{\mathcal{P}})$. Figure 6.6.2 shows the optimal flow solutions $f_P(\epsilon)$ and $f_{\mathcal{P}}(\epsilon)$. Figure 6.6.3 shows the optimal value functions $v_P(\epsilon)$, $v_{\mathcal{P}}(\epsilon)$.

6.7 Synthesis

We studied the continuity of branch problem $(Q_P(\epsilon))$ under affine linear latencies and showed that the feasible set, optimal value function and optimal solution set corresponding to this problem are (Lipschitz) continuous with respect to perturbations in the indifference band. In the next chapter we try to generalize the results of problem $(Q_P(\epsilon))$ to the Best-case BRUE problem $(Q(\epsilon))$.



(a) Optimal route flow solution $f_P(\epsilon)$ corresponding to $(Q_P(\epsilon))$.



(b) Optimal route flow solution $f_{\mathcal{P}}(\epsilon)$ corresponding to $(Q_{\mathcal{P}}(\epsilon))$

Figure 6.6.2: Optimal flow solutions $f(\epsilon)$ corresponding to the network in Figure 6.6.1. *The dots indicate borders between two stability sets.*

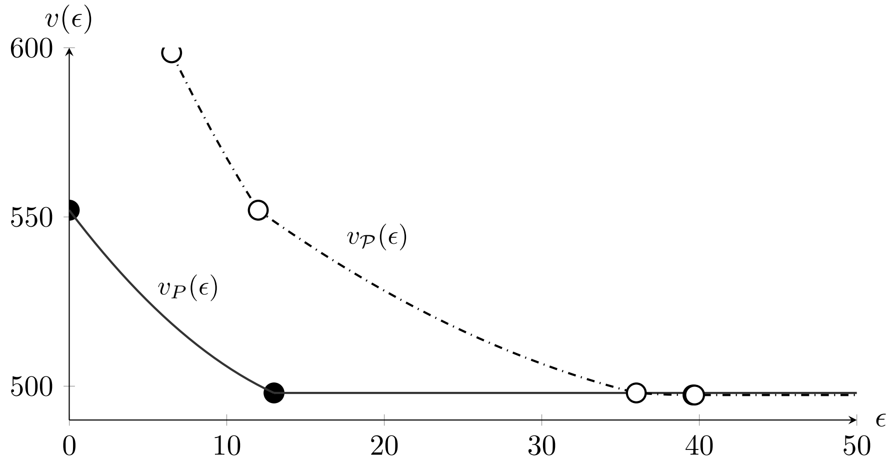


Figure 6.6.3: Piecewise quadratic optimal value function $v_P(\epsilon)$ and $v_{\mathcal{P}}(\epsilon)$ corresponding to the network of Figure 6.6.1.

Chapter 7

Behavior of $(Q(\epsilon))$

This chapter tries to generalize the results of problem $(Q_P(\epsilon))$ to $(Q(\epsilon))$. We treat in detail continuity of feasible set $\mathcal{F}(\epsilon)$, optimal value function $v(\epsilon)$ and optimal solution set $\mathcal{S}(\epsilon)$ with respect to perturbations in ϵ. We look at an equivalent formulation of $(Q(\epsilon))$:

$$\min_x s(x) \qquad \text{s.t.} \qquad x \in \mathcal{F}^x(\epsilon). \tag{7.0.1}$$

Throughout this chapter, Assumption 2 holds. So, we only consider the case in which the latency functions of the edges are affine linear and strict monotone.

For each $\epsilon \geq 0$ the feasible set $\mathcal{F}^x(\epsilon)$ in problem (7.0.1) is the projection of $\mathcal{F}(\epsilon)$ onto $\mathbb{R}^{|E|}$, i.e.:

$$\mathcal{F}^x(\epsilon) = \left\{ x \in \mathbb{R}^{|E|} \mid \exists (f, \lambda) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|\mathcal{K}|} \text{ s.t. } (f, x, \lambda) \in \mathcal{F}_0 \cap \mathcal{F}_u(\epsilon) \cap \mathcal{F}_l \right\},$$

equivalently defined by

$$\mathcal{F}^x(\epsilon) = \bigcup_{P \subseteq \mathcal{P}} \mathcal{F}^x_P(\epsilon),$$

which is a union of closed polyhedra and $\mathcal{F}^x(\epsilon)$ is therefore a compact and nonempty set for each $\epsilon \geq 0$. The corresponding optimal value function and optimal solution set are respectively defined as

$$v(\epsilon) := \min_{x \in \mathcal{F}^x(\epsilon)} s(x), \qquad \text{and} \qquad \mathcal{S}^x(\epsilon) := \arg\min_{x \in \mathcal{F}^x(\epsilon)} s(x).$$

Before we treat continuity of the mentioned (multi-)functions in detail, it is important to notice that the statements of previous chapter do not easily be generalize to problem $(Q(\epsilon))$. For instance $\mathcal{F}^x_P(\epsilon)$ being continuous with respect to $\operatorname{dom}(\mathcal{F}_P)$ for all $P \subseteq \mathcal{P}$ does not imply that $\mathcal{F}^x(\epsilon)$ is continuous at all $\epsilon \geq 0$. Similar claims hold with respect to $v(\epsilon)$ and $\mathcal{S}^x(\epsilon)$.

7.1 Behavior of $\mathcal{F}(\epsilon)$

We turn our attention to the feasible set $\mathcal{F}^x(\epsilon)$. Recall that the feasible set is the union of the feasible sets of the branches, i.e. $\mathcal{F}^x(\epsilon) = \bigcup_{P \subseteq \mathcal{P}} \mathcal{F}^x_P(\epsilon)$. Using this statement, we arrive at the following lemma:

Lemma 7.1.1 ($\mathcal{F}^x(\epsilon)$ closed and upper semicontinuous). *The multifunction $\mathcal{F}^x(\epsilon)$ is*

(i) a closed mapping at all $\epsilon \geq 0$.

(ii) upper Lipschitz semicontinuous relative to its domain, i.e. for all $\epsilon^0 \geq 0$, exists $K > 0$ and $\delta > 0$ so that

$$\mathcal{F}^x(\epsilon^1) \subseteq U_{K\|\epsilon^1-\epsilon^0\|}\left(\mathcal{F}^x(\epsilon^0)\right), \qquad \textit{for all } \epsilon^1 \in U_\delta(\epsilon^0). \tag{7.1.1}$$

Proof.

(i) Let $\epsilon^0 \geq 0$ be arbitrary. Let $\epsilon^l \to \epsilon^0$, $x^l \to x^0$ and $x^l \in \mathcal{F}^x(\epsilon^l)$. We show that $x^0 \in \mathcal{F}^x(\epsilon^0)$. By Lemma 5.5.1 exists for each $l \in \mathbb{N}$, path set $P^l \subseteq \mathcal{P}$ so that $x^l \in \mathcal{F}_{P^l}(\epsilon^l)$. We can choose a subsequence x^{l_j} so that $\epsilon^{l_j} \to \epsilon^0$, $x^{l_j} \to x^0$ and $x^{l_j} \in \mathcal{F}^x_{P^0}(\epsilon^{l_j})$, i.e. $P^{l_j} = P^0$ for all ϵ^{l_j}. By closedness of $\mathcal{F}^x_{P^0}(\epsilon)$ we get that $x^0 \in \mathcal{F}^x_{P^0}(\epsilon^0)$.

(ii) Let ϵ^0 be given. We show that there exists $\delta > 0$ so that for all ϵ^1 sufficiently close to ϵ^0, i.e. $\|\epsilon^1 - \epsilon^0\| < \delta$ we have $\epsilon^1 \in \operatorname{dom}(\mathcal{F}_P) \Rightarrow \epsilon^0 \in \operatorname{dom}(\mathcal{F}_P)$ for any $P \subseteq \mathcal{P}$. Otherwise there would exist sequence $\epsilon^1 \to \epsilon^0$, for which $\epsilon^1 \in \operatorname{dom}(\mathcal{F}_P)$ but $\epsilon^0 \notin \operatorname{dom}(\mathcal{F}_P)$. This contradicts the fact $\operatorname{dom}(\mathcal{F}_P)$ is closed. So, take $\delta > 0$ as above and $x^1 \in \mathcal{F}^x(\epsilon^1)$. There exists $P \subseteq \mathcal{P}$ so that $x^1 \in \mathcal{F}^x_P(\epsilon^1)$. By applying Corollary 6.3.1 we find (7.1.1).

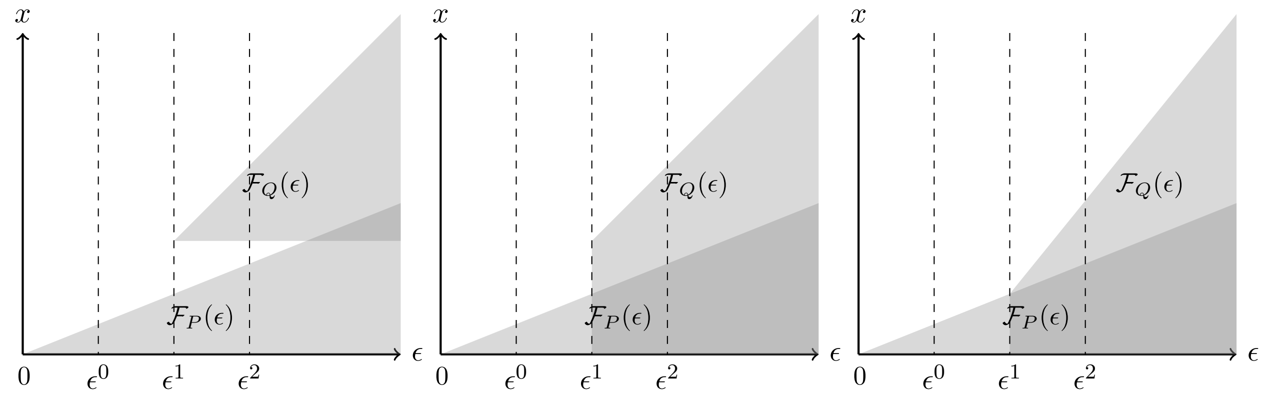
□

Upper semicontinuity of $\mathcal{F}(\epsilon)$ under affine linear latency functions was also proven by Di et al. (2016).

In general, we cannot expect that $\mathcal{F}(\epsilon)$ is Lipschitz continuous with respect to perturbations in ϵ. The 'discontinuity' may arise for example if the feasible sets for different subsets of paths are distinct. By that we mean that

$$\operatorname{dom}(\mathcal{F}_P) \neq \operatorname{dom}(\mathcal{F}_Q), \qquad \text{and} \qquad \mathcal{F}_P(\epsilon) \neq \mathcal{F}_Q(\epsilon), \qquad \text{for } Q \neq P.$$

Figure 7.1.1 illutsrates this behavior graphically. Here, we assume ϵ to be one-dimensional, i.e. $\epsilon = te_{|\mathcal{K}|}$, in which $t \in \mathbb{R}_+$ and $e_{|\mathcal{K}|}$ is the *all-ones vector* of size $|\mathcal{K}|$. Note that in our illustration $\mathcal{F}(\epsilon) = \mathcal{F}_P(\epsilon) \cup \mathcal{F}_Q(\epsilon)$ and we see that $\operatorname{dom}(\mathcal{F}_P) = \{\epsilon \mid \epsilon \geq 0\}$ and $\operatorname{dom}(\mathcal{F}_Q) = \{\epsilon \mid \epsilon \geq \epsilon^1\}$.



(a) $\mathcal{F}(\epsilon)$ upper semicontinuous but not lower semicontinuous at ϵ^1.

(b) $\mathcal{F}(\epsilon)$ connected but not lower semicontinuous at ϵ^1.

(c) $\mathcal{F}(\epsilon)$ continuous at all ϵ.

Figure 7.1.1: Illustrations of continuity of $\mathcal{F}(\epsilon) = \mathcal{F}_Q(\epsilon) \cup \mathcal{F}_P(\epsilon)$.

Figure 7.1.1a indicates that although $\mathcal{F}_P(\epsilon)$ and $\mathcal{F}_Q(\epsilon)$ are both Lipschitz continuous relative to their domain, $\mathcal{F}(\epsilon)$ is upper semicontinuous at all $\epsilon \geq 0$ but it is not lower semicontinuous at ϵ^1. For an arbitrary $x^1 \in \mathcal{F}_Q(\epsilon^1)$ there does not exist a nearby $x^0 \in \mathcal{F}(\epsilon^0) = \mathcal{F}_P(\epsilon^0)$ so that the *distance* between these two points is bounded from above. Indeed, the set $\mathcal{F}_Q(\epsilon)$ implodes after perturbing ϵ^1 to ϵ^0. However, if we choose indifference band $\epsilon^2 \geq \epsilon^1$ this problem will not arise since $\mathcal{F}_Q(\epsilon^2)$ is nonempty and then Corollary 6.3.1 holds.

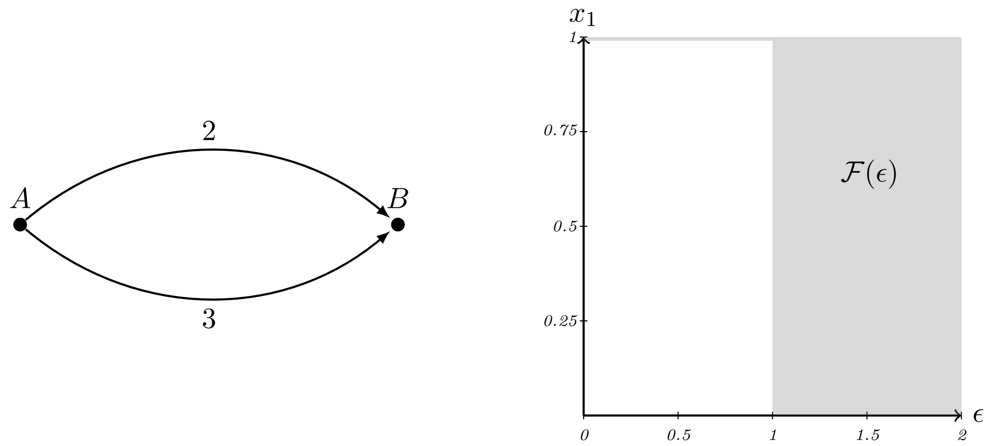
If $\mathcal{F}(\epsilon)$ is connected, as in Figure 7.1.1b, we can construct a *path* from any (ϵ^0, x^0) to any (ϵ^1, x^1) without leaving the graph. Intuitively, all nonempty subsets $\mathcal{F}_P(\epsilon)$ should then overlap. Figure 7.1.1b and Figure 7.1.1c shows that $\mathcal{F}(\epsilon)$ being connected at all $\epsilon \geq 0$ is not a sufficient condition to assure continuity of $\mathcal{F}(\epsilon)$. Still, perturbing ϵ^1 to ϵ^0 leads to an implosion of feasible set $\mathcal{F}(\epsilon)$ in Figure 7.1.1b. However, connectedness plays a role in stability of the boundedly rational user equilibrium in a day-to-day setting (Di et al., 2014b).

The implosion does not occur if we consider the setting in Figure 7.1.1c. Here, $\mathcal{F}(\epsilon)$ is continuous. These illustrations support our earlier statement that we cannot easily generalize the continuity results of previous chapter.

Example 7.1.1. *We show later that if f is uniquely determined out of x, $\mathcal{F}(\epsilon)$ behaves continuously. In this example, based on Di et al. (2016), we show that if we relax Assumption 1, we cannot expect feasible set $\mathcal{F}(\epsilon)$ to be lower semicontinuous.*

Consider the two-link network in Figure 7.1.2a. Two edges connect commodity (A, B)

in the network and each link has a linear latency function. We send demand $d = 1$ from node A to B.



(a) Example Network. **(b)** $\mathcal{F}(\epsilon)$ is only upper semicontinuous at $\epsilon = 1$.

Figure 7.1.2: Example, relaxing Assumption 1, in which $\mathcal{F}(\epsilon)$ is not lower semicontinuous.

Figure 7.1.2b shows the corresponding feasible set $\mathcal{F}(\epsilon)$ for varying ϵ. Obviously, $\mathcal{F}(\epsilon)$ is only upper semicontinuous at $\epsilon = 1$. Indeed, for $\epsilon \geq 1$ any flow distribution that meets the demand is a feasible flow. For $\epsilon < 1$, only $x_1 = 1$ is a feasible BRUE flow.

7.2 Behavior of $v(\epsilon)$

It is straightforward to see that the optimal value function is the minimum of $v_P(\epsilon)$ over all possible subsets of $\mathcal{P}$. In other words:

$$v(\epsilon) = \min_{P \subseteq \mathcal{P}: \mathcal{F}_P(\epsilon) \neq \emptyset} v_P(\epsilon), \qquad \text{for all } \epsilon \geq 0. \tag{7.2.1}$$

The optimal value function $v(\epsilon)$ for $\epsilon = 0$ equals the total travel time of the PRUE flow distribution. Hence, $v(0) = v_{P^n}(0)$ in which P^n consists of all the paths having minimum cost in the PRUE (see (5.5.1)). From our earlier discussion, the optimal value function shows convex (in fact, piecewise quadratic) behavior over $[0, \epsilon^0]$ as long as the optimal solution lies in $\mathcal{F}^x_{P^n}(\epsilon)$ for all $\epsilon \in [0, \epsilon^0]$. The optimal value function will make a *jump* at ϵ^0 in case the global minimizer at ϵ^0 lies in $\mathcal{F}_P(\epsilon^0) \setminus \mathcal{F}_{P^n}(\epsilon^0)$, $P \neq P^n$. This jump makes that we cannot expect $v(\epsilon)$ to behave continuously relative ϵ.

Figure 7.2.1 illustrates this property of $v(\epsilon)$ graphically. Here, we introduce a quadratic objective function and depict its level set. In addition, we show the behavior of the corresponding objective function $v(\epsilon)$ as defined in (7.2.1). Since $\mathcal{F}^x(\epsilon)$ is in general only upper semicontinuous (Figure 7.2.1a) we cannot expect $v(\epsilon)$ to behave continuously

(Figure 7.2.1b). In this particular case, $v(\epsilon)$ makes a jump at ϵ^1. Connectedness of $\mathcal{F}^x(\epsilon)$ as in Figure 7.2.1c is not a sufficient condition for continuity of $v(\epsilon)$. Indeed, the optimal value function Figure 7.2.1d depicts is not different from the optimal value function in Figure 7.2.1b. The behavior of $v(\epsilon)$ is then said to be *piecewise continuous*, i.e. $v(\epsilon)$ is a single-valued function that is discontinuous but can be partitioned so that $v(\epsilon)$ behaves continuously over each partition. In addition, $v(\epsilon)$ is a decreasing function with respect to increasing ϵ (Proposition 6.4.1).

The next lemma formalizes previous discussion and indicates that $v(\epsilon)$ is a lower semicontinuous function. So, $v(\epsilon)$ cannot become significantly smaller by perturbing ϵ.

Lemma 7.2.1 ($v(\epsilon)$ lower semicontinuous). *For all $\epsilon^0 \geq 0$ exists a Lipschitz modulus $LK > 0$ and $\delta > 0$ so that:*

$$v(\epsilon^0) - v(\epsilon^1) \leq LK\|\epsilon^1 - \epsilon^0\|, \qquad \textit{for all } \epsilon^1 \in V_\delta(\epsilon^0).$$

Proof. Let ϵ^0 be arbitrary. Define $\delta > 0$ as in Lemma 7.1.1 and take $\epsilon^1 \in V_\delta(\epsilon^0)$. Suppose $\bar{x}(\epsilon^1) \in \mathcal{F}^x(\epsilon^1)$ and $\bar{x}(\epsilon^0) \in \mathcal{F}^x(\epsilon^0)$ are the optimal link flow solutions corresponding to $v(\epsilon^1), v(\epsilon^0)$ respectively. We have from Lemma 7.1.1 (i) that under the assumptions of the Lemma for $\bar{x}(\epsilon^1) \in \mathcal{F}^x(\epsilon^1)$ exists an $x(\epsilon^0) \in \mathcal{F}^x(\epsilon^0)$ (which is the closest in the Euclidean norm) so that the distance is bounded by $K\|\epsilon^1 - \epsilon^0\|$:

$$\begin{aligned} v(\epsilon^0) - v(\epsilon^1) &= s(\bar{x}(\epsilon^0)) - s(\bar{x}(\epsilon^1)) \\ &\leq s(x(\epsilon^0)) - s(\bar{x}(\epsilon^1)) \\ &\leq L\|x(\epsilon^0) - \bar{x}(\epsilon^1)\| \\ &\leq LK\|\epsilon^1 - \epsilon^0\|. \end{aligned}$$

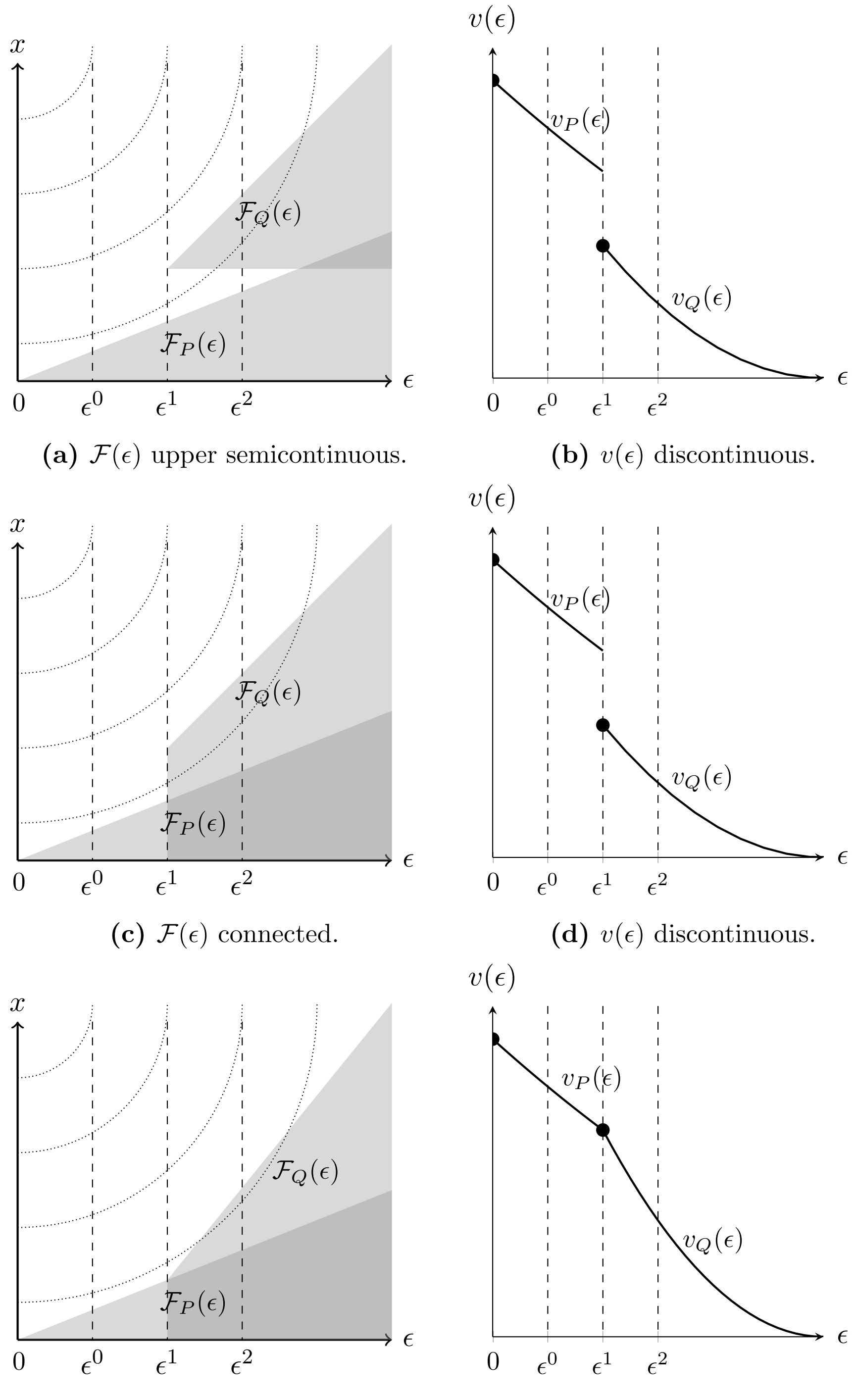
The second inequality follows from the Lipschitz continuity of the (convex) objective function on the Euclidean space. □

Figure 7.2.1e and Figure 7.2.1f show that continuity of $\mathcal{F}^x(\epsilon)$ is a sufficient condition for $v(\epsilon)$ to be continuous.

Theorem 7.2.1 ($v(\epsilon)$ Lipschitz continuous if $\mathcal{F}^x(\epsilon)$ is Lipschitz continuous). *Suppose $\mathcal{F}^x(\epsilon)$ is Lipschitz continuous relative to its domain, then $v(\epsilon)$ is Lipschitz continuous (in fact, piecewise quadratic) at all $\epsilon \geq 0$.*

Proof. The proof is similar to the one given in Lemma 7.2.1. □

Let us interpret the result of this section. We showed that $v(\epsilon)$ is not necessarily continuous with respect to perturbations in ϵ under Assumption 2. Hence, from a sensitivity analysis point of view, a small perturbation in ϵ may lead to a significant change in per-



(a) $\mathcal{F}(\epsilon)$ upper semicontinuous. **(b)** $v(\epsilon)$ discontinuous.

(c) $\mathcal{F}(\epsilon)$ connected. **(d)** $v(\epsilon)$ discontinuous.

(e) $\mathcal{F}(\epsilon)$ continuous. **(f)** $v(\epsilon)$ continuous.

Figure 7.2.1: Continuity of $\mathcal{F}(\epsilon)$ and implications for continuity of $v(\epsilon)$.

formance of the system. This possibly occurs if the parameter ϵ is not in accordance with the real-life indifference band, i.e. if ϵ is not perfectly calibrated.

Remark 7.2.1. *If $\mathcal{F}^x(\epsilon)$ is Lipschitz continuos, $v(\epsilon)$ is Lipschitz continuous as well rela-*

tive to $\epsilon \geq 0$. In fact, $v(\epsilon)$ is then a piecewise quadratic function because $v_P(\epsilon)$ is piecewise quadratic relative to dom$(\mathcal{F}_P)$ for all $P \subseteq \mathcal{P}$. However, although $v_P(\epsilon)$ is convex relative to dom$(\mathcal{F}_P)$, $v(\epsilon)$ is then not necessarily convex (Figure 7.2.1f).

7.3 Behavior of $\mathcal{S}(\epsilon)$

We indicated that the continuity of $\mathcal{F}^x(\epsilon)$ is a sufficient condition for $v(\epsilon)$ to be continuous relative to its domain. In this section, we consider the continuity of the optimal solution set $\mathcal{S}(\epsilon)$.

Assumptions 1 and 2 assure that the optimal link flow solution is unique for problem $(Q_P(\epsilon))$, $P \subseteq \mathcal{P}$, for all indifference vectors $\epsilon \geq 0$ in dom$(\mathcal{F}_P)$. Moreover, treating this optimal solution $x_P(\epsilon)$ as a function with respect to perturbations in ϵ, the optimal link flow solution is piecewise affine and thus Lipschitz continuous relative to dom$(\mathcal{F}_P)$ (Theorem 6.5.2).

Similar claims for $\mathcal{S}^x(\epsilon)$ are unlikely. The optimal solution set $\mathcal{S}^x(\epsilon)$ is in general not continuous. A look at Figure 7.2.1a and Figure 7.2.1b supports this claim. Here, $\mathcal{F}(\epsilon)$ is not continuous and therefore we cannot expect that $v(\epsilon)$ behaves continuously. Obviously, having a continuous objective function makes that if the optimal value function is not continuous, the optimal solution set cannot be continuous. On the other hand, if it turns out that $\mathcal{F}(\epsilon)$ is continuous, then we have that the optimal value function $v(\epsilon)$ is continuous but in general we do not have that $\mathcal{S}^x(\epsilon)$ is continuous.

Lemma 7.3.1 ($v(\epsilon)$ continuous $\Rightarrow$ $\mathcal{S}^x(\epsilon)$ upper semicontinuous). *Suppose $v(\epsilon)$ is continuous relative to its domain, then $\mathcal{S}^x(\epsilon)$ is upper semicontinuous at all $\epsilon \geq 0$.*

Proof. Using the same proof as in Theorem 6.5.1, we find that $\mathcal{S}^x(\epsilon)$ is a closed mapping and it is contained in the compact set $\mathcal{F}_0$. Assume that $\mathcal{S}^x(\epsilon)$ is not upper semicontinuous at ϵ^0. Now, exists sequence $\epsilon^l \to \epsilon^0$, $x^l \in \mathcal{S}^x(\epsilon^l)$ and $x^l \to x^0$ but $x^0 \notin \mathcal{S}^x(\epsilon^0)$. Since $x^0 \notin \mathcal{S}^x(\epsilon^0)$ we have that $s(x^0) > v(\epsilon^0)$. Then,

$$\lim_{l\to\infty} v(\epsilon^l) = \lim_{l\to\infty} s(x^l) = s(x^0) > v(\epsilon^0),$$

contradicting the continuity of $v(\epsilon)$ at ϵ^0. $\square$

Main observation is that in general even $\mathcal{F}^x(\epsilon)$ being continuous is not sufficient for our optimal solution set $\mathcal{S}^x(\epsilon)$ to behave continuously relative to $\epsilon \geq 0$. For an authority this behavior is undesirable. For instance, if ϵ is not well-calibrated, the route choice behavior of travelers for this indifference band may be completely different compared to the setting in which the ϵ is perfectly mirrors real-life decisions. Obviously, it is questionable whether the indifference band can be perfectly calibrated in practice.

We finish this section by a numerical example in which $v(\epsilon)$ is continuous but the optimal solution set is not.

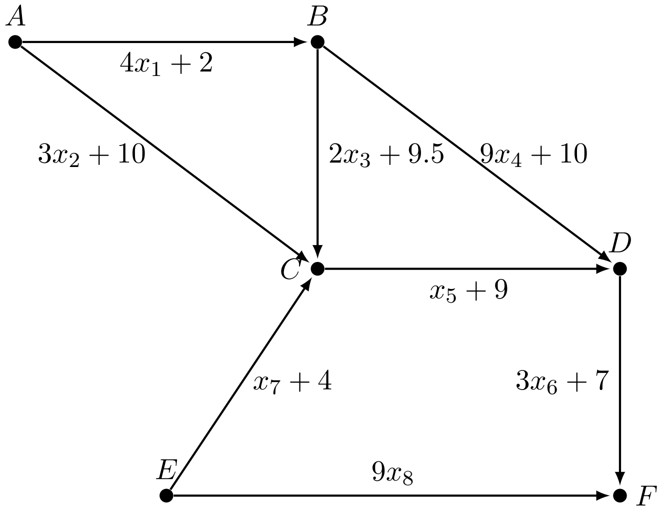
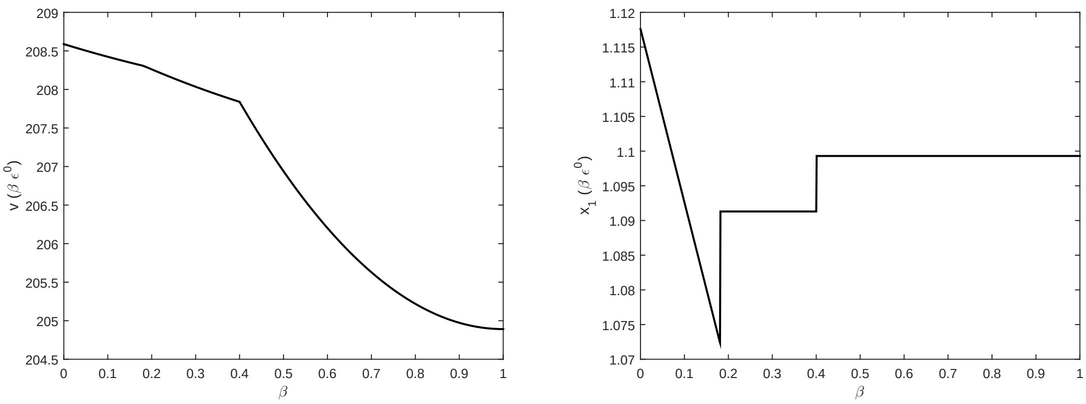


Figure 7.3.1: Network corresponding to Example 7.3.1

Example 7.3.1 (Discontinuous optimal link flows). *Figure 7.3.1 shows a test network including the latency functions for all links. There are two commodities:* (A, F) *and* (E, F)*, demand for both commodities is* 3*. There are three paths that connect* (A, F) *and two paths which connect* (E, F)*. The route flow* f *is uniquely determined out of* x*. Indeed, rank* $\left(\begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix}\right) = 5$*. We will see in a later chapter that the full column rank of this matrix implies continuity of the feasible set.*



(a) $v(\epsilon)$ **(b)** $x_1(\epsilon)$

Figure 7.3.2: Optimal value function and optimal solution for different ϵ in the Network of Figure 7.3.1.

We perform numerical experiments for $\beta\epsilon^0$ *in which* ϵ^0 *is the vector of maximum travel* *time differences in the system-optimal flow distribution and β ranges from* 0 *to* 1*. We found the Best-case BRUE flow distribution using a branch-and-bound algorithm (Section 10.1). We see that the optimal value function $v(\epsilon)$ behaves continuously and piecewise quadratic with respect to ϵ (Figure 7.3.2a). However, the optimal link flow solution $x_1(\epsilon)$ for edge* 1 *does not behave continuous with respect to an increasing ϵ (Figure 7.3.2b). We notice that on the pieces on which $v(\epsilon)$ behaves quadratic, $x_1(\epsilon)$ behaves affine linear.*

7.4 Synthesis

We generalized the results from the branch problem $(Q_P(\epsilon))$ to the Best-case BRUE problem $(Q(\epsilon))$. We showed that even under affine linear link costs, the problem is difficult and lacks favorable properties. By that we mean that the feasible set, optimal value function and optimal solution set are in general not continuous. In the next chapter we propose a bilevel reformulation to generalize these results to more general latency functions.

Chapter 8

The Bilevel Reformulation

In previous chapters we showed that the Best-case BRUE problem ($Q(\epsilon)$) is possibly not continuous with respect to perturbations in ϵ, even if we limit ourselves to affine linear link cost functions. In this chapter, we reformulate the problem as a bilevel optimization problem which gives us the opportunity to consider problem ($Q(\epsilon)$) under higher order polynomial link cost functions given Assumption 1 (e.g. the BPR-function). We see that the same difficulties arise as in the previous chapters. However, the bilevel formulation we present in this chapter allows us to formulate an algorithm which finds an approximate Best-case BRUE flow distribution for a given $\epsilon \geq 0$ without any additional assumptions on the link cost functions. Chapter 9 discusses this algorithm.

8.1 The Bilevel Problem

The construction of the feasible set $\mathcal{F}(\epsilon)$ of all BRUE flow distributions requires the identification of the paths $p \in \mathcal{P}_k, k \in \mathcal{K}$, for which $c_p(x) - \lambda_k - \epsilon_k \leq 0$ holds. To decompose the feasible set by branches, we require a method to identify the acceptable paths. However, we can construct $\mathcal{F}(\epsilon)$ by solving a set of mathematical programs without any explicit knowledge about these paths beforehand.

Consider the following optimization problem for any $0 \leq \rho \leq \epsilon$ (formally: $0 \leq \rho \leq \Lambda^T \epsilon$, but we abuse notation for convenience):

$$\min_{(f,x)} z(f,x,\rho) := \sum_{e \in E} \int_0^{x_e} l_e(\tau) d\tau - \rho^T f \qquad \text{s.t.} \qquad (f,x) \in \mathcal{F}_0. \qquad (q(\rho))$$

We will refer to problem ($q(\rho)$) as the *lower-level problem.*

Before we continue, let us introduce the optimal value function $\phi(\rho)$ corresponding to ($q(\rho)$):

$$\phi(\rho) = \min_{(f,x) \in \mathcal{F}_0} z(f,x,\rho),$$

and the optimal solution set $\psi(\rho)$ given $0 \leq \rho \leq \epsilon$:

$$\psi(\rho) = \left\{ \ (f,x) \ \Big| \ (f,x) \in \arg\min \phi(\rho), (f,x) \in \mathcal{F}_0 \ \right\}.$$

Under Assumption 1, the objective function $z(f,x,\rho)$ is continuous on $\mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} \times \mathbb{R}^{|\mathcal{P}|}$ and is convex in (ρ, f, x). Thus, $(q(\rho))$ is a convex optimization problem. We see that any solution of $(q(\rho))$, provided $0 \leq \rho \leq \epsilon$, is a flow distribution which is contained in $\mathcal{F}(\epsilon)$. Since for convex optimization problems it holds that the KKT conditions are both sufficient and necessary for a solution to be globally optimal, we can formulate the next lemma.

Lemma 8.1.1. *(Di et al., 2013). Let $\epsilon \geq 0$ be fixed. $(f,x) \in \mathcal{F}(\epsilon)$ if and only if exists ρ, $0 \leq \rho \leq \epsilon$, so that $(f,x) \in \psi(\rho)$.*

Proof. We consider the system of KKT conditions corresponding to $(q(\rho))$. For that, we introduce Langrange multiplier vector $(\beta, \lambda, \gamma) \in \mathbb{R}^{|E|} \times \mathbb{R}^{|\mathcal{K}|} \times \mathbb{R}^{|\mathcal{P}|}$ and then any $(f,x) \in \mathcal{F}_0$ is a global optimal solution of $(q(\rho))$ if and only if (f,x) satisfies the system with Lagrange multipliers (β, λ, γ), $\gamma \geq 0$:

$$\begin{array}{rrr} l(x) - \beta = 0 & (d - \Lambda f)^T \lambda = 0 & (\Gamma f - x)^T \beta = 0 \\ f^T \gamma = 0 & \Gamma^T \beta - \gamma - \Lambda^T \lambda - \rho = 0 & (f,x) \in \mathcal{F}_0 \end{array} \tag{8.1.1}$$

We find that (we substitute $\beta_e = l_e(x_e)$ for all $e \in E$):

$$\sum_{e \in p} l_e(x_e) - \gamma_p - \lambda_k - \rho_p = 0, \qquad \text{for all } p \in \mathcal{P}_k, k \in \mathcal{K}.$$

Since $\gamma_p, \rho_p \geq 0$ for all $p \in \mathcal{P}$ we obtain:

$$\sum_{e \in p} l_e(x_e) \geq \lambda_k + \gamma_p + \rho_p \geq \lambda_k, \qquad \text{for all } p \in \mathcal{P}_k, k \in \mathcal{K}.$$

Whereas this inequality holds for all paths for this commodity, we interpret λ_k as the minimum cost to travel for commodity $k \in \mathcal{K}$:

$$\min_{p \in \mathcal{P}_k} \sum_{e \in p} l_e(x_e) \geq \lambda_k, \qquad \text{for all } k \in \mathcal{K}.$$

For any $p \in \mathcal{P}_k, k \in \mathcal{K}$, which carries flow (i.e. $f_p > 0$), we obtain from the complementarity condition $f_p \gamma_p = 0$ that

$$f_p > 0 \Rightarrow \sum_{e \in p} l_e(x_e) = \lambda_k + \rho_p \leq \lambda_k + \epsilon_k \leq \min_{q \in \mathcal{P}_k} \sum_{e \in q} l_e(x_e) + \epsilon_k,$$

which is exactly the BRUE condition (4.1.1).

To prove that for any $(f,x) \in \mathcal{F}(\epsilon)$ exists $0 \leq \rho \leq \epsilon$ so that $(f,x) \in \psi(\rho)$, we again use the KKT conditions in (8.1.1) and use a similar approach as in Di et al. (2013). Let $(f,x) \in \mathcal{F}(\epsilon)$ be arbitrary (obviously, $(f,x) \in \mathcal{F}_0$) and define $\beta_e = l_e(x_e)$ for each $e \in E$. Furthermore define for all $k \in \mathcal{K}$:

$$\lambda_k := \min_{p \in \mathcal{P}_k} l_e(x_e).$$

and

$$\rho_p := \begin{cases} \sum_{e\in p} l_e(x_e) - \lambda_k & \text{if } \sum_{e\in p} l_e(x_e) \leq \lambda_k + \epsilon_k; \\ \epsilon_k & \text{otherwise.} \end{cases}$$

It directly follows that $0 \leq \rho \leq \epsilon$. Define γ_p for each $p \in \mathcal{P}_k, k \in \mathcal{K}$ as follows

$$\gamma_p := \begin{cases} 0 & \text{if } \sum_{e\in p} l_e(x_e) \leq \lambda_k + \epsilon_k; \\ \sum_{e\in p} l_e(x_e) - \epsilon_k - \lambda_k & \text{otherwise.} \end{cases}$$

The complementarity condition $f^T\gamma$ is satisfied and $\gamma_p \geq 0$. Then:

$$\sum_{e\in p} l_e(x_e) - \gamma_p - \lambda_k - \rho_p = 0, \qquad \text{for all } p \in \mathcal{P}_k, k \in \mathcal{K}.$$

By our choices for the Lagrange multiplier vectors and $0 \leq \rho \leq \epsilon$, we have that $(f, x, \beta, \lambda, \gamma)$ satisfies the system in (8.1.1). □

The relation between mathematical program $(q(\rho))$ and set $\mathcal{F}(\epsilon)$ has become clear. Based on Lemma 8.1.1, the following holds:

$$\mathcal{F}(\epsilon) = \bigcup_{0\leq\rho\leq\epsilon} \psi(\rho). \tag{8.1.2}$$

Since $(q(\rho))$ is a convex problem for all $0 \leq \rho \leq \epsilon$, the lower-level problem is easy to solve (in polynomial time if we consider affine linear link cost functions). In general, we can apply optimization tools for traffic engineering (e.g. convex combinations method). The solution leads to a BRUE flow distribution(s) (for given $0 \leq \rho \leq \epsilon$) and the union of all possible BRUE flow distributions (as expressed in (8.1.2)) serves as the feasible set in the Best-/Worst-case BRUE problem (5.0.1).

Based on our analysis, we reformulate problem $(Q(\epsilon))$ for given $\epsilon \geq 0$ as a bilevel optimization problem:

$$
\begin{array}{ll}
\min_{(\rho,f,x)} s(x) & \\
\quad\quad \text{s.t.} & 0 \le \rho \le \epsilon, \text{ and } (f,x) \text{ solves} \\
& \min_{(f,x)} z(f,x,\rho) \\
& (f,x) \in \mathcal{F}_0
\end{array}
\tag{8.1.3}
$$

Rather than considering the program in (8.1.3) we study a single-level reformulation of the bilevel problem:

$$
\begin{array}{ll}
\min_{(\rho,f,x)} s(x) & \\
\quad\quad \text{s.t.} & 0 \le \rho \le \epsilon \\
& (f,x) \in \psi(\rho)
\end{array}
\tag{$\tilde{Q}(\epsilon)$}
$$

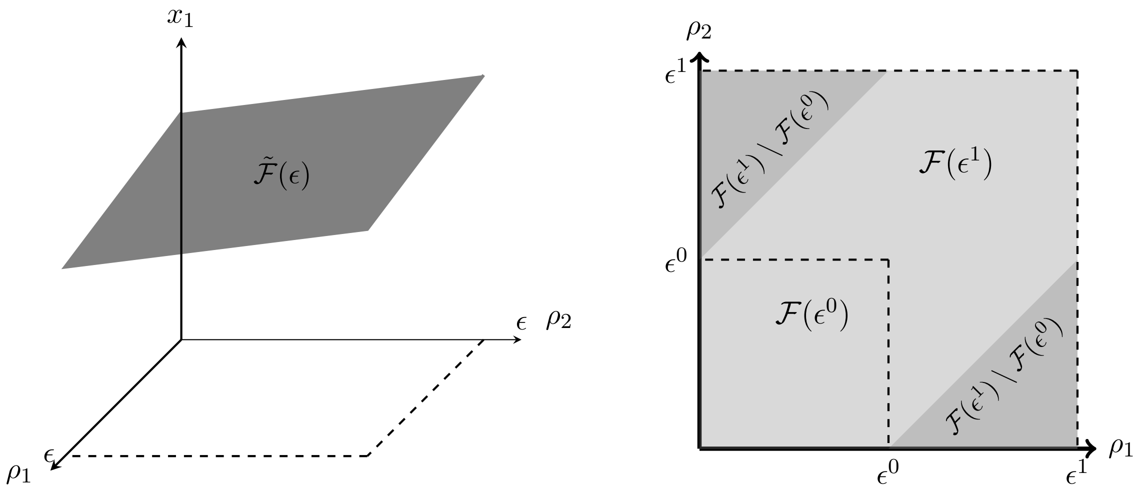
($\tilde{Q}(\epsilon)$) is the problem to find the Best-case BRUE flow distribution: the *leader* chooses variable $0 \le \rho \le \epsilon$ so that the realized BRUE in the lower level (Lemma 8.1.1) leads to the minimum total travel time. Although problem ($q(\rho)$) is easy to solve, problem ($\tilde{Q}(\epsilon)$) is a bilevel optimization problem and NP-hard in general (Bard, 1998). Furthermore, the equivalent formulation of ($\tilde{Q}(\epsilon)$) that considers KKT conditions of the lower-level problem does not satisfy a constraint qualification and thus we cannot find for each descent direction a feasible curve into this direction. Consequently, we cannot simply apply the KKT conditions to the bilevel problem (Section 5.2)

A similar approach would adopt the *inertial user equilibrium* (IUE) (Zhang and Yang, 2015) in the lower-level. Zhang and Yang (2015) showed that any BRUE flow can be constructed by an IUE and that any IUE is contained in $\mathcal{F}(\epsilon)$ for a certain ϵ. However, we prefer the formulation in ($q(\rho)$) due to its similarites with Beckmann's formulation for the PRUE.

Remark 8.1.1. *We note that the feasible set of problem ($\tilde{Q}(\epsilon)$) is defined on the space $\mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|}$, while ($Q(\epsilon)$) is defined on the space $\mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|}$. Figure 8.1.1a illustrates this behavior abstractly for Example 4.2.1.*

Important observation is that the mapping $\psi : 0 \le \rho \le \epsilon \to \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|}$ is not injective. Figure 8.1.1b shows the projection of $\tilde{\mathcal{F}}(\epsilon)$ onto the (two-dimensional) ρ-space for Example 4.2.1. We depict $\mathcal{F}(\epsilon^0) = \cup_{0 \le \rho \le \epsilon^0} \psi(\rho)$ and $\mathcal{F}(\epsilon^1) = \cup_{0 \le \rho \le \epsilon^1} \psi(\rho)$. We indicate how $\mathcal{F}(\epsilon^0)$ and $\mathcal{F}(\epsilon^1)$ are related. By that we mean that for any $(\rho^0, f^0, x^0) \in \tilde{\mathcal{F}}(\epsilon^0)$ any $\rho^1 = \rho^0 + e_2\delta$, $\delta \in \mathbb{R}$, has the same corresponding flow distribution (f^0, x^0). (e_2 is an all-ones vector of size 2.) Hence, $(\rho^1, f^0, x^0) \in \tilde{\mathcal{F}}(\epsilon^1)$ if $0 \le \rho^1 \le \epsilon^1$. In particular, any $0 \le \rho \le \epsilon^1$ in the light-gray area corresponds to a flow distribution (also contained) in $\mathcal{F}(\epsilon^0)$.

Remark 8.1.2 (Variational inequality formulation). *The formulation of ($q(\rho)$) allows us to define a (more general) variational inequality of the BRUE flow distribution. In fact, $(f,x) \in \mathcal{F}_0$ is a BRUE flow as in (4.1.1) if and only if exists $0 \le \rho \le \epsilon$ so that*



(a) Abstract illustration of $\tilde{\mathcal{F}}(\epsilon)$. **(b)** Projection of $\tilde{\mathcal{F}}(\epsilon)$ onto $\mathbb{R}^{|\mathcal{P}|}$.

Figure 8.1.1: Illustration of feasible set $\tilde{\mathcal{F}}(\epsilon)$ corresponding to Example 4.2.1 .

$(f, x) \in \mathcal{F}_0$ *solves the variational inequality*

$$(c(f) - \rho)\,(\bar{f} - f) \geq 0, \qquad \text{for all } (\bar{f}, \bar{x}) \in \mathcal{F}_0.$$

Bilevel optimization problems often use a variational inequality formulation. A variational inequality that defines a boundedly rational flow distribution in the dynamic environment is given by Han et al. (2015). Their definition can be reduced to the static setting as well.

8.2 Behavior of the Lower-Level Problem

We turn the attention to lower-level problem $(q(\rho))$. We indicated that we can construct $\mathcal{F}(\epsilon)$ by solving $(q(\rho))$ for all $0 \leq \rho \leq \epsilon$. Even under linear latencies, $\mathcal{F}(\epsilon)$ does not meet a constraint qualification (Section 5.2). In this section we study the easy lower-level problem $(q(\rho))$ and we will generalize results to $\mathcal{F}(\epsilon)$. Therefore we consider $(q(\rho))$ as a parametric optimization problem under perturbations in ρ.

Remark 8.2.1. $\mathcal{F}_0$ *is compact and nonempty. Moreover, it does not depend* ρ.

Since $z(f, x, \rho)$ is continuous, we state he following lemma.

Lemma 8.2.1 (Existence lower-level solution). *Problem* $(q(\rho))$ *attains its minimum for all* $\rho \geq 0$, *i.e.* $\psi(\rho) \neq \emptyset$. *Moreover,* $\phi(\rho) > -\infty$.

Proof. Let $\rho^0 \geq 0$ be fixed. Since $z(f, x, \rho^0)$ is a continuous function and $\mathcal{F}_0$ is compact, there exists a point $(f^0, x^0) \in \mathcal{F}_0$ at which this minimum is attained. So $\phi(\rho^0) = z(\rho^0, f^0, x^0)$ is finite. $\square$

Similar to the PRUE and system-optimal flow distribution, the solution set $\psi(\rho)$ for fixed $\rho \geq 0$ is unique with respect to link flow x. This follows from the strict convexity of the objective function $z(f, x, \rho)$ with respect to x. We formalize the claim.

Lemma 8.2.2 (Unique x in lower-level solution). *For any $\rho \geq 0$, $(f, x) \in \psi(\rho)$ is unique with respect to link flow x.*

Proof. Suppose $\rho^0 \geq 0$ to be fixed. Let $(f^0, x^0), (f^1, x^1) \in \psi(\rho^0)$ be two optimal solutions of problem $(q(\rho^0))$ with $x^0 \neq x^1$. We construct $(f^\beta, x^\beta) = (\beta f^0+(1-\beta)f^1, \beta x^0+(1-\beta)x^1)$ for all $\beta \in [0, 1]$. Since $z(f, x, \rho)$ is strictly convex with respect to x under Assumption 1 (by the second derivative test), we get for all $0 < \beta < 1$:

$$\begin{aligned}
z(f^\beta, x^\beta, \rho^0) &= z(\beta f^0 + (1-\beta)f^1, \beta x^0 + (1-\beta)x^1, \rho^0) \\
&= \sum_{e \in E} \int_0^{\beta x_e^0+(1-\beta)x_e^1} l_e(\tau)d\tau - {\rho^0}^T(\beta f^0 + (1-\beta)f^1) \\
&< \beta \left(\sum_{e \in E} \int_0^{x_e^0} l_e(\tau)d\tau - {\rho^0}^T f^0 \right) + (1-\beta) \left(\sum_{e \in E} \int_0^{(1-\beta)x_e^1} l_e(\tau)d\tau - {\rho^0}^T f^1 \right) \\
&= \beta z(f^0, x^0, \rho^0) + (1-\beta)z(f^1, x^1, \rho^0).
\end{aligned}$$

This would imply that $z(f^\beta, x^\beta, \rho^0)$ has a (strict) lower objective value, which contradicts the optimality of (f^0, x^0) and (f^1, x^1). $\square$

Remark 8.2.2. *The result of Lemma 8.2.2 implies that there may exist several BRUE route flow distributions f corresponding to the same link flow x. This problem is well-studied for the PRUE (see e.g. Bar-Gera (2006); Borchers et al. (2015)) and most studies imposed additional conditions on f to force uniqueness. For problem $(q(\rho))$ the set $\mathcal{H}(x, \rho)$ of route flows corresponding to BRUE link flow distribution x for $0 \leq \rho \leq \epsilon$ is defined as*

$$\mathcal{H}(x, \rho) = \left\{ \; f \;\middle|\; \Gamma f = x, \Lambda f = d, f \geq 0, \rho^T f = \omega \; \right\}.$$

Here, ω is the value of $\omega = \rho^T f^0$ for an arbitrary $(f^0, x^0) \in \psi(\rho)$. We emphasize that an additional condition should be fulfilled for any f to be a BRUE route flow distribution compared to the PRUE flow distribution (see Remark 3.2.1 and Example 4.2.3).

As was earlier indicated, any PRUE flow distribution is also a BRUE flow distribution for any $\epsilon \geq 0$. Indeed, if $\rho = 0$, the lower-level problem reduces to

$$\min_{(f,x)} \sum_{e \in E} \int_0^{x_e} l_e(\tau)d\tau \qquad \text{s.t} \qquad (f, x) \in \mathcal{F}_0. \tag{8.2.1}$$

This is exactly Beckmann's formulation (Beckmann et al., 1956) of the problem of finding a PRUE flow distribution.

Although the behavior of the optimal value function $\phi(\rho)$ of the lower-level problem under perturbations of ρ is not of interest perse, the continuity of $\phi(\rho)$ is used to study the continuity of the optimal solution set mapping $\psi(\rho)$ with respect to ρ.

Lemma 8.2.3 ($\phi(\rho)$ continuous). *$\phi(\rho)$ is a continuous function at all $\rho \geq 0$.*

Proof. The proof is given in Theorem 6.4.1. $\square$

We proved continuity of $\phi(\rho)$ since continuity of $\phi(\rho)$ turns out to be (more than) sufficient for $\psi(\rho)$ to be a closed mapping.

Lemma 8.2.4 ($\psi(\rho)$ closed mapping). *$\psi(\rho)$ is a closed mapping at all $\rho \geq 0$.*

Proof. We define $\psi(\rho)$ equivalently as

$$\psi(\rho) = \mathcal{F}_0 \cap \{(f, x) \mid z(f, x, \rho) - \phi(\rho) \leq 0\}.$$

By this definition of $\psi(\rho)$, we can use the proof that is given in the first part of Theorem 6.5.1. $\square$

Obviously, $\psi(\rho)$ is a closed mapping and a compact set for each $\rho \geq 0$. Moreover, since $\mathcal{F}_0$ contains $\psi(\rho)$, we find that $\psi(\rho)$ is an upper semicontinuous mapping. We can, however, solely guarantee continuity if $\psi(\rho)$ is unique for each ρ. We emphasize that this does not hold in general since there may exist several route flows f that all correspond to the same link flow x (Remark 8.2.2).

Lemma 8.2.5 ($\psi(\rho)$ upper semicontinuous). *$\psi(\rho)$ is upper semicontinuous at all $\rho \geq 0$.*

Proof. The proof is analagous to the proof of Lemma 7.3.1. $\square$

As mentioned, we can prove the continuity of the lower-level solution set if the solution is unique with respect to x and f for any $0 \leq \rho \leq \epsilon$. The proof for this case is similar as the proof of Theorem 6.5.1 and therefore omitted here.

8.3 Behavior of the Bilevel problem

We showed that the lower-level problem $(q(\rho))$ is a convex optimization problem. The corresponding feasible set and optimal value function are convex. We try to generalize results to the bilevel problem $(\tilde{Q}(\epsilon))$. ϵ is the parameter vector, similar as in the analysis of $(Q(\epsilon))$.

We introduce for each $\epsilon \geq 0$, the feasible set $\tilde{\mathcal{F}}(\epsilon)$ that corresponds to $(\tilde{Q}(\epsilon))$:

$$\tilde{\mathcal{F}}(\epsilon) = \left\{ \ (\rho, f, x) \ \Big| \ 0 \leq \rho \leq \epsilon, (f, x) \in \psi(\rho) \ \right\}.$$

By this definition, we can regard $\mathcal{F}(\epsilon)$ as the projection of $\tilde{\mathcal{F}}(\epsilon)$ onto $\mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|}$. The optimal value function $v(\epsilon)$ of the problem is defined as:

$$v(\epsilon) := \min_{(\rho,f,x)\in\tilde{\mathcal{F}}(\epsilon)} s(x),$$

and the optimal solution set $\tilde{\mathcal{S}}(\epsilon)$ as

$$\tilde{\mathcal{S}}(\epsilon) := \operatorname*{arg\,min}_{(\rho,f,x)\in\tilde{\mathcal{F}}(\epsilon)} v(\epsilon).$$

We regard $\mathcal{S}(\epsilon)$ as the projection of $\tilde{\mathcal{S}}(\epsilon)$ onto the (f, x)-space.

We study continuity of the feasible set $\tilde{\mathcal{F}}(\epsilon)$ with respect to perturbations in ϵ. Obviously, $\tilde{\mathcal{F}}(\epsilon)$ is an increasing set, i.e. for all $\epsilon^1 \geq \epsilon^0 \geq 0$ we have that (note the equivalence with Lemma 5.5.1 (v))

$$\tilde{\mathcal{F}}(\epsilon^0) \subseteq \tilde{\mathcal{F}}(\epsilon^1).$$

Furthermore, if $\epsilon = 0$, $\mathcal{F}(0) = \{(f^n, x^n)\}$ in which (f^n, x^n) is a PRUE flow distribution. Now, $\tilde{\mathcal{F}}(0) = \{(0, f^n, x^n)\}$. It directly follows that $\tilde{\mathcal{F}}(0) \subseteq \tilde{\mathcal{F}}(\epsilon)$ for any $\epsilon \geq 0$. The problem in $(\tilde{Q}(\epsilon))$ reduces to finding a system optimum if $\epsilon \geq \epsilon^{so}$. Indeed, the *leader* in the upper level is free to choose ρ and does so such that in the lower-level a system-optimal flow is realized.

Lemma 8.3.1 ($\tilde{\mathcal{F}}(\epsilon)$ closed). *$\tilde{\mathcal{F}}(\epsilon)$ is a closed mapping at all $\epsilon \geq 0$ and compact for all $\epsilon \geq 0$.*

Proof. Let $\epsilon^0 \geq 0$ be arbitrary. Let sequences $\epsilon^l \to \epsilon^0$ be given and $(\rho^l, y^l) \in \tilde{\mathcal{F}}(\epsilon^l)$ for all $l \in \mathbb{N}$. Then, $(\rho^l, y^l) \to (\rho^0, y^0)$ implies that $(\rho^0, y^0) \in \mathcal{F}(\epsilon^0)$ since $0 \leq \rho^0 \leq \epsilon^0$ by continuity, and $y^0 \in \psi(\rho^0)$ whereas $\psi(\rho^0)$ is closed. $\square$

We assure existence of a BRUE flow distribution under the same mild assumptions Beckmann et al. (1956) used to assure existence and uniqueness of a PRUE flow distribution. The statement in the next lemma is similar to the one in Lemma 8.2.1. Moreover, the optimal value function $v(\epsilon)$ is positive for all $\epsilon \geq 0$ since the travel time is positive for each link and the total demand is positive.

Trivially, $v(\epsilon)$ is decreasing with respect to ϵ. That is, for $\epsilon^1 > \epsilon^0$ we have $v(\epsilon^1) \leq v(\epsilon^0)$.

Lemma 8.3.2. *For any $\epsilon \geq 0$, $v(\epsilon) > 0$ and $\tilde{\mathcal{S}}(\epsilon)$ is nonempty. Moreover, $\tilde{\mathcal{F}}(\epsilon)$, $v(\epsilon)$ are upper and lower semicontinuous relative to $\epsilon geq 0$, respectively.*

Proof. We prove that $\tilde{\mathcal{S}}(\epsilon)$ is nonempty. Consider problem $(Q(\epsilon))$. Then, $\mathcal{F}(\epsilon) = \cup_{P\subseteq\mathcal{P}}\mathcal{F}_P(\epsilon)$. Obviously the inequalities that define $\mathcal{F}_P(\epsilon)$ are continuous and hence $\mathcal{F}_P(\epsilon)$ is a compact set for each $P \subseteq \mathcal{P}$. Since a finite union of compact sets is compact, $\mathcal{F}(\epsilon)$ is compact set for each $\epsilon \geq 0$. By the extreme value theorem it follows that there exists $(f, x, \lambda) \in \mathcal{F}(\epsilon)$ which minimizes $s(x)$. We can easily find a ρ which gives $(f, x) \in \psi(\rho)$.

We prove that $\tilde{\mathcal{F}}(\epsilon)$ is upper semicontinuous at all $\epsilon \geq 0$ using the proof of Still (2006). Suppose, for the sake of contradiction, $\tilde{\mathcal{F}}(\epsilon)$ is not upper semicontinuous at ϵ^0. Then, for some sequence $\epsilon^l \to \epsilon^0$ and $(\rho^l, y^l) = (\rho^l, f^l, x^l) \in \mathcal{F}(\epsilon^l)$ and some $\delta > 0$ so that for all l

$$\|(\rho^l, y^l) - (\rho, y)\| \geq \delta, \qquad \text{for all } (\rho, y) \in \tilde{\mathcal{F}}(\epsilon^0).$$

However, since $\tilde{\mathcal{F}}(\epsilon)$ is contained in a compact set, we have that there exists a converging subsequence $(\rho^{l_j}, y^{l_j}) \to (\rho^0, y^0)$. Since $\tilde{\mathcal{F}}(\epsilon)$ is closed, $(\rho^0, y^0) \in \tilde{\mathcal{F}}(\epsilon^0)$. So, $\tilde{\mathcal{F}}(\epsilon)$ is upper semicontinuous at ϵ^0.

Given $\epsilon^0 \geq 0$, we show that to any $\delta > 0$ exists $\tau > 0$ so that

$$v(\epsilon^0) \leq v(\epsilon) + \delta, \qquad \text{for all } \|\epsilon - \epsilon^0\| < \tau.$$

Assume for the sake of contradiction that this does not hold. Then exists $\delta_0 > 0$ and sequence $\epsilon^l \to \epsilon^0$ so that

$$v(\epsilon^0) > v(\epsilon^l) + \delta_0, \tag{8.3.1}$$

for all $l \in \mathbb{N}$. Take to ϵ^l corresponding $(\rho^l, y^l) = (\rho^l, f^l, x^l) \in \tilde{\mathcal{F}}(\epsilon^l)$ with $y^l \in \psi(\rho^l)$. Without loss of generality (taking a subsequence) we may assume $(\rho^l, y^l) \to (\rho^0, y^0)$. By closedness of $\tilde{\mathcal{F}}(\epsilon)$ it follows that $(\rho^0, y^0) \in \tilde{\mathcal{F}}(\epsilon^0)$ and thus

$$v(\epsilon^0) \leq s(x^0) = \lim_{l\to\infty} s(x^l) = \lim_{l\to\infty} v(\epsilon^l)$$

contradicting (8.3.1). □

For general latency functions under Assumption 1 the same problems arise as for problem $(Q(\epsilon))$ under Assumption 2. The feasible set does not explode after a small perturbation in ϵ but it can, on the other hand, implode. The optimal value function $v(\epsilon)$ cannot become significantly smaller by a change the parameter, but is in general not upper semicontinuous.

The discussed characteristics are undesirable from a traffic engineering perspective. As discussed for the problem under affine linear latencies, the tolerance vector ϵ should be perfectly calibrated. A small perturbation of ϵ can lead - from a behavioral perspective - to completely different choice behavior. Since in practice it is difficult to pertfectly calibrate ϵ, for the NDP this may lead to the implementation of a policy measure which is not necessarily optimal in practice.

8.3.1 Connectedness of $\mathcal{F}(\epsilon)$

This section concerns the connectedness of $\mathcal{F}(\epsilon)$, i.e. the set of feasible BRUE flow distributions. First, we present the necessary definitions.

Definition 8.3.1 (Connected set). *Set X is said to be connected if X is not a union of two nonempty separated sets, i.e. there exist no nonempty open sets Y and Z for which*

$$Y \cup Z = X, \qquad \text{and} \qquad \bar{Y} \cap Z = Y \cap \bar{Z} = \emptyset$$

holds. Here, $\bar{Y}$ ($\bar{Z}$) stands for the closure of set Y (Z).

In some cases it is more appealing to use the definition of *path-connectedness* (or *arcwise connectedness*) of a set. Intuitively, in a path-connected set X we can construct a path from any point $x \in X$ to any other point $y \in X$ without leaving X. A path-connected set X is also connected.

Definition 8.3.2 (Path-connected set). *Set X is said to be path-connected if for any $x, y \in X$ exists path γ connecting x and y, i.e. exists continuous mapping $\gamma : [0, 1] \to X$ so that $\gamma(0) = x$ and $\gamma(1) = y$.*

For general latency functions, Han et al. (2015) showed that under assumptions $\mathcal{F}(\epsilon)$ consists of a union of connected sets but is not necessarily connected itself. In Di et al. (2014b) it is claimed that $\mathcal{F}(\epsilon)$ is connected in case the latency functions in the network are affine linear. However, the proof utilizes a strong assumption which is difficult to check beforehand. We study the connectedness of $\mathcal{F}(\epsilon)$ in more detail. Note that $\psi(\rho)$ is a polyhedron (and thus connected) for each ρ. It appears that $\mathcal{F}(\epsilon)$ is connected for any $\epsilon \geq 0$ provided that Assumption 1 holds.

Theorem 8.3.1 ($\mathcal{F}(\epsilon)$ connected). *$\mathcal{F}(\epsilon)$ is connected for each $\epsilon \geq 0$.*

Proof. Let us fix ϵ and define

$$E(\epsilon) = \{0 \leq \rho \leq \epsilon\}.$$

Because $E(\epsilon)$ is a convex set for all $\epsilon \geq 0$, it is also connected for all $\epsilon \geq 0$. Indeed, take $\rho^0, \rho^1 \in E(\epsilon)$ and construct path γ via $\gamma(\beta) = \beta\rho^1 + (1 - \beta)\rho^0$. Obviously, $\gamma(0) = \rho^0$ and $\gamma(1) = \rho^1$ and the path is linear and thus continuous.

Let us define $\mathcal{F}(\epsilon) = \cup_{\rho \in E(\epsilon)} \psi(\rho)$ and suppose, for the sake of contradiction, that $\mathcal{F}(\epsilon)$ is not connected. Then there are two open sets A, B so that $\mathcal{F}(\epsilon) = A \cup B$ and

$\bar{A} \cap B = A \cap \bar{B} = \emptyset$. Denote by $\rho(A), \rho(B)$ the following projections:

$$\rho(A) = \{\rho \in E(\epsilon) \mid \psi(\rho) \subseteq A\};$$
$$\rho(B) = \{\rho \in E(\epsilon) \mid \psi(\rho) \subseteq B\}.$$

We may assume that both $\rho(A), \rho(B)$ are nonempty. Furthermore, $\rho(A) \cap \rho(B) = \emptyset$ since otherwise exists $\bar{\rho} \in \rho(A) \cap \rho(B)$ and $\psi(\bar{\rho}) \in A \cap B$ contradicting $\bar{A} \cap B = \emptyset$.

By the definition of upper semicontinuity of $\psi(\rho)$ at any $\rho^0 \in E(\epsilon)$ for each τ_{ρ^0} exists $\delta_{\rho^0} > 0$ so that

$$\psi(\rho) \subseteq U_{\tau_{\rho^0}}(\psi(\rho^0)), \qquad \text{for all } \rho \in U_{\delta_{\rho^0}}(\rho^0)$$

holds. For any $\rho \in \rho(A)$ take τ_ρ and corresponding δ_ρ so that $U_{\tau_\rho}(\psi(\rho)) \subset A$, then we obtain that

$$\bigcup_{\rho \in \rho(A)} U_\delta(\rho) \cap E = \rho(A).$$

The union of the open balls $\bigcup_{\rho \in \rho(A)} U_{\delta_\rho}(\rho)$ is obviously open. So we find that $\rho(A)$ is relatively open in $E(\epsilon)$. Furthermore, $\rho(B)$ is relatively open in $E(\epsilon)$ as well. Then, $E(\epsilon)$ can be separated by a pair of relatively open sets in $E(\epsilon)$. This contradicts the connectedness of $E(\epsilon)$. This proof appeared earlier in Zgurovsky et al. (2012). □

8.3.2 Unique Lower-Level Response

It turns out that the continuity of the feasible set and optimal value function highly depend on the continuity of the optimal solution set in the lower level. In general, we cannot expect $\psi(\rho)$ being continuous. However, in an upcoming chapter we study a regularized problem for which this property holds. Therefore, we include the claims in which the lower-level response is assumed to be unique for each choice of ρ by the upper level. We add a remark to discuss the conditions under which $\psi(\rho)$ behaves continuously with respect to ρ. By a unique lower-level response we mean that for each upper-level choice of ρ, the solution of the lower-level response (f, x) is unique with respect to f and x. It turns out that a unique lower-level response is a sufficient condition for $\tilde{\mathcal{F}}(\epsilon)$ and $v(\epsilon)$ to be continuous.

Theorem 8.3.2 ($\tilde{\mathcal{F}}(\epsilon)$ continuous). *Let the lower-level response be unique for each $0 \leq \rho \leq \epsilon$, then $\tilde{\mathcal{F}}(\epsilon)$ is continuous with respect to ϵ. In fact, $\tilde{\mathcal{F}}(\epsilon)$ is Lipschitz-continuous relative to its domain under Assumption 2.*

Proof. First, we prove the claim of the theorem under Assumption 2. We proved earlier that $\tilde{\mathcal{F}}(\epsilon)$ is a closed mapping. Let $\epsilon^0, \epsilon^1 \geq 0$ be arbitrary. Then for any $(\rho^0, y^0) =$

$(\rho^0, f^0, x^0) \in \tilde{\mathcal{F}}(\epsilon^0)$ we have that there exists $(\rho^1, y^1) \in \tilde{\mathcal{F}}(\epsilon^1)$ so that

$$\begin{aligned} \|(\rho^1, y^1) - (\rho^0, y^0)\| &\leq C\|\rho^0 - \rho^1\| \\ &\leq CN\|\epsilon^0 - \epsilon^1\|, \end{aligned}$$

in which the latter inequality follows from Proposition 6.3.1. The first inequality follows from the optimality conditions of the objective function over a polyhedron for a strictly convex function (Theorem 9.1.2).

Let us consider the case which relaxes Assumption 2. Since we have that $\tilde{\mathcal{F}}(\epsilon)$ is upper semicontinuous (Lemma 8.3.2) we solely prove that $\tilde{\mathcal{F}}(\epsilon)$ is lower semicontinuous relative to ϵ. Recall that $\psi(\rho)$ is continuous relative to ρ, and thus lower semicontinuous.

So, given ρ^0 for any $y^0 = (f^0, x^0) \in \psi(\rho^0)$ and sequence $\rho^l \to \rho^0$, exists a sequence y^l so that $y^l \in \psi(\rho^l)$ for l large and $\|y^l - y^0\| \to 0$. We emphasize that this sequence is unique. Define $E(\epsilon) = \{0 \leq \rho \leq \epsilon\}$. This multifunction is continuous at each ϵ.

Suppose $\tilde{\mathcal{F}}(\epsilon)$ is not lower semicontinuous at $\epsilon^0 \geq 0$. Then exists $(\rho^0, y^0) \in \tilde{\mathcal{F}}(\epsilon^0)$ and $\epsilon^l \to \epsilon^0$, then for all (ρ^l, y^l) so that $(\rho^l, y^l) \in \tilde{\mathcal{F}}(\epsilon^n)$, for l large, $\|(\rho^l, y^l) - (\rho^0, y^0)\| > \delta$ for some $\delta > 0$.

However, by continuity of $E(\epsilon)$ for each ϵ^0 and $\epsilon^l \to \epsilon^0$ exists $\rho^l \in E(\epsilon^l)$, for l large, and $\|\rho^l - \rho^0\| \to 0$. By continuity of $\psi(\rho)$, exists $y^l \in \psi(\rho^l)$ so that $\|y^l - y^0\| \to 0$. Hence, $\tilde{\mathcal{F}}(\epsilon)$ is lower semicontinuous. $\square$

The result states that the feasible set cannot explode or implode after perturbing ϵ. Consequently, the optimal value function cannot become sufficiently smaller or larger by perturbing ϵ. If $\psi(\rho)$ is continuous with respect to perturbations in ρ (i.e. any $\rho \geq 0$ has a unique lower level response (f, x)), the optimal value function $v(\epsilon)$ is continuous relative to ϵ.

Theorem 8.3.3 ($v(\epsilon)$ continuous). *If $\psi(\rho)$ is continuous with respect to perturbations in ρ, $v(\epsilon)$ is continuous at all $\epsilon \geq 0$. In fact, if $\psi(\rho)$ is Lipschitz continuous with respect to ρ then $v(\epsilon)$ is Lipschitz continuous relative to ϵ.*

Proof. First, we prove this statement for the case in which the latency functions are linear. Suppose $(\bar{\rho}^0, \bar{y}^0)$, $(\bar{\rho}^1, \bar{y}^1)$ are the optimal solutions corresponding to $v(\epsilon^0), v(\epsilon^1)$ respectively. Under the assumptions of the theorem, for all $(\rho^0, y^0) = (\rho^0, f^0, x^0) \in \tilde{\mathcal{F}}(\epsilon^0)$ exists $(\rho^1, y^1) \in \tilde{\mathcal{F}}(\epsilon^1)$ so that the distance is bounded by $N\|\epsilon^1 - \epsilon^0\|$ (by Lipschitz continuity of $\tilde{\mathcal{F}}(\epsilon)$, see Theorem 8.3.2):

$$\begin{aligned} v(\epsilon^1) - v(\epsilon^0) &= s(\bar{x}^1) - s(\bar{x}^0) \\ &\leq s(x^1) - s(\bar{x}^0) \\ &\leq \Phi\|x^1 - \bar{x}^0\| \\ &\leq \Phi N\|\epsilon^1 - \epsilon^0\|. \end{aligned}$$

The second inequality follows from the Lipschitz continuity of the (convex) objective function. $x^1 \in \tilde{\mathcal{F}}(\epsilon^1)$ is chosen so that it is closest to x^0 with respect to the ∞-norm. Similarly, we find that $v(\epsilon^0) - v(\epsilon^1) \leq \Phi N\|\epsilon^1 - \epsilon^0\|$. These two results lead to the claim.

Let us relax Assumption 2. We know by Lemma 8.3.2 that $v(\epsilon)$ is lower semicontinuous. Left is to prove that $v(\epsilon)$ is upper semicontinuous. We know that $\tilde{\mathcal{F}}(\epsilon)$ is continuous, in particular lower semicontinuous.

Since $\tilde{\mathcal{F}}(\epsilon^0)$ is lower semicontinuous, for any $(\rho^0, y^0) \in \tilde{\mathcal{F}}(\epsilon^0)$ and sequence $\epsilon^l \to \epsilon^0$ there exists a sequence (ρ^l, y^l) so that $(\rho^l, y^l) \in \tilde{\mathcal{F}}(\epsilon^l)$, for l large, and $\|(\rho^l, y^l) - (\rho^0, y^0)\| \to 0$.

Take $(\rho^0, y^0) \in \tilde{\mathcal{S}}(\epsilon^0)$, i.e. $s(y^0) = v(\epsilon^0)$. Then we have for the mentioned sequence, for l large,

$$v(\epsilon^l) \leq s(y^l) \leq s(y^0) + \delta = v(\epsilon^0) + \delta,$$

for some $\delta > 0$, by continuity of $s(y)$. $v(\epsilon)$ is thus upper semicontinuous. $\square$

The result of this theorem is in line with our earlier results. By the assumptions of the theorem $v(\epsilon)$ behaves piecewise continuous relative to $\epsilon \geq 0$.

Remark 8.3.1 (Conditions under which lower-level response is unique). *Full column rank of* $\begin{bmatrix} \Lambda \\ \Gamma \end{bmatrix}$ *guarantees continuity of the optimal solution set* $\psi(\rho)$ *of the lower problem. Then, any* $0 \leq \rho \leq \epsilon$ *leads to a unique solution in* (f, x) *and the optimal solution set is continuous. The feasible set* $\mathcal{F}(\epsilon)$ *and optimal value function* $v(\epsilon)$ *are now both continuous as well.*

The question remains whether this condition is natural. We answer this question negatively. Borchers et al. (2015) showed a range of basic examples for which the condition was not satisfied.

The difficulty of the problem $(\tilde{Q}(\epsilon))$ lies in the non-unique lower-level response for a given $0 \leq \rho \leq \epsilon$ choice of the leader. It is well known that these problems are difficult to solve (see Dempe (2003)). Therefore, we propose a regularized approach in the next chapter to approximate the Best-case BRUE flow distribution.

8.4 Synthesis

We reformulated the Best-case BRUE problem as bilevel optimization problem. This reformulation allowed us to evaluate continuity of $(Q(\epsilon))$ under more general link cost functions. It turns out that the same difficulties arise as under linear latency functions. However, if the lower-level response is unique for each $0 \leq \rho \leq \epsilon$ the feasible set and optimal value function behave continuously under perturbations in ϵ. The next chapter therefore proposes a perturbed lower-level problem that forces a unique lower-level response for each ρ.

Chapter 9

A Regularized Approach

In previous chapter we saw that if x uniquely determines f, the feasible set and optimal value function behave continuously for any link cost function under Assumption 1. However, f is generally not unique for each x. In this chapter we propose an approach to handle the non-unique lower-level response by perturbing the objective function of the lower-level problem. Now, we find a regularized lower-level problem which we use for algorithm development.

9.1 The Regularized Bilevel Problem

Let us consider, as suggested by e.g. Dempe (2000), the *regularized lower-level problem* $(q^\alpha(\rho))$ corresponding to $(q(\rho))$. We define the regularized problem, given $0 \leq \rho \leq \epsilon$ and $\alpha \in \mathbb{R}_+$, as:

$$\min_{(f,x)\in\mathcal{F}_0} \tilde{z}(f,x,\rho,\alpha) := z(f,x,\rho) + \alpha\|f\|^2 \qquad \text{s.t.} \qquad (f,x) \in \mathcal{F}_0. \qquad (q^\alpha(\rho))$$

Trivially, if $\alpha = 0$ problem $(q^\alpha(\rho))$ reduces to $(q(\rho))$. For any positive α we have that the objective function $\tilde{z}(f,x,\rho,\alpha)$ is strictly convex since the Hessian matrix of this function with respect to variables (f,x) becomes:

$$\nabla^2_{(f,x)}\tilde{z}(f,x,\rho,\alpha) = \begin{bmatrix} 2\alpha I_{|\mathcal{P}|} & 0 \\ 0 & \nabla_x l(x) \end{bmatrix},$$

in which $I_{|\mathcal{P}|}$ is the identity matrix of size $|\mathcal{P}|$. This matrix is positive definite by positive definiteness of $2\alpha I_{|\mathcal{P}|}$ and $\nabla_x l(x)$. From Lemma 8.2.2 it follows that the solution of $(q^\alpha(\rho))$ is unique both in f and x. However, we cannot guarantee that the solution of $(q^\alpha(\rho))$ is a solution to $(q(\rho))$. Therefore, we choose $\alpha > 0$ close to zero to assure that the solution is close to the solution of the BRUE lower-level problem $(q(\rho))$. We say that we approximate a BRUE flow distribution.

Let $\psi^\alpha(\rho)$ denote the optimal solution set corresponding to problem $(q^\alpha(\rho))$ given $0 \leq \rho \leq \epsilon$. The regularized bilevel problem $(\tilde{Q}(\epsilon))$, given $\alpha > 0$, becomes:

$$\begin{aligned} \min_{(\rho,f,x)} \; & s(x) \\ \text{s.t.} \quad & 0 \leq \rho \leq \epsilon \\ & (f,x) \in \psi^\alpha(\rho) \end{aligned} \tag{$\tilde{Q}^\alpha(\epsilon)$}$$

As mentioned, the lower level has a unique response for any $\alpha > 0$. Furthermore, let us introduce the corresponding feasible set

$$\tilde{\mathcal{F}}^\alpha(\epsilon) = \left\{ \; (\rho, f, x) \;\middle|\; 0 \leq \rho \leq \epsilon, (f,x) \in \psi^\alpha(\rho) \; \right\},$$

and optimal value function

$$v^\alpha(\epsilon) := \min_{(\rho,f,x)\in\tilde{\mathcal{F}}^\alpha(\epsilon)} s(x).$$

Lastly, the optimal solution set of problem $(q^\alpha(\rho))$:

$$\tilde{\mathcal{S}}^\alpha(\epsilon) := \operatorname*{arg\,min}_{(\rho,f,x)\in\tilde{\mathcal{F}}^\alpha(\epsilon)} v^\alpha(\epsilon).$$

We prove the continuity properties of the regularized lower problem $(q^\alpha(\rho))$ for fixed $\alpha > 0$. We emphasize that any continuity property for the feasible set, optimal value function and optimal solution set of problem $(\tilde{Q}(\epsilon))$ holds in particular for the feasible set, optimal value function and optimal solution set respectively of problem $(\tilde{Q}^\alpha(\epsilon))$.

Theorem 9.1.1 ($\psi^\alpha(\rho)$ continuous). *For any $\alpha > 0$, $\psi^\alpha(\rho)$ is continuous relative to ρ.*

Proof. The same proof as in Theorem 6.5.1 holds. $\square$

It turns out that the optimal solution set $\psi^\alpha(\rho)$ behaves Lipschitz continuously with respect to changes in ρ if Assumption 2 holds and is locally $\frac{1}{2}$-Hölder continuous without any additional assumption. Consider the case in which Assumption 2 holds. In other words, the latency functions of the edges in the network are affine linear.

Theorem 9.1.2 ($\psi^\alpha(\rho)$ Lipschitz continuous). *Let Assumption 2 hold. Suppose $\alpha > 0$ to be fixed. Suppose $y(\rho) = (f(\rho), x(\rho)) \in \psi^\alpha(\rho)$ for $\rho \geq 0$, then there exists $C > 0$ so that for all $\rho^0, \rho^1 \geq 0$,*

$$\|y(\rho^1) - y(\rho^0)\| \leq C\|\rho^1 - \rho^0\|$$

holds.

Proof. If we consider the case in which the latency functions are affine linear, the objective function $\tilde{z}(f, x, \rho, \alpha)$ becomes quadratic for any $\alpha \geq 0$. Let us consider this regularized lower problem for any $\alpha > 0$, $\rho \geq 0$:

$$\min_{(\rho,f,x)} \tilde{z}(f,x,\rho,\alpha) = \frac{1}{2}x^T Ax + b^T x - \rho^T f + \alpha\|f\|^2 \qquad \text{s.t.}(f,x) \in \mathcal{F}_0 \qquad (\tilde{q}^\alpha(\rho))$$

Obviously, the matrix

$$\begin{bmatrix} 2\alpha I_{|\mathcal{P}|} & 0 \\ 0 & A \end{bmatrix}$$

is diagonal and positive definite. Consequently, the solution (f,x) of $(\tilde{q}^\alpha(\rho))$ is unique. For each $\rho \geq 0$, the KKT conditions of $(\tilde{q}^\alpha(\rho))$ become:

$$\begin{aligned} Ax + b + \beta &= 0 & \Lambda f &= d \\ \Gamma^T\beta + 2\alpha f - \rho - \Lambda^T\lambda - \gamma &= 0 & \Gamma f - x &= 0 \\ \gamma^T f &= 0 & \gamma, f &\geq 0 \end{aligned} \tag{9.1.1}$$

with corresponding multipliers (β,λ,γ). By Assumption 2, (f,x) solves $(\tilde{q}^\alpha(\rho))$ if and only (f,x) satisfies (9.1.1) with the corresponding Lagrange multipliers. As in the proof of Lemma 6.5.1 we can choose $\mathcal{I} \subseteq \{1,\ldots,|\mathcal{P}|\}$ and consider solution $(\rho,f,x,\beta,\lambda,\gamma)$ of (9.1.1) so that (f,γ) satisfies:

$$\begin{aligned} f_p &= 0, & p &\in \mathcal{I}; \\ \gamma_p &= 0, & p &\notin \mathcal{I}; \end{aligned} \tag{9.1.2}$$

and define stability set $\Omega(\mathcal{I})$ as

$$\Omega(\mathcal{I}) = \{\ \rho \ \Big|\ \text{solution of (9.1.1) satisfies (9.1.2) given } \mathcal{I}\ \}.$$

We can apply the same proof as in Lemma 6.5.1 and Theorem 6.5.2 to prove Lipschitz continuity of the optimal solution set $\psi^\alpha(\rho)$. □

In general, we cannot expect $\psi^\alpha(\rho)$ to be Lipschitz continuous. For higher order polynomial link cost functions under Assumption 1, $\psi^\alpha(\rho)$ is locally $\frac{1}{2}$-Hölder continuous with respect to perturbations in ρ.

For the upcoming result we require the optimality conditions of nonlinear programming. We can apply the next theorem since a constraint qualification holds for the constraints in $\mathcal{F}_0$. We introduce the Lagrange function corresponding to problem $(q^\alpha(\rho))$, $\alpha > 0$, for $\rho \geq 0$:

$$\mathbb{L}(f,x,\rho,\beta,\lambda,\gamma) = \tilde{z}(f,x,\rho,\alpha) + (d - \Lambda f)^T\lambda + (\Gamma f - x)^T\beta - f^T\gamma$$

in which (β,λ,γ) is the vector of KKT multipliers.

Theorem 9.1.3 (Second-order sufficient optimality conditions over a polyhe-

dron). *Let $\rho^0 \geq 0$. If (f^0, x^0) satisfies the first-order stationarity conditions (note that $\nabla = \nabla_{(f,x)}$)*

$$\nabla \mathbb{L}(f^0, x^0, \rho^0, \beta^0, \lambda^0, \gamma^0) = 0, \qquad \textit{and} \qquad {\gamma^0}^T f^0 = 0$$

with multipliers $(\beta^0, \lambda^0, \gamma^0)$, $\gamma^0 \geq 0$ in combination with

$$\tau^T \nabla^2 \mathbb{L}(f^0, x^0, \rho^0, \beta^0, \lambda^0, \gamma^0)\tau > 0, \qquad \textit{for all} \qquad \tau \in C(f^0, x^0) \textit{ where } \tau \neq 0,$$

then there exists a neighborhood $U(f^0, x^0) \subseteq \mathbb{R}^{|E|} \times \mathbb{R}^{|\mathcal{P}|}$ of (f^0, x^0) and $\zeta_{\rho^0} > 0$ so that for all $(f, x) \in \mathcal{F}_0 \cap U(f^0, x^0)$:

$$\tilde{z}(f, x, \rho^0, \alpha) - \tilde{z}(f^0, x^0, \rho^0, \alpha) \geq \zeta_{\rho^0} \|(f, x) - (f^0, x^0)\|^2. \tag{9.1.3}$$

Here, $C(f^0, x^0)$ is the set of feasible directions at (f^0, x^0):

$$C(f,x) := \left\{ \tau = (\tau^f, \tau^x) \in \mathbb{R}^{|\mathcal{P}|+|E|} \;\middle|\; \begin{array}{ll} \nabla \tilde{z}(f, x, \rho^0, \alpha)^T \tau \leq 0 & \\ \Gamma \tau^f - \tau^x = 0 & \\ \Lambda \tau^f = 0 & \\ \tau_p^f \geq 0 & \textit{for all } p \in \mathcal{P} : f_p = 0 \end{array} \right\}$$

Proof. (The proof is based on Faigle et al. (2013) and Still and Streng (1996)). For notational purposes, we drop ρ and α in the notation of our analysis. Suppose a flow (f^0, x^0) satisfies the assumptions of the theorem with KKT multipliers $(\beta^0, \lambda^0, \gamma^0)$. Define the compact set of feasible directions at (f^0, x^0) as

$$K(f^0, x^0) = \left\{ \tau \in C(f^0, x^0) \;\middle|\; \|\tau\| = 1 \right\}.$$

Put ζ as

$$\zeta := \min_{\tau \in K(f^0, x^0)} \frac{1}{2} \tau^T \nabla^2 \mathbb{L}(f^0, x^0, \beta^0, \lambda^0, \gamma^0)\tau. \tag{9.1.4}$$

The minimization problem in (9.1.4) is well-defined. Indeed, $K(f^0, x^0)$ is compact, the objective function is continuous, and

$$\tau^T \nabla^2 \mathbb{L}(f^0, x^0, \beta^0, \lambda^0, \gamma^0)\tau = \tau^T \nabla^2 \tilde{z}(f^0, x^0)\tau > 0$$

holds for all $\tau \neq 0$. Thus, $\zeta > 0$. By any choice of τ, due to the linear constraints, τ is a feasible direction. That is: for each τ exists a feasible curve through (f^0, x^0) into feasible direction τ. We show that (9.1.3) holds for each $\zeta_{\rho^0} < \zeta$.

Suppose that $\phi < \zeta$ arbitrarily and suppose (f^0, x^0) is not a second order minimizer as in (9.1.3) with $\zeta_{\rho^0} = \phi$. Then, exists an infinite sequence of points $(f^l, x^l) \to (f^0, x^0)$ satisfying $(f^l, x^l) \neq (f^0, x^0)$ and $\tilde{z}(f^l, x^l) - \tilde{z}(f^0, x^0) \leq \phi \|(f^l, x^l) - (f^0, x^0)\|^2$, $l \in \mathbb{N}$. We can write $(f^l, x^l) = (f^0, x^0) + \tau^l t^l$, $\|\tau^l\| = 1$, $t^l > 0$. Now $t^l \to 0$, $\tau^l \to \tau$ and $\|\tau\| = 1$ (for a subsequence). For this sequence it holds that

$$\tilde{z}(f^l, x^l) - \tilde{z}(f^0, x^0) \leq \phi \|(f^l, x^l) - (f^0, x^0)\|^2.$$

We apply the second-order Taylor expansion:

$$\begin{aligned}
\phi(t^l)^2 &\geq \tilde{z}(f^l, x^l) - \tilde{z}(f^0, x^0) = t^l \begin{bmatrix} -\rho + 2\alpha f^0 \\ l(x^0) \end{bmatrix}^T \tau^l + \frac{1}{2} {t^l}^2 {\tau^l}^T \begin{bmatrix} 2\alpha I_{|\mathcal{P}|} & 0 \\ 0 & \nabla_x l(x^0) \end{bmatrix} \tau^l \\
0 &= \Gamma f^l - x^l = t^l \begin{bmatrix} \Gamma \\ -I_{|E|} \end{bmatrix}^T \tau^l \\
0 &= \Lambda f^l - d = t^l \Lambda^T \tau^{f,l} \\
0 &\geq -f^l_P = t_l - \nabla_f {f^0_P}^T \tau^{f,l}
\end{aligned} \tag{9.1.5}$$

where $P = \{p \in \mathcal{P} \mid f^0_p = 0\}$.

Multiplying with the respective KKT multipliers and adding, we obtain for (f^l, x^l):

$$\phi(t^l)^2 \geq t^l \nabla \mathbb{L}(f^0, x^0, \beta^0, \lambda^0, \gamma^0) \tau^l + \frac{1}{2} {t^l}^2 {\tau^l}^T \nabla^2 \mathbb{L}(f^0, x^0, \beta^0, \lambda^0, \gamma^0) \tau^l \tag{9.1.6}$$

By the assumptions of the theorem we find that $\nabla \mathbb{L}(f^0, x^0, \beta^0, \lambda^0, \gamma^0) = 0$. Dividing (9.1.6) by ${t^l}^2$ we obtain:

$$\zeta > \phi \geq \frac{1}{2} {\tau^l}^T \nabla^2 \mathbb{L}(f^0, x^0, \beta^0, \lambda^0, \gamma^0) \tau^l > 0 \tag{9.1.7}$$

Recall that $\|\tau^l\| = 1$. In case $\tau \in K(f^0, x^0)$ we reach a contradiction since we chose ζ so that it was the minimum in (9.1.4). We have $\lim_{l \to \infty} (f^l, x^l) = (f^0, x^0)$, and, hence, $\lim_{l \to \infty} t^l = 0$. We divide the equations in (9.1.5) by t^l and let $l \to \infty$:

$$\begin{aligned}
0 &\geq \begin{bmatrix} -\rho + 2\alpha f^0 \\ l(x^0) \end{bmatrix}^T \tau \\
0 &= \Gamma \tau^f - \tau^x \qquad\qquad , \\
0 &= \Lambda \tau^f \\
0 &\geq -\tau^f_P
\end{aligned}$$

which are exactly the (in)equalities defining $C(f^0, x^0)$. So, $\tau \in K(f^0, x^0)$ which contradicts $\phi < \zeta$. $\square$

We use the second order optimality conditions to prove that the feasible set of the regularized bilevel problem is locally $\frac{1}{2}$-Hölder-continuous.

Theorem 9.1.4 ($\psi^\alpha(\rho)$ locally Hölder). *Given $\alpha > 0$, the mapping $\psi^\alpha(\rho)$ is locally $\frac{1}{2}$-Hölder continuous relative to its domain, i.e. for all $\rho^0 \geq 0$ exists a $\theta, \delta > 0$ so that for all $\rho^1 \in U_\delta(\rho^0)$ and $y^0 \in \psi^\alpha(\rho^0)$, $y^1 \in \psi^\alpha(\rho^1)$ the following holds:*

$$\|y^1 - y^0\| \leq \theta \|\rho^1 - \rho^0\|^{\frac{1}{2}}.$$

Proof. Let $\alpha > 0$ and $\rho^0 \geq 0$. Since $\psi^\alpha(\rho)$ is continuous (Theorem 9.1.1) we may assume that if ρ^1 is sufficiently close to ρ^0, the corresponding optimal solution $y^1 \in \psi^\alpha(\rho^1)$ is sufficiently close to $y^0 \in \psi^\alpha(\rho^0)$. Formally, the second order optimality conditions say that there exists a $\gamma > 0$ so that for all flows in the neighborhood $U_\gamma(\psi^\alpha(\rho^0))$ satisfy (9.1.3). By continuity of $\psi^\alpha(\rho)$: for all $\tau > 0$ exists $\delta > 0$ so that

$$\psi^\alpha(\rho) \subseteq U_\tau(\psi^\alpha(\rho^0)) \qquad \text{for all } \rho \in U_\delta(\rho^0),$$

and selecting $\tau = \gamma$. So, denoting $y^1 \in \psi^\alpha(\rho^1)$ as the optimal solution for ρ^1 and $y^0 \in \psi^\alpha(\rho^0)$ in which $\rho^1 \in U_\delta(\rho^0)$ we find:

$$\begin{aligned}
\|y^1 - y^0\|^2 &\leq \frac{\|\tilde{z}(f^1, x^1, \rho^0) - \tilde{z}(f^0, x^0, \rho^0)\|}{\zeta} \\
&\leq \frac{\|\tilde{z}(f^1, x^1, \rho^0) - \tilde{z}(f^1, x^1, \rho^1) + \phi^\alpha(\rho^1) - \phi^\alpha(\rho^1)\|}{\zeta} \\
&\leq \frac{D^1\|\rho^1 - \rho^0\| + \|\phi^\alpha(\rho^1) - \phi^\alpha(\rho^0)\|}{\zeta}.
\end{aligned}$$

Then Lipschitz continuity of the lower level optimal value function $\phi^\alpha(\rho)$ is sufficient to prove the claim of the theorem. Let y^1, y^0 be chosen so that $\tilde{z}(y^1, \rho^1) = \phi^\alpha(\rho^1)$ and $\tilde{z}(y^0, \rho^0) = \phi^\alpha(\rho^0)$. Notice that $y^1, y^0 \in \mathcal{F}_0$. We get:

$$\begin{aligned}
\phi^\alpha(\rho^1) - \phi^\alpha(\rho^0) &\leq \tilde{z}(y^0, \rho^1) - \tilde{z}(y^0, \rho^0) \\
&\leq D^2\|\rho^1 - \rho^0\|.
\end{aligned}$$

Analogously, we find that

$$\begin{aligned}
\phi^\alpha(\rho^0) - \phi^\alpha(\rho^1) &\leq \tilde{z}(y^1, \rho^0) - \tilde{z}(y^1, \rho^1) \\
&\leq D^2\|\rho^1 - \rho^0\|.
\end{aligned}$$

The last inequality follows since the function $\rho^T f$ is Lipschitz continuous with respect to both ρ and f. Here, D^1, D^2 denote Lipschitz constants. Then,

$$\|y^1 - y^0\|^2 \leq \frac{D^1 + D^2}{\zeta}\|\rho^1 - \rho^0\|.$$

So,

$$\|y^1 - y^0\| \le \theta \|\rho^1 - \rho^0\|^{\frac{1}{2}},$$

by defining

$$\theta = \left(\frac{D^1 + D^2}{\zeta}\right)^{\frac{1}{2}}.$$

□

The continuity of the feasible set $\tilde{\mathcal{F}}^\alpha(\epsilon)$ follows directly. From Theorem 8.3.2 and Theorem 8.3.3 we conclude the following.

Corollary 9.1.1 ($\tilde{\mathcal{F}}^\alpha(\epsilon), v^\alpha(\epsilon)$ continuous). *For any $\alpha > 0$, $\tilde{\mathcal{F}}^\alpha(\epsilon)$, $v^\alpha(\epsilon)$ are continuous relative to $\epsilon \ge 0$.*

If we assume affine linear latencies, i.e. Assumption 2 holds, then we can strengthen our result.

Corollary 9.1.2 ($\tilde{\mathcal{F}}^\alpha(\epsilon), v^\alpha(\epsilon)$ Lipschitz continuous). *Let Assumption 2 hold. For any $\alpha > 0$, $\tilde{\mathcal{F}}^\alpha(\epsilon)$ and $v^\alpha(\epsilon)$ are Lipschitz continuous relative to $\epsilon \ge 0$.*

We proposed a regularized problem to overcome the difficulties of the lower-level problem. The objective function of the lower-level problem is strictly convex in (f, x), assuring a unique result $(f, x) \in \psi(\rho)$ for each $0 \le \rho \le \epsilon$. However, the solution of this lower problem is not necessarily a BRUE flow distribution as in (4.1.1).

9.2 Descent Algorithm

The mathematical optimization problem for the Best-case BRUE flow distribution requires complex and computationally expensive algorithms. Due to this complexity, these approaches are less appropriate for larger network instances. Therefore, this section proposes descent method to approximate an optimal solution of the regularized bilevel problem. We use the necessary optimality conditions of a bilevel convex optimization problem without upper-level constraints. Although there may exist more efficient approaches for this particular context, we consider this approach as it easily extends to the context of e.g. second-best toll pricing (11.2.1). Before we continue, let us rewrite the problem in $(q^\alpha(\rho))$ to reduce the number of variables. Here, $\mathcal{F}_0^f$ is the projection of $\mathcal{F}_0$ onto the f-space.

$$\min_f \sum_{e \in E} \int_0^{(\Gamma f)_e} l_e(\tau) d\tau - \rho^T f + \alpha \|f\|^2 \qquad \text{s.t.} \qquad f \in \mathcal{F}_0^f = \left\{ f \;\middle|\; \Lambda f = d, f \ge 0 \right\}. \tag{9.2.1}$$

Based on this formulation, we make some necessary observations:

- For any ρ, the lower-level solution exists and is unique with respect to (f, x).

- It is easy to show that the *Linear Independence constraint qualification* (LICQ) holds. For an arbitrary solution f of (9.2.1) consider for each $k \in \mathcal{K}$,

$$J_k(f) = \{p \in \mathcal{P}_k \mid f_p = 0\}.$$

  Then, the vectors

$$\Lambda_{k,.}, \qquad \nabla_f f_p, \ p \in J_k,$$

  are linearly independent for any $k \in \mathcal{K}$ since there exists for each $k \in \mathcal{K}$ a path $p \in \mathcal{P}_k$ for which $f_p > 0$ holds. The active constraint gradients are linearly independent and the KKT multipliers of the lower-level problem are unique.

- The second-order sufficienct optimality conditions of the lower-level problem hold for each $0 \leq \rho \leq \epsilon$ by strict convexity of the objective function in the lower level (see Theorem 9.1.3).

For the sake of our method, based on Friesz et al. (1990), we reduce the regularized bilevel problem to a bilevel problem in which no upper-level constraints are present:

$$\begin{array}{ll} \min_{(\rho,f)} \tilde{s}(f) := (\Gamma f)^T l(\Gamma f) & \\ \qquad\qquad \text{s.t.} & f \text{ solves} \\ & \min_f \sum_{e\in E} \int_0^{(\Gamma f)_e} l_e(\tau)d\tau - \rho^T f + \alpha\|f\|^2 \\ & f \in \mathcal{F}_0^f \end{array} \tag{9.2.2}$$

We try to set up a descent method for computing a local minimum of problem (9.2.2) as follows. Suppose at (ρ^0, f^0), with solution $f^0 = f(\rho^0)$ of the lower-level problem in (9.2.2), we are able to compute the directional derivative of $f(\rho)$ into direction d (assuming this limit exists):

$$w(f^0; d) = \lim_{t\to 0^+} \frac{f(\rho^0 + td) - f(\rho^0)}{t}. \tag{9.2.3}$$

We obtain using (9.2.3):

$$\begin{aligned} \tilde{s}(f(\rho^0 + td)) - \tilde{s}(f(\rho^0)) &= \nabla_f \tilde{s}(f^0)(f(\rho^0 + td) - f(\rho^0)) \\ &= t\nabla_f \tilde{s}(f^0) w(f^0; d). \end{aligned}$$

So, any direction d that satisfies

$$\nabla_f \tilde{s}(f^0) w(f^0; d) < 0,$$

yields a descent direction d for the upper-level problem.

We use the following theorem, which was presented in a more general form in Savard and Gauvin (1994).

Theorem 9.2.1 (Necessary optimality conditions bilevel problem). *Let (ρ^0, f^0) be an optimal solution for (9.2.2). Then for any upper-level direction $d \in \mathbb{R}^{|\mathcal{P}|}$ at ρ^0, the directional derivative of the objective function of the upper-level problem satisfies*

$$\tilde{s}'(f^0; d) = \nabla_f \tilde{s}(f^0) w(\rho^0; d) \geq 0.$$

The descent direction $w(\rho^0, d)$ of $f(\rho)$ at ρ^0 in direction d is the optimal solution for $\rho = \rho^0$ of the quadratic program ($f^0 = f(\rho^0)$):

$$\begin{aligned} \min_w \quad & (d^T, w^T) \nabla^2_{\rho,f} \mathcal{L}(\rho, f)(d, w) \\ s.t. \quad & \Lambda w = 0 \\ & w_p \geq 0 \qquad \text{for all } p : f_p^0 = 0 \\ & \left(\Gamma^T l(\Gamma f) - \rho + 2\alpha f\right)^T w = 0 \end{aligned} \tag{9.2.4}$$

To prove that the directional derivative equals the optimal value corresponding to the solution of quadratic program (9.2.4) we refer to e.g. Gauvin and Janin (1988).

The necessary optimality conditions gives us a straightforward approach to find a descent direction. Indeed, as long as there exists a solution (d^0, w^0) of the linear-quadratic bilevel problem:

$$\min_{(d,w)} \nabla_f \tilde{s}(f^0) w \qquad \text{s.t.} \qquad (w, d) \in \mathcal{G} = \left\{ (w, d) \;\middle|\; \begin{array}{ll} -1 \leq d \leq 1 & \\ w \text{ solves (9.2.4)} & \\ d_p \leq 0 & \rho_p^0 = \epsilon_k, p \in \mathcal{P}_k, k \in \mathcal{K} \\ d_p \geq 0 & \rho_p^0 = 0, p \in \mathcal{P} \end{array} \right\}, \tag{9.2.5}$$

for which the optimal value is negative, we can move in a direction to lower the total travel time. We emphasize that the linear-quadratic problem (9.2.5) reduces to a linear problem with complementarity constraints if we consider the single-level reformulation of (9.2.5). That is, problem (9.2.5) in which we replace lower-level problem (9.2.4) with its KKT optimality conditions. We apply a simple branch-and-bound technique to solve this problem (see Bard (1998)).

Let us assume that we find at iteration i a direction vector d^i and step size Δ^i, then ρ^{i+1} is updated out of ρ^i by $\rho^{i+1} = \rho^i + \Delta^i d^i$. Nevertheless, in the above analysis we considered the problem in which no upper-level constraints are present. Since $0 \leq \rho \leq \epsilon$ assures that the solution of the problem is a BRUE flow as in (4.1.1) for a fixed ϵ, we need to make sure that in our algorithm the upper-level constraints are met. Therefore, we update ρ^{i+1} via

$$\rho_p^{i+1} = \min\{\max\{0, \rho_p^i + \Delta_p^i d_p^i\}, \epsilon_k\} \tag{9.2.6}$$

for each $p \in \mathcal{P}_k$, $k \in \mathcal{K}$, as suggested by Friesz et al. (1990).

Descent Algorithm

Step 1. Determine initial value $0 \leq \rho^0 \leq \epsilon$ and $\delta > 0$, set $i = 0$.

Step 2. Solve (9.2.1) and obtain f^i. Identify path set $P^i = \{p \in \mathcal{P} \mid f_p^i = 0\}$, and solve (9.2.5) using P^i to find descent direction vector d^i.

Step 3. Calculate stepsize Δ^i.

Step 4. Determine ρ^{i+1} using (9.2.6).

Step 5. If $\|\rho^{i+1} - \rho^i\| < \delta$, then terminate, otherwise let $i = i + 1$ and go to Step 2.

Let us take a look at Step 3. Although we can apply a basic line search procedure, we use a predetermined convergent step size. Hence,

$$\Delta^i = \frac{\Lambda^T \epsilon}{i}$$

for all $i \in \mathbb{N}$.

We emphasize that this descent method is beneficial since during the steps of the algorithm only convex problems should be solved. Indeed, at Step 2 we first solve the convex lower-level problem and then we solve a linear-quadratic bilevel problem. A branch-and-bound algorithm globally solves the linear-quadratic problem. This branch-and-bound algorithm only requires to solve linear programs. However, the number of linear programs to be solved can be exponential with respect to $|\mathcal{P}|$. For numerical issues, we need to implement a tolerance value to check whether the complementarity condition is satisfied. We consider $ab \leq 1 \times 10^{-3}$ to be zero, for $a, b \in \mathbb{R}_+$.

9.3 Synthesis

We proposed a regularized bilevel problem in which the lower-level objective function is perturbed to assure that the lower-level response is unique. Moreover, this regularized problem has a continuous feasible set and optimal value function with respect to perturbations in the indifference band ϵ. Based on the regularized bilevel problem we proposed a descent method to approximate the Best-case BRUE flow distribution. The next chapter applies this descent method on a test network.

Chapter 10

Numerical Experiments

Appraisal of transport policies is a major concern for transportation engineers (Sheffi, 1985) and requires accurate prediction of traffic flows. In the previous chapters we discussed the behavior of the Best-case BRUE problem $(Q(\epsilon))$ under perturbations in the indifference band ϵ and we proposed a descent method in approximating the Best-case BRUE flow distribution. This chapter reports on some numerical experiments in a test network. First, we propose a relaxation algorithm to find the best-performing BRUE flow distribution with respect to the total travel time.

10.1 A Relaxation Algorithm

Let us once again take a look at the complementarity reformulation of $(Q(\epsilon))$:

$$\begin{aligned} \min s(x) & \\ & -c(x) + \Lambda^T \lambda \leq 0 \\ & f_p(c_p(x) - \lambda_k - \epsilon_k) \leq 0 \qquad \text{for all } p \in \mathcal{P}_k, k \in \mathcal{K} \\ & (f, x) \in \mathcal{F}_0 \end{aligned} \tag{10.1.1}$$

Difficulty is that (10.1.1) not necessarily satisfies a constraint qualification. We showed that we can overcome this issue by identifying path set $P \subseteq \mathcal{P}$ beforehand (Section 5.5). If we assume that affine linear latency functions are present in a network (i.e. Assumption 2 holds), the problem (10.1.1) is relatively straightforward to solve. The course of action is, for a fixed $\epsilon \geq 0$, to evaluate $(Q_P(\epsilon))$ for all possible choices of $P \subseteq \mathcal{P}$. For this particular problem, we consider two approaches in identifying the proper path set P so that the convex problem $(Q_P(\epsilon))$ finds the global optimum of $(Q(\epsilon))$.

- The *two-stage* approach. This approach consists of two steps: choice set generation and traffic assignment. First, we generate a choice set and then we assign traffic over these possibly used paths in a way such that the total travel time is minimized. That is, generate for a given $\epsilon \geq 0$ a family of path set $\tilde{\mathcal{P}}(\epsilon) = \{P \subseteq \mathcal{P} \mid \epsilon \in \mathrm{dom}\,(\mathcal{F}_P)\}$

and then solve $(Q_P(\epsilon))$ for all $P \in \tilde{\mathcal{P}}(\epsilon)$. Such an approach was applied by Di et al. (2016). They constructed this family by assuming that the path set is *monotonically non-decreasing*: given $f^1 \in \mathcal{F}(\epsilon^1)$ then for any $f^0 \in \mathcal{F}(\epsilon^0)$, $\epsilon^1 > \epsilon^0$, we have that $P^0 \subseteq P^1$. Here,

$$P^0 = \{p \in \mathcal{P} \mid c_p(f^0) \leq \min_{q\in\mathcal{P}} c_q(f^0)+\epsilon\}, \qquad \text{and} \qquad P^1 = \{p \in \mathcal{P} \mid c_p(f^1) \leq \min_{q\in\mathcal{P}} c_q(f^1)+\epsilon\}.$$

We show via a counterexample that this claim is not necessarily true:

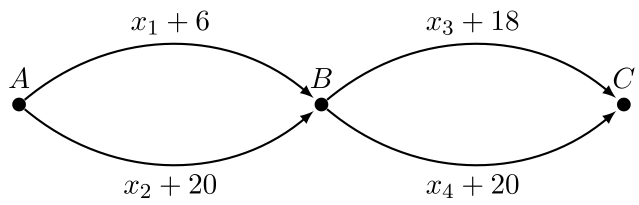


Figure 10.1.1: Example Network.

Example 10.1.1. *Consider the network with a single commodity (A, C) in Figure 10.1.1. Demand $d = 20$. We solve for the PRUE which leads to the PRUE link flow x^n:*

$$x^n = (17, 3, 11, 9).$$

The route cost is $c_p(x^n) = 52$ for all $p \in \mathcal{P}$. From the route cost vector $c(x^n)$, we identify the used paths:

$$P^n = \{a, b, c, d\},$$

in which

$$a = \{1, 3\}, \qquad b = \{1, 4\}, \qquad c = \{2, 3\}, \qquad d = \{2, 4\}.$$

In other words, we can find a route flow distribution in which all the paths carry flow. If we solve for a BRUE for $\epsilon^0 = 6.5$, we get the link flow vector

$$x^0 = (14, 6, 10.5, 9.5),$$

with corresponding path cost vector

$$c(x^0) = (48.5, 49.5, 54.5, 55.5).$$

The used paths are then

$$P^0 = \{a, b, c\}.$$

Obviously, $P^0 \subset P^n$.

- The *enumeration* or *branch-and-bound* approach. We consider the following program, given $P, R \subseteq \mathcal{P}$, for given $\epsilon \geq 0$:

$$\min_{(f,x,\lambda)} s(x) \qquad \text{s.t.} \qquad (f,x,\lambda) \in \mathcal{F}_{(P,R)}(\epsilon), \tag{10.1.2}$$

in which $\mathcal{F}_{(P,R)}(\epsilon)$ is defined as

$$\mathcal{F}_{(P,R)}(\epsilon) = \left\{ (f,x,\lambda) \;\middle|\; \begin{array}{ll} (f,x) \in \mathcal{F}_0 & \\ c_p(x) - \lambda_k - \epsilon_k \leq 0 & \text{for all } p \in \mathcal{P}_k, k \in \mathcal{K} \\ (f,x,\lambda) \in \mathcal{F}_l & \\ f_R = 0 & \end{array} \right\}.$$

This method starts by solving for the system optimum (Q_s), i.e. solve (10.1.2) with $P = R = \emptyset$. If the optimal solution to this problem happens to satisfy the complementarity condition in (4.1.1) then we are done. Otherwise we have at least one path $p \in \mathcal{P}_k, k \in \mathcal{K}$ for which $f_p(c_p(x) - \lambda_k - \epsilon_k) > 0$. We add two subproblems. One subproblem has $P = P \cup \{p\}$, $R = R$, the other subproblem considers $R = R \cup \{p\}$ and $P = P$. We solve (10.1.2) for the updated P, R. We continue to add subproblems until the solution of (10.1.2) satisfies (4.1.1) for all $p \in \mathcal{P}_k, k \in \mathcal{K}$. If we find a feasible BRUE flow, we compare the corresponding objective value with the best BRUE solution thus far. We update the optimal value function or discard our new BRUE flow. If the objective function is worse, we discard a complete subtree. We also discard a subtree if the objective value of a solution (not necessarily satisfying (4.1.1)) is worse than the best solution satisfying (4.1.1) thus far. We compare and update the optimal value function and continue until we find the global optimal solution. This algorithm terminates in a finite number of iterations with the global optimum of $(Q(\epsilon))$.

It is well-known that a branch-and-bound approach in a general setting possibly ends up evaluating $2^{|\mathcal{P}|}$ convex problems. Some preliminary tests with this algorithm showed that also in practice numerous convex optimization problems should be solved in our context.

Remark 10.1.1. *The two-stage approach is solely applicable under strong conditions. If the path set is monotonically non-decreasing under Assumption 2 we solve a sequence of integer programs to find all $P \subseteq \mathcal{P}$ for which $\epsilon \in dom(\mathcal{F}_P)$. If we consider parallel-link networks under the same assumption, we can efficiently find the acceptable path set. Hence, in this particular case, we find the Best-case BRUE in polynomial time.*

Under more general latency functions, $l(x)$ might be highly nonlinear. This makes problem (10.1.1) an optimization problem with a set of nonlinear constraints. The enumeration approach, as discussed, possibly evaluates a combinatorial amount of choices of

P. If $l(x)$ is nonlinear, it solves numerous nonlinear optimization problems which leads to high computation times. We provide an alternative approach based on Ban et al. (2006).

Let us introduce parameter vector $0 < \alpha \in \mathbb{R}_+^{|P|}$ and define the perturbed complementarity constraints

$$f_p(c_p(x) - \lambda_k - \epsilon_k) \leq \alpha_p \qquad \text{and} \qquad -c_p(x) + \lambda_k \leq 0, \tag{10.1.3}$$

for all $p \in \mathcal{P}_k, k \in \mathcal{K}$. A similar analysis as in Section 5.2 shows that for any $\alpha > 0$ the system in (10.1.3) satisfies a constraint qualification for any feasible point in the feasible set (Figure 10.1.2).

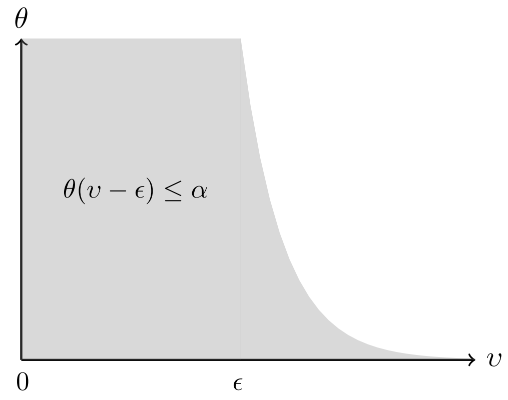


Figure 10.1.2: Abstract one-dimensional illustration of the feasible set corresponding to the system of inequalities (10.1.3) for abitray $\epsilon > 0$ and $\alpha > 0$. Here, $\theta = f_p$, $\upsilon = c_p(x) - \lambda_k$.

The system in (10.1.1) becomes for $\alpha \geq 0$:

$$\begin{aligned} \min s(x) & \\ & -c(x) + \Lambda^T \lambda \leq 0 \\ & f_p(c_p(x) - \lambda_k - \epsilon_k) \leq \alpha_p, \qquad \text{for all } p \in \mathcal{P}_k, k \in \mathcal{K} \\ & (f, x) \in \mathcal{F}_0 \end{aligned} \tag{10.1.4}$$

Then, letting $\alpha \to 0$ will provide a solution to $(Q(\epsilon))$. The problem in (10.1.4) is still a nonlinear optimization problem, but there are ample approaches to solve these type of problems, e.g. the Sequential Quadratic Programming (SQP) algorithm.

Optimization software requires an initial choice (f^0, x^0). Therefore, α reduces every iteration and we use the solution of the previous iteration as input for the new iteration. Overall, the algorithm to solve $(Q(\epsilon))$ for local minimizers is similar to the one presented in Ban et al. (2006) and is as follows:

Relaxation Algorithm

Step 1. Choose initial parameter vector $\alpha^0 > 0$. Set iteration limit M and update factor

$0 < \sigma < 1$. Set $i = 0$.

Step 2. Solve (10.1.4) using α^i. If $i \leq M$, set $\alpha^{i+1} = \sigma\alpha^i$. Moreover, set $i = i + 1$ and repeat Step 2. Otherwise, go to Step 3.

Step 3. Solve the exact problem $(Q(\epsilon))$, if it is successful, we obtain an exact solution. Otherwise, an approximate solution is achieved.

The algorithm requires to solve $M + 2$ nonlinear programs. Step 2 is performed $M + 1$ times while we need an additional problem in Step 3. We overcome the combinatorial curse of $(Q(\epsilon))$ and we solve a predetermined number of nonlinear programs. The time needed to perform such a nonlinear program highly depends on the network instance and mainly determines the actual CPU time of the algorithm.

The proposed relaxation algorithm differs from the descent approach as follows. The relaxation approach can be seen as a dual approach while the descent approach is a primal method. In the descent algorithm we start with a feasible solution and seek for an improvement while remaining feasible, in the relaxation algorithm, on the other hand, we start with an infeasible solution and tighten the constraints so that we find a solution which is feasible for $(Q(\epsilon))$.

This algorithm was originally constructed by Ban et al. (2006) to solve the Continuous Network Design Problem (CNDP). In fact, frequent algorithms meant for the CNDP also apply to the Best-case BRUE problem.

Remark 10.1.2. *In all the discussed algorithms, we assume that we know the matrix* Γ *beforehand. The formulation in* $(\tilde{Q}(\epsilon))$ *requires this matrix since we define* ρ *per path in the path set. We stress that it might be difficult to construct the whole path set beforehand. Therefore, Lou et al. (2010) proposed a link-based definition of the BRUE flow distribution. This definition is more restrictive since for paths that only consist of links with positive flows, the link-based BRUE definition requires its travel time to be within the indifference band, even when the path carries no flow (Lou et al., 2010). Another possible approach is to construct a subset of the complete path set via e.g. the k-shortest path algorithm. Although this approach is more restrictive, it is not necessarily unrealistic. Large detours that connect a commodity are not included in this set but are also not likely to be used in practice.*

10.2 Network Instances

We study the impact of a varying ϵ on the Best-case BRUE link flow distribution. Therefore, we adopt a test network. Although the test network cannot be considered realistic, it suits the purpose of this section to compare performance of the provided algorithms.

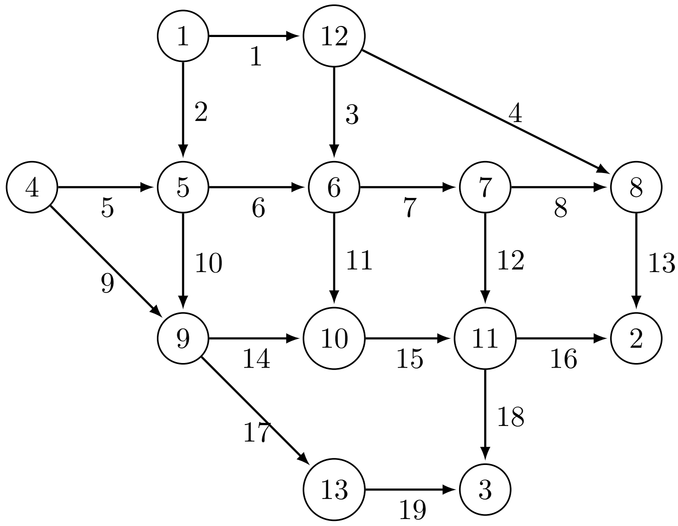


Figure 10.2.1: Nguyen & Dupuis Network

Nguyen & Dupuis Network

Figure 10.2.1 depicts the network which was introduced by Nguyen and Dupuis (1984). The network has four origin-destination pairs, namely $(1, 2)$, $(1, 3)$, $(4, 2)$ and $(4, 3)$. We identified all possible paths ($|\mathcal{P}| = 25$) for these commodities. The settings for this network were taken from Ohazulike et al. (2013).

Table 10.2.1 lists the demand for each commodity.

Table 10.2.1: Demand for the Nguyen & Dupuis Network

origin	destination	demand
1	2	200
1	3	400
4	2	200
4	3	150

We adopted the BPR travel time function for all edges in the network:

$$l_e(x_e) = \kappa t_e \left(1 + 0.15 \left(\frac{x_e}{c_e}\right)\right)^4,$$

in which t_e is the free flow travel time on link e, c_e is the capacity of the edge e and κ is the value of time. As mentioned, the BPR function satisfies Assumption 1.

We adopted some basic parameter settings for both algorithms (Table 10.2.2).

For both methods, we used the MATLAB software tool for implementation. In particular, for any nonlinear program we used the SQP algorithm to find an optimal solution. The interior-point algorithm (with termination tolerance 1×10^{-5}) is used to solve a linear problem in the descent method. In general, we used the basic settings for all these

Table 10.2.2: Algorithm settings

	Δ	ρ^0	α^0	M	δ
Descent	$\Lambda^T \epsilon / i$	0	1×10^{-3}	-	1×10^{-2}
Relaxation	0.1	-	10	10	-

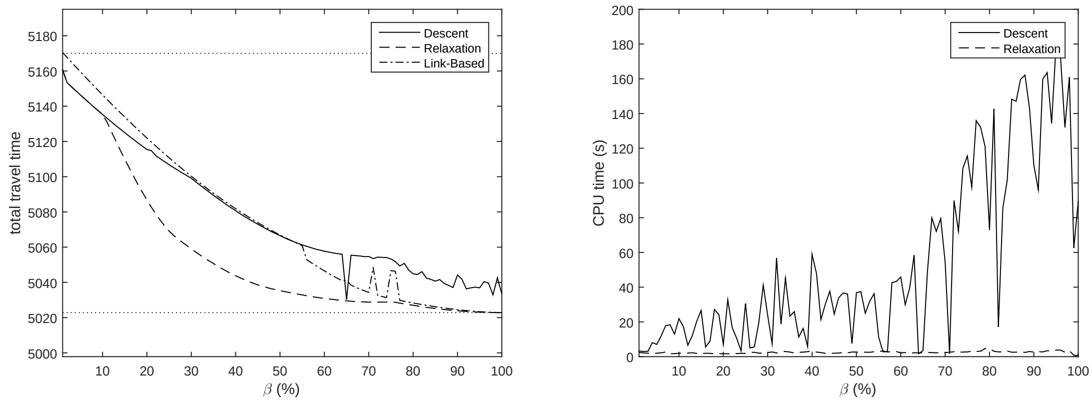
algorithms. Nevertheless, preliminary tests showed that we had to manually increase MATLAB's default value for the maximum number of iterations allowed. Furthermore, we introduced a numerical lower and upper bound on w in problem (9.2.5) and multiplied the constraints and objective function by factor 10. Both algorithms are tested on the same computer. We applied the manifold suboptimization algorithm (Lawphongpanich and Yin, 2010) to solve the Best-case link-based BRUE flow distribution as articulated by Lou et al. (2010). We used the SNOPT solver (Gill et al., 2005) in AIMMS to solve this problem. As initial solution for this algorithm we used the PRUE flow distribution. We note that the manifold algorithm requires to solve nonlinear programs. Moreover, Lou et al. (2010) indicated that only a subset of the BRUE flow distributions is considered in this setting. On the other hand, the link-based BRUE formulation does not require matrix Γ (Remark 10.1.2).

10.3 Performance of the Algorithms

Let us compare both algorithms with respect to the outcome (i.e. total travel time of the BRUE flow distribution) and CPU time. We tested the network instances for $\beta\epsilon^0$ in which $\beta \in \{0.01, 0.02, \ldots, 1\}$ and ϵ_k^0 is the maximum travel time difference between two routes for each commodity $k \in \mathcal{K}$ in the system-optimal flow distribution.

Figure 10.3.1a depicts the system objective corresponding to the flow distributions of the relaxation algorithm and the descent approach. It also shows performance with respect to total travel time of the link-based BRUE. The upper vertical line in Figure 10.3.1a gives the total travel time of the PRUE flow while the bottom line corresponds to the global performance for the system-optimal flow. Trivially, the total travel time is bounded from below and above by the travel time attached to the system-optimal and PRUE flow respectively (*cf.* Lou et al. (2010)).

Figure 10.3.1a shows that the relaxation method outperforms the descent approach with respect to global performance. Indeed, for any choice of β, the total travel time of the flow found by the relaxation method is less than or equal to the performance of the flow distribution found by the descent method. Moreover, the relaxation approach attains the system optimum for $\beta = 1$, which is line with the behavior of the Best-case BRUE problem (see Section 5.3). For small values of β, the descent method finds a BRUE flow distribution with equal travel time.



(a) System performance

(b) CPU time

Figure 10.3.1: Algorithm performance in Nguyen & Dupuis Network.

A possible explanation for the inferior performance of the descent method is as follows. The descent approach is based on the necessary optimality conditions of a relaxed problem. Thus even for this relaxed problem only a local minimum can be found. We note that due to the complexity of the problem, we cannot check whether a Best-case BRUE is actually obtained.

We see that our relaxation method outperforms the link-based BRUE: for all values of β, we find a BRUE flow distribution with less travel time. However, for large values of β (i.e. $\beta \geq 0.8$), the link-based BRUE approaches the global performance of the relaxation algorithm. When comparing the descent method with the link-based BRUE, we see that the descent method performs better than the link-based approach for small values of β.

Figure 10.3.1b analyzes the CPU time for both approach and shows that the relaxation approach outperforms the descent method. These results are somewhat surprising since the relaxation method requires nonlinear programs to be solved. However, the optimization tool handles these programs quite efficiently. Moreover, for each iteration, the descent method requires to solve a branch-and-bound scheme. This might lead to solving an exponential number of linear programs with respect to the number of paths in the network. The CPU time for the relaxation method is bounded from above by 5 seconds, while the descent approach might require almost 200 seconds to find a flow distribution.

We discuss some apparent anomalies in Figure 10.3.1a. We see that for some values of β, the total travel time corresponding to the descent approach worsens compared to the performance for a smaller indifference band. This behavior is due to some numerical issues in the optimization tool. We emphasize that the relaxation method was able to solve the exact problem for any β except $\beta = 0.26, 0.34, 0.35, 0.95, 0.97$. In these cases, it terminated with an approximate solution.

All in all, we conclude that the relaxation algorithm outperforms both the descent method as the link-based BRUE approach. Nonetheless, we can improve performance of the latter two approaches. Therefore we should use a different initial solution or evaluate the same instance for multiple initial solutions. Moreover, the CPU time of the descent approach improves if we look for any descent direction rather than the steepest.

10.4 Policy Impact

The purpose of this section is twofold. First, we study the implications of adopting a bounded rationality traffic assignment for scenario evaluation. We answer the question whether implementation of the (Best-case) BRUE assignment is different from the PRUE assignment whereas other bottlenecks may be identified. Since the BRUE flow distribution adds complexity to the traffic assignment problem, a negligible difference in flows implies that the adoption of a BRUE assignment is unnecessary complex. Second, we study the implications for the NDP. If the Best-case BRUE substantially improves performance compared to the PRUE, it might be beneficial to steer the network for a given ϵ towards the Best-case BRUE. Whereas the global performance of the PRUE flow gives a lower bound on performance of the Worst-case BRUE, our analysis here provides a lower bound on the maximum improvement that is obtained by steering to the Best-case BRUE. From the assumptions of bounded rational choice behavior, the Best-case BRUE improves overall performance while no individual has the incentive to change routes.

We assume that the indifference band is fixed and does not differ among commodities. Furthermore, ϵ ranges from 0.1 to 0.6. We implement this setting to capture a more realistic setting than in the previous section, which solely had the purpose to compare computational performance of the algorithms. For comparison, we also evaluate the outcomes of the PRUE and the system-optimal flow.

Table 10.4.1: Nguyen & Dupuis network results for different values of ϵ.

ϵ	Obj.		RMSE	MaxAbs	Paths
0	5173				8
0.1	5112	(-1.18%)	32.35	72.31	8
0.2	5068	(-2.04%)	50.96	112.52	8
0.3	5036	(-2.65%)	65.90	145.07	8
0.4	5029	(-2.79%)	71.77	142.86	6
0.5	5023	(-2.90%)	74.83	169.39	8
0.6	5023	(-2.91%)	77.00	175.49	10
ϵ^{so}	5023	(-2.91%)	76.99	175.49	10

Table 10.4.1 lists the results, generated using the relaxation algorithm. In addition, we used the by-product route flow distribution. We report on differences with respect to the *objective value* (total travel time of the solution), the *maximum absolute difference* (the maximum absolute difference in link flow), *root mean square error* (RMSE) and the

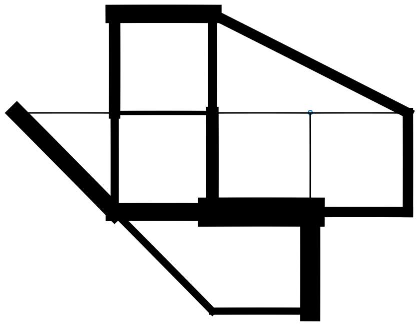
number of paths carrying flow. The RMSE gives the expected difference in link flow for an arbitrary edge. For policy evaluation, we discuss the differences between the link flow distributions. Large differences between the PRUE and BRUE assignment with respect to the link flow imply that new or different bottlenecks might be identified for the same scenario under different behavioral assumptions.

The results show that boundedly rational behavior of users leads to a different traffic assignment. By that we mean that we observe large differences in the link flow solution. Even for $\epsilon = 0.1$, we expect a link flow difference on an arbitrary link of more than 30 (total demand is 950), with a maximum absolute difference over 70. Hence, traffic distributes differently over the network. If we further increase the indifference band ϵ, we observe even more extreme results. Figure 10.4.2 plots the flow distributions and shows that traffic distributes more evenly over the network. We emphasize that for $\epsilon = 0.2$ the algorithm terminated with an approximate solution.

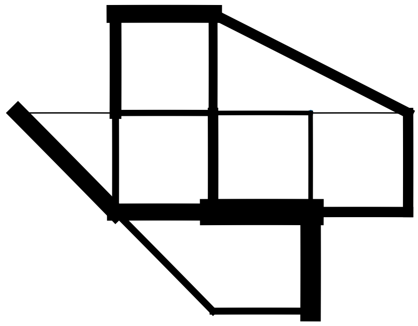
We conclude that traffic management measures which steer the network towards the Best-case BRUE flow distribution are powerful to minimize congestion and thereby preserve fairness. Assume ϵ to be fixed and assume an arbitrary BRUE flow distribution is realized in practice. Table 10.4.1 shows that even for small values of ϵ, the Best-case distribution leads to a major improvement with respect to the total transit time. For example, assuming that users have an indifference band of $\epsilon = 0.1$, more than 30% of the benefit of the system optimum with respect to total travel time is obtained by the Best-case BRUE distribution. For $\epsilon = 0.2$, this percentage further increases to 70%.

It follows that the indifference vector ϵ should be a well-calibrated parameter. Earlier, we indicated that the optimal solution set is not necessarily continuous with respect to perturbations in ϵ. The numerical experiments confirm this claim. A small change in ϵ leads - from a behavioral perspective - to completely different route choices. Han et al. (2015) made the same claim in the dynamic setting.

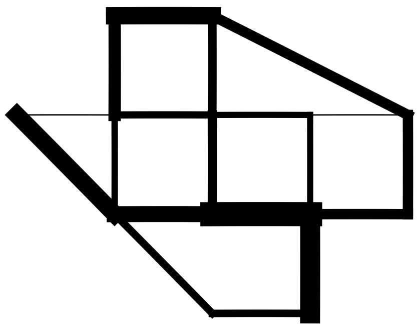
To reduce complexity of the problem $(Q(\epsilon))$, we suggest to identify for given ϵ, even under nonlinear latencies, an appropriate path set $P_k \subseteq \mathcal{P}_k$, $k \in \mathcal{K}$. Table 10.4.1 lists therefore the number of paths that carry flow for this instance. Hence, if we solve $(Q_P(\epsilon))$ under nonlinear latencies with the listed paths, we find the same flow distribution as the relaxation algorithm. We emphasize that, although not directly clear, an ad-hoc identification of P_k would not lead to this flow distribution. In other words: $P \neq P^n$ and $P \neq P^{so}$ for all tested $0 < \epsilon < \epsilon^{so}$. We need to further investigate whether this ad-hoc identification is a possible way to approximate the Best-case BRUE and to limit CPU time.



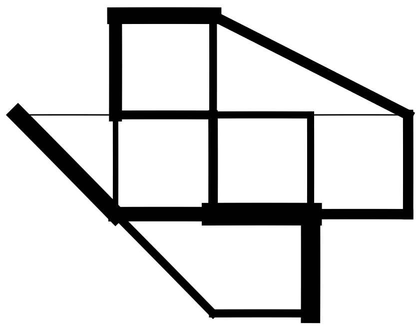
(a) $\epsilon = 0$



(b) $\epsilon = 0.1$



(c) $\epsilon = 0.2$



(d) $\epsilon = 0.3$

Figure 10.4.1: Flow distributions in Nguyen & Dupuis Network under different ϵ.

10.4.1 Route Guidance with User Constraints

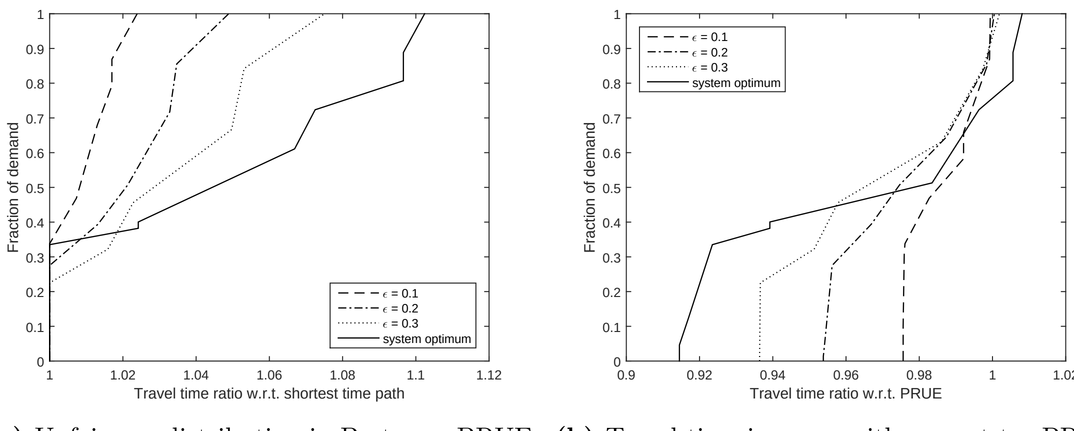
We recap the setting of the route guidance system with user constraints as introduced in Section 4.3. Route guidance systems often face unfairness issues, which makes implementation difficult in practice. We therefore consider these issues for the Best-case BRUE flow distribution. That is, we assume that an authority implements a route guidance system which steers the network towards this BRUE flow. The numerical results for the Nguyen & Dupuis network imply that, assuming a PRUE to be realized in absence of coordination, a route guidance measure with user constraints is beneficial.

Figure 10.4.2a shows the cumulative distribution of travel time among travelers (unfairness) in a Best-case BRUE flow distribution for varying ϵ. These results show that in our instance a Best-case BRUE flow steers at least 20% of all travelers along the shortest time path. A perfect rationality perspective assumes that the remaining demand has an incentive to switch routes. On the other hand, the bounded rationality perspective says that the travel time improvement is only marginal and therefore induces no switch. In the system-optimal flow distribution, some travelers experience 10% longer travel times compared to other users. This percentage reduces to 2.5%, 5%, 7.5% for $\epsilon = 0.1, 0.2$ and 0.3 respectively.

Figure 10.4.2b compares the transit time of users in a route guidance setting to the travel time in a PRUE. We observe that some travelers face an increase in travel time in order to achieve a system optimum. In fact 20% of all users experience more delay compared to the delay in the PRUE. In practice, drivers may not notice this increase. In other words, they are indifferent to this change (Vreeswijk et al., 2013a). On the other hand, 10% of the demand ends up on a route which is 9% faster.

These preliminary results show that a route guidance system which steers the network to a Best-case BRUE flow distribution is promising. That is: we can improve travel time, while we limit the unfairness and the additional experienced delay. For this particular instance, the Best-case BRUE for $\epsilon = 0.1$ is the flow distribution to aim for. Indeed, no traveler experiences additional delay, a route switch only marginally improves the transit time (at most 2.5%) and the distribution achieves 30% of the benefit of the system optimum with respect to travel time.

We emphasize that the experiment should be repeated in more realistic networks. For instance, the travel time increase in a system optimum compared to a PRUE was shown to be 25% in a realistic network considered by Jahn et al. (2005). Furthermore, the travel time ratio with respect to the shortest time path was in this instance of Jahn et al. (2005) more than 2. A future analysis should also include the distribution of impacts among OD-pairs.



(a) Unfairness distribution in Best-case BRUE **(b)** Travel time increase with respect to. PRUE

Figure 10.4.2: Unfairness distributions in Nguyen & Dupuis Network under different ϵ.

10.5 Synthesis

In order to incorporate the BRUE in the NDP, we proposed in previous chapter a descent method in approximating the Best-case BRUE flow distribution. In this chapter, we proposed a relaxation algorithm, which solves a sequence of nonlinear programs to find this best-performing flow. Both algorithms were tested in a numerical example and compared based on outcome and CPU time. For both aspects, the relaxation algorithm outperforms the descent method.

Furthermore, we investigated the implications for policy evaluation in the bounded rationality traffic assignment. In the test network, the BRUE differs substantially from the PRUE and system-optimal flow distributions. Lastly, we found that steering the network towards the Best-case BRUE may lead to major benefits in terms of total travel time. Indeed, 30% of the benefit of the system-optimal assignment was obtained by the Best-case BRUE for a small indifference band.

Chapter 11

Conclusions

This study incorporated a behavioral view on route choice decision making in a traffic assignment model to increase the validity of the Network Design Problem. The objective of the research was *to develop and analyze a static traffic assignment model that incorporates bounded rationality.*

We draw the following main conclusions:

- *(Choice Behavior and the Network Design Problem).* The adoption of boundedly rational decision making of travelers in the static traffic assignment leads to a non-unique lower-level response in the NDP. It adds a new dimension to the NDP since it is unknown which lower-level flow is most likely to be realized in practice. In order to cope with the uncertainty, we recommend to consider the best and worst-performing flow distribution under bounded rationality.

- *(Mathematical Analysis).* We formulated a bilevel optimization problem for which the solutions correspond to the Best-case BRUE conditions. This optimization problem is a mathematical program with complementarity constraints, which are difficult to solve numerically and different local minimizers may co-exist. As a solution, we offered an approach that replaces the problem by a union of easy to solve problems. Parametric analysis showed that the problem lacks favorable continuity properties: only a strong and unnatural condition guarantees continuity of the feasible set and optimal value function with respect to perturbations in the indifference band.

- *(Usability for Scenario Analysis).* Due to the complex mathematical structure of the problem, we constructed an optimization problem that finds an approximate Best-case BRUE flow distribution. The feasible set and optimal value function of this problem are continuous with respect to perturbations in the indifference band. We provided a relaxation algorithm and descent approach to find this flow distribution. We recommend authorities to consider adopting these algorithms to evaluate scenarios: numerical experiments showed that the BRUE assignment substantially differs from the PRUE assignment.

We used a parametric optimization approach in our study. We elaborate on the conclusions.

Choice Behavior and the Network Design Problem

Traffic assignment models assume users to be perfectly rational. In other words, travelers evaluate all possible route alternatives and choose the minimum cost route. In the NDP, any intervention by an authority in the upper level leads to a unique perfectly rational response in the lower level. However, empirical studies suggest that the assumption of perfectly rational decision making is naive. In practice, users are boundedly rational and choose suboptimal routes. We incorporated the definition of bounded rationality from Mahmassani and Chang (1987) in the static traffic assignment: users solely change paths if the new route improves the travel time by more than the indifference band.

This behavioral aspect in the static traffic assignment has serious consequences for the NDP. Even in simple 2×2 matrix games difficulties arise: there is a whole set of BRUE flows that differ in system performance. This leads to an additional dimension in the lower level of the NDP and thereby induces uncertainty among upper-level performance. It unknown which lower-level flow is most likely to be realized in practice, and it becomes unclear which upper-level measure will lead to optimal performance of the system. Hence, we need to make an additional assumption on the lower-level response. We showed that this might lead to adverse effects: a pricing scheme designed under best-case assumptions can lead to an inferior performance of the system in practice.

Authorities should therefore focus on the extremes of possible network performances. That is: assess performance of BRUE flow distributions with best and worst-case performance with respect to the total travel time.

Mathematical Analysis

We formulated a mathematical optimization problem for which the optimality conditions correspond to the BRUE condition. We were particularly interested in finding the best-performing flow with respect to total travel time among all BRUE flows. It turned out that this best-performing flow is difficult to calculate since the feasible set is not convex and violates a regularity condition. Consequently, standard optimization tools do not apply.

We adopted a parametric optimization approach whereas the Best-case BRUE problem reduces to simple convex problems for certain values of the indifference band. Indeed, the problem is equivalent to the PRUE problem in case the indifference band is zero and reduces the system-optimal traffic assignment when the indifference band is sufficiently large.

To handle the non-convexity and violation of the regularity constraint, we proposed to decompose the problem - assuming affine linear latencies - into branches of optimization problems. These branches are easy to solve and the union of the feasible sets of all branches corresponds to the feasible set of the Best-case BRUE problem. Each branch behaves continuously with respect to perturbations in the indifference band. By that we mean that the feasible set, optimal value function and optimal solution set are continuous (multi-)functions with respect to perturbations in the band.

The Best-case BRUE problem, on the other hand, is not necessarily continuous. Even under affine linear link latencies, the feasible set, optimal value function and optimal solution set are generally not continuous relative to perturbations in the parameter. However, mild assumptions guarantee the existence of a BRUE flow.

A bilevel reformulation of the Best-case BRUE flow distribution problem allowed us to analyze the assignment under general latency functions. We showed that this bilevel problem lacks favorable continuity properties. The reformulation indicated that under the condition that the link flow uniquely determines the route flow, continuity of the set of BRUE flows and optimal value function is guaranteed. Nevertheless, this condition is very strong and does not occur regularly.

Usability for Scenario Analysis

The mathematical optimization problem for the Best-case BRUE flow distribution requires complex and computationally expensive algorithms. Due to this complexity, these approaches are less appropriate for large networks. We provided two algorithms that approach local minima of the Best-case BRUE problem in limited time. The relaxation algorithm approaches the Best-case BRUE by solving a sequence of nonlinear programs while the descent method solely uses convex and linear optimization problems to approach a minimum.

We conducted numerical experiments on a test network to indicate the usability of these algorithms for scenario analysis. The relaxation method outperforms the descent approach in terms of outcome with respect to total travel time and CPU time. Both methods allow integration in the NDP and authorities should consider to adopt these approaches.

A preliminary analysis showed the impact for the NDP. If a Best-case BRUE flow distribution is realized rather than the PRUE, the distribution leads to different results for scenario analysis: link flows and total travel time differ significantly. Furthermore, we showed that the Best-case BRUE flow distribution is a desired state for authorities since this flow limits unfairness among users and substantially improves the travel time.

11.1 Discussion

Some aspects of this research are subject for discussion.

We made some necessary assumptions. Assumption 1 forces the travel time function to be monotonically increasing and separable with respect to the link flow. Although this assumption is common in static traffic assignment literature, we recognize that this assumption does not mirror a realistic link cost function. Major drawback of the basic setting is that the flow on a link can exceed capacity and an excess of demand does not lead to a traffic queue. Simple extension of the model (by adding a linear constraint that does not allow the flow to exceed capacity) partially reduces this problem. Mathematically, this will not induce any new problems compared to the setting we adopted. In a realistic setting we suggest to adopt e.g. elastic demand and asymmetric link cost functions. Other basic extensions are likely to improve realism as well. For instance, we can integrate the BRUE assignment in a stochastic setting. Then, the travel time includes a random term. In this context, a BRUE assumes travelers to switch routes when they think they can improve travel time.

We made some basic assumptions with respect to the indifference band ϵ. As earlier mentioned, the indifference band is an individual-specific threshold (Vreeswijk et al., 2013b) and we prefer to consider this band as a stochastic variable. Also, the indifference band depends on the traffic situation, which suggests that the indifference band is endogenous. This setting is considered by Han et al. (2015). Both extensions increase complexity and were therefore not considered in our study.

The numerical experiments were conducted on the Nguyen & Dupuis network to indicate performance of the proposed algorithms. Obviously, this network is not realistic. We should further study the integration of the assignment model for scenario analysis and the algorithms should be tested with respect to their outcome on multiple realistic instances. In particular, an efficient method should be adopted to generate a path set. However, the analysis still holds if not a complete path set is generated. In this perspective, we prefer the link-based approach of Lou et al. (2010) . They formulated a link-based reformulation of the BRUE problem to avoid constructing this set. However, this approach only considers a subset of all possible BRUE flows.

In total, three algorithms were compared with respect to the outcome with regard to global performance. However, the performance of the descent approach and manifold suboptimization algorithm highly depend on the initial solution, as was also indicated by Lou et al. (2010). We adopted basic settings for these algorithms.

11.2 Future Research

In our research, we mainly focused on the Best-case BRUE flow distribution. As suggested in Example 4.2.2, from an authorities' perspective, it might be more appropriate to consider the Worst-case BRUE problem in the NDP. Under linear latency assumptions, any branch problem turns into a concave minimization problem over a convex set. These problems are known to be difficult to solve. However, scenario analysis requires an algorithm to approach the Worst-case BRUE. A single-level reformulation of the Worst-case BRUE, similar as in Section 8, allows us to apply the manifold suboptimization algorithm (Lawphongpanich and Yin, 2010).

In a traffic management context, authorities want to steer the network to a system optimum. A possible measure implements road pricing. We could adapt our problem to the program of finding the first- or second-best tolling scheme under boundedly rational choice behavior. Let $y \in \mathbb{R}_+^{|E|}$ be the toll vector and $y^{\max} \in \mathbb{R}_+^{|E|}$ be the toll vector of the maximum allowed tolls. The optimization problem corresponding to the second-best tolling scheme becomes:

$$\begin{aligned} &\min_{(\rho,y,f,x)} s(x) \\ &\qquad \text{s.t.} \quad 0 \le \rho \le \epsilon \\ &\qquad\qquad 0 \le y \le y^{\max}, \text{ and } (f,x) \text{ solves} \\ &\qquad\qquad \min_{(f,x)} \textstyle\sum_{e \in E} \int_0^{x_e} (l_e(\tau) + y_e)\, d\tau - \rho^T f \\ &\qquad\qquad (f,x) \in \mathcal{F}_0 \end{aligned} \tag{11.2.1}$$

It is clear that, mathematically, this setting does not induce additional issues. However, this toll-setting is said to be 'optimistic'. We need further research on the Worst-case BRUE flow distribution to find a toll setting which is robust. By that we mean a toll setting which minimizes the Worst-case travel time, i.e. replace the upper-level objective in (11.2.1) by

$$\min_{y} \max_{(\rho,f,x)} s(x).$$

This context was considered in Di et al. (2016) under linear latency assumptions. A traffic management setting in which an authority steers the network from a given BRUE state to the Best-case BRUE is also topic for future research.

We focused ourselves on a static model which gives certain restrictions compared to a day-to-day or fully dynamic model. The static model suffices to predict impacts on a strategic level and is favorable in terms of computational effort and understandability compared to its dynamic counterparts. On the other hand, we recognize that the static assignment is a steady state of a day-to-day model. Other research (e.g. Guo and Liu (2011)) investigated a day-to-day model in which individuals are assumed to repeat their choice from day to day unless an alternative becomes available which substantially improves travel time. This behavior of individuals can be 'exploited' for traffic management purposes as was shown in a simple numerical example in Vreeswijk et al. (2013a).

In a fully dynamic setting, the route guidance system as considered in Section 4.3 is particularly applicable in case of non-recurrent congestion. In case of incidents, users tend to switch routes as their intended route alternative shows unexpected bad performance (Chorus et al., 2006; Mahmassani and Liu, 1999). In particular, a route guidance with user constraints could reduce congestion. Since these non-recurrent events account for 26% of all congestion on the Dutch highway network (Rijkswaterstaat, 2015), improvement of global performance leads to substantial economic benefits. However, the computational effort to calculate the Best-case BRUE is important whereas optimization occurs in a dynamic setting and ongoing fashion. Therefore we underline the necessity to use heuristics in this context. In particular genetic algorithms may serve this purpose.

A whole range of applications can adopt a similar setting. The route guidance setting describes an optimization problem that both concerns overall performance and equity. For instance, a similar problem arises in personnel scheduling, in which workload should be equally distributed among employees (Karsu and Morton, 2015).

Bibliography

Ban, J. X., Liu, H. X., Ferris, M. C., Ran, B. (2006). A general MPCC model and its solution algorithm for continuous network design problem. *Mathematical and Computer Modelling*, 43(5-6):493–505.

Ban, X. J., Lu, S., Ferris, M., Liu, H. X. (2009). Risk averse second best toll pricing. In *Transportation and Traffic Theory 2009: Golden Jubilee*, pages 197–218. Springer.

Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K. (1983). *Non-linear parametric optimization*. Birkhäuser Verlag.

Bar-Gera, H. (2006). Primal method for determining the most likely route flows in large road networks. *Transportation Science*, 40(3):269–286.

Bard, J. F. (1998). *Practical bilevel optimization: algorithms and applications*. Springer Science & Business Media.

Beckmann, M., McGuire, C. B., Winsten, C. B. (1956). Studies in the Economics of Transportation.

Ben-Akiva, M. E., Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Borchers, M., Breeuwsma, P., Kern, W., Slootbeek, J., Still, G., Tibben, W. (2015). Traffic user equilibrium and proportionality. *Transportation Research Part B: Methodological*, 79:149–160.

Braess, D., Nagurney, A., Wakolbinger, T. (2005). On a paradox of traffic planning. *Transportation science*, 39(4):446–450.

Brands, T. (2015). *Multi-objective optimisation of multimodal passenger transportation networks*. PhD thesis.

Brands, T., van Berkum, E. C. (2014). Performance of a genetic algorithm for solving the multi-objective, multimodal transportation network design problem. *International Journal of Transportation*, 2(1).

Cho, H.-J., Smith, T. E., Friesz, T. L. (2000). A reduction method for local sensitivity analyses of network equilibrium arc flows. *Transportation Research Part B: Methodological*, 34(1):31–51.

Chorus, C. G., Arentze, T. a., Molin, E. J. E., Timmermans, H. J. P., Van Wee, B. (2006). The value of travel information: Decision strategy-specific conceptualizations and numerical examples. *Transportation Research Part B: Methodological*, 40(6):504–519.

Chorus, C. G., Timmermans, H. J. P. (2009). Measuring user benefits of changes in the transport system when traveler awareness is limited. *Transportation Research Part A: Policy and Practice*, 43(5):536–547.

Christodoulou, G., Koutsoupias, E., Spirakis, P. G. (2011). On the performance of approximate equilibria in congestion games. *Algorithmica*, 61(1):116–140.

Chung, B. D., Cho, H. J., Friesz, T. L., Huang, H., Yao, T. (2014). Sensitivity Analysis of User Equilibrium Flows Revisited. *Networks & Spatial Economics*, 14(2):183–207.

Ciscal-Terry, W., Dell'Amico, M., Hadjidimitriou, N. S., Iori, M. (2016). An analysis of drivers route choice behaviour using GPS data and optimal alternatives. *Journal of Transport Geography*, 51:119–129.

Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature*, 34(2):669–700.

Correa, J. R., Schulz, A. S., Stier-Moses, N. E. (2007). Fast, Fair, and Efficient Flows in Networks. *Operations Research*, 55(2):215–225.

Dempe, S. (2000). Bundle algorithm applied to bilevel programming problems with non-unique lower level solutions. *Computational Optimization and Applications*, 15(2):145–166.

Dempe, S. (2003). Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints. *Optimization*, 52(1):333–359.

Di, X., He, X., Guo, X., Liu, H. X. (2014a). Braess paradox under the boundedly rational user equilibria. *Transportation Research Part B: Methodological*, 67:86–108.

Di, X., Liu, H. X. (2016). Boundedly rational route choice behavior: A review of models and methodologies. *Transportation Research Part B: Methodological*, 85:142–149.

Di, X., Liu, H. X., Ban, X. J. (2016). Second best toll pricing within the framework of bounded rationality. *Transportation Research Part B: Methodological*, 83:74–90.

Di, X., Liu, H. X., Ban, X. J. X., Yu, J. W. (2014b). Submission to the DTA 2012 Special Issue: On the Stability of a Boundedly Rational Day-to-Day Dynamic. *Networks and Spatial Economics*, 15(3):537–557.

Di, X., Liu, H. X., Pang, J. S., Ban, X. J. (2013). Boundedly rational user equilibria (BRUE): Mathematical formulation and solution sets. *Transportation Research Part B: Methodological*, 57:300–313.

Faigle, U., Kern, W., Still, G. (2013). *Algorithmic principles of mathematical programming*, volume 24. Springer Science & Business Media.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., Rashidi, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, 229(2):281–302.

Friesz, T. L., Tobin, R. L., Cho, H. J., Mehta, N. J. (1990). Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. *Mathematical Programming*, 48(1-3):265–284.

Gauvin, J., Janin, R. (1988). Directional behaviour of optimal solutions in nonlinear mathematical programming. *Mathematics of Operations Research*, 13(4):629–649.

Gill, P. E., Murray, W., Saunders, M. A. (2005). SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM review*, 47(1):99–131.

Guo, X., Liu, H. X. (2011). Bounded rationality and irreversible network change. *Transportation Research Part B: Methodological*, 45(10):1606–1618.

Han, K., Szeto, W. Y., Friesz, T. L. (2015). Formulation, existence, and computation of boundedly rational dynamic user equilibrium with fixed or endogenous user tolerance. *Transportation Research Part B: Methodological*, 79:16–49.

Hato, E., Asakura, Y. (2000). Incorporating Bounded Rationality Concept into Route Choice Model for Transport Network Analysis. In *European Transport Conference.*

Jahn, O., Möhring, R. H., Schulz, A. S., Stier-Moses, N. E. (2005). System-optimal routing of traffic flows with user constraints in networks with congestion. *Operations research*, 53(4):600–616.

Jan, O., Horowitz, A., Peng, Z. (2000). Using GPS data to understand variations in path choice. *Transportation Research Record*, 1706(00):145–151.

Karsu, O., Morton, A. (2015). Inequity averse optimization in operational research. *European Journal of Operational Research*, 245(2):343–359.

Kolata, G. (1990). What if they closed 42nd Street and nobody noticed. http://www.nytimes.com/1990/12/25/health/what-if-they-closed-42d-street-and-nobody-noticed.html.

Lawphongpanich, S., Yin, Y. (2010). Solving the Pareto-improving toll problem via manifold suboptimization. *Transportation Research Part C: Emerging Technologies*, 18(2):234–246.

Lou, Y., Yin, Y., Lawphongpanich, S. (2010). Robust congestion pricing under boundedly rational user equilibrium. *Transportation Research Part B: Methodological*, 44(1):15–28.

Lu, S., Nie, Y. (2010). Stability of user-equilibrium route flow solutions for the traffic assignment problem. *Transportation Research Part B: Methodological*, 44(4):609–617.

Luo, Z.-Q., Pang, J.-S., Ralph, D. (1996). *Mathematical programs with equilibrium constraints*. Cambridge University Press.

Mahmassani, H. S., Chang, G.-L. (1987). On Boundedly Rational User Equilibrium in Transportation Systems. *Transportation Science*, 21(2):89–99.

Mahmassani, H. S., Liu, Y.-H. (1999). Dynamics of commuting decision behaviour under advanced traveller information systems. *Transportation Research Part C: Emerging Technologies*, 7(2):91–107.

Mallard, G. (2012). Modelling cognitively bounded rationality: An evaluative taxonomy. *Journal of Economic Surveys*, 26(4):674–704.

Mangasarian, O. L., Shiau, T.-H. (1987). Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *SIAM Journal on Control and Optimization*, 25(3):583–595.

Nguyen, S., Dupuis, C. (1984). An efficient method for computing traffic equilibria in networks with asymmetric transportation costs. *Transportation Science*, 18(2):185–202.

Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V. V. (2007). *Algorithmic game theory*, volume 1. Cambridge University Press Cambridge.

Ohazulike, A. E., Still, G., Kern, W., van Berkum, E. C. (2013). An origin-destination based road pricing model for static and multi-period traffic assignment problems. *Transportation Research Part E: Logistics and Transportation Review*, 58:1–27.

Patriksson, M. (2004). Sensitivity analysis of traffic equilibria. *Transportation Science*, 38(3):258–281.

Peters, H. (2008). *Game theory: A Multi-leveled approach*. Springer.

Rijkswaterstaat (2015). Publieksrapportage Rijkswegennet, 2e periode 2015. Technical report.

Rockafellar, R. T., Wets, R. J.-B. (2009). *Variational analysis*, volume 317. Springer Science & Business Media.

Roughgarden, T. (2002). How unfair is optimal routing? *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 203–204.

Roughgarden, T. (2005). *Selfish Routing and the Price of Anarchy.* MIT Press, Cambridge.

Roughgarden, T., Tardos, E. (2002). How bad is selfish routing? *Journal of the ACM*, 49(2):236–259.

Savard, G., Gauvin, J. (1994). The steepest descent direction for the nonlinear bilevel programming problem. *Operations Research Letters*, 15(5):265–272.

Sheffi, Y. (1985). *Urban Transportation Networks.*

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69:99–118.

Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason*, volume 3. MIT press.

Smith, M. J. (1979). The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological*, 13(4):295–304.

Still, G. (2006). Lectures on parametric optimization: An introduction. *University of Twente*, pages 1–19.

Still, G., Streng, M. (1996). Optimality conditions in smooth nonlinear programming. *Journal of optimization theory and applications*, 90(3):483–515.

Sun, L., Karwan, M. H., Kwon, C. (2015). Incorporating Driver Behaviors in Network Design Problems: Challenges and Opportunities. *Transport Reviews*, 36(4):454–478.

Thomas, T., Tutert, S. (2008). Revealed route choice from a license plate registration study in Enschede.

Tobin, R. L., Friesz, T. L. (1988). Sensitivity analysis for equilibrium network flow. *Transportation Science*, 22(4):242–250.

van Essen, M., Thomas, T., van Berkum, E., Chorus, C. (2016). From user equilibrium to system optimum: a literature review on the role of travel information, bounded rationality and non-selfish behaviour at the network and individual levels. *Transport Reviews*, 36(4):527–548.

van Wee, B., Annema, J. A., Banister, D. (2013). *The transport system and transport policy: an introduction*. Edward Elgar Publishing.

Vreeswijk, J., Bie, J., van Berkum, E., van Arem, B. (2013a). Effective traffic management based on bounded rationality and indifference bands. *IET Intelligent Transport Systems*, 7(3):265–274.

Vreeswijk, J., Thomas, T., van Berkum, E., van Arem, B. (2013b). Drivers' Perception of Route Alternatives as Indicator for the Indifference Band. *Transportation Research Record: Journal of the Transportation Research Board*, 2383:10–17.

Wardrop, J. G. (1952). Some Theoretical Aspects of Road Traffic Research. *ICE Proceedings: Engineering Divisions*, 1:767–768.

Wu, J., Sun, H., Wang, D. Z. W., Zhong, M., Han, L., Gao, Z. (2013). Bounded-rationality based day-to-day evolution model for travel behavior analysis of urban railway network. *Transportation Research Part C: Emerging Technologies*, 31:73–82.

Yang, H. (1997). Sensitivity analysis for the elastic-demand network equilibrium problem with applications. *Transportation Research Part B: Methodological*, 31(I):55–70.

Yang, H., Huang, H.-J. (2005). *Mathematical and economic theory of road pricing.*

Zgurovsky, M. Z., Kasyanov, P., Kapustyan, O. V., Valero, J., Zadoianchuk, N. V. (2012). *Evolution Inclusions and Variation Inequalities for Earth Data Processing III: Long-time Behavior of Evolution Inclusions Solutions in Earth Data Analysis*, volume 27. Springer Science & Business Media.

Zhang, J., Yang, H. (2015). Modeling route choice inertia in network equilibrium with heterogeneous prevailing choice sets. *Transportation Research Part C: Emerging Technologies*, 57:42–54.

Zhao, C. L., Huang, H. J. (2016). Experiment of boundedly rational route choice behavior and the model under satisficing rule. *Transportation Research Part C: Emerging Technologies*, 68:22–37.

Zhu, S., Levinson, D. (2010). Do people use the shortest path? An empirical test of Wardrop's first principle. *91st annual meeting of the Transportation Research Board, Washington. Vol. 8*, pages 1–22.