European Journal of Operational Research 299 (2022) 195–207

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing, Transportation and Logistics

Improving the performance of a traffic system by fair rerouting of travelers

Oskar A. L. Eikenbroek[a,*], Georg J. Still[b], Eric C. van Berkum[a]

[a] Transport Engineering & Management, University of Twente, P.O. Box 217, AE Enschede 7500, The Netherlands
[b] Discrete Mathematics and Mathematical Programming, University of Twente, P.O. Box 217, AE Enschede 7500, The Netherlands

ARTICLE INFO

Article history:
Received 13 May 2020
Accepted 16 June 2021
Available online 25 June 2021

Keywords:
Traffic
Transportation
Assignment
System optimum
Social routing

ABSTRACT

Some traffic management measures route drivers towards socially-desired paths in order to achieve the system optimum: the traffic state with minimum total travel time. In previous attempts, the behavioral response to route advice is often not accounted for since some drivers are advised to take significantly longer paths for the system's benefit. Hence, these drivers may not comply with such advice and the optimal state will not be achieved. In this paper, we propose a social routing strategy to approach the optimal state while accounting for fairness in the resulting state. This routing strategy asks travelers to take a limited detour in order to improve efficiency. We show that the best possible paths (in terms of efficiency) to be proposed by a service adopting this strategy can be found by solving a bilevel optimization problem with a non-unique lower-level solution. We use techniques from parametric analysis to show that the directional derivative of the lower-level link flows however exists. This derivative is the optimal solution of a quadratic optimization problem with a suitable route flow solution as parameter. We use the derivative in a descent algorithm to solve the bilevel problem. Numerical experiments in a realistic environment show that the routing strategy only asks a small fraction of the drivers to take a limited detour and thereby substantially improves the performance of the traffic system.

1. Introduction

Transport authorities face the daily challenge to reduce congestion. Traditionally, this was solved by increasing road capacity through building new or expanding existing infrastructure. However, the construction of infrastructure is costly, and may also lead to an increase in demand. Nowadays, authorities implement management measures alongside to improve utilization of existing roads.

The need for policy measures in general stems from the observation that individuals typically behave selfishly, i.e., travelers are mainly concerned with their own utility when making decisions. The resulting traffic state (i.e., flow distribution) with respect to route choice, the *user equilibrium*, does mostly not correspond to the *system optimum*: the traffic state with minimum (total or average) travel time (Wardrop, 1952). Without intervention, in particular with the increasing use of real-time routing apps, the real-world traffic state is likely to be closer to the inefficient user equilibrium than to the system optimum (Klein, Levy, & Ben-Elia, 2018). In the user equilibrium, travelers with the same origin-destination pair have equal travel times. The system optimum, on the other hand, is 'unstable' since it is unfair: some drivers may travel longer than others for the same origin-destination pair. Hence, we can characterize the system optimum as (perfectly) efficient but unfair, while the user equilibrium is inefficient and perfectly fair.

Recently, traffic management measures, e.g., social routing, have been proposed that steer or nudge travelers towards socially-desired routes. The 'pure' system optimum is difficult to achieve (Klein et al., 2018) and maintain over time, because only some travelers use and comply with advice from information systems, and the individual intra- (within the system optimum) and inter-state (compared to the user equilibrium) travel time differences might be substantial (Jahn, Möhring, Schulz, & Stier-Moses, 2005; van Essen, Eikenbroek, Thomas, & van Berkum, 2020). Hence, any social routing strategy should in essence anticipate user responses and persuade travelers to comply with socially-oriented advice.

Empirical evidence (e.g., Djavadian, Hoogendoorn, van Arem, & Chow, 2014) shows that some (travelers) are receptive for advice that proposes reasonable routes for the system's benefit. A possible explanation is that individuals have a so-called *indifference*

* Corresponding author.
E-mail addresses: o.a.l.eikenbroek@utwente.nl (O.A.L. Eikenbroek), e.c.vanberkum@utwente.nl (E.C. van Berkum).

https://doi.org/10.1016/j.ejor.2021.06.036

band (Simon, 1997), which means in our context that when a route is only slightly longer than the best one, it is still acceptable to use (Vreeswijk et al., 2015). A social routing strategy can 'exploit' the indifference band and propose acceptable routes (possibly, *sub-optimal* from an individual's perspective) to *receptive drivers* (those that use and comply with advice from the service), and thereby potentially steer the network to a state close to the system optimum. Compared to the system optimum, the resulting distribution is easier to achieve and maintain over time.

In this paper, we propose and evaluate a centrally coordinated social routing strategy that improves overall efficiency, while we explicitly account for the above-mentioned practical requirements. The routing strategy incorporates user-induced constraints in the sense that travel time differences in the resulting state are explicitly limited, and only a fraction of the travelers is asked to take an acceptable detour to the system's benefit. We note that a routing service adopting the strategy, in practice, offers a single route advice using a personalized information device to its users before departure.

1.1. Research contribution

Although empirical research has shown that social routing has great potential in real life, there is not yet a corresponding routing strategy that improves efficiency while explicitly incorporating user responses to advice in terms of route choice behavior. Route choice behavior is crucial for the strategy's performance in practice. Compliance is expected to be much higher when the advised route is only slightly longer than the shortest route. Behavioral responses influence the travel times, and should thus be anticipated in order to advise routes that are acceptable with respect to travel time.

In this study, we propose a social routing strategy that explicitly accounts for behavioral responses to a routing service. In fact, changes in route choice may occur from travelers that comply with the advice but also from those that do not comply, but are now confronted with altered travel times on routes as a result of behavioral changes by others. We introduce a bilevel optimization problem that calculates the best possible paths (with respect to efficiency) with a limited (realized) detour to be proposed to the compliant travelers. Although in this paper we limit ourselves to a static environment, the bilevel problem is already highly challenging to solve. Many of the theoretical difficulties that occur in our case, also apply to a real-world social route guidance service in which limited detours are suggested in a dynamic fashion. Hence, before considering such a guidance system we should address the theoretical challenges and potential impact in a static traffic assignment first. In particular, the service as proposed in this paper can serve as a proof-of-concept for a dynamic variant.

Related social routing approaches

We discuss related social routing approaches from literature. Jahn et al. (2005) proposed a routing strategy that limits the 'normal length' difference before and after implementation, assuming that the normal length is independent of the traffic flow. This mechanism was numerically evaluated on realistic network instances, and showed performance (with respect to efficiency) close to the system optimum. The intra-state time differences, however, were not explicitly limited. A related approach by Angelelli, Arsik, Morandi, Savelsbergh, & Speranza (2016) considers a mathematical program that tries to achieve an optimal flow with a regularization term to minimize the 'total inconvenience' alongside. Here, the travel time is assumed to be independent of the flow. Both studies assume a full market penetration of the routing service. Bagloee, Sarvi, Patriksson, & Rajabifard (2017); van Essen et al. (2020); Zhang & Nie (2018) proposed systems to route a fraction of the demand onto social routes. We refer to Li, Liu, & Nie (2018) and Zhou, Xu, Meng, & Huang (2017), for dynamic (day-to-day) variations on such routing policies.

In contrast to the above-mentioned studies, we propose a routing strategy that steers the network to a system optimum while explicitly limiting the intra-state time differences whereas we argue that travelers evaluate acceptability of routes in terms of realized travel time rather than free-flow travel time or distance. This necessary user-induced constraint makes the accompanying optimization problem substantially harder to solve, which might be a reason that a majority of the studies relax this real-life constraint or introduce heuristic approaches (e.g., Angelelli, Morandi, & Speranza, 2019; Roughgarden, 2005). Recently, Angelelli, Morandi, & Speranza (2020) studied a similar setting, with a so-called 'constrained system optimum'. They used an integer linear program and matheuristic to formulate and solve the corresponding optimization problem, respectively. In contrast to our study, they do not incorporate the route choices of travelers that do not comply with route advice. Angelelli, Morandi, Savelsbergh, & Speranza (2021) proposed a fast heuristic to find the constrained system optimum and use a piecewise linearization of the travel time function. In our paper, we formulate the problem as a continuous optimization problem and keep the nonlinearity of the travel time function.

We note that our optimization problem is a generalized case of finding the *boundedly rational user equilibrium* (BRUE) with minimum travel time. Although there is a body of literature on BRUE (e.g., Di, Liu, Pang, & Ban, 2013; Lou, Yin, & Lawphongpanich, 2010), a thorough quantitative analysis of this problem is still lacking. Thus far, analyses have been based on relatively strong assumptions which reduce the complexity of the problem but might not hold in practice.

Bilevel problem

The success of the social routing strategy, as discussed, hinges on the (travel time of the) paths suggested to the drivers. We show that best possible paths can be found by solving a bilevel optimization problem. Our bilevel problem (see Section 2.2) can be seen as a game between a leader (authority) and a follower (travelers) (Josefsson & Patriksson, 2007). The leader chooses the paths to be proposed, while the travelers update their route choice based on this advice. The compliant travelers follow the advice if the travel time differences (based on the route choice of the travelers) compared to the fastest paths are limited, while non-compliant travelers find the cheapest paths available. These dynamics should be anticipated to find the best possible advice in terms of total travel time.

Bilevel problems are typically difficult to solve directly, and therefore often reformulated as single-level problems. In this paper, we use an implicit reformulation, and require parametric analysis of the lower-level problem to describe the behavior of the corresponding solution set as a function of the upper-level variable. Parametric analysis is either quantitative or qualitative in nature (Fiacco & Ishizuka, 1990b). The qualitative analysis is mainly concerned with the continuity of the optimal solution set. Here, we require a - local - quantitative analysis that focuses on the estimation of generalized derivatives of the optimal solution set, e.g., to be used in numerical procedures. We refer to Eikenbroek, Still, van Berkum, & Kern (2018) for the qualitative analysis of this problem (the mentioned paper's setting is however different).

Techniques from variational analysis are used to study the quantitative behavior of the lower-level problem. We refer to Luo, Pang, & Ralph (1996); Mordukhovich (2018); Rockafellar & Wets (2009) for an overview of theoretical results. Many of these results, however, are presented in general form and require relatively strong conditions when applying an implicit reformulation. A critical issue in our case is that the lower-level solution is not unique for a given upper-level variable. This leads to practical and theoretical challenges, since the desired lower-level solution might not be realized, and small changes in the parameter might lead to major changes in the solution (Dempe, 2002). Hence, at first sight, many of the algorithms designed for bilevel problems do not apply in our context.

In this paper, we theoretically assess the lower-level problem and use techniques from variational analysis to show that we can guarantee the existence and calculation of a generalized derivative of the lower-level solution projected onto a subspace. The generalized derivative of the solution of the lower-level problem contributes to the understanding of the optimization problem finding the best possible paths. Indeed, the theoretical analysis in this paper allows one to formulate the necessary optimality conditions of the bilevel problem. Moreover, the derivative can be used in exact numerical procedures to find descent directions. Hence, not only can the generalized derivative be used in standard algorithms to solve bilevel programs, it can also support the assessment of heuristic procedures (e.g., Angelelli et al. (2020)). Although a comparative analysis of algorithms that solve the formulated program is beyond the scope of our paper, we provide nonetheless a numerical procedure and refer to related algorithms that could be applied.

In a static traffic assignment context, bilevel problems are well-known, mainly in *Network Design Problems* (NDPs) in which optimal network settings (e.g., link tolls) are determined. Parametric analysis has been topic of a body of literature in this context (Do Chung, Cho, Friesz, Huang, & Yao, 2014; Josefsson & Patriksson, 2007; Lu, 2008; Lu & Nie, 2010; Outrata, 1997; Patriksson, 2004; Patriksson & Rockafellar, 2002; 2003; Qiu & Magnanti, 1989; Robinson, 2006; Tobin & Friesz, 1988; Yin, Madanat, & Lu, 2009). Mainly, these papers concern perturbations that occur in the (parameters of the) link cost function and/or demand vector. Our paper is different from the aforementioned studies since we basically consider perturbations in a path-dependent parameter. It turns out that the analysis and the computational results rely on the choice of a suitable route flow corresponding to a link flow solution. This forces us to study the behavior of the (multi-valued) route flow solution set in dependence of the parameter. In the context of perturbations in the parameter of the demand vector, the analysis of Qiu & Magnanti (1989) also depends on the choice of a specific route flow solution. However, Patriksson & Rockafellar (2002) show that the results of Qiu & Magnanti (1989) are actually independent of a specific choice. In our context, this does not hold (as we will show in Example 1). Some of our findings show similarities to the results for parametric optimization problems with a unique minimizer but non-unique multipliers (Dempe, 1989; 1993; Ralph & Dempe, 1995). There, a generalized derivative of the optimal solution can be calculated by choosing a suitable multiplier (which might be difficult to find). In our case, we consider a setting with a non-unique optimal solution.

Considering the practical application, we assess a possible implementation of the social routing strategy. Specifically, we evaluate the interaction among compliance rate, acceptable travel time differences, and network-wide performance in a static setting. The numerical experiments provide insight which minimum penetration rate and indifference band might be required to substantially lower the total travel time in the network, and, thus, how much some travelers have to sacrifice for the network's benefit. These experiments substantiate the opportunities for a real-life implementation of a social routing service or guidance mechanism.

Summarizing, the main contributions of this paper are as follows:

- We propose a social routing strategy that steers the traffic network towards an efficient but also fair, and therefore achievable and maintainable, traffic state. We show that the best possible paths to be proposed by a social routing service can be found by solving a bilevel program that explicitly accounts for behavioral responses to the service;
- We use parameteric analysis to prove that the generalized derivative of the lower-level link flow solution problem exists and can be calculated efficiently. The generalized derivative can be used to find descent directions and to formulate optimality conditions of the bilevel problem;
- We use the generalized derivative in a descent algorithm to solve the bilevel problem and numerically evaluate our proposed social routing strategy in test networks. Here, only a small fraction of the travelers need to take a limited detour to substantially improve the traffic system's performance.

The remainder of our paper is organized as follows. We formally introduce our social routing strategy in Section 2. In Section 3, we analyze qualitatively the 'behavior' of the optimization problem that relates to our social routing strategy. In Section 4, we investigate the existence and calculation of the directional derivative of the link flows, which we use in Section 5 in a descent algorithm for solving the bilevel problem. Section 6 reports on numerical experiments and management implications. Section 7 draws the conclusions.

2. Problem formulation

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

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 $(f, x) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} = (f_p, p \in \mathcal{P}; x_e, e \in E)$ so that

$$\Lambda f = d, \quad \Delta f - x = 0, \quad f \geq 0. \tag{1}$$

The matrix $\Lambda \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{P}|}$ is the *OD-path incidence matrix* with $\Lambda_{kp} = 1$ if $p \in \mathcal{P}_k$, and $\Lambda_{kp} = 0$ otherwise. $\Delta \in \mathbb{R}^{|E|\times|\mathcal{P}|}$ denotes the *link-path incidence matrix*: $\Delta_{ep} = 1$ if edge e is in route p, and $\Delta_{ep} = 0$ otherwise. For each edge, $e \in E$, $l_e(x_e)$ is the non-negative, continuous, and non-decreasing link cost (or: travel time) function for a given flow x_e on that edge. The cost of a route $c_p(f)$, $p \in \mathcal{P}$, is the sum of travel costs of all edges in that path, $c_p(f) = \sum_{e\in p} l_e(x_e)$.

Throughout our paper we make the following (natural) assumption regarding the travel time function (we refer to Patriksson & Rockafellar (2002) for a study that relaxes this assumption).

Assumption 1. We assume throughout the paper that the travel time functions $l_e(x_e)$ are continuous, convex, and strictly monotone: $l_e(x_e) < l_e(x_e^0)$, for $x_e < x_e^0$, for all $e \in E$.

2.1. A social routing strategy

We consider the setting in which a central authority asks travelers to take a small detour for the system's benefit (see Section 1). The *social travelers* comply with such an advice if the alternative route is reasonable, i.e., the route is not perceived to be substantially worse (in terms of travel time) compared to the fastest path. The remaining drivers do not comply with travel advice and behave in a selfish manner, i.e., choose the fastest path available.

The demand vector $d^s \in \mathbb{R}_+^{|\mathcal{K}|}$ ($d_k^s \le d_k$ for all $k \in \mathcal{K}$) denotes the travelers that receive and comply with a route advice from the authority (superscript s refers to the social travelers). The remaining demand $d^n \in \mathbb{R}^{|\mathcal{K}|}$, so that $d = d^s + d^n$, behaves selfishly. (Superscript n refers to Nash equilibrium - see (2b) below: a driver cannot improve travel time by changing strategy (route)).

We define $\mathcal{F}$ as the set of feasible flows. Formally,

$$\mathcal{F} = \left\{ (g,h,x) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} \;\middle|\; \begin{array}{l} \Lambda g = d^s, g \ge 0, \\ \Lambda h = d^n, h \ge 0, \\ \Delta(g+h) - x = 0 \end{array} \right\}.$$

Obviously, any $(g,h,x) \in \mathcal{F}$ is a flow as in (1) for $f = g + h$.

The advised routes to compliant travelers d^s have to be fair in the sense that the realized (i.e., traffic flow-dependent) travel time differences are limited. We assume that social travelers accept any travel time difference (compared to the shortest path for the same OD pair) with a maximum of $\varepsilon_k \ge 0, k \in \mathcal{K}$. Hence, the resulting state in the network is so that no social traveler for OD pair $k \in \mathcal{K}$ can improve travel time with more than ε_k by unilaterally changing routes. At the same time, the selfish travelers choose the fastest path. The following definition (Definition 1) formalizes our notion of the resulting state among social (receptive) and selfish travelers. We refer to this state as a *mixed equilibrium*.

Definition 1 (Mixed equilibrium)**.** Given $\varepsilon \in \mathbb{R}_+^{|\mathcal{K}|}$, a traffic flow $(g,h,x) \in \mathcal{F}$ with corresponding path costs $c(f)$, $f = g + h$, is called a mixed equilibrium among social and selfish travelers if for all $k \in \mathcal{K}$, the following conditions are satisfied for all $p \in \mathcal{P}_k$:

$$g_p > 0 \Rightarrow c_p(f) \le \min_{q \in \mathcal{P}_k} c_q(f) + \varepsilon_k \tag{2a}$$

$$h_p > 0 \Rightarrow c_p(f) = \min_{q \in \mathcal{P}_k} c_q(f) \tag{2b}$$

Assuming only selfish demand, in a traffic state in user equilibrium as in (2b), travelers with the same OD pair share travel times. However, it is well-known that this state does not necessarily minimize total travel time $\sum_{e \in E} x_e l_e(x_e)$. The traffic state (f,x) as in (1) which minimizes the total travel time, is referred to as the system optimum (Wardrop, 1952). Typically, it may be assumed that in practice, without intervention, a state close to a user equilibrium arises.

Condition (2a) gives a range of acceptable travel times for a receptive user. We assume that any social traveler that is routed onto an *acceptable path* (i.e., any route $p \in \mathcal{P}_k, k \in \mathcal{K}$ for which $c_p(f) \le \min_{q \in \mathcal{P}_k} c_q(f) + \varepsilon_k$) complies with such an advice although the user might be aware that it is not necessarily the fastest path available. The condition as defined in (2a) is equivalent to the BRUE condition (see Section 1). The mixed equilibrium as in (2), i.e., (2a) and (2b), has the user equilibrium as a special case and does not correspond (even if $\varepsilon \to \infty$) to a mixed user equilibrium and system-optimal flow, e.g., as in Yang, Zhang, & Meng (2007).

In (2a), we model the band ε as being *additive*. In particular for shorter travel times, an additive indifference band is more appropriate compared to a multiplicative one as in, e.g., Roughgarden (2005). In combination with ε_k, $k \in \mathcal{K}$, being OD-pair dependent, we allow a range of scenarios regarding the maximum detour to be modeled using the condition in (2a).

The mixed-equilibrium conditions (2) do not provide a unique state (yet all travelers are satisfied with their route), which is key for the social routing strategy. We exploit this range of allowed distributions to find one which is the best for the system. That is, our routing strategy is designed so that we achieve - among all $(g,h,x) \in \mathcal{F}$ that satisfy (2) - the one with the minimum total travel time. Hence, the optimal strategy can be found by solving the following optimization program for a known $\varepsilon \ge 0$:

$$\min_{(g,h,x) \in \mathcal{F}} \varphi(x) \qquad \textit{s.t.} \qquad (g,h,x) \text{ satisfies } (2), \tag{3}$$

where $\varphi(x) = \sum_{e \in E} x_e l_e(x_e)$ is the total travel time.

For a routing service, the optimal solution of (3) with respect to g is typically the variable of interest, since g represents the distribution of the social travelers over the different acceptable paths. The selfish demand basically responds to the choices of the social demand in the sense of (2b). In fact, selfish travelers are confronted with a change in travel times on routes due to the choices of others. When determining the best distribution g (with condition (2a)), the authority needs to anticipate the travel times depending on the route choices of both the social and selfish demand. This *Stackelberg* mechanism is implicitly in (3). After solving (3), the route to be suggested to a social traveler can be extracted from solution g.

One should note that, in principle, while solving (3), one is free to choose any (g,h,x) satisfying (2). In practice, for a given g, the distribution h is a result of the route choice behavior of the selfish travelers and cannot be precisely predicted (if there are multiple h satisfying (2b)). However, as we will see in Theorem 1, the response to g with respect to the link flows x is uniquely determined. Since x is the only variable appearing in the objective function, it is therefore not necessary to consider a pessimistic variant of (3).

2.2. Bilevel reformulation

The optimization problem in (3) is difficult to solve. Indeed, Eikenbroek et al. (2018) and Lou et al. (2010) show that the feasible set corresponding to (3) is in general not convex, does not satisfy a regularity condition, and different local minimizers can coexist. We use the following proposition (Proposition 1) to reformulate our problem. In the remainder of the analysis we drop parameter ε in the notation: we assume it is known and fixed. During the experiments (Section 5 and Section 6), we numerically investigate the impact of a varying ε.

Proposition 1 (Di et al. (2013); Eikenbroek et al. (2018))**.** *The following are equivalent for* (g,h,x)*:*

1. $(g,h,x) \in \mathcal{F}$ *is a mixed equilibrium as in (2);*
2. *There exists*

   $$\rho \in \Xi := \left\{ \rho \in \mathbb{R}^{|\mathcal{P}|} \;\middle|\; 0 \le \rho \le \Lambda^T \varepsilon \right\}$$

   such that (g,h,x) *solves the convex optimization problem*

   $$Q(\rho): \quad \min_{(g,h,x)} z(\rho,g,x) = z_0(x) + \rho^T g \quad \textit{s.t.} \quad (g,h,x) \in \mathcal{F}, \tag{4}$$

   where $z_0(x) = \sum_{e \in E} \int_0^{x_e} l_e(\omega) d\omega$.

We omit the proof, which is a generalization of Proposition 2.2 in Di et al. (2013) or Proposition 1 in Eikenbroek et al. (2018). These references use objective function $z_0(x) - \tilde{\rho}^T g$, but the two problems are equivalent by selecting $\tilde{\rho} = \Lambda^T \varepsilon - \rho$. We prefer our objective function in (4) whereas it eases the upcoming analysis. We note that ρ does not necessarily have an intuitive interpretation.

Problem (3) is a *mathematical program with equilibrium constraints*. According to Proposition 1, we can rewrite (3) as a bilevel problem. We use the following reformulation, which eases the parametric analysis in Section 3 and 4 (Eikenbroek et al., 2018):

$$(BL): \qquad \min_{(g,h,x,\rho)} \varphi(x) \qquad \textit{s.t.} \qquad \begin{array}{l} \rho \in \Xi \\ (g,h,x) \text{ solves } Q(\rho). \end{array}$$

(BL) is a technical reformulation of the bilevel problem in which the leader finds the best possible paths to be proposed, while anticipating route choices (see Section 1.1). Basically, $Q(\rho)$ describes the route choice behavior of both the social and selfish travelers for a given ρ.

In the remainder, we refer to *parametric optimization problem* $Q(\rho)$ as the *lower-level problem*. Here, ρ is a parameter in the lower-level problem but a variable in the *upper-level problem*. Note that in (*BL*) both lower-level variables (g, h, x) as well as upper-level variable ρ appear as variables. Even in case there is no upper bound with respect to ρ, i.e., the social travelers can be routed onto any path, the problem (*BL*) might be difficult to solve. In the upcoming sections we rewrite and (numerically) solve (*BL*) as a single-level optimization problem.

3. Parametric analysis

Based on reformulation (*BL*) of previous section, one basically needs to find an appropriate $\rho \in \Xi$ so that the corresponding (g, h, x) that solves $Q(\rho)$ minimizes total travel time $\varphi(x)$. In this paper, we apply parametric analysis with respect to problem $Q(\rho)$, i.e., we investigate the 'behavior' of (g, h, x) that solves $Q(\rho)$ under perturbations in ρ.

The purpose of the analysis is, from a computational perspective, as follows. The parametric analysis provides an estimate for the rate of change in the lower-level solution as the lower-level parameter (which is an upper-level variable) changes (Patriksson, 2004). Then, we use this estimate to move into a direction that decreases the total travel time. In this and next section (Section 4), we provide the parametric analysis of the lower-level problem. The results of these sections are used to reformulate and solve (*BL*) as single-level optimization problem (Section 5).

3.1. Notation, definitions and preliminary results

We introduce notations that correspond to lower-level problem $Q(\rho)$ (see (4)) with parameter ρ:

$$v(\rho) = \min\{z(\rho, g, x) \mid (g, h, x) \in \mathcal{F}\},$$
$$S(\rho) = \{(g, h, x) \mid (g, h, x) \text{ is a global minimizer of } Q(\rho)\}.$$

We refer to $\mathcal{F}$ as the *feasible set*, $v(\rho)$ as the *optimal value function*, and to $S(\rho)$ as the *solution set* at ρ.

To study the parametric problem $Q(\rho)$, we introduce definitions that describe the behavior of functions. In this paper, we consider both single and multi-valued functions (or: mappings). A multi-valued function F assigns to each $\varepsilon \in X \subseteq \mathbb{R}^n$ a possibly empty subset $F(\varepsilon) \subseteq Y \subseteq \mathbb{R}^m$. We denote by $\mathrm{dom}(F) := \{\varepsilon \in X | F(\varepsilon) \neq \emptyset\}$ the domain of multifunction F. We further define for $\tau > 0$, $\delta > 0$, the neighborhoods $U_\tau(F(\varepsilon^0)) := \left\{x \in \mathbb{R}^m | \|x - x'\| < \tau \text{ for some } x' \in F(\varepsilon^0)\right\}$ and $U_\delta(\varepsilon) := \{x \in \mathbb{R}^n | \|x - \varepsilon\| < \delta\}$.

We use the following definitions (Bank, Guddat, Klatte, Kummer, & Tammer, 1983; Robinson, 1982):

Definition 2. A multifunction $F(\varepsilon)$ is said to be:

1. *closed* at ε^0 if for any sequences ε^l, x^l, $l \in \mathbb{N}$, with $\varepsilon^l \to \varepsilon^0$, $x^l \in F(\varepsilon^l)$, the condition $x^l \to x^0$ implies $x^0 \in F(\varepsilon^0)$;
2. *upper/outer semicontinuous* at ε^0, if for any $\tau > 0$, exists $\delta > 0$ such that
$$F(\varepsilon) \subseteq U_\tau(F(\varepsilon^0)), \qquad \text{for all } \varepsilon \in U_\delta(\varepsilon^0);$$
3. *lower/inner semicontinuous* at ε^0, if for any $\tau > 0$, exists $\delta > 0$ such that
$$F(\varepsilon^0) \subseteq U_\tau(F(\varepsilon)), \qquad \text{for all } \varepsilon \in U_\delta(\varepsilon^0);$$
4. *(locally) upper Lipschitz continuous* at ε^0 if there exists a $\delta > 0$ and Lipschitz constant $L < \infty$ such that
$$F(\varepsilon) \subseteq F(\varepsilon^0) + L\|\varepsilon - \varepsilon^0\|\mathbb{B}, \qquad \text{for all } \varepsilon \in U_\delta(\varepsilon^0),$$
where $\mathbb{B} := \{x \in \mathbb{R}^m \mid \|x\| \le 1\}$;
5. *(locally) Lipschitz continuous* at ε^0 if there exists a $\delta > 0$ and Lipschitz constant $L < \infty$ such that
$$F(\varepsilon) \subseteq F(\varepsilon') + L\|\varepsilon - \varepsilon'\|\mathbb{B}, \qquad \text{for all } \varepsilon, \varepsilon' \in U_\delta(\varepsilon^0).$$

The following results are from Eikenbroek et al. (2018). Here, $S^x(\rho), S^g(\rho), S^h(\rho)$ denote the projections of $S(\rho)$ onto the x, g, and h-space, respectively.

Theorem 1 (Eikenbroek et al. (2018))**.**

1. *$S(\rho^0) \neq \emptyset$ for all $\rho^0 \in \Xi$;*
2. *$S(\rho^0)$, $S^g(\rho^0)$, and $S^h(\rho^0)$ are (polyhedral) convex sets for each $\rho^0 \in \Xi$;*
3. *$S^x(\rho^0)$ is a singleton for each $\rho^0 \in \Xi$, i.e., $S^x(\rho^0) = \{x(\rho^0)\}$, and $x(\rho)$ is a continuous function on Ξ, i.e., $x(\rho)$ is upper and lower semicontinuous at each $\rho^0 \in \Xi$. Moreover,*
$$\psi(\rho) := \left\{ \ \rho^T g \ \middle| \ g \in S^g(\rho) \ \right\}$$
is uniquely determined at each $\rho^0 \in \Xi$;
4. *The mappings $S(\rho)$, $S^g(\rho)$, and $S^h(\rho)$, are upper semicontinuous at each $\rho^0 \in \Xi$;*
5. *The mapping $S(\rho)$ is not injective, i.e., different $\rho^0 \neq \rho^1 \in \Xi$ might have a common solution $(g^0, h^0, x^0) \in S(\rho^0) \cap S(\rho^1)$.*

We underline that in our setting we cannot expect Theorem 1 to be stronger in the sense that $S^g(\rho)$ is also lower semicontinuous at each ρ^0. The route flow set

$$S^g(\rho) = \left\{ g \in \mathbb{R}^{|\mathcal{P}|} \ \middle| \ \exists h, \begin{array}{l} \Lambda g = d^s,\ \Lambda h = d^n,\ \Delta(g+h) = x(\rho), \\ g \ge 0,\ h \ge 0,\ \rho^T g = \psi(\rho) \end{array} \right\},$$

is a polyhedral convex set at each $\rho^0 \in \Xi$. So, although the x-part of the solution to $Q(\rho)$ is uniquely determined, there might be multiple route flow solutions that correspond to a single link flow solution $x(\rho)$. In the context of perturbations of a parameter in the link-cost and/or demand vector, the route flow set is a continuous mapping relative to its domain (Lu & Nie, 2010), given that the link flow changes continuously. We demonstrate later (Section 4) it is in fact the absence of lower semicontinuity of $S^g(\rho)$ at some $\rho^0 \in \Xi$ that causes the practical difficulties for the calculation of the directional derivative $x'(\rho^0; r)$ of $x(\rho)$ at ρ^0 in direction $r \in \mathbb{R}^{|\mathcal{P}|}$.

Remark 1. To improve readability, we assume for now that $d^s = d$ (i.e., $d^n = 0$). We prove in Section 4.4 that we can extend the results to the more general case $d^n \neq 0$.

3.2. Directional derivative of the optimal value function

This subsection covers the parametric analysis of the optimal value function $v(\rho)$. We show that the directional derivative $v'(\rho^0; r)$ of $v(\rho)$ exists for any $\rho^0 \in \Xi$ and direction r, $\|r\| = 1$, and we use - in Section 4.3 - the sensitivity of the optimal value function to find a specific route flow.

Definition 3 (Directional derivative)**.** A function $f(\rho)$ is said to be directionally differentiable at $\rho^0 \in \mathrm{dom}(f)$ in direction r, $\|r\| = 1$, if

$$f'(\rho^0; r) := \lim_{t \to 0^+} \frac{f(\rho^0 + tr) - f(\rho^0)}{t}$$

exists.

The following proposition (Proposition 2) demonstrates that the optimal value function $v(\rho)$ is directionally differentiable at any ρ^0 for any direction r, $\|r\| = 1$. This is a well-known result in parametric optimization (see Fiacco & Ishizuka, 1990a), but the accompanying proof (provided in the Appendix) is easier in our case.

Proposition 2. *The optimal value function $v(\rho)$ is directionally differentiable at each $\rho^0 \in \Xi$ and in each direction $r \in \mathbb{R}^{|\mathcal{P}|}$, $\|r\| = 1$. In fact, $v'(\rho^0; r)$ is the optimal value that corresponds to a solution of the parametric linear program*

$$P(r): \qquad \min r^T g \quad s.t. \quad g \in S^g(\rho^0).$$

In this section, we proved that the directional derivative $v'(\rho^0; r)$ of $v(\rho)$ exists for any ρ^0 and direction r ($\|r\| = 1$). In the upcoming section, we treat the (existence and calculation of the) directional derivative of the link flows $x(\rho)$. The sensitivity analysis of $v(\rho)$ can also be used to formulate a single-level problem, see, e.g., Dempe & Zemkoho (2012) and Mordukhovich (2018).

4. Parametric analysis of the optimal solution

Intuitively, directional derivative $x'(\rho^0; r)$ is the rate of change of the optimal solution $x(\rho)$ at ρ^0 along r. This section investigates the existence and calculation of the directional derivative, which we use in Section 5 to formulate a solution method for bilevel program (BL).

In the remainder, we repeatedly use the following assumption (Assumption 2), which states that the Jacobian of the link cost function is a positive definite matrix. This assumption is stronger than necessary for some upcoming results, and that it does not follow directly from Assumption 1 (e.g., when using the Bureau of Public Roads-function (Bureau of Public Roads, 1964) with $x_e = 0$, for some $e \in E$). See Lu (2008) for conditions that can replace Assumption 2.

Assumption 2. Assumption 2 is said to hold at x^0 if $\nabla_x^2 z_0(x)(= \nabla_x l(x))$ is a positive definite matrix at x^0.

Let $\rho^0 \in \Xi$ be in the remainder of this section a *reference value* and we consider *reference point* (ρ^0, x^0), with $x^0 \in S^x(\rho^0)$.

We prove that the *Karush-Kuhn-Tucker* (KKT)-set mapping corresponding to $Q(\rho)$ is an upper Lipschitz continuous multifunction at ρ^0, given that Assumption 2 holds at x^0. Consider therefore the system of KKT optimality conditions for $Q(\rho)$. For each ρ, this system can be written as

$$\begin{array}{ll} l(x) - \beta = 0 & g^T \gamma = 0 \\ \Delta^T \beta - \gamma - \Lambda^T \lambda + \rho = 0 & (g, x) \in \mathcal{F}, \end{array} \tag{5}$$

with accompanying Lagrange multiplier vector $\phi := (\beta, \lambda, \gamma), \gamma \geq 0$. The KKT-set mapping $S_{KKT}(\rho)$ is the function that maps ρ onto the set of (g, x, ϕ) that satisfies (5), i.e., for $\rho \in \Xi$:

$$S_{KKT}(\rho) = \left\{(g, x, \phi) \;\middle|\; (g, x, \phi) \text{ satisfies } (5)\,, \gamma \geq 0\right\}.$$

In our context, the Lagrange multiplier vector ϕ is uniquely determined at ρ^0. Indeed, for each fixed ρ^0, $x^0 = S^x(\rho^0)$ is a singleton, which implies that $l(x^0)$ and thus β^0 are uniquely determined (with $(g^0, x^0, \phi^0) \in S_{KKT}(\rho^0)$). Whereas ρ^0 is fixed, and there exists at least one $p \in \mathcal{P}_k$, for which $\gamma_p^0 = 0$ (which is true by $d_k > 0$) for all $k \in \mathcal{K}$, it follows that also λ^0 (and thus γ^0) are uniquely determined given ρ^0.

We state the main result of this section (Theorem 2): $S_{KKT}(\rho)$ is (locally) upper Lipschitz continuous at ρ^0. We moved the (rather technical) proof to the Appendix.

Theorem 2. *Let Assumption 2 hold at x^0, the multifunction $S_{KKT}(\rho)$ is upper Lipschitz continuous at $\rho^0 \in \Xi$.*

We need the auxiliary result of this section in the upcoming subsections to prove existence of the directional derivative $x'(\rho^0; r)$, under Assumption 2 at x^0.

4.1. Directional derivative of the link flow solution

This and upcoming subsections (Section 4.2 and 4.3) are devoted to treat the existence and calculation of the directional derivative

$$x'(\rho^0; r) = \lim_{t \to 0^+} \frac{x(\rho^0 + tr) - x^0}{t},$$

with $x^0 = x(\rho^0)$, since in particular the link flows are of interest for authorities (i.e., the upper-level objective function $\varphi(x)$ in (BL) is a function of x). Some of our arguments are taken from Dempe (1993) and Pang & Ralph (1996).

Let ρ^0 be the reference value and $r \in \mathbb{R}^{|\mathcal{P}|}$, $\|r\| = 1$, is an arbitrary direction. Let $t^k > 0$, $k \in \mathbb{N}$, so that $t^k \to 0$. From previous analysis (Theorem 2), we know that, if Assumption 2 holds at x^0, for each

$$(g^k, x^k, \phi^k) \in S_{KKT}(\rho^k), \quad \rho^k := \rho^0 + t^k r,$$

exists

$$(\tilde{g}^k, x^0, \phi^0) \in S_{KKT}(\rho^0) \tag{6}$$

so that

$$\frac{(g^k, x^k, \phi^k) - (\tilde{g}^k, x^0, \phi^0)}{t^k} \tag{7}$$

is a bounded sequence, and thus has (for a certain subsequence) a limit point $w = (w^g, w^x, w^\phi)$. We investigate whether w^x of w is unique and independent of the choices of t^k and $\tilde{g}^k$.

The complexity of the analysis lies in the fact that $S^g(\rho)$ is only upper semicontinuous at ρ^0. Intuitively, for some $\rho^k \to \rho^0$, not all $g \in S^g(\rho^0)$ can be *reached* by some (sub)sequence $g^k \in S^g(\rho^k)$. We follow the strategy of Dempe (1993), and introduce *reachable set* $V(S^g(\rho^0); r)$ of $S^g(\rho)$ at $\rho^0 \in \Xi$ into direction r:

$$\begin{aligned} V(r) &= V(S^g(\rho^0); r) \\ &= \left\{ g \in \mathbb{R}^{|\mathcal{P}|} \;\middle|\; \begin{array}{l} \text{exists sequence } t^k > 0, k \in \mathbb{N}, t^k \to 0, \\ \text{and } g^k \in S^g(\rho^k) \text{ so that } g^k \to g \end{array} \right\}. \end{aligned}$$

We first show that $V(r)$ is nonempty, and that it is a subset of $SP(r)$ (and thus $S^g(\rho^0)$) (*cf.* Dempe, 1993). $SP(r)$ is the solution set corresponding to problem $P(r)$ with parameter r, i.e.,

$$SP(r) = \left\{g \in S^g(\rho^0) \;\middle|\; g \text{ solves } P(r)\right\}.$$

Lemma 1. *For arbitrary direction r, $\|r\| = 1$:*

$$\emptyset \neq V(r) \subseteq SP(r) \subseteq S^g(\rho^0). \tag{8}$$

Proof. We prove the lemma in two parts. First, we prove that $\emptyset \neq V(r)$, and then we prove that $V(r) \subseteq SP(r)$. It is trivial that $SP(r) \subseteq S^g(\rho^0)$.

($\emptyset \neq V(r)$). Consider ρ^k, $k \in \mathbb{N}$, so that ρ^k converges to ρ^0. Choose $g^k \in S^g(\rho^k)$. Since $\|g^k\|$ is bounded, there exists subsequence g^{k_j} of g^k so that g^{k_j} converges to some g^0. $S^g(\rho)$ is a closed mapping at ρ^0, and thus $g^0 \in S^g(\rho^0)$. So, $V(r) \neq \emptyset$.

($V(r) \subseteq SP(r)$). Choose any $g^0 \in V(r)$. By definition, there exists $g^k \in S^g(\rho^k)$ so that $g^k \to g^0 \in S^g(\rho^0)$. In the proof of Proposition 2, we established that

$$r^T g^0 \geq v'(\rho^0; r) = \lim_{k \to \infty} \frac{v(\rho^k) - v(\rho^0)}{t^k} \geq \lim_{k \to \infty} r^T g^k = r^T g^0$$

So, $r^T g^0 = v'(\rho^0; r) = \min_{g \in S^g(\rho^0)} r^T g$. That is, $g^0 \in SP(r)$. □

In general, it holds that $V(r)$ is a proper subset of $S^g(\rho^0)$ (as we show in Example 1 in Section 4.3), and $V(r) = S^g(\rho^0)$ follows if $S^g(\rho)$ is lower semicontinuous at ρ^0. $S^g(\rho)$ is lower semicontinuous relative to its domain if ρ is a parameter in the link cost function (see Lu & Nie, 2010).

Lemma 2. *Let Assumption 2 hold at x^0. For direction r, $\|r\| = 1$, any limit point w of (7) satisfies the following system:*

$$\begin{array}{l} 0 = \Delta^T(\nabla_x l(x^0)w^x) + \rho^0 - \Lambda^T w^\lambda - \sum_{p\in I(g^0)}(w_p^\gamma)\mathbb{1}_p;\\ \Delta w^g - w^x = 0, \Lambda w^g = 0;\\ w_p^g = 0, p \in I(g^0) : \gamma_p^0 > 0;\\ w_p^g \geq 0, p \in I^0;\\ w_p^\gamma \geq 0, p : \gamma_p^0 = 0;\\ w_p^\gamma w_p^g = 0, p \in I^0, \end{array} \tag{9}$$

for some $I^0 \subseteq I(g^0)$, with $g^0 \in V(r)$.

Here, $\mathbb{1}_p \in \{0,1\}^{|\mathcal{P}|}$ is the indicator vector. $I(g) \subseteq \mathcal{P}$ denotes the active index set at $g \in \mathcal{F}^g$:

$$I(g) = \{\ p \in \mathcal{P}\ |\ g_p = 0\ \}.$$

Proof. We prove this lemma in three parts. In the first part of the proof, we prove that for $g^0 \in V(r)$, $\tilde{g}^k$ of (6) converges to g^0. In Part 2, we prove that the limit point w satisfies the first equality of (9), that w satisfies the (in)equalities of (9) is proven in Part 3.

(Part 1). Note that we can assume (by passing to a subsequence) that $g^k \to g^0$, i.e., $g^0 \in V(r)$. We prove that for $g^0 \in V(r)$, $\tilde{g}^k$ as in (6) converges to g^0. Let $g^0 \in V(r)$. By definition there exists a sequence $t^k > 0$, with $t^k \to 0$, and $g^k \in S^g(\rho^k)$ so that $g^k \to g^0$. Since

$$\|(g^k, x^k, \phi^k) - (\tilde{g}^k, x^0, \phi^0)\| \to 0,$$

and $g^k \to g^0$ as $k \to \infty$, it follows that $\tilde{g}^k \to g^0$.

(Part 2). Consider the set $S_{KKT}(\rho^k)$ for each $k \in \mathbb{N}$. Recall, for each $k \in \mathbb{N}$, and $g^k \in S^g(\rho^k)$ exists unique (x^k, ϕ^k) so that $(g^k, x^k, \phi^k) \in S_{KKT}(\rho^k)$. That is, for each k, (g^k, x^k, ϕ^k) satisfies the KKT conditions that correspond to $Q(\rho^k)$, i.e., with $\gamma^k \geq 0$,

$$\begin{array}{ll} l(x^k) - \beta^k = 0 & (g^k)^T\gamma^k = 0\\ \Delta^T\beta^k + \rho^0 + t^k r - \Lambda^T\lambda^k - \gamma^k = 0 & (g^k, x^k) \in \mathcal{F}. \end{array} \tag{10}$$

Since $g^0 \in V(r)$, $g^k \to g^0$ with $g^0 \in S^g(\rho^0)$. Hence, $g_p^0 > 0$ implies $g_p^k > 0$ for sufficiently large k, and thus $I(g^k) \subseteq I(g^0)$ for these k. Now, we can rewrite the first three KKT conditions in (10) as

$$0 = \Delta^T l(x^k) + \rho^0 + t^k r - \Lambda^T\lambda^k - \sum_{p\in I(g^0)} \gamma_p^k \mathbb{1}_p. \tag{11}$$

Taylor's expansion of $l(x)$ around x^0 says that

$$l(x^k) = l(x^0) + \nabla_x l(x^0)(x^k - x^0) + o(\|x^k - x^0\|), \tag{12}$$

where $o(\|x^k - x^0\|)/t^k$ converges to zero for $k \to \infty$.

We repeat a similar argument for $(\tilde{g}^k, x^0, \phi^0)$. We have that $\tilde{g}^k$ converges to g^0, and thus $I(\tilde{g}^k) \subseteq I(g^0)$ for large k. The KKT conditions of $Q(\rho^0)$ say that $(\tilde{g}^k, x^0, \phi^0) \in S_{KKT}(\rho^0)$ satisfies (at least) the following condition for sufficiently large k (using the uniqueness of γ^0):

$$0 = \Delta^T l(x^0) + \rho^0 - \Lambda^T\lambda^k - \sum_{p\in I(g^0)} \gamma_p^0 \mathbb{1}_p. \tag{13}$$

Subtracting (13) from (11), and using the Taylor expansion (12), we obtain

$$\begin{aligned} 0 &= \Delta^T(\nabla_x l(x^0)(x^k - x^0)) + t^k r - \Lambda^T(\lambda^k - \lambda^0)\\ &\quad - \sum_{p\in I(g^0)} (\gamma_p^k - \gamma_p^0)\mathbb{1}_p + o(\|x^k - x^0\|). \end{aligned} \tag{14}$$

We divide (14) by t^k, using that the quotient in (7) is bounded by upper Lipschitz continuity of $S_{KKT}(\rho)$ at ρ^0, then the limit point w of (7) satisfies (at least) the following equation:

$$0 = \Delta^T(\nabla_x l(x^0)w^x) + r - \Lambda^T w^\lambda - \sum_{p\in I(g^0)} (w_p^\gamma)\mathbb{1}_p. \tag{15}$$

(Part 3). The last KKT condition in (5) for ρ^k and ρ^0 says that for each $k \in \mathbb{N}$, $(g^k, x^k) \in \mathcal{F}$ and $(\tilde{g}^k, x^0) \in \mathcal{F}$. Therefore,

$$\Delta w^g - w^x = 0, \quad \text{and} \quad \Lambda w^g = 0.$$

Also, for any $p \in \mathcal{P}$, we find that (for a subsequence)

$$\left(\frac{g^k - \tilde{g}^k}{t^k}\right)_p \to w_p^g \begin{cases} = 0, & \text{if } \gamma_p^0 > 0\\ \geq 0, & \text{if } \gamma_p^0 = 0 \text{ and s.t. exist infinitely many } k \text{ with } \tilde{g}_p^k = 0, \end{cases}$$

which yields

$$w_p^g \begin{cases} = 0, & p \in I(g^0) : \gamma_p^0 > 0\\ \geq 0, & p \in I^0, \end{cases} \tag{16}$$

for some $I^0 \subseteq I(g^0)$.

By the non-negativity constraint with respect to multiplier γ in (5), in combination with the fact that γ is a singleton for each ρ, we have that for $p \in \mathcal{P}$,

$$\left(\frac{\gamma^k - \gamma^0}{t^k}\right)_p \to w_p^\gamma \quad \geq 0, \quad \text{if } \gamma_p^0 = 0.$$

Finally, note that also a complementarity condition arises:

$$w_p^\gamma w_p^g = 0, \quad \text{for all } p \in I^0. \tag{17}$$

□

We recall that, in order to determine whether directional derivative $x'(\rho^0; r)$ exists, we have to show that the limit point w^x of $\frac{x^k - x^0}{t^k}$ does not depend on choices of $t^k, g^k, \tilde{g}^k$, and $I^0 = I^0(\tilde{g}^k)$. Based on the result as presented in Lemma 2, even in the case that $V(r)$ is a singleton, different choices of I^0 could possibly lead to different solutions w^x of (9). In the following section, we present a method that finds $x'(\rho^0; r)$ without the trouble finding an appropriate I^0.

4.2. A quadratic program reformulation

Recall reference point (ρ^0, x^0). As mentioned, even if $V(r)$ is a singleton ($V(r) = \{g^0\}$), different $I^0 \subseteq I(g^0)$ in (16), (17), might be possible, which makes it difficult to calculate (a) limit point w. In this subsection, we demonstrate that, under the assumption that $V(r) = \{g^0\}$, w^x is actually independent of I^0 and can be found efficiently by solving a convex optimization problem.

Before we continue, we define $T_{\mathcal{F}}(g, x)$ as the *tangent cone* to $\mathcal{F}$ at $(g, x) \in \mathcal{F}$, i.e.,

$$T_{\mathcal{F}}(g, x) = \left\{ (w^g, w^x) \in \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} \;\middle|\; \begin{array}{ll} \Lambda w^g = 0 & \\ \Delta w^g - w^x = 0 & \\ w_p^g \geq 0 & p \in I(g) \end{array} \right\}.$$

We introduce the following parametric (convex) quadratic optimization problem (with parameter (g^0, r), and for now $V(r) = \{g^0\}$):

$$QP(g^0, r): \ \min_w \frac{1}{2}(w^x)^T A w^x + r^T w^g \quad s.t. \quad (w^g, w^x) \in \mathcal{C}(g^0, x^0, \phi^0),$$

where $A := \nabla_x l(x^0) = \nabla_x^2 z_0(x^0)$, and

$$\mathcal{C}(g^0, x^0, \phi^0) := T_{\mathcal{F}}(g^0, x^0) \cap T_{\mathcal{D}(\phi^0)}(g^0)$$

is the *critical cone* to $\mathcal{F}$ at (g^0, x^0, ϕ^0). Here,

$$T_{\mathcal{D}(\phi^0)}(g^0) = \left\{ w^g \in \mathbb{R}^{|\mathcal{P}|} \;\middle|\; w_p^g = 0, p \in \mathcal{P} : \gamma_p^0 > 0 \right\}$$

is the tangent cone to

$$\mathcal{D}(\phi^0) = \left\{ g \in \mathbb{R}^{|\mathcal{P}|} \;\middle|\; g_p = 0, p \in \mathcal{P} : \gamma_p^0 > 0 \right\},$$

at g^0. Under Assumption 2 at x^0, $QP(g^0, r)$ is a convex problem (strictly convex in w^x).

Lemma 3. *Let Assumption 2 hold at x^0. For direction r, $\|r\| = 1$, for which $V(r) = \{g^0\}$, w^x of any limit point w of (7) is the (global) optimal solution of $QP(g^0, r)$.*

Proof. Consider a limit point w of (7). We prove that the w^x-part of w is the optimal solution of $QP(g^0, r)$ with $V(r) = \{g^0\}$. Therefore, we first show that w^x, with accompanying w^g, is a feasible solution of $QP(g^0, r)$, then we prove (w^g, w^x) is a global optimal solution of $QP(g^0, r)$.

(Feasibility). For given direction r and $V(r) = \{g^0\}$, with $g^k \to g^0$, we note that for sufficiently large k, $(g^k, x^k) \in S(\rho^k)$ are also optimal solutions to

$$\tilde{Q}(\rho): \qquad \min z_0(x) + \rho^T g \quad s.t. \quad (g, x) \in \tilde{\mathcal{F}} := \mathcal{F} \cap \mathcal{D}(\phi^0),$$

with $\rho = \rho^k$. So, $x^k \in \tilde{\mathcal{F}}^x$ (the projection of $\tilde{\mathcal{F}}$ onto the x-space) for all these k, and therefore w^x of any limit point w of (7) satisfies $w^x \in T_{\tilde{\mathcal{F}}^x}(x^0)$. Since $\tilde{\mathcal{F}}^x$ is a polyhedral set (the projection of a polyhedral set is a polyhedral set), $x^1 = x^0 + \alpha w^x \in \tilde{\mathcal{F}}^x$ for some $\alpha > 0$. Hence, exists $g^1 \in \tilde{\mathcal{F}}^g$ so that $\Delta g^1 = x^1$ (see Rockafellar & Wets, 2009, Theorem 6.43). Now, let $\bar{w}^g = \frac{g^1 - g^0}{\alpha}$, then $\bar{w}^g \in T_{\tilde{\mathcal{F}}}(g^0)$, since $g^1 = g^0 + \alpha \bar{w}^g \in \tilde{\mathcal{F}}^g$. In particular, it holds that $\bar{w}^g_p \geq 0$ for all $p \in I(g^0)$. Thus, limit point w^x of (7) is in the feasible set $\mathcal{C}^x(g^0, x^0, \phi^0)$. We underline that $\bar{w}^g$ is different from the w^g-part of w in (7), i.e., it might hold that $\bar{w}^g \neq w^g$.

(Optimality). We showed that w^x of the limit point of (7) with an accompanying $\bar{w}^g$ is a feasible solution of $QP(g^0, r)$. Now, we demonstrate that $(\bar{w}^g, w^x)$ is the optimal solution of $QP(g^0, r)$.

Note from (15) that there exists w^ϕ so that (w^x, w^ϕ) satisfies

$$\Delta^T (\nabla_x l(x^0) w^x) + r - \Lambda^T w^\lambda - \sum_{p \in I(g^0)} (w^\gamma_p) \mathbb{1}_p = 0$$

Then let $(u^g, u^x) \in \mathcal{C}(g^0, x^0, \phi^0)$ be arbitrary, we find that

$$\begin{aligned} 0 &= (\Delta^T (\nabla_x l(x^0) w^x) + r)^T u^g - (\Lambda^T w^\lambda)^T u^g \\ &\quad - \sum_{p \in I(g^0)} ((w^\gamma_p) \mathbb{1}_p)^T u^g \\ &\leq (\Delta^T (\nabla_x l(x^0) w^x) + r)^T u^g, \end{aligned}$$

which is exactly the first-order optimality condition of convex problem $QP(g^0, r)$. Note that the latter inequality holds whereas $w^\gamma_p < 0$ for some $p \in I(g^0)$ implies that $\gamma^0_p > 0$, and thus $u^g_p = 0$. Since $\mathcal{C}(g^0, x^0, \phi^0)$ is a polyhedral cone, and by strict convexity of the objective function in $QP(g^0, r)$ with respect to w^x, the limit point w^x is contained in the optimal solution w (unique with respect to w^x) of $QP(g^0, r)$. We show in the proof accompanying Lemma 4 that the optimal solution is bounded. □

For any given direction r ($\|r\| = 1$), in combination with the extra assumptions that $|V(r)| = 1$ and Assumption 2 holds at x^0, we proved that the directional derivative $x'(\rho^0; r)$ exists. This directional derivative is the optimal solution (with respect to w^x) of $QP(g^0, r)$ with $V(r) = \{g^0\}$. An opportunity to force uniqueness of $V(r)$ (and also $S^g(\rho^0)$) is to include a regularization term in the objective function of the lower-level problem.

4.3. $V(r)$ Not a singleton

The more interesting case occurs when $V(r)$ is not a singleton. Note that only a finite number of different $I(g^0)$, $g^0 \in V(r)$, can occur, and, under Assumption 2 at x^0, finitely many w^x exist.

The previous analysis in Section 4.2 relied on the choice of $g^0 \in V(r)$. One might ask the question whether we can choose any $g^0 \in S^g(\rho^0)$, and solve $QP(g^0, r)$ to obtain directional derivative w^x, if it exists. The following example illustrates that an arbitrary $g^0 \in S^g(\rho^0)$ may lead to an unbounded solution of $QP(g^0, r)$.

Example 1 (Unbounded Solutions)**.** In this example, we show that optimization program $QP(g^0, r)$ with $g^0 \in S^g(\rho^0) \setminus V(r)$, may have a corresponding unbounded solution.

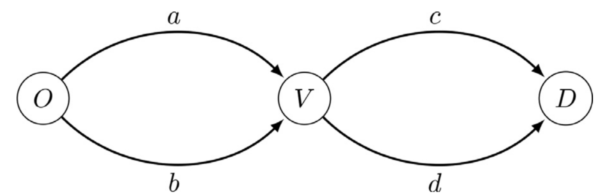


Fig. 1. Example traffic network.

Fig. 1 shows the single OD pair ($|\mathcal{K}| = 1$) network we consider. The network has 4 links with travel time function $l_e(x_e) = x_e$ for all $e \in E$. Demand for the OD pair is 1. The paths

$$p_1 = \{a, c\}, \quad p_2 = \{a, d\}, \quad p_3 = \{b, c\}, \quad \text{and} \quad p_4 = \{b, d\},$$

connect the OD pair (O, D). Define $\rho = (\rho_{p_1}, \rho_{p_2}, \rho_{p_3}, \rho_{p_4})$, and let

$$\rho(t) = t \cdot r, \quad \text{with } r = (1, 0, 0, 0), \quad \text{and } t \in [0, 1],$$

for the sake of this example. We solve $Q(\rho(0))$: the traditional user equilibrium problem (Beckmann, McGuire, & Winsten, 1956). We denote this solution with respect to x by x^n and find

$$x^n = (x^n_a, x^n_b, x^n_c, x^n_d) = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right).$$

Since the link cost functions are strictly increasing, we find the optimal solution vector $x(\rho(t))$ as a function of t: $x(\rho(t)) = x^n, t \in [0, 1]$. Consider $S^{g_{p_1}}(\rho(t))$, the route flow solution g on path p_1, as a multifunction of t:

$$S^{g_{p_1}}(\rho(t)) = \begin{cases} [0, \frac{1}{2}] & \text{if } t = 0; \\ 0 & \text{if } t \in (0, 1]. \end{cases}$$

It is clear that $S^{g_{p_1}}(\rho(t))$ is not a lower semicontinuous function at $t = 0$. Moreover, choose $g^0 \in S^g(\rho(0))$ so that $g^0_{p_1} > 0$. It is easy to check that $QP(g^0, r)$ gives an unbounded solution for $r = (1, 0, 0, 0)$. In fact, $g^0 \notin V(r)$ and observe that g^0 is not a solution of $P(r)$.

Example 1 illustrates the practical difficulties calculating the directional derivative. In fact, if we choose $g^0 \in S^g(\rho^0)$ arbitrarily, we might not be able calculate $x'(\rho^0; r)$ using $QP(g^0, r)$ (even if it exists - see Theorem 3). We should select therefore $g^0 \in S^g(\rho^0)$ carefully. From a practitioner's perspective, this result is undesirable since some $g^0 \in S^g(\rho^0)$ is often a by-product of the algorithm that solves $Q(\rho^0)$. In the upcoming analysis, we prove that $g^0 \in S^g(\rho^0)$ could be selected so that $g^0 \in SP(r)$.

Lemma 4. *Let Assumption 2 hold at x^0. For arbitrary r, $\|r\| = 1$, $QP(g^0, r)$, with $g^0 \in SP(r)$, has a bounded solution w which is unique in w^x.*

Proof. Let $g^0 \in SP(r)$, and $(g^0, x^0, \phi^0) \in S_{KKT}(\rho^0)$. From Corollary 2.1 in Lee, Tam, & Yen (2006) it follows that $QP(g^0, r)$ has a solution if and only if

$$\left.\begin{array}{l} (u^g, u^x), (w^g, w^x) \in \mathcal{C}(g^0, x^0, \phi^0) \\ (u^x)^T A u^x = 0 \end{array}\right\} \Rightarrow (u^x)^T A w^x + r^T u^g \geq 0. \quad (18)$$

By Assumption 2, A is a positive definite matrix, and $(u^x)^T A u^x = 0$ implies $u^x = 0$ and it automatically follows that $(u^x)^T A w^x = 0$. Suppose now that the right-hand side of (18) is not satisfied, i.e., $r^T u^g < 0$ for some $(u^g, u^x) \in \mathcal{C}(g^0, x^0, \phi^0)$. Note that

$$\nabla_x z_0(x^0)^T u^x + (\rho^0)^T u^g = 0, \quad (19)$$

for any $(u^g, u^x) \in \mathcal{C}(g^0, x^0, \phi^0)$ (see Luo et al., 1996, p. 225). Since $u^x = 0$, by (19), $(\rho^0)^T u^g = 0$. So, for small $t > 0$, $(g^0 + t u^g) \in S^g(\rho^0)$ and $r^T(g^0 + t u^g) < r^T g^0$, which contradicts that $g^0 \in SP(r)$. The uniqueness of w with respect to w^x can then be concluded from the fact that A is positive definite matrix and that $\mathcal{C}(g^0, x^0, \phi^0)$ is a polyhedral cone. □

Hence, selecting $g^0 \in SP(r)$ makes that the issue as described in Example 1 cannot occur. Now, we prove the main result of the paper. For direction r, rather than explicitly using $V(r)$, we can choose an arbitrary $g^0 \in SP(r)$ to calculate directional derivative $x'(\rho^0; r)$ of $x(\rho)$ at ρ^0.

Theorem 3. *Let Assumption 2 hold at x^0. For arbitrary direction r, $\|r\| = 1$, $x'(\rho^0; r)$ exists and is the optimal solution (with respect to w^x) of optimization problem $QP(g^0, r)$, $g^0 \in SP(r)$.*

Proof. Based on Lemma 3 and 4, we only need to prove that for any r the solution w^x of w that corresponds to $QP(g^0, r)$ is independent of the choice $g^0 \in SP(r)$. Assume r to be fixed, and let $g^1 \neq g^2 \in SP(r)$. Suppose $(w^{g,1}, w^{x,1})$ solves $QP(g^1, r)$, and $(w^{g,2}, w^{x,2})$ solves $QP(g^2, r)$, but $w^{x,1} \neq w^{x,2}$. Note that both problems have an optimal solution by Lemma 4.

We may assume, without loss of generality, that

$$\frac{1}{2}(w^{x,1})^T A(w^{x,1}) + r^T w^{g,1} \leq \frac{1}{2}(w^{x,2})^T A(w^{x,2}) + r^T w^{g,2}.$$

Since $w^{x,1} \neq w^{x,2}$, and the optimal solution of $QP(g^2, r)$ is unique with respect to $w^{x,2}$, we have

$$\frac{1}{2}(w^{x,1})^T A(w^{x,1}) + r^T w^{g,1} < \frac{1}{2}(w^{x,1})^T A(w^{x,1}) + r^T \bar{w}^{g,2}, \tag{20}$$

for all $\bar{w}^{g,2}$ so that $(\bar{w}^{g,2}, w^{x,1}) \in \mathcal{C}(g^2, x^0, \phi^0)$. It directly follows from (20) that $r^T w^{g,1} < r^T \bar{w}^{g,2}$ for all such $\bar{w}^{g,2}$, given that there exist such $(\bar{w}^{g,2}, w^{x,1}) \in \mathcal{C}(g^2, x^0, \phi^0)$.

Note that for all sufficiently small $\alpha > 0$, $g^1 + \alpha w^{g,1} \in \tilde{\mathcal{F}}$. Hence, for any such α, let

$$\bar{w}^{g,2} = \frac{g^1 + \alpha w^{g,1} - g^2}{\alpha}.$$

Then, $g^2 + \alpha \bar{w}^{g,2} \in \tilde{\mathcal{F}}$, hence $(\bar{w}^{g,2}, w^{x,1}) \in \mathcal{C}(g^2, x^0, \phi^0)$. Since $r^T(g^1 + \alpha w^{g,1}) = r^T(g^2 + \alpha \bar{w}^{g,2})$, it follows that $r^T g^1 > r^T g^2$, which contradicts that $g^1 \in SP(r)$. □

Theorem 3 proves that $x'(\rho^0; r)$ exists for any ρ^0 in any direction r, $\|r\| = 1$, provided that Assumption 2 holds globally (i.e., for all $x(\rho^0)$ with $\rho^0 \in \Xi$). Now, for $\rho^0 \in \Xi$, we can estimate $x(\rho^1) \approx x^0 + t x'(\rho^0; r)$, with $\rho^1 = \rho^0 + tr$, $t > 0$ small, and $\|r\| = 1$. To do so, we have to choose $g^0 \in SP(r)$, and subsequently solve $QP(g^0, r)$. We use this result to formulate an optimization method for (BL) in Section 5.

We compare the result of Theorem 3 with Theorem 2 in Ralph & Dempe (1995). There, the directional derivative of a solution of a parametric nonlinear program (with a locally unique minimizer) can be calculated (under a constraint qualification) by selecting a suitable KKT multiplier as a solution of auxiliary program. In our case, we have a non-unique solution, and need a linear program to find directional derivative $x'(\rho^0; r)$ of the link flows $x(\rho)$ at ρ^0.

4.4. General results

In previous sections, we assumed $d^n = 0$. We extend the results to the case $d^n \neq 0$. We omit the corresponding proofs which are straightforward extensions of the proofs in previous sections.

For $\rho^0 \in \Xi$ and r, $\|r\| = 1$, arbitrary, and $(g^0, h^0, x^0) \in S(\rho^0)$, the linear program

$$P(r): \quad \min_{g,h} r^T g \quad s.t. \quad (g, h) \in S^{(g,h)}(\rho^0),$$

finds $(g^0, h^0) \in SP(r)$. The quadratic (convex) optimization problem to find directional derivative $x'(\rho^0; r)$ of $x(\rho)$ at ρ^0 in direction r corresponds to

$$QP(g^0, h^0, r): \quad \min \frac{1}{2} w^{x^T} A w^x + r^T w^g \quad s.t. \quad (w^g, w^h, w^x) \in \mathcal{C}(g^0, h^0, x^0),$$

with

$$\mathcal{C}(g^0, h^0, x^0) = \left\{ \begin{array}{l|ll} (w^g, w^h, w^x) \in \mathbb{R}^{|\mathcal{P}|} & w_p^g \geq 0 & p \in \mathcal{P}_{g,1} \\ & w_p^g = 0 & p \in \mathcal{P}_{g,2} \\ & w_p^h \geq 0 & p \in \mathcal{P}_{h,1} \\ \times \mathbb{R}^{|\mathcal{P}|} \times \mathbb{R}^{|E|} & w_p^h = 0 & p \in \mathcal{P}_{h,2} \\ & \Delta(w^g + w^h) - w^x = 0 & \\ & \Lambda w^g = 0 & \\ & \Lambda w^h = 0 & \end{array} \right\}.$$

Here, we decompose path set $I(g^0) \subseteq \mathcal{P}$, $I(h^0) \subseteq \mathcal{P}$, as follows

$$\begin{aligned} \mathcal{P}_{g,1} &= \{p \in \mathcal{P}_k, k \in \mathcal{K} \mid p \in I(g_0), (c_p(x^0) + \rho_p^0) - \min_{q \in \mathcal{P}_k}(c_q(x^0) + \rho_q^0) = 0\}, \\ \mathcal{P}_{g,2} &= I(g^0) \setminus \mathcal{P}_{g,1}, \\ \mathcal{P}_{h,1} &= \{p \in \mathcal{P}_k, k \in \mathcal{K} \mid p \in I(h_0), c_p(x^0) - \min_{q \in \mathcal{P}_k} c_q(x^0) = 0, \} \\ \mathcal{P}_{h,2} &= I(h^0) \setminus \mathcal{P}_{h,1}. \end{aligned}$$

$\mathcal{P}_{g,2}, \mathcal{P}_{h,2}$ are the path sets that consist of the paths with an accompanying positive multiplier. Note that $QP(g^0, h^0, r)$ can be interpreted as a traffic assignment problem with a restricted path set (*cf.* Patriksson (2004)). In comparison with $Q(\rho)$, the link cost function is linear, some paths might carry negative flows, and each OD pair has zero demand.

5. Algorithm and numerical experiments

Thus far, we proved the existence of the directional derivative of the link flows under perturbations in the parameter, and found a constructive method to calculate it. In this section, we solve optimization problem (BL) using a feasible descent method. The algorithm is so that we solely need to solve convex optimization problems and, thus, it can be implemented in standard optimization toolboxes.

5.1. Algorithm

Consider optimization problem (BL). We can reformulate it as (BL'), a *nonsmooth* optimization program in which x is an implicit function of ρ, i.e.,

$$(BL'): \quad \min_{\rho} \varphi(x(\rho)) \quad s.t. \quad \rho \in \Xi.$$

Consider ρ^0 with solution $x^0 = x(\rho^0)$ of lower-level problem $Q(\rho^0)$. We proved that the directional derivative $x'(\rho^0; r)$ into direction r exists, i.e., for $t > 0$ small,

$$\begin{aligned} \varphi(x(\rho^0 + tr)) - \varphi(x^0) &= \nabla_x \varphi(x^0)^T (x(\rho^0 + tr) - x^0) \\ &= t \nabla_x \varphi(x^0)^T x'(\rho^0; r). \end{aligned} \tag{21}$$

So, any direction r, $\|r\| = 1$, that satisfies $\nabla_x \varphi(x^0)^T x'(\rho^0; r) < 0$ yields a descent direction for (BL'). This allows us to formulate the necessary optimality conditions for (BL').

The calculation of a steepest descent direction r is difficult and is the optimal solution of a linear-quadratic optimization problem, which can be found using an expensive branch-and-bound technique (Bard, 1998). To reduce computational intensity and to enhance application by traffic engineers, we use an algorithm that assumes that $x(\rho)$ is differentiable at any ρ^0, i.e., $\nabla_\rho x(\rho^0)$ exists (see Josefsson & Patriksson, 2007). Algorithms that explicitly use the nonsmoothness of the objective function in (BL') can be found in Outrata, Kocvara, & Zowe (2013).

At every iteration $i \in \mathbb{N}$, with iteration point $\rho^i \in \Xi$, we find a feasible descent direction by solving convex optimization problem

$$\min_{\nu} \frac{1}{2} \| -\nabla_\rho \varphi(x(\rho^i)) - \nu \|^2 \quad s.t. \quad \nu \in D(\rho^i), \tag{22}$$

with feasible cone

$$D(\rho^i) = \left\{ \nu \in \mathbb{R}^{|\mathcal{P}|} \;\middle|\; \begin{array}{ll} \nu_p \geq 0 & p \in P^1 = \{p \in \mathcal{P} \mid \rho^i_p = 0\} \\ \nu_p \leq 0 & p \in P^2 = \{p \in \mathcal{P}_k, k \in \mathcal{K} \mid \rho^i_p = \varepsilon_k\} \end{array} \right\}.$$

Summarizing, the algorithm is as follows (based on Faigle, Kern, & Still, 2013; Josefsson & Patriksson, 2007):

Step 0 Initialize $\rho^0 \in \Xi$, $\eta > 0$ small, Armijo line search factor $\tau > 0$ and multiplier κ, set $i := 0$;
Step 1 Solve $Q(\rho^i)$ to obtain $x(\rho^i)$;
Step 2 Construct the *approximate Jacobian*, $\nabla_\rho x(\rho^i)$ by solving for each $p \in \mathcal{P}$:
- (a) let $r = \mathbb{1}_p$;
- (b) find $(g, h) \in SP(r)$, i.e., solve $P(r)$.
- (c) find w that solves $QP(g, h, r)$;
- (d) let $(\nabla x(\rho^i))_p = w^x$.

Step 3 Solve (22) to find ν^i;
Step 4 Use the inexact Armijo line search (using κ) to find $m \geq 0$ that satisfies :

$$\varphi(x(p^i)) \leq \varphi(x(\rho^i)) - \tau m((\nu^i)^T(\nabla_\rho \varphi(x(\rho^i))), \tag{23}$$

where p^i is the projection of $(\rho^i + m\nu^i)$ onto Ξ, let $\rho^{i+1} = p^i$ and $i := i + 1$, goto Step 1. If there is no such m, terminate.

5.2. Implementation and settings

We implemented our method in MATLAB, and adapted a path-based algorithm to solve $Q(\rho)$ for a fixed ρ. Therefore, we used an adapted version of the gradient projection method, with a quadratic approximation line search (Gentile, 2014; Perederieieva, Ehrgott, Raith, & Wang, 2015). We used the built-in linear programming method of MATLAB to solve $\tilde{P}(r)$ rather than $P(r)$. Here,

$$\tilde{P}(r): \qquad \min r^T g \quad s.t. \qquad (g, h) \in \tilde{S}^{(g,h)}(\rho),$$

where $\tilde{S}^{(g,h)}$ is equivalent to $S^{(g,h)}$, except that we replace

$$\rho^T g = \psi(\rho) \qquad \text{with} \qquad \rho^T g \in [\psi(\rho) - \delta, \psi(\rho) + \delta],$$

in which $\delta > 0$. To solve $QP(g, h, r)$, given (g, h, r), we use the algorithm as described by Josefsson & Patriksson (2007). In order to apply our algorithm based on sensitivity analysis, one needs to solve $Q(\rho)$ with high accuracy. Therefore, we introduced the following metric to measure accuracy (for simplicity, here assuming $d^n = 0$):

$$Acc = \frac{\sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}_k} g_p((c_p(f) + \rho) - \min (c_p(f) + \rho))}{\sum_{k\in\mathcal{K}} \sum_{p\in\mathcal{P}_k} g_p(c_p(f) + \rho)},$$

and stopped when an accuracy of 10^{-12}, or a maximum number of iterations, was achieved. In the remainder, we assumed $\delta = 5 \times 10^{-4}$ in $\tilde{P}(r)$, and used $\tau = 0.1$ and $\kappa = 0.5$ in the backtracking line search.

Two networks are implemented to provide insight into the potential of social routing in practice. We use the network of Nguyen & Dupuis (1984) ($|\mathcal{K}| = 4$), with the settings of Ohazulike, Still, Kern, & van Berkum (2013) and the demand scenario of the latter paper of 400, 800, 600, and 200, respectively. To assess performance in larger networks, we used the Sioux Falls network (Transportation Networks for Research Core Team, 2019), with $|\mathcal{K}| = 528$. For the first network the path set is known a priori, in the latter network the path set needs to be constructed iteratively while solving the bilevel problem. Therefore, we add (if necessary) the k-shortest paths ($k = 2$) for each commodity every time we accept the Armijo condition (23). To initialize the path set, we used the path set generated while solving the user equilibrium and system optimum.

The main computational burden of the presented algorithm - compared to approaches solving NDPs - is the construction of an approximate Jacobian $\nabla_\rho x(\rho^i)$, which requires $P(r)$ and $QP(g, h, r)$ to be solved $|\mathcal{P}|$ times for each *outer iteration*. In particular for dense networks with many OD pairs this might lead to increasing run times. For example, for the Sioux Falls network, we ended with about 2050 paths in the path set. Therefore, we limited the outer iterations to 25. For practical purposes, one might relieve the computation time by aggregating zones.

6. Results and management implications

We explore the potential network impacts of a social routing service adopting the proposed strategy: we apply the algorithm (Section 5.1) to two test networks (see Section 5.2). In Section 6.2, we draw some preliminary conclusions about social routing for traffic management purposes.

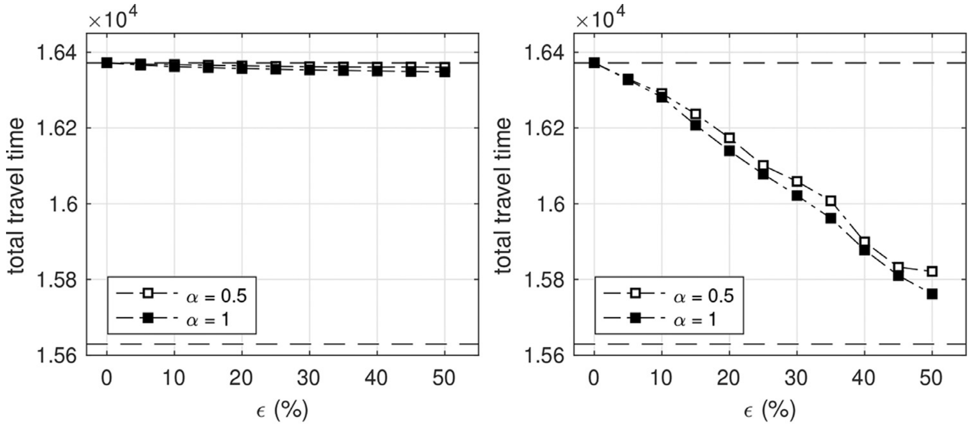
6.1. Network impact

We provide insight in the potential network efficiency, by assuming varying social trip rates d^s, and acceptable travel time differences ε. In these experiments, we assume that only a portion of the travelers is receptive for advice. Receptive drivers might be unequally distributed over the network, and, therefore, we consider for each network eight social demand scenarios. We assume that 25%, 50%, 75% or 100% of the largest OD pairs (in terms of trips) can be reached or targeted by a social routing service. Furthermore, only a portion of this demand is assumed to comply with the advice, hence we assume $d^s = \alpha d$ ($d^n = (1 - \alpha)d$)) for these OD pairs, with $\alpha \in \{\frac{1}{2}, 1\}$. To allow comparison with the unfair system optimum, we express the OD-pair dependent maximum detour ε as a percentage of the maximum detour needed in the system optimum (for the same OD pair). For each scenario we determine the distribution of social demand over the network by solving problem (BL').

Fig. 2 and 3 show the performance of the routing service (in terms of total travel time) for the Nguyen & Dupuis and Sioux Falls network, respectively, under different scenarios. In each figure, the upper and lower dashed lines depict the total travel time in user equilibrium and system optimum, respectively. In general, a larger share of social trips, and a less equitable (i.e., larger values of ε) routing strategy leads to a better performance. in terms of total travel time.

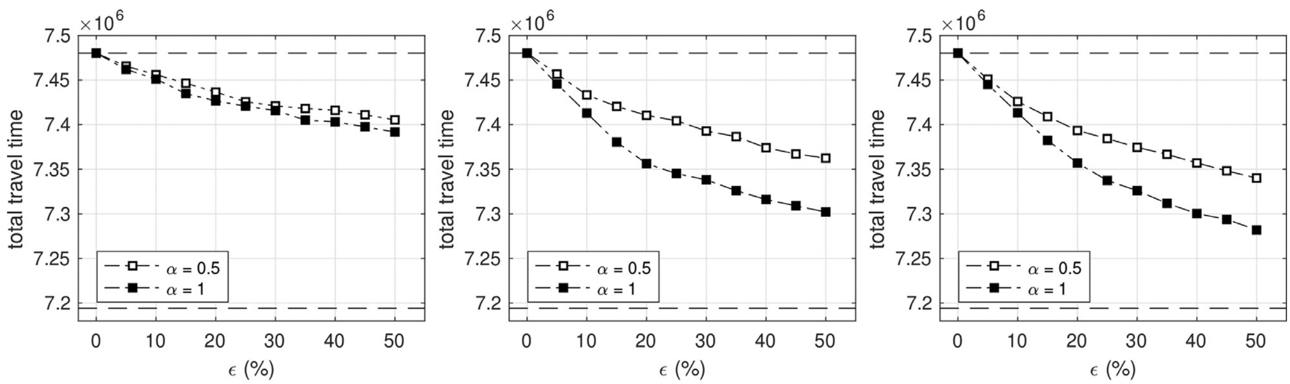
When analyzing the results for the Nguyen & Dupuis network (Fig. 2), we observe that the routing strategy is able to approach the performance of the system optimum (Fig. 2b). However, targeting the right (amount of) OD pairs is crucial, since we see in Fig. 2a almost no travel time improvement with only one OD pair reached. This can be explained by the minor detour in the system optimum for this OD pair. Further increasing the social trip rate to 75% and 100% does not substantially change performance and the corresponding results are therefore not shown. Interestingly, the compliance rate α has only limited impact on the results.

In the Sioux Falls network, the total travel time improvement is 2.7% compared to the user equilibrium (Fig. 3); the system optimum shows an improvement of 3.8%. With a compliance rate of 50%, the strategy has a maximum improvement of 1.9% in total travel time. The results with 100% of the OD pairs targeted are comparable to the results as depicted in Fig. 3c and therefore not shown. If only 25% of the largest OD pairs can be targeted by a



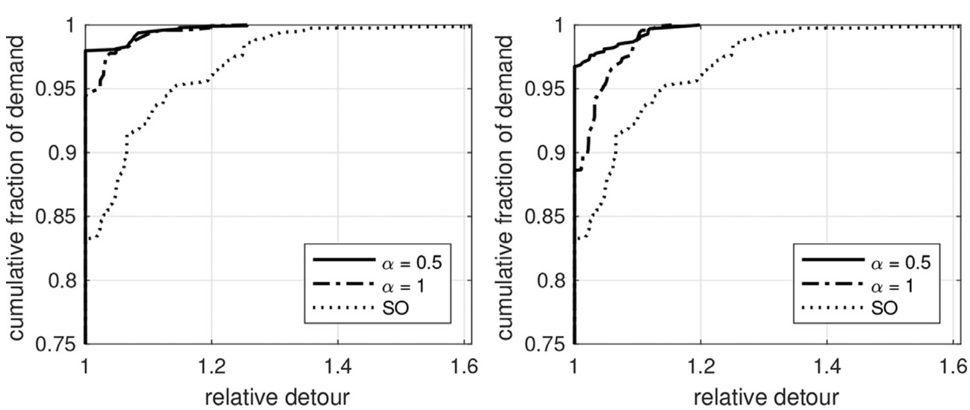
(a) 25% of the OD pairs targeted (b) 50% of the OD pairs targeted

Fig. 2. Impact varying social demand, acceptable travel time difference ε, and compliance rate α with respect to system performance in the Nguyen & Dupuis network.



(a) 25% of the OD pairs targeted (b) 50% of the OD pairs targeted (c) 75% of the OD pairs targeted

Fig. 3. Impact varying social demand, acceptable travel time difference ε, and compliance rate α with respect to system performance in the Sioux Falls network.



(a) 25% of OD pairs targeted (b) 50% of OD pairs targeted

Fig. 4. Cumulative distribution of relative travel time detours compared to the fastest paths in the Sioux Falls network. Fig. 4a corresponds to the scenario of Fig. 3a with $\varepsilon = 50\%$, Fig. 4b corresponds to the demand scenario of Fig. 3b with $\varepsilon = 50\%$.

routing service, improvements drop (Fig. 3a). Again, we observe only a minor change in total travel time when OD pairs targeted increase above 50% (compare Fig. 3b and 3 c).

In Fig. 4, we depict the cumulative distributions of the detours (in travel time, relative to the shortest path available) in the resulting states (assuming $\varepsilon = 50\%$) for the different demand scenarios. We also show the distribution of detours in the system optimum (SO). We note that in user equilibrium, all travelers take the fastest path (i.e., no detour) - see (2b). Here, we see that - although more than 50% of the drivers receive advice - only about 12% of the drivers need to take a small detour to obtain 2.7% total travel time improvement (Fig. 4b), i.e., a major share of the social travelers is still advised to take the shortest route. At the same time, the detours, if advised, are less than 26% worse compared to the fastest path. For a system-optimal assignment, detours might potentially take 60% longer. Fig. 4a shows that here only a very small fraction (2.1% of all trips) of social trips is needed to obtain already 1% improvement in total travel time.

6.2. Management implications

A real-life implementation of a social routing system adopting the proposed strategy requires a travel information service, using, e.g., a smartphone application. Based on the market penetration rate, (expected) compliance rate, and acceptable travel time differences, a central system calculates the paths for each user by solving (*BL*). These paths, provided to the drivers, are the best possible ones for the traffic system while meeting user constraints alongside. Based on the results of Section 6.1, we provide some preliminary management implications.

The numerical experiments show that a social routing system is a potential powerful measure to improve efficiency, and preserve fairness at the same time. Even if a small portion of travelers can be rerouted onto social routes, the resulting traffic state might show a major improvement in total travel time compared to the user equilibrium.

We note that the spatial distribution of the social travelers, in combination with the maximum acceptable travel time difference of users, might highly impact the strategy's performance. In the experiments, advised detours are usually fairly limited which is expected to lead to high compliance rates. In addition, travelers can be motivated to take a detour, e.g., by providing rewards. Obviously, also autonomous vehicles might be routed onto such paths (Speranza, 2018).

Even for the relatively simple setting we considered in this study, finding the optimal solution of the bilevel problem is highly complex. The algorithm as proposed in Section 5.1 finds an improving solution over the iterations. This procedure is however time-consuming. Evaluating the potential of the strategy on real-world network instances requires therefore an alternative procedure. The theoretical analysis and algorithm can nonetheless be used to assess the quality of faster heuristics that find a good solution of (*BL*).

An application of the social routing system in real life requires further research. First, we only considered fairness of the resulting state, but one might also evaluate the inter-state travel time differences, i.e., before and after implementation of the service (see Jahn et al., 2005). Second, we used a relatively simple procedure to construct the path set. In practice, one might consider column generation that further explores the path set while solving the bilevel problem. Finally, we focused ourselves to the equilibrium state in an assignment with static demand. Developing a similar routing strategy for the dynamic case is much more complex, in particular since a range of possible behavioral responses should be accounted for.

7. Conclusion

In this paper, we consider a social routing strategy that explicitly accounts for the route choice behavior of drivers. The routing strategy asks a portion of the travelers to take a small detour for the system's benefit. Recent empirical research proved that such a strategy is implementable in a routing system in real life.

We showed that the best possible routes (with respect to efficiency) to be proposed by a routing system can be found by solving a bilevel optimization problem that anticipates the route choice behavior of compliant and non-compliant travelers. We used parametric analysis to study the behavior of the solution set of the lower-level problem as a function of the upper-level variable. Under mild conditions, we can efficiently calculate the directional derivative of the lower-level link flow solution by solving a convex quadratic optimization problem. A numerical procedure uses this directional derivative to find the paths to be proposed. The numerical experiments show the potential efficiency gain of such a system in practice. Indeed, only a small portion of the travelers need to take a fairly limited detour to achieve a substantial travel time improvement.

This paper assumed a static setting, but finding the best possible paths to be proposed to the receptive travelers is already difficult. Nonetheless, the paper introduces a strategy (and proves it potential) worth considering for application in a general traffic engineering context. For instance, in the case of incidents, authorities can particularly apply a similar routing strategy to mitigate the impact on the network with respect to the total travel time, but at the same time limit the detour of individual drivers.

Acknowledgment

This work was supported by the Netherlands Organisation for Scientific Research (NWO), project number 439.16.103.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ejor.2021.06.036

References

Angelelli, E., Arsik, I., Morandi, V., Savelsbergh, M., & Speranza, M. (2016). Proactive route guidance to avoid congestion. *Transportation Research Part B: Methodological, 94*, 1–21.
Angelelli, E., Morandi, V., Savelsbergh, M., & Speranza, M. G. (2021). System optimal routing of traffic flows with user constraints using linear programming. *European journal of operational research, 293*(3), 863–879.
Angelelli, E., Morandi, V., & Speranza, M. G. (2019). A trade-off between average and maximum arc congestion minimization in traffic assignment with user constraints. *Computers & Operations Research, 110*, 88–100.
Angelelli, E., Morandi, V., & Speranza, M. G. (2020). Minimizing the total travel time with limited unfairness in traffic networks. *Computers & Operations Research, 123*, 105016.
Bagloee, S. A., Sarvi, M., Patriksson, M., & Rajabifard, A. (2017). A mixed user-equilibrium and system-optimal traffic flow for connected vehicles stated as a complementarity problem. *Computer-Aided Civil and Infrastructure Engineering, 32*(7), 562–580.
Bank, B., Guddat, J., Klatte, D., Kummer, B., & Tammer, K. (1983). *Non-linear parametric optimization*. Birkhäuser Verlag.
Bard, J. F. (1998). *Practical bilevel optimization: Algorithms and applications*. Springer Science & Business Media.
Beckmann, M., McGuire, C., & Winsten, C. B. (1956). *Studies in the economics of transportation*. New Haven: Yale University Press.
Bureau of Public Roads (1964). Traffic assignment manual. US Department of Commerce.
Dempe, S. (1989). On the directional derivative of the optimal solution mapping without linear independence constraint qualification. *Optimization, 20*(4), 401–414.
Dempe, S. (1993). Directional differentiability of optimal solutions under slater's condition. *Mathematical programming, 59*(1–3), 49–69.
Dempe, S. (2002). *Foundations of bilevel programming*. Springer Science & Business Media.
Dempe, S., & Zemkoho, A. B. (2012). Bilevel road pricing: Theoretical analysis and optimality conditions. *Annals of operations research, 196*(1), 223–240.
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.
Djavadian, S., Hoogendoorn, R. G., van Arem, B., & Chow, J. Y. (2014). Empirical evaluation of drivers' route choice behavioral responses to social navigation. *Transportation research record, 2423*(1), 52–60.
Do Chung, B., Cho, H.-J., Friesz, T. L., Huang, H., & Yao, T. (2014). Sensitivity analysis of user equilibrium flows revisited. *Networks and Spatial Economics, 14*(2), 183–207.
Eikenbroek, O. A. L., Still, G. J., van Berkum, E. C., & Kern, W. (2018). The boundedly rational user equilibrium: A parametric analysis with application to the network design problem. *Transportation Research Part B: Methodological, 107*, 1–17.
Faigle, U., Kern, W., & Still, G. (2013). *Algorithmic principles of mathematical programming*: 24. Springer Science & Business Media.
Fiacco, A. V., & Ishizuka, Y. (1990a). Sensitivity and stability analysis for nonlinear programming. *Annals of operations research, 27*(1), 215–235.
Fiacco, A. V., & Ishizuka, Y. (1990b). Suggested research topics in sensitivity and stability analysis for semi-infinite programming problems. *Annals of operations research, 27*(1), 65–76.

Gentile, G. (2014). Local user cost equilibrium: A bush-based algorithm for traffic assignment. *Transportmetrica A: Transport Science, 10*(1), 15–54.

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.

Josefsson, M., & Patriksson, M. (2007). Sensitivity analysis of separable traffic equilibrium equilibria with application to bilevel optimization in network design. *Transportation Research Part B: Methodological, 41*(1), 4–31.

Klein, I., Levy, N., & Ben-Elia, E. (2018). An agent-based model of the emergence of cooperation and a fair and stable system optimum using atis on a simple road network. *Transportation research part C: emerging technologies, 86*, 183–201.

Lee, G. M., Tam, N. N., & Yen, N. D. (2006). *Quadratic programming and affine variational inequalities: A qualitative study*: 78. Springer Science & Business Media.

Li, R., Liu, X., & Nie, Y. M. (2018). Managing partially automated network traffic flow: efficiency *vs*. stability. *Transportation Research Part B: Methodological, 114*, 300–324.

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. (2008). Sensitivity of static traffic user equilibria with perturbations in arc cost function and travel demand. *Transportation science, 42*(1), 105–123.

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.

Mordukhovich, B. S. (2018). *Variational analysis and applications*. Springer.

Nguyen, S., & Dupuis, C. (1984). An efficient method for computing traffic equilibria in networks with asymmetric transportation costs. *Transportation Science, 18*(2), 185–202.

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.

Outrata, J. (1997). On a special class of mathematical programs with equilibrium constraints. In *Recent advances in optimization* (pp. 246–260). Springer.

Outrata, J., Kocvara, M., & Zowe, J. (2013). *Nonsmooth approach to optimization problems with equilibrium constraints: Theory, applications and numerical results*: 28. Springer Science & Business Media.

Pang, J.-S., & Ralph, D. (1996). Piecewise smoothness, local invertibility, and parametric analysis of normal maps. *Mathematics of Operations Research, 21*(2), 401–426.

Patriksson, M. (2004). Sensitivity analysis of traffic equilibria. *Transportation Science, 38*(3), 258–281.

Patriksson, M., & Rockafellar, R. T. (2002). A mathematical model and descent algorithm for bilevel traffic management. *Transportation Science, 36*(3), 271–291.

Patriksson, M., & Rockafellar, R. T. (2003). Sensitivity analysis of aggregated variational inequality problems, with application to traffic equilibria. *Transportation Science, 37*(1), 56–68.

Perederieieva, O., Ehrgott, M., Raith, A., & Wang, J. Y. (2015). A framework for and empirical study of algorithms for traffic assignment. *Computers & Operations Research, 54*, 90–107.

Qiu, Y., & Magnanti, T. L. (1989). Sensitivity analysis for variational inequalities defined on polyhedral sets. *Mathematics of Operations Research, 14*(3), 410–432.

Ralph, D., & Dempe, S. (1995). Directional derivatives of the solution of a parametric nonlinear program. *Mathematical programming, 70*(1–3), 159–172.

Robinson, S. M. (1982). Generalized equations and their solutions, part ii: applications to nonlinear programming. In *Optimality and stability in mathematical programming* (pp. 200–221). Springer.

Robinson, S. M. (2006). Strong regularity and the sensitivity analysis of traffic equilibria: A comment. *Transportation Science, 40*(4), 540–542.

Rockafellar, R. T., & Wets, R. J.-B. (2009). *Variational analysis*: 317. Springer Science & Business Media.

Roughgarden, T. (2005). *Selfish routing and the price of anarchy*: 174. MIT press Cambridge.

Simon, H. A. (1997). *Models of bounded rationality: Empirically grounded economic reason*: 3. MIT press.

Speranza, M. G. (2018). Trends in transportation and logistics. *European journal of operational research, 264*(3), 830–836.

Tobin, R. L., & Friesz, T. L. (1988). Sensitivity analysis for equilibrium network flow. *Transportation Science, 22*(4), 242–250.

Transportation Networks for Research Core Team (2019). Transportation networks for researchhttps://github.com/bstabler/transportationnetworks/. Accessed September 11, 2020.

van Essen, M., Eikenbroek, O., Thomas, T., & van Berkum, E. (2020). Travelers' Compliance with social routing advice: impacts on road network performance and equity. *IEEE Transactions on Intelligent Transportation Systems, 21*(3), 1180–1190.

Vreeswijk, J. D., Landman, R. L., van Berkum, E. C., Hegyi, A., Hoogendoorn, S. P., & van Arem, B. (2015). Improving the road network performance with dynamic route guidance by considering the indifference band of road users. *IET Intelligent Transport Systems, 9*(10), 897–906.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *ICE Proceedings: Engineering Divisions, 1*, 767–768.

Yang, H., Zhang, X., & Meng, Q. (2007). Stackelberg games and multiple equilibrium behaviors on networks. *Transportation Research Part B: Methodological, 41*(8), 841–861.

Yin, Y., Madanat, S. M., & Lu, X.-Y. (2009). Robust improvement schemes for road networks under demand uncertainty. *European journal of operational research, 198*(2), 470–479.

Zhang, K., & Nie, Y. M. (2018). Mitigating the impact of selfish routing: An optimal-ratio control scheme (orcs) inspired by autonomous driving. *Transportation Research Part C: Emerging Technologies, 87*, 75–90.

Zhou, B., Xu, M., Meng, Q., & Huang, Z. (2017). A day-to-day route flow evolution process towards the mixed equilibria. *Transportation Research Part C: Emerging Technologies, 82*, 210–228.