# Sensitivity analysis of separable traffic equilibrium equilibria, with application to bilevel optimization in network design

Magnus Josefsson* and Michael Patriksson†

November 22, 2005

**Abstract**

We provide a sensitivity analysis of separable traffic equilibrium models with travel cost and demand parameters. We establish that while equilibrium link flows may not always be directionally differentiable (even when the link travel costs are strictly increasing), travel demands and link costs are; this improves the general results of Patriksson (2004). The new results contradict common belief that equilibrium cost and demand sensitivities hinge on that of equilibrium flows.

The paper by Tobin and Friesz (1988) brought the classic nonlinear programming subject of sensitivity analysis to transportation science. Theirs is still the most widely used device by which "gradients" of traffic equilibrium solutions are calculated, for use in bilevel transportation planning applications such as network design, origin–destination (OD) matrix estimation and problems where link tolls are imposed on the users in order to reach a traffic management objective. However, it is not widely understood that the regularity conditions proposed by them are stronger than necessary. Also, users of their method sometimes misunderstand its limitations and are not aware of the computational advantages offered by more recent methods. In fact, a more often applicable formula was proposed already in 1989 by Qiu and Magnanti, and Bell and Iida (1997) describe one of the cases in practice in which the formula by Tobin and Friesz would not be able to generate sensitivity information, because one of their regularity conditions fails to hold. This paper provides an overview of this formula, and illustrates by means of examples that there are several cases where it is not applicable. Our findings are illustrated with small numerical examples, as are our own analysis.

The findings of this paper are hoped to motivate replacing the previous approach with the more often applicable one, not only because of this fact but equally importantly because it is intuitive and also can be much more efficiently utilized: the sensitivity problem that provides the directional derivative is a linearized traffic equilibrium problem, and the sensitivity information can be generated efficiently by only slightly modifying a state-of-the-art traffic equilibrium solver. This is essential for bringing the use of sensitivity analysis in transportation planning beyond the solution of only toy problems. We finally utilize a new sensitivity solver in the preliminary testing of a simple heuristic for bilevel optimization in continuous traffic network design, and compare it favourably to previous heuristics on known small-scale problems.

## Introduction and contributions

Performing a sensitivity analysis of traffic equilibria means evaluating the directions and rate of change that occur in the flows and travel costs as parameters in the cost and demand functions change. A sensitivity analysis is particularly useful in control and pricing applications since, if we can anticipate the effects of a change in, say, the traffic infrastructure, on the behaviour of the travellers, then we can utilize this knowledge to optimize these changes according to some goal fulfillment, like a reduction in flows or delays, a higher revenue from congestion tolls, etc. Such problems constitute instances of *bilevel optimization* problems, or *mathematical programs with equilibrium constraints* (MPEC), which is the scientific field within operations research and mathematical programming that is associated with hierarchical optimization problems, and which also includes the origin–destination (OD) matrix estimation problem. It is without question extremely important that the sensitivity analysis computations are correct, since otherwise an algorithm for an MPEC problem that utilizes it might terminate at an arbitrarily bad solution.

*M. Sc. in Engineering Physics, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. E-mail: f98majf@dd.chalmers.se

†Professor in Applied Mathematics, Department of Mathematics, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. E-mail: mipat@math.chalmers.se

Recently, the second author has been involved in a project having the goal to provide a precise sensitivity analysis of elastic and fixed demand traffic equilibrium problems, focusing on general models involving possibly non-separable and non-invertible link cost and demand functions; cf. Patriksson and Rockafellar (2002, 2003) and Patriksson (2004). The present paper provides several contributions, which we briefly outline below.

The analysis is specialized to the case of separable link cost and demand functions, the latter also being invertible, in which case we can work directly on optimization formulations. This fact makes it possible to provide a less technical presentation than in the above publications. We show how to perform a sensitivity analysis in practice by using a modification of state-of-the-art traffic equilibrium software.

With a specialization to separable link costs and demands, the results that can be obtained are, in one important respect, much stronger. In general, local Lipschitz continuity, which implies the directional differentiability of the equilibrium link flows (as well as of all other unique entities, like the travel costs and demands), follows under a second-order matrix condition. However, in the case of convex optimization problems, local Lipschitz continuity of the travel costs and demands can be shown to always follow when link costs are strictly increasing. The application to traffic equilibrium problems then states that although there are cases where the equilibrium link flow is not directionally differentiable in certain directions, travel costs and demands always are. The chain-rule like calculus formulas for directional derivatives of travel costs and demands that will emerge in this paper are hence valid even when the link flow perturbation term is not possible to interpret as a directional derivative.

In 1988 Tobin and Friesz did the transportation science community the great service of bringing to it the nonlinear programming topic of sensitivity analysis, with their publication. Their analysis is quite accessible to practitioners; for example, they utilize the rather intuitive Implicit Function Theorem in their analysis. It also remains the most popular tool for producing sensitivity analysis information in traffic equilibrium problems. We illustrate through examples how their formula is however less applicable in several ways. Moreover, it relies on direct matrix calculations, and therefore in general cannot be applied to large-scale networks. Our sensitivity analysis problem is however quite structured and need not involve matrix calculations at all; it amounts to solving a perturbed, affine traffic equilibrium problem, which is no more difficult to solve than the original one. It is in fact easier than the original problem because the route set is fixed and the link cost and demand functions are affine.

We utilize the sensitivity analysis in order to create a heuristic solution methodology for a classic bilevel optimization model in transportation science—the equilibrium network design problem. Apart from providing a description of how such a method can be devised with traffic equilibrium and sensitivity analysis instruments at hand, our work on implementing and testing it has enabled us to reach some conclusions that may have a main importance for a larger class of bilevel optimization models in the field. For example, our tests show that it is vital that the equilibrium solutions are determined accurately in order to ensure that the bilevel optimization method does not terminate too early; the classic tool for traffic equilibrium—the Frank–Wolfe algorithm—is definitely not up to par when used in tandem with this class of more general traffic models. In comparisons with classic methods on known networks, the new tool always wins. The findings are illustrations only; we do not propose to actually use such an algorithm in practice, as it ignores the possible non-differentiability of an optimal solution to the bilevel problem. Our algorithm can, however, be extended such that it generate subgradients (cf. Patriksson, 2004), thus turning it into a subgradient/bundle algorithm for the non-differentiable problem at hand.

# 1 The traffic model

Let $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ be a transportation network, where $\mathcal{N}$ and $\mathcal{L}$ are the sets of nodes and directed links, respectively. For certain ordered pairs of nodes, $(p, q) \in \mathcal{C}$, where node $p$ is an origin, node $q$ is a destination, and $\mathcal{C}$ is a subset of $\mathcal{N} \times \mathcal{N}$, there is a transport demand, which may be given by a function of the travel cost. We assume that the network is strongly connected, that is, that at least one route joins each origin–destination (OD) pair.

Wardrop's user equilibrium principle states that for every OD pair $(p, q) \in \mathcal{C}$, the travel costs of the routes utilized are equal and minimal for each individual user. We denote by $\mathcal{R}_{pq}$ the set of simple (loop-free) routes for OD pair $(p, q)$, by $h_r$ the flow on route $r \in \mathcal{R}_{pq}$, and by $c_r$ the travel cost on the route as experienced by an individual user.

We introduce the parameter to be present in the sensitivity analysis: it is denoted $\rho$, and is assumed to be of dimension $d$. This parameter could be present in one or both of the travel cost and demand functions. We assume that the travel cost function has the form $c(\rho,\cdot) : \mathbb{R}_+^{|\mathcal{R}|} \to \mathbb{R}^{|\mathcal{R}|}$ given a value of $\rho$, where $|\mathcal{R}|$ denotes the total number of routes in the network. Further, the demand function is given by $g(\rho,\cdot) : \mathbb{R}^{|\mathcal{C}|} \to \mathbb{R}_+^{|\mathcal{C}|}$. (We introduce the notation $\mathbb{R}_+ := \{\, x \in \mathbb{R} \mid x \geq 0 \,\}$ and $\mathbb{R}_{++} := \{\, x \in \mathbb{R} \mid x > 0 \,\}$.)

In an application to OD estimation, $d$ is in the order of $|\mathcal{C}|$, while $d \approx |\mathcal{L}|$ holds in equilibrium network design, pricing and control models.

We also introduce the matrix $\Gamma \in \{0,1\}^{|\mathcal{R}|\times|\mathcal{C}|}$, which is the route–OD pair incidence matrix (i.e., the element $\gamma_{rk}$ is 1 if route $r$ joins OD pair $k = (p,q) \in \mathcal{C}$, and 0 otherwise). Then, demand-feasibility is described by the conditions that $h \in \mathbb{R}_+^{|\mathcal{R}|}$ and

$$\Gamma^{\mathrm{T}} h = g(\rho, \pi) \tag{1}$$

holds, while the Wardrop equilibrium conditions for the route flows are that

$$h_r > 0 \implies c_r(\rho,h) = \pi_{pq}, \qquad r \in \mathcal{R}_{pq}, \qquad (p,q) \in \mathcal{C}, \tag{2a}$$
$$h_r = 0 \implies c_r(\rho,h) \geq \pi_{pq}, \qquad r \in \mathcal{R}_{pq}, \qquad (p,q) \in \mathcal{C}, \tag{2b}$$

where the value of $\pi_{pq} := \pi_{pq}(\rho,h)$ is the minimal (i.e., equilibrium) route cost in OD pair $(p,q)$. By the non-negativity of the route flows, the system (1)–(2) can more compactly be written as the mixed complementarity problem (MCP)

$$0^{|\mathcal{R}|} \leq h \perp (c(\rho,h) - \Gamma\pi) \geq 0^{|\mathcal{R}|}, \tag{3a}$$
$$\Gamma^{\mathrm{T}} h = g(\rho,\pi), \tag{3b}$$

where $a \perp b$, for two arbitrary vectors $a, b \in \mathbb{R}^n$, means that $a^{\mathrm{T}} b = 0$. (By nonnegativity, this implies that $a_j b_j = 0$ for all $j$.)

As we are interested in the sensitivity of link flows, we will assume that the route cost is additive. For each link $l \in \mathcal{L}$, the travel cost has the form $t_l(\rho, v_l)$, where $v \in \mathbb{R}^{|\mathcal{L}|}$ is the vector of link flows. The route and link travel costs and flows are related through a route–link incidence matrix, $\Lambda \in \{0,1\}^{|\mathcal{L}|\times|\mathcal{R}|}$, whose element $\lambda_{lr}$ equals one if route $r \in \mathcal{R}$ utilizes link $l \in \mathcal{L}$, and zero otherwise. Route $r$ has an additive route cost $c_r(\rho,h)$ if it is the sum of the costs of using all the links defining it. In other words, $c_r(\rho,h) = \sum_{l\in\mathcal{L}} \lambda_{lr} t_l(\rho, v_l)$. In short, then, $c(\rho,h) = \Lambda^{\mathrm{T}} t(\rho, v)$. Also, implicit in this relationship is the assumption that the pair $(h, v)$ is consistent, in the sense that $v$ equals the sum of the route flows: $v = \Lambda h$. We shall use the representation in terms of $v$, as it is an entity for which we can introduce conditions ensuring that uniqueness holds at equilibrium.

As could be noted above, the link travel cost was assumed to be separable. The same assumption is made with respect to the demand function, which is supposed to be of the form $g_k(\rho, \pi_k)$, $k \in \mathcal{C}$.

In order to be able to work with an optimization formulation, which furthermore admits a unique solution $(v^*, d^*)$ for the given value of $\rho^*$ and is such that we can apply sensitivity analysis theory, we introduce the following assumption, which is supposed to hold throughout the paper:

Assumption 1 (properties of the network model).

(a) For each $l \in \mathcal{L}$, the link travel cost function $t_l(\cdot,\cdot)$ is continuously differentiable, and strictly increasing in its second argument.

(b) For each $k \in \mathcal{C}$, the demand function $g_k(\cdot,\cdot)$ is continuously differentiable, non-negative, upper bounded, and strictly decreasing in its second argument. The function $g_k(\rho,\cdot)$ is therefore invertible, and has a single-valued inverse, $\xi_k(\rho,\cdot)$, which also is continuously differentiable and strictly decreasing. □

The optimization formulation that we will work with is the following standard one for elastic demand traffic assignment (cf., e.g., Beckmann et al., 1956; Sheffi, 1985; Patriksson, 1994):

$$\underset{(v,d)}{\text{minimize}}\ \phi(v,d) := \sum_{l\in\mathcal{L}} \int_0^{v_l} t_l(\rho,s)\,ds - \sum_{k\in\mathcal{C}} \int_0^{d_k} \xi_k(\rho,s)\,ds, \tag{4a}$$
$$\text{subject to } \Gamma^{\mathrm{T}} h = d, \tag{4b}$$
$$v = \Lambda h, \tag{4c}$$
$$h \geq 0^{|\mathcal{R}|}. \tag{4d}$$

For future use, let $C$ denote the polyhedral set of feasible solutions to (4) in $(h, v, d)$, that is,

$$C = \{ (h, v, d) \in \mathbb{R}_+^{|\mathcal{R}|} \times \mathbb{R}^{|\mathcal{L}|} \times \mathbb{R}^{|\mathcal{C}|} \mid \Gamma^{\mathrm{T}} h = d; \quad v = \Lambda h \}.$$

The variational inequality problem, which characterizes the solution $(h^*, v^*, d^*)$ to this problem, is stated as that of finding $(h^*, v^*, d^*) \in C$ such that

$$t(\rho, v^*)^{\mathrm{T}}(v - v^*) - \xi(\rho, d^*)^{\mathrm{T}}(d - d^*) \geq 0, \qquad (h, v, d) \in C. \tag{5}$$

To see that this expression characterizes the Wardrop conditions stated earlier in (3), we notice that (5) is equivalent to $(h^*, v^*, d^*)$ solving the following linear program:

$$\begin{aligned} &\underset{(v,d)}{\text{minimize}} \; t(\rho, v^*)^{\mathrm{T}} v - \xi(\rho, d^*)^{\mathrm{T}} d, && \text{(6a)} \\ &\text{subject to } \Gamma^{\mathrm{T}} h - d = 0^{|\mathcal{C}|}, && \text{(6b)} \\ &\qquad v - \Lambda h = 0^{|\mathcal{L}|}, && \text{(6c)} \\ &\qquad h \geq 0^{|\mathcal{R}|}. && \text{(6d)} \end{aligned}$$

Its LP dual is to

$$\begin{aligned} &\underset{(\pi,\alpha)}{\text{maximize}} \; 0, && \text{(7a)} \\ &\text{subject to } \Gamma \pi - \Lambda^{\mathrm{T}} \alpha \leq 0^{|\mathcal{R}|}, && \text{(7b)} \\ &\qquad -\pi = -\xi(\rho, d^*), && \text{(7c)} \\ &\qquad \alpha = t(\rho, v^*), && \text{(7d)} \end{aligned}$$

where $\pi$ and $\alpha$ are, respectively, the LP dual variables for the constraints (6b) and (6c). The dual variable $\alpha$ is eliminated by using (7d). The complementarity conditions between the two LP problems can then be written as

$$0^{|\mathcal{R}|} \leq h^* \perp (\Lambda^{\mathrm{T}} t(\rho, v^*) - \Gamma \pi^*) \geq 0^{|\mathcal{R}|}, \tag{8}$$

which is identical to the Wardrop condition (3a). The condition (3b) is obtained as follows: from (6b) and (7c), $\Gamma^{\mathrm{T}} h^* = d^* = g(\rho, \pi^*)$. As $t(\rho, \cdot)$ and $-g(\rho, \cdot)$ both are strictly monotone, the objective function of (4) is strictly convex; therefore, the solution in $(v^*, d^*)$ to (4), and equivalently to the variational inequality (5) and to the Wardrop conditions (3), is unique. We see that from (7c)–(7d), also the dual entities $(\pi^*, \alpha^*)$ are unique.

## 2 The basis for our sensitivity analysis

The basis of our sensitivity analysis is a result which is stated for a general variational inequality problem with a differentiable mapping $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ in the parameters $\rho \in \mathbb{R}^d$ and variables $x \in \mathbb{R}^n$: find $x^* \in X$ such that $f(\rho, x^*)^{\mathrm{T}}(x - x^*) \geq 0$ holds for all $x \in X$, where $X \subseteq \mathbb{R}^n$ is a polyhedral set. Equivalently, we can write it in a more natural form as follows:

$$-f(\rho, x^*) \in N_X(x^*), \tag{9}$$

where $N_X(x)$ denotes the normal cone to $X$ at $x$:

$$N_X(x) := \begin{cases} \{ v \in \mathbb{R}^n \mid v^{\mathrm{T}}(y - x) \leq 0, \quad y \in X \}, & \text{if } x \in X, \\ \emptyset & \text{otherwise.} \end{cases}$$

We let $S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ denote the mapping that assigns to each vector $\rho \in \mathbb{R}^d$ the set $S(\rho)$ of solutions to the problem (9):

$$S(\rho) := \{ x^* \in X \mid -f(\rho, x^*) \in N_X(x^*) \}, \qquad \rho \in \mathbb{R}^d. \tag{10}$$

(The notation " $\rightrightarrows$ " signifies that the mapping $S$ in general is a point-to-set mapping.) Letting $\rho = \rho^*$ be the current value of the parameter vector, we are interested in the direction and rate of change of the solution $x^*$ as $\rho^*$ is perturbed along a direction $\rho'$. This directional derivative of $S$ is the solution to an auxiliary variational inequality, which has the following form: find $x' \in K$ such that

$$-r(\rho', x') \in N_K(x'), \tag{11a}$$

where

$$K := T_X(x^*) \cap f(\rho^*, x^*)^\perp, \tag{11b}$$

and

$$r(\rho', x') := \nabla_\rho f(\rho^*, x^*)\rho' + \nabla_x f(\rho^*, x^*)x'. \tag{11c}$$

We let $DS(\rho^*|x^*) : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ denote the mapping that assigns to each perturbation $\rho' \in \mathbb{R}^d$ the set $DS(\rho^*|x^*)(\rho')$ of solutions to this problem:

$$DS(\rho^*|x^*)(\rho') := \{\, x' \in K \mid -r(\rho', x') \in N_K(x') \,\}, \qquad \rho' \in \mathbb{R}^d.$$

The set $K$ denotes the set of variations around $x^*$ that, roughly speaking, retains feasibility and optimality to the first order. $T_X$ denotes the tangent cone to $X$, which means that if $X$ is defined by linear constraints, we have that $X = \{\, x \in \mathbb{R}^n \mid Ax \geq b;\ Bx = d \,\}$ implies that $T_X(x^*) = \{\, z \in \mathbb{R}^n \mid \bar{A}z \geq 0;\ Bz = 0 \,\}$, where $\bar{A}$ consists of the rows $A_i$ of $A$ corresponding to the binding inequality constraints at $x^*$, that is, the indices $i$ with $A_i x^* = b_i$. Further, for any vector $z \in \mathbb{R}^n$, $z^\perp := \{\, y \in \mathbb{R}^n \mid z^{\mathrm{T}} y = 0 \,\}$ is the orthogonal subspace associated with the vector $z$. The mapping $r$ is a linearization of $f$ around $(\rho^*, x^*)$; it is also an affine mapping in $x'$. We remark that the two polyhedral cones $T_X(x)$ and $N_X(x)$ in fact are polar to each other: $T_X(x) := (N_X(x))^* := \{\, w \in \mathbb{R}^n \mid w^{\mathrm{T}} v \leq 0,\ v \in N_X(x)$. This classic result in polyhedral theory lies behind some of the development of the results of this section.

Suppose now that $f(\rho, \cdot)$ is monotone on $X$ around $\rho = \rho^*$, and that the parameterization is such that $\operatorname{rank} \nabla_\rho f(\rho^*, x^*) = n$. (The rank property can always be ensured by including enough dummy parameters.) According to a result by Dontchev and Rockafellar (2001),

$$S \text{ is single-valued and Lipschitz continuous near } \rho^* \tag{12a}$$

$$\Longleftrightarrow$$

$$DS(\rho^*|x^*) \text{ is single-valued.} \tag{12b}$$

Moreover, then the unique solution $x'$ to (11) is the directional derivative of the solution $x^*$ to (9) at $\rho^*$, in the direction of $\rho'$. A sufficient condition for the property (12b) to hold is, by Kyparisis (1988, Lemma 2.1), that

$$\nabla_x f(\rho^*, x^*) \text{ is positive definite on } (K - K). \tag{13}$$

We refer to this as a sufficient *second-order* condition. A stronger result than directional differentiability can also be obtained under additional assumptions: according to a result by Kyparisis (1990), under the assumption (12a),

$$S \text{ is differentiable at } \rho^* \tag{14a}$$

$$\Longleftrightarrow$$

$$DS(\rho^* \mid x^*)(\rho') \in -K, \qquad \rho' \in \mathbb{R}^d \qquad \text{(i.e., } DS(\rho^* \mid x^*) \text{ is linear).} \tag{14b}$$

Moreover, if further $K$ is a subspace, that is, if $K = K \cap (-K)$, then the gradient can be represented as

$$\nabla_\rho x(\rho^*) = -Z\left[Z^{\mathrm{T}} \nabla_x f(\rho^*, x^*) Z\right]^{-1} Z^{\mathrm{T}} \nabla_\rho f(\rho^*, x^*), \tag{15}$$

for any $n \times \ell$ matrix $Z$ such that $Z^{\mathrm{T}} Z$ is nonsingular and $z \in K \cap (-K)$ if and only if $z = Zy$ for some $y \in \mathbb{R}^\ell$, where $\ell$ is the dimension of $K \cap (-K)$. This differentiability result is a kind of implicit function theorem; the relationship in (12) shows how it naturally extends to more general cases.

We refer to this latter property not because we will establish sufficient conditions for its application in the present context (this has already been done in Patriksson, 2004), but to remark that the sensitivity analysis that is developed in the paper by Tobin and Friesz (1988) and its follow-up by Cho, Smith, and Friesz (2000) strives to utilize (15). Unfortunately, not only does the property $DS(\rho^* \mid x^*)(\rho') \in -K$ fail to hold in many cases (cf. Patriksson, 2004, as well as in the examples below), but also there may not exist a nonsingular matrix of the kind that is referred to above (cf. Bell and Iida, 1997).

# 3 Digestible sensitivity analysis examples

We illustrate the results of the previous section by performing a complete sensitivity analysis of two one-dimensional problems.

## 3.1 An unconstrained, parameterized quadratic program

The problem studied first is that to

$$\underset{x \in \mathbb{R}}{\text{minimize}}\ \phi(\rho^*, x) := \frac{1}{2}x^2 - [(\rho^*)^2 - 1]x, \tag{16}$$

where $\rho^* \in \mathbb{R}$. This problem has a unique optimal solution, $x^* = x(\rho^*) = (\rho^*)^2 - 1$. We plot the function $\rho^* \mapsto x(\rho^*)$ in Figure 1. Clearly, it is differentiable, with $\nabla x(\rho^*) = 2\rho^*$.

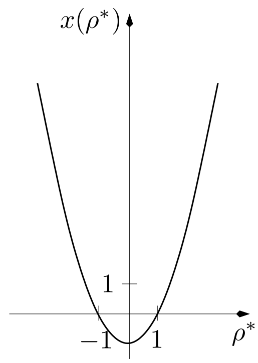


Figure 1: The first example.

In order to put this problem into the framework of Section 2, we put

$$f(\rho^*, x) := \nabla_x \phi(\rho^*, x) = x - [(\rho^*)^2 - 1]; \qquad C = \mathbb{R}.$$

Noting that the original problem corresponds to setting $f(\rho^*, x) = 0$, the sensitivity problem solved in order to produce the derivative of the solution is a linearization of this problem. Hence, the cost function is the first-order Taylor expansion of $f$ [note that the zero:th order term is zero at $(\rho^*, x^*)$]:

$$r(\rho', x') := \nabla_\rho f(\rho^*, x^*)\rho' + \nabla_x f(\rho^*, x^*)x' = -2\rho^*\rho' + x'.$$

Similarly, the feasible set of the linearized problem is a linearization of the feasible set of the original problem. In the unconstrained case, it comes as no surprise that $K := T_C(x^*) \cap f(\rho^*, x^*)^\perp = \mathbb{R}$, since the tangent cone to $\mathbb{R}$ is, again, $\mathbb{R}$ and $f(\rho^*, x^*) = 0$. The sensitivity problem, which provides the directional derivative $x' = x'(\rho^*; \rho')$ of $x^*$ at $\rho^*$ in the direction of $\rho'$ therefore is to solve a linear equation:

$$-r(\rho', x') \in N_K(x') \quad \Longleftrightarrow \quad r(\rho', x') = 0 \quad \Longleftrightarrow \quad x' = 2\rho^*\rho'.$$

(Note that the normal cone operator to $\mathbb{R}$ is identically zero, by polarity: $N_{\mathbb{R}}(x') = \{0\}$ for every $x' \in \mathbb{R}$.) We see that the directional derivative $x' := x'(\rho^*; \rho') = 2\rho^*\rho'$ is linear in $\rho'$ and therefore $x(\cdot)$ is differentiable at $\rho^*$ [cf. (14)]. In particular, we notice that the derivative of $x^*$ at $\rho^*$ is $x'(\rho^*; 1) = 2\rho^*$, which we already knew from the above development.

The above development is also produced by an application of the implicit function theorem: for $\rho' \in \mathbb{R}$,

$$r(\rho', x') = 0 \iff \nabla_\rho f(\rho^*, x^*)\rho' + \nabla_x f(\rho^*, x^*)x' = 0 \iff x' = -\nabla_x f(\rho^*, x^*)^{-1}\nabla_\rho f(\rho^*, x^*)\rho' \iff$$
$$\nabla x(\rho^*) = -\nabla_x f(\rho^*, x^*)^{-1}\nabla_\rho f(\rho^*, x^*) = (-1)(-2\rho^*) = 2\rho^*.$$

Next, we introduce a non-negativity constraint into the problem (16).

### 3.2 A sign constrained, parameterized quadratic program

Our second problem is to

$$\underset{x \geq 0}{\text{minimize}}\ \phi(\rho^*, x) := \frac{1}{2}x^2 - [(\rho^*)^2 - 1]x. \tag{17}$$

This problem also has a unique solution:

$$\begin{aligned} x^* = x(\rho^*) &= \begin{cases} (\rho^*)^2 - 1, & \text{if } (\rho^*)^2 \geq 1, \\ 0, & \text{otherwise} \end{cases} \\ &= \max\{0, (\rho^*)^2 - 1\}. \end{aligned}$$

Figure 2 plots this function.

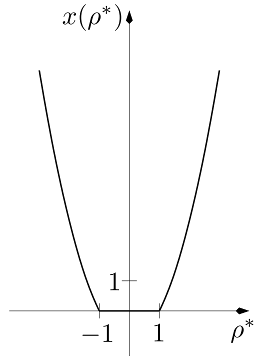


Figure 2: The second example.

The non-differentiability of $x(\cdot)$ at $\rho^* = \pm 1$ is evident from this picture. We have that

$$\nabla x(\rho^*) = \begin{cases} 2\rho^*, & \text{if } (\rho^*)^2 > 1, \\ 0, & \text{if } (\rho^*)^2 < 1; \end{cases}$$

at $\rho^* = -1$, the left [respectively, right] derivative of $x(\cdot)$ is $x'(\rho^*; -1) = -2$ [respectively, $x'(\rho^*; 1) = 0$]; similarly for $\rho^* = 1$.

How does the derivative structure emerge from the sensitivity framework of Section 2? The setting now is

$$f(\rho^*, x) := \nabla_x \phi(\rho^*, x) = x - [(\rho^*)^2 - 1]; \qquad C = \mathbb{R}_+.$$

The variational problem corresponding to the optimization problem (17) is the linear complementarity problem to find $x^* \in \mathbb{R}$ such that

$$0 \leq x^* - [(\rho^*)^2 - 1] \perp x^* \geq 0.$$

In all cases of choices of value of the parameter $\rho^*$ and perturbation $\rho'$, $r(\rho', x') := \nabla_\rho f(\rho^*, x^*)\rho' + \nabla_x f(\rho^*, x^*)x' = -2\rho^*\rho' + x'$ holds. The critical cone $K$ is studied next in detail.

$$T_C(x^*) = T_{\mathbb{R}_+}(x^*) = \{\, x' \in \mathbb{R} \mid x' \geq 0 \text{ if } x^* = 0 \,\} = \begin{cases} \mathbb{R}, & \text{if } \rho^* < -1, \\ \mathbb{R}_+, & \text{if } -1 \leq \rho^* \leq 1, \\ \mathbb{R}, & \text{if } 1 < \rho^*. \end{cases}$$

Further,

$$f(\rho^*, x^*)^\perp = \{\, x' \in \mathbb{R} \mid f(\rho^*, x^*)x' = 0 \,\} = \begin{cases} \mathbb{R}, & \text{if } \rho^* \leq -1, \\ \{0\}, & \text{if } -1 < \rho^* < 1, \\ \mathbb{R}, & \text{if } 1 \leq \rho^* : \end{cases}$$

in the open interval $\rho^* \in (-1, 1)$, the problem (17) does not have an optimal solution where the gradient is zero.

We clearly then have five possible cases of the critical cone, that is, the set of feasible solutions in the sensitivity problem, for (a) $\rho^* < -1$, (b) $\rho^* = -1$, (c) $\rho^* \in (-1, 1)$, (d) $\rho^* = 1$, and (e) $\rho^* > 1$. We analyze them, and the corresponding derivatives, next.

(a) For $\rho^* < -1$, we have that $K = \mathbb{R}$. This leads to the sensitivity problem to find $x' \in \mathbb{R}$ with $r(\rho', x') := x' - 2\rho^*\rho' = 0$; that is, $x' = x'(\rho^*; \rho') = 2\rho^*\rho'$. This is linear in $\rho'$, whence we have that for $\rho^* < -1$, $\nabla x(\rho^*) = 2\rho^*$.

(b) For $\rho^* = -1$, we have that $K = \mathbb{R}_+$. This leads to the sensitivity problem to find $x' \in \mathbb{R}$ with

$$0 \leq x' + 2\rho' \perp x' \geq 0.$$

Its solution is

$$x' = x'(\rho^*; \rho') = \begin{cases} 2\rho', & \text{if } \rho' \leq 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is not linear in $\rho'$, whence $\nabla x(\rho^*)$ does not exist.

(c) For $\rho^* \in (-1, 1)$, we have that $K = \{0\}$. The resulting sensitivity problem has only one feasible solution ($x' = 0$); hence, $x' = x'(\rho^*; \rho') = 0$ for every $\rho' \in \mathbb{R}$. This is linear in $\rho'$, whence we have that for $-1 < \rho^* < 1$, $\nabla x(\rho^*) = 0$.

(d) For $\rho^* = 1$, we have that $K = \mathbb{R}_+$. This leads to the sensitivity problem to find $x' \in \mathbb{R}$ with

$$0 \leq x' - 2\rho' \perp x' \geq 0.$$

Its solution is

$$x' = x'(\rho^*; \rho') = \begin{cases} 2\rho', & \text{if } \rho' \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is not linear in $\rho'$, whence $\nabla x(\rho^*)$ does not exist.

(e) For $\rho^* > 1$, we have that $K = \mathbb{R}$. This leads to the sensitivity problem to find $x' \in \mathbb{R}$ with $r(\rho', x') := x' - 2\rho^*\rho' = 0$; that is, $x' = x'(\rho^*; \rho') = 2\rho^*\rho'$. This is linear in $\rho'$, whence we have that for $\rho^* > 1$, $\nabla x(\rho^*) = 2\rho^*$.

Summarizing these five cases completes the sensitivity analysis of the problem at hand; we can directly compare these derivative properties with the above analysis and the graph in Figure 2.

We see that at most points the solution $x^*$ is in fact differentiable; in the case of the open intervals $\rho^* < -1$ and $\rho^* > 1$ the analysis reduces to the Implicit Function Theorem through the analysis of a linear equation, while for the open interval $\rho^* \in (-1, 1)$ we utilized that the critical cone reduces to a very special linear subspace (namely the set $\{0\}$!). For the remaining two points ($\rho^* = \pm 1$) the sensitivity analysis was made through the formulation and solution of special linear complementarity problems.

It deserves to be mentioned here, finally, that in the actual use of the sensitivity analysis, one would focus on one particular value of the parameter $\rho^*$, and probably also on a particular value of its perturbation $\rho'$; the complete analysis made in this section then serves to illustrate the possible outcomes of such an analysis, depending on the values of $\rho^*$ and $\rho'$. In common for all of them is that the sensitivity problem is of the same type as the original one, and it is in many cases also simpler: it is a linear complementarity problem equivalent to a convex quadratic program.

# 4 Sensitivity analysis of separable traffic equilibria

We first identify the sensitivity problem in our notation. Let

$$x = \begin{pmatrix} h \\ v \\ d \end{pmatrix}; \qquad f(\rho, x) = \begin{pmatrix} 0^{|\mathcal{R}|} \\ t(\rho, v) \\ -\xi(\rho, d) \end{pmatrix}; \qquad X = C.$$

Then, we can identify the sensitivity problem through the following identifications:

$$K = \{\, (h', v', d') \in \mathbb{R}^{|\mathcal{R}|} \times \mathbb{R}^{|\mathcal{L}|} \times \mathbb{R}^{|\mathcal{C}|} \mid \Gamma^{\mathrm{T}} h' = d'; \quad v' = \Lambda h'; \quad h' \in H' \,\},$$

where

$$H' = \left\{ h' \in \mathbb{R}^{|\mathcal{R}|} \;\middle|\; \begin{array}{l} h'_r \text{ free if } h^*_r > 0 \\ h'_r \geq 0 \text{ if } h^*_r = 0 \text{ and } c_r(\rho^*, h^*) = \pi^*_k \\ h'_r = 0 \text{ if } h^*_r = 0 \text{ and } c_r(\rho^*, h^*) > \pi^*_k \\ [r \in \mathcal{R}_k,\ k \in \mathcal{C}] \end{array} \right\},$$

and

$$r(\rho', x') = \begin{pmatrix} 0^{|\mathcal{R}|} \\ \nabla_\rho t(\rho^*, v^*)\rho' + \nabla_v t(\rho^*, v^*)v' \\ -[\nabla_\rho \xi(\rho^*, d^*)\rho' + \nabla_d \xi(\rho^*, d^*)d'] \end{pmatrix}.$$

By the monotonicity and separability of $t$ and $-\xi$, the resulting sensitivity variational inequality can be equivalently written as the following convex quadratic optimization problem to

$$\begin{aligned} \underset{(v',d')}{\text{minimize}}\ \phi'(v', d') := [\nabla_\rho t(\rho^*, v^*)\rho']^{\mathrm{T}} v' + \frac{1}{2} \sum_{l \in \mathcal{L}} \frac{\partial t_l(\rho^*, v^*_l)}{\partial v_l} (v'_l)^2 \\ - [\nabla_\rho \xi(\rho^*, d^*)\rho']^{\mathrm{T}} d' - \frac{1}{2} \sum_{k \in \mathcal{C}} \frac{\partial \xi_k(\rho^*, d^*_k)}{\partial d_k} (d'_k)^2, \end{aligned} \tag{18a}$$

$$\text{subject to } \Gamma^{\mathrm{T}} h' = d', \tag{18b}$$

$$v' = \Lambda h', \tag{18c}$$

$$h \in H'. \tag{18d}$$

The derivation follows the same pattern as that in Patriksson and Rockafellar (2002, 2003) and Patriksson (2004), and is therefore not repeated. The sensitivity problem is closely related to the original model. Two main differences are notable: the link cost and demand functions are replaced by their linearizations, and the sign restrictions on $h$ are replaced by individual restrictions on the route flow perturbations $h'_r$ that depend on whether the route in question was used at equilibrium or not, cf. the set $H'$. Although the appearance of $H'$ depends on the choice of route flow solution $h^*$, it is an interesting fact that the possible choices of $v'$ in $K$ does *not*; this is a general consequence of aggregation, which was also utilized in Patriksson and Rockafellar (2002, 2003) and Patriksson (2004), as well as in several previous papers on the subject (Qiu and Magnanti, 1989; Yen, 1995; Outrata, 1997); their relations to the present analysis is analyzed in detail in Patriksson (2004). In summary, the sensitivity problem can be solved using software similar to those for the original traffic equilibrium model, provided of course that route flow information can be extracted. This paper describes such a software.

The main theoretical result of this paper is the application of the result (12) in the previous section. This result states that the sensitivity problem provides directional derivatives, provided that the solution is unique. So, under what circumstances will the optimal solution to (18) be unique? Similarly, which entities in the solution to the problem (4) [flow, travel cost, demand] have directional derivatives?

Clearly, we cannot apply the theory of the previous section to the problem stated in $(h, v, d)$, since $h^*$ is not unique. As $(v^*, d^*)$ is unique, we could project the problem onto this space. This is simply accomplished by redefining

$$C^{\text{proj}} = \{\, (v, d) \in \mathbb{R}^{|\mathcal{L}|} \times \mathbb{R}^{|\mathcal{C}|} \mid \exists h \text{ with } (h, v, d) \in C \,\};$$

further, we would let

$$x = \begin{pmatrix} v \\ d \end{pmatrix}; \qquad f(\rho, x) = \begin{pmatrix} t(\rho, v) \\ -\xi(\rho, d) \end{pmatrix}.$$

We stress that this type of projection of the problem is only valid thanks to the particular relationships between the link and route flows; our projection is the same as eliminating $h$ through an affine transformation. Normally, a projection such as the one above does not preserve the regularity properties we are utilizing. (In Qiu and Magnanti, 1989, and Yen, 1995, the route flow variables are gotten rid of by choosing a particular value of them, namely that which minimizes, over the equilibrium set of route flows, the route flow vector's Euclidean norm.)

As we are also interested in the sensitivity of the travel costs, we will, for the first time, introduce yet another modification: we introduce a dummy variable, $s \in \mathbb{R}^{|\mathcal{L}|}$, which will take on the (negative) value $s^* = -t(\rho^*, v^*)$ of the link travel cost at equilibrium, and likewise a dummy variable, $\pi_- \in \mathbb{R}^{|\mathcal{C}|}$, to take on the (negative) value $\pi^*_- = -\xi(\rho^*, d^*)$ of the equilibrium OD travel costs. [Note that $\pi^*_- = -\pi^*$, where

$\pi^*$ is given in (8).] In the sensitivity problem, then, $s'$ and $\pi'_-$ will equal the (negative of the) link and OD travel cost perturbation, respectively.

The problem which will be analyzed is the following: in (9), let

$$x = \begin{pmatrix} v \\ d \\ s \\ \pi_- \end{pmatrix}; \qquad f(\rho, x) = \begin{pmatrix} t(\rho, v) \\ -\xi(\rho, d) \\ s + t(\rho, v) \\ -\pi_- - \xi(\rho, d) \end{pmatrix}; \qquad X = C^{\text{proj}} \times \mathbb{R}^{|\mathcal{L}|} \times \mathbb{R}^{|\mathcal{C}|}.$$

The variational inequality corresponding to (9) states that, at $\rho^*$,

$$\begin{aligned} t(\rho^*, v^*)^{\mathrm{T}}(v - v^*) - \xi(\rho^*, d^*)^{\mathrm{T}}(d - d^*) \geq 0, \qquad (v, d) \in C^{\text{proj}}, \\ s^* = -t(\rho^*, v^*), \\ \pi^*_- = -\xi(\rho^*, d^*), \end{aligned}$$

so it is entirely equivalent to the VIP in (5). The reason for introducing the last two rows of the problem, that is, the extra variables $(s, \pi_-)$, is that by doing so, we have direct access to the sensitivity of the travel costs, through the corresponding elements $(s', \pi'_-)$ of $x'$.

The sensitivity problem has the form of (11), with

$$r(\rho', x') = \begin{pmatrix} \nabla_\rho t(\rho^*, v^*)\rho' + \nabla_v t(\rho^*, v^*)v' \\ -[\nabla_\rho \xi(\rho^*, d^*)\rho' + \nabla_d \xi(\rho^*, d^*)d'] \\ s' + \nabla_\rho t(\rho^*, v^*)\rho' + \nabla_v t(\rho^*, v^*)v' \\ -\pi'_- - [\nabla_\rho \xi(\rho^*, d^*)\rho' + \nabla_d \xi(\rho^*, d^*)d'] \end{pmatrix}, \tag{19}$$

and

$$K = \{ (v', d', s', \pi'_-) \in \mathbb{R}^{|\mathcal{L}|} \times \mathbb{R}^{|\mathcal{C}|} \times \mathbb{R}^{|\mathcal{L}|} \times \mathbb{R}^{|\mathcal{C}|} \mid \exists h' \in H' \text{ with } \Gamma^{\mathrm{T}} h' = d'; \quad v' = \Lambda h' \}.$$

The associated sensitivity optimization problem in $(v', d')$ is (18), and the value of $(s', \pi'_-)$ is

$$\begin{aligned} s' &= -[\nabla_\rho t(\rho^*, v^*)\rho' + \nabla_v t(\rho^*, v^*)v'], \\ \pi'_- &= -[\nabla_\rho \xi(\rho^*, d^*)\rho' + \nabla_d \xi(\rho^*, d^*)d'], \end{aligned}$$

that is, the cost perturbations are given by a kind of chain rule. What we are going to establish is that this chain rule provides uniquely given values of $(s', \pi'_-)$ even when $v'$ is not unique!

THEOREM 2 (sensitivity of separable traffic equilibrium problems). *Let Assumption 1 hold, and consider an arbitrary vector $\rho^* \in \mathbb{R}^d$. Then, the solution $(v^*, d^*)$ to* (4) *is unique, and so are the (negative) travel cost entities $(s^*, \pi^*_-) = -(t(\rho^*, v^*), \xi(\rho^*, d^*))$. Let $\rho' \in \mathbb{R}^d$ be an arbitrary perturbation.*

(a) *In the solution to* (18), *the travel cost perturbations $(s', \pi'_-)$ are unique; therefore, the values*

$$\begin{aligned} -s' &= \nabla_\rho t(\rho^*, v^*)\rho' + \nabla_v t(\rho^*, v^*)v', \\ -\pi'_- &= \nabla_\rho \xi(\rho^*, d^*)\rho' + \nabla_d \xi(\rho^*, d^*)d', \end{aligned}$$

*are the directional derivatives of, respectively, the equilibrium link and OD travel costs, at $\rho^*$, in the direction of $\rho'$.*

(b) *Assume that the link travel cost function $t(\rho^*, \cdot)$ is such that*

$$\frac{\partial t_l(\rho^*, v_l^*)}{\partial v_l} > 0, \qquad l \in \mathcal{L}. \tag{20}$$

*Assume further that the demand function $g(\rho^*, \cdot)$ is such that*

$$\frac{\partial g_k(\rho^*, \pi_k^*)}{\partial \pi_k} < 0, \qquad k \in \mathcal{C}. \tag{21}$$

*Then, in the solution to* (18), *the values of the link flow and demand perturbation $v'$ and $d'$ are unique; therefore, the value $v'$ (respectively, $d'$) is the directional derivative of the equilibrium link flow (respectively, demand), at $\rho^*$, in the direction of $\rho'$.*

Proof. That $(v^*, d^*)$ is unique follows from the strict convexity of the objective function $\phi$ in (4), and the convexity of the feasible set.

(a) Since $t(\rho^*, \cdot)$ and $-\xi(\rho^*, \cdot)$ both are monotone and differentiable, their Jacobians at equilibrium are positive semi-definite. By separability, they are moreover diagonal matrices with non-negative entries $\frac{\partial t_l(\rho^*, v_l^*)}{\partial v_l}$ and $-\frac{\partial \xi_k(\rho^*, d_k^*)}{\partial d_k}$. The objective function $\phi'$ in (18) is therefore convex, and since the set $K$ is polyhedral, the sensitivity problem is a convex, differentiable program.

The gradient of the objective function $\phi'$ in (18) is $r$, as given in the first two rows of (19); the last two rows are essentially a repetition of the first two. By a result in Burke and Ferris (1991), the value of $r$ then is unique. In particular, $(s', \pi'_{-})$ is unique. The interpretation of this tuple as a directional derivative then follows from (12).

(b) Under the additional assumptions, the objective function $\phi'$ in (18) is strictly convex in $(v', d')$, and therefore $x'$ is unique at the solution. The rest follows as in (a), by referring to (12). □

Note that the second-order condition (21) is equivalent to the condition that $\frac{\partial \xi_k(\rho^*, d_k^*)}{\partial d_k} < 0$ for all $k \in \mathcal{C}$. Obviously, this condition on the demand function derivative (or its inverse) is not needed in the case when we consider fixed demands.

An interesting aspect of this result is that the cost perturbations $(s', \pi'_{-})$, which are related to the perturbations $(v', d')$ through a kind of chain rule, are *not* dependent on the perturbations $(v', d')$ to be unique. This is in contrast to the type of analysis offered by Tobin and Friesz (1988)—see also Bell and Iida (1997, Section 5.4)—where the sensitivity of the costs is considered an implication of that of the flows and demands. A perhaps even more striking aspect of this result is that it is not possible to immediately apply it when the link travel cost is described by the use of the fourth-order polynomials known as the BPR (Bureau of Public Roads) functions; if a link has zero flow, then its cost derivative is also zero, so the condition (20) is not satisfied.

# 5 Illustrative examples

Our previous papers on sensitivity analysis issues do include illustrative examples. The ones that we shall provide in this section however complement them first by being simpler and thus possible to analyze in greater detail, and second by showing exactly when the sensitivity analysis works and when it does not, according to the theory just presented.

## 5.1 The Braess network

The network of Braess (1968) (see Figure 3) is classic in the analysis of system optimal solutions. We utilize it here for our purposes of providing a first example of a sensitivity analysis solution; it will also be utilized later, providing a test example of an equilibrium network design problem.

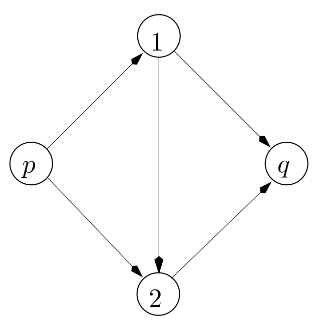


Figure 3: Braess' traffic network.

For this problem [where link $(1, 2)$ has a travel cost that is deliberately chosen so that one of the routes is not used even though it has minimum cost], we have the data given in Table 1.

The data corresponds to an instance of the fixed (and unperturbed) demand traffic equilibrium prob-

| *Link* | $t_{ij}(\rho, v_{ij})$ | *OD pair* | $d_{pq}$ |
|---|---|---|---|
| 1: $(p, 1)$ | $10v_{p1}$ | 1: $(p, q)$ | 6 |
| 2: $(p, 2)$ | $50 + v_{p2}$ | | |
| 3: $(1, q)$ | $50 + v_{1q}$ | | |
| 4: $(2, q)$ | $10v_{2q}$ | | |
| 5: $(1, 2)$ | $23 + \rho + v_{12}$ | | |

Table 1: Network data.

lem, which can be written as that to

$$\underset{v}{\text{minimize}}\ \phi(v) := \sum_{l\in\mathcal{L}} \int_0^{v_l} t_l(\rho, s)\,ds, \tag{22a}$$

$$\text{subject to } \Gamma^{\mathrm{T}} h = d, \tag{22b}$$

$$v = \Lambda h, \tag{22c}$$

$$h \geq 0^{|\mathcal{R}|}, \tag{22d}$$

where $d \in \mathbb{R}^{|\mathcal{C}|}_{++}$ is the vector of demands.

Solving the fixed demand traffic equilibrium problem with $\rho = \rho^* = 0$, we obtain the link flow solution $v^* = (3, 3, 3, 3, 0)^{\mathrm{T}}$. The cost of the three routes $\{1, 3\}$, $\{2, 4\}$, and $\{1, 5, 4\}$, are in fact the same, namely 83, but the route flows are 3 on each of the first two, and zero on the third. (This unique route flow solution is hence non-strictly complementary.)

Since the parameter $\rho$ is present on link $(1, 2)$, which has link flow zero at equilibrium, it looks clear that a positive value of $\rho'$ should lead to no changes [since the flow on link $(1, 2)$ cannot be negative] whereas a negative value of $\rho'$ should imply that $v'_{12}$ is positive. Indeed, that is the case.

In this special case where demand is fixed and unperturbed, the sensitivity problem is that to

$$\underset{v'}{\text{minimize}}\ \phi'(v') := [\nabla_\rho t(\rho^*, v^*)\rho']^{\mathrm{T}} v' + \frac{1}{2}\sum_{l\in\mathcal{L}} \frac{\partial t_l(\rho^*, v^*_l)}{\partial v_l}(v'_l)^2, \tag{23a}$$

$$\text{subject to } \Gamma^{\mathrm{T}} h' = 0^{|\mathcal{C}|}, \tag{23b}$$

$$v' = \Lambda h', \tag{23c}$$

$$h \in H'. \tag{23d}$$

For any $\rho' \in \mathbb{R}$, therefore, we have that $\phi'(v') = \rho' v'_{12} + 5(v'_{p1})^2 + \frac{1}{2}(v'_{p2})^2 + \frac{1}{2}(v'_{1q})^2 + 5(v'_{2q})^2 + \frac{1}{2}(v'_{12})^2$, and the constraints specify that

$$\begin{aligned}
h'_1 + h'_2 + h'_3 &= 0,\\
-v'_{p1} + h'_1 + h'_3 &= 0,\\
-v'_{p2} + h'_2 &= 0,\\
-v'_{1q} + h'_1 &= 0,\\
-v'_{2q} + h'_2 + h'_3 &= 0,\\
-v'_{12} + h'_3 &= 0,\\
h'_3 &\geq 0.
\end{aligned}$$

Letting $\rho' = 1$, the optimal solution is $h' = 0^{\mathrm{T}}$ and $v' = 0^{\mathrm{T}}$; letting $\rho' = -1$, the optimal solution is $h' = \frac{1}{26}(-1, -1, 2)^{\mathrm{T}}$ and $v' = \frac{1}{13}(1, -1, -1, 1, 2)^{\mathrm{T}}$. This is therefore also a case where the traffic equilibrium solution is non-differentiable, since clearly the directional derivative mapping is not linear.

## 5.2 A further investigation into the analysis conditions

**A directional derivative exists** Consider the network in Figure 5.2. The demand is supposed to be fixed, with one unit from node $p$ to node $q$ defining the OD pair's demand.

Let $t_1(\rho^*, v_1) := 9 + v_1 + \rho^*$ and $t_2(v_2) = 10 + v_2^2$, respectively, be the two links' costs. With $\rho^* = 0$, the (unique) equilibrium link flow is $v^* = (1, 0)^{\mathrm{T}}$. Suppose that the perturbation of $\rho^*$ is given by $\rho' = 1$.

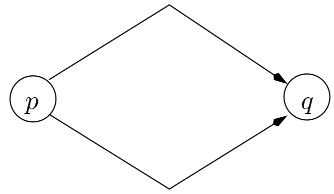


Figure 4: A two-link network.

While the second link has no flow at $\rho^* = 0$, it clearly has the same travel cost, so we have a case of a non-strictly complementary equilibrium solution. The sensitivity problem's feasible set defines a flow circulation, where the first link is possible to use in both directions, while the second one is only valid in the forward direction. The link cost in this problem is given by $\nabla_\rho t(\rho^*, v^*)^{\mathrm{T}}\rho' + \nabla_v t(\rho^*, v^*)^{\mathrm{T}} v' = (1 + v_1', 0)^{\mathrm{T}}$. The sensitivity problem therefore is to

$$\underset{v'}{\text{minimize}}\ \phi'(v') := v_1' + \frac{1}{2}(v_1')^2, \tag{24a}$$

$$\text{subject to } v_1' + v_2' = 0, \tag{24b}$$

$$v_2' \geq 0. \tag{24c}$$

Since for every choice of $\rho' \in \mathbb{R}$ the sensitivity problem has a unique solution, the (unique) solution to the above problem has the interpretation of the directional derivative.[1] The solution for $\rho' = 1$ is $v' = (-1, 1)^{\mathrm{T}}$. To relate this result to the analysis of the equilibrium solution as $\rho^*$ varies from 0 to 1, we consider the equilibrium flow trajectory as a function of $\rho^*$ in the interval $[0, 1]$: the simultaneous fulfillment of $9 + v_1^* + \rho^* = 10 + (v_2^*)^2$ and $v_1^* + v_2^* = 1$ yields the non-negative solution $v^*(\rho^*) = (\frac{3-\sqrt{1+\rho^*}}{2}, \frac{\sqrt{1+\rho^*}-1}{2})^{\mathrm{T}}$. (We see in particular that the derivative of $v^*(\rho^*)$ at $\rho^* = 0$ has the same direction as $v'$.) We also remark that the cost sensitivity, $\pi'$, is given by the Lagrange multiplier for the demand perturbation constraint; in our case, its value is zero.

**The sensitivity problem has a non-unique solution** Consider a situation similar to the previous one, but where we have three parallel links:

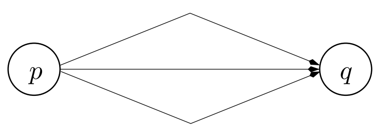


Figure 5: A three-link network.

In this network we have one OD pair, with demand 1, and the three links have travel costs $t_1(\rho, v_1) := 9 + \rho + v_1^2$, $t_2(v_2) := 10 + v_2$, and $t_3(v_3) := 10 + v_3$. With $\rho = \rho^* = 0$, the traffic equilibrium solution is that $v_1^* = 1$ and $v_2^* = v_3^* = 0$. Notice that the travel costs at equilibrium are 10 for all three links, again, therefore, a case of a non-strictly complementary solution.

Let $\rho' = 1$, and consider solving the sensitivity problem. Then, the problem becomes that to

$$\underset{v'}{\text{minimize}}\ \phi'(v') := v_1' + \frac{1}{2}(v_1')^2, \tag{25a}$$

$$\text{subject to } v_1' + v_2' + v_3' = 0, \tag{25b}$$

$$v_2', v_3' \geq 0. \tag{25c}$$

This problem has the following set of solutions: $v' = (-\frac{1}{2}, \alpha, \frac{1}{2} - \alpha)^{\mathrm{T}}$, for $\alpha \in [0, 1/2]$; that is, it is non-unique. The interpretation is that only the first link has a unique perturbation, that is, a directional

[1]Notice, however, that the sufficient, second-order, matrix condition in (13) is not satisfied. This confirms that it is indeed only a sufficient condition.

derivative may exist in the space of flows for the first link; the data of the problem is however such that we cannot predict how the change will be on the other two links. This is clearly a case where Theorem 2(b) is not applicable, because the last two links have a zero cost derivative at equilibrium.

It is important to notice, for this as well as all the other examples, that the value of $v'$ is that of a *perturbation* from the current value: it is not a flow in itself but it is a *direction and rate of movement* from the current equilibrium given by the direction of movement of the parameter values in $\rho$. For example, for the first link we have that $v_1' = -1/2$, while $v_1^* = 1$. This means that the flow on this link will be reduced from its initial value of 1 by some constant times 1/2 after perturbation, while, similarly, link 2's flow will be increased from its initial value of 0 by the same constant times $\alpha$.

**The sensitivity problem has no solution** Things can go even more wrong if an uncongested link is perturbed. Look at the first example again, but let instead the travel costs be given by $t_1(\rho^*, v_1) := 10 + \rho^*$ and $t_2(v_2) = 10 + v_2^2$, respectively. With $\rho^* = 0$, the (unique) equilibrium flow is $v^* = (1, 0)^{\mathrm{T}}$. Notice that this solution is not strictly complementary. Suppose that the perturbation of $\rho^*$ is given by $\rho' = 1$. The effect of making the first link more expensive then ought to be that the second link gets a positive flow; indeed, if we solve the equilibrium problem with $\rho^* = 1$, the result is that the link flow becomes $v^* = (0, 1)^{\mathrm{T}}$, where both links have travel cost 11. But at $\rho^* = 0$, the equilibrium solution has no directional derivative in the direction of $\rho' = 1$, so derivative information cannot be used to predict this change in the equilibrium flows! The link cost in the sensitivity problem is given by $\nabla_\rho t(\rho^*, v^*)^{\mathrm{T}} \rho' + \nabla_v t(\rho^*, v^*)^{\mathrm{T}} v' = (1, 0)^{\mathrm{T}}$. Hence, the sensitivity problem is the linear program to

$$\underset{v'}{\text{minimize}}\ \phi'(v') := v_1', \tag{26a}$$

$$\text{subject to } v_1' + v_2' = 0, \tag{26b}$$

$$v_2' \geq 0. \tag{26c}$$

This problem has an unbounded solution, in which flow can be sent in the negative cycle described by $(v_1', v_2') = (-\alpha, \alpha)$, with $\alpha \to \infty$. This is hence an example where the directional derivative does not exist. We can also notice this if we do as above, and express the equilibrium flow as a function of $\rho^*$ in the interval $\rho^* \in [0, 1]$: $v^*(\rho^*) = (1 - \sqrt{\rho^*}, \sqrt{\rho^*})^{\mathrm{T}}$. Clearly, this trajectory has no derivative at $\rho^* = 0$.

# 6 A dissection of the sensitivity analysis of Tobin and Friesz

We show, by means of both analytical and numerical tools, some examples in which the sensitivity analysis presented in Tobin and Friesz (1988) requires too strong assumptions.

Their analysis is performed on a variational inequality extension of the problem (22), but where the fixed demand is perturbed. In order to ensure local uniqueness, they introduce the following condition:

(Condition 1—strong monotonicity) $t(\rho, \cdot)$ is strongly monotone in a neighbourhood of $\rho^*$.

This condition is stronger than necessary, as we have already seen.

## 6.1 The strict complementarity condition

The analysis is based on first selecting a particular equilibrium route flow solution. Among the conditions stated, the route flow is supposed to be strictly complementary. The definition is however not the one commonly used, the common one being the following: a route flow solution $h^*$ is *strictly complementary* if and only if that it is complementary (that is, that $0 \leq h_r^* \perp [c_r(\rho^*, h^*) - \pi_{pq}(\rho^*)] \geq 0$ holds for all $r \in \mathcal{R}_{pq}$, $(p, q) \in \mathcal{C}$), and

$$h_r^* + [c_r(\rho^*, h^*) - \pi_{pq}(\rho^*)] > 0, \qquad r \in \mathcal{R}_{pq}, \quad (p, q) \in \mathcal{C}. \tag{27}$$

In other words, our use of the term strict complementarity means that for an arbitrary route $r \in \mathcal{R}_{pq}$, it is either used ($h_r^* > 0$) or it is more expensive than the least costly route used in the OD pair $[c_r(\rho^*, h^*) > \pi_{pq}(\rho^*)]$.

Tobin and Friesz state a definition of traffic equilibrium in terms of *total* link flows $v$ only, and which unfortunately is not consistent with the standard definition, given in (2). Their definition of a user equilibrium in terms of the vector $v$ is that there exists a vector $\lambda^* \in \mathbb{R}^{|\mathcal{N}|}$ such that for every link $l = (i, j) \in \mathcal{L}$,

$$\begin{aligned} v_l^* = 0 \quad &\Longrightarrow \quad t_l(v^*) \geq \lambda_j^* - \lambda_i^*, \\ v_l^* > 0 \quad &\Longrightarrow \quad t_l(v^*) = \lambda_j^* - \lambda_i^*. \end{aligned}$$

A problem with this definition is that the *aggregated* potential differences, $\lambda_j^* - \lambda_i^*$, are not consistent with our node price vectors $\pi_k$, $k \in \mathcal{C}$, associated with the shortest route problem for OD pair $k$ at $v^*$; in other words, at a traffic equilibrium, $\pi_{jk}^* - \pi_{ik}^* \neq \pi_{j\kappa}^* - \pi_{i\kappa}^*$ may hold for two OD pairs $k$ and $\kappa$. (For example, it can happen as soon as link $(i, j)$ lies on a shortest route in the OD pair $k$ but not in $\kappa$.)

Not only is it possible that the vector $\lambda^*$ of "node numbers" does not exist, an equilibrium solution can also be non-differentiable when there is an unused route equally cheap as the used ones—even if all links on the route are used by traffic on other routes. We conclude that the equilibrium condition *must* incorporate one potential difference vector for each OD pair, which we have done.

Based on their equilibrium definition, however, the authors continue by defining a strict complementarity criterion for the perturbed problem:

(Condition 2—strict complementarity) For each link $l = (i, j) \in \mathcal{L}$, $v_l^* = 0 \Longrightarrow t_l(\rho^*, v^*) > \lambda_j^* - \lambda_i^*$ holds.

So, whenever the total link flow vector $v^*$ is positive, this condition is satisfied. Clearly, it is therefore not compatible with the strict complementarity condition (27).

In any case, the strict complementarity condition is not a necessary condition for the differentiability of the traffic equilibrium solution, although our strict complementarity condition is *sufficient*. An example below will illustrate this fact.

## 6.2 The linear independence condition

Next, we are asked to restrict the network $\mathcal{G}$ to $\mathcal{G}_+ = (\mathcal{N}, \mathcal{L}_+)$, where $l \in \mathcal{L}_+$ if and only if $v_l^* > 0$, that is, to the network corresponding to the links having a positive flow given $\rho^*$. Consequently, there are possibly some routes that will be removed as well. The + notation to follow reflects this restriction.

Under the assumptions stated so far, the set $H_+^*(\rho^*)$ of equilibrium route flows is a bounded polyhedron. The next condition states that an equilibrium route flow vector $h_+^*$ is selected such that it is a "non-degenerate extreme point" of $H_+^*(\rho^*)$ satisfying Condition 2:

(Condition 3—linear independence) An equilibrium route flow $h_+^*$ is chosen such that it is an extreme point of $H_+^*(\rho^*)$ which has exactly as many routes with a positive flow as the rank of the matrix $[\Lambda_+^{\mathrm{T}} \mid \Gamma_+]$.

The rank of this matrix is never higher than the number of links with a positive flow at $v^*$ plus $|\mathcal{C}|$. The authors state an LP problem that can be used to generate such a point, but also remark in their Theorem 6 that the sensitivity values do not depend on this choice, as long as it satisfies Condition 3.

A final restriction is then made, such that we remove all the indices in the vector $h_+^*$ for which the flow is zero. (We do not change the notation to reflect this restriction.) The sensitivity problem is then finally set up as follows:

$$\begin{pmatrix} \nabla_\rho h_+ \\ \nabla_\rho \pi \end{pmatrix} = \begin{pmatrix} \nabla_h c_+(\rho^*, h_+^*) & -\Lambda_+^{\mathrm{T}} \\ \Lambda_+ & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\nabla_\rho c_+(\rho^*, h_+^*) \\ \nabla_\rho g(\rho^*) \end{pmatrix}. \tag{28}$$

## 6.3 Examples

### 6.3.1 A case of differentiability where Condition 2 fails

To show that strict complementarity is not necessary for differentiability, we consider the network depicted in Figure 6.

There are two OD pairs, $(1, 2)$ and $(4, 2)$, with a fixed and unperturbed demand of 2 and 1 units of flow, respectively. The link cost functions are given by

$$t_1(v_1, \rho) := 2v_1 + \rho;\ t_2(v_2) := v_2;\ t_3(v_3) := 1;\ t_4(v_4) := v_4 + 2;\ t_5(v_5) = v_5.$$

We have four routes: $\{1\}$, $\{2, 3\}$, $\{4\}$, and $\{5, 3\}$, two for each OD pair.

With $\rho^* = 0$, the unperturbed traffic equilibrium solution is $v^* = (1, 1, 1, 1, 1)^{\mathrm{T}}$. The route flow is unique: $h^* = (1, 1, 0, 1)^{\mathrm{T}}$. We see that the travel cost on route 3 is 2, as is the case for route 4, so this is

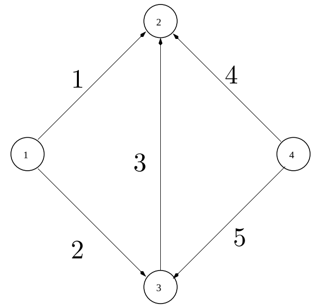


Figure 6: A traffic network concerning Condition 2.

a non-strictly complementary equilibrium solution. Since it is the unique route flow, we do not comply with our conditions (27) for strict complementarity; we do satisfy Condition 2.

In order to check if the solution $v^*$ is nevertheless differentiable at $\rho^* = 0$, we solve the sensitivity problem for both $\rho' := 1$ and $\rho' := -1$. For $\rho' = 1$, we obtain the following unique solution to the sensitivity problem (23), thus being the directional derivative of $v^*$ with respect to the direction $\rho' = 1$ at $\rho^* = 0$: $v' = \frac{1}{3}(-1, 1, 1, 0, 0)^{\mathrm{T}}$. The effect, as we can see, of perturbing link 1's cost such that it becomes more expensive, is that of sending flow in the cycle $\{-1, 2, 3\}$, where the minus reflects that flow is sent backwards on link 1. When solving the sensitivity problem for $\rho' := -1$, we obtain the directional derivative $v' = \frac{1}{3}(1, -1, -1, 0, 0)^{\mathrm{T}}$, that is, the negative of the directional derivative of $v^*$ in the direction of $\rho' := 1$. This proves that the directional derivative mapping is linear, and thus that the derivative of $v^*$ with respect to $\rho'$ at $\rho^* = 0$ equals $\frac{\mathrm{d}\, v^*}{\mathrm{d}\, \rho} = (-\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)^{\mathrm{T}}$.

At the same time, we have here shown that the sufficient matrix condition (13) indeed is only sufficient; it is not satisfied here because the set of feasible route flow perturbations is the entire space and the partial Jacobian of $t$ with respect to $v$ at the pair $(v^*, \rho^*)$ is the non-positive definite diagonal matrix with diagonal entries $(2, 1, 0, 1, 1)$; yet the equilibrium solution is even differentiable.

This is then an example where the formula (28) is not applicable, although the solution is differentiable.

#### 6.3.2 A case of differentiability where Condition 3 fails

Consider the network shown in Figure 7.

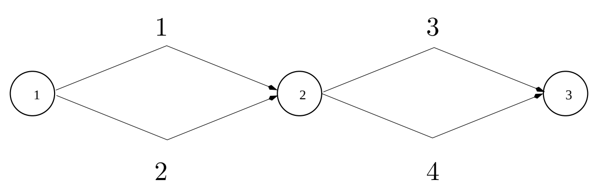


Figure 7: A traffic network concerning Condition 3.

There is a single OD pair, $(1, 3)$, with a fixed demand of 2 units of flow. The link cost functions are

$$t_1(v_1, \rho) := v_1 + \rho; \quad t_2(v_2) := v_2; \quad t_3(v_3) := v_3; \quad t_4(v_4) := v_4.$$

We have four routes: $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, and $\{2, 4\}$.

With $\rho^* = 0$, the unperturbed traffic equilibrium solution is $v^* = (1, 1, 1, 1)^{\mathrm{T}}$. We can easily see that the solution is differentiable; it is strictly complementary even. The derivative with respect to $\rho$ at $\rho^*$ moreover is $(-\frac{1}{2}, 0, \frac{1}{2}, 0)^{\mathrm{T}}$. This is intuitive: if the value of $\rho$ increases, then the flow on link 1 should decrease, whence link 2 must increase its flow with the same amount. If, on the other hand, the value of $\rho$ decreases, the reverse should happen.

Consider then the workings of the formula (28) outlined above. We obviously fulfill Condition 1 on the travel cost functions. We also satisfy Condition 2, because $v^* > 0^{|\mathcal{L}|}$. Also, then, $\mathcal{G}_+ = \mathcal{G}$. We last try to comply with the linear independence Condition 3, by choosing the right equilibrium route flow solution. Note then that

$$[\Lambda^{\mathrm{T}} \mid \Gamma] = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix},$$

which has rank 3. So, we should find a route flow solution, $h^*$, in which exactly 3 routes have a positive flow. This is however impossible; the only alternatives are 2 or 4. To see why, let's suppose that the flow on the first route, $\{1, 3\}$, is $\alpha \in [0, 1]$. Then, the flows on the routes $\{1, 4\}$ and $\{2, 3\}$ must both be $1 - \alpha$, in order to comply with the total flow on the links. This implies that the flow on route $\{2, 4\}$ is $\alpha$. This shows that for any value of $\alpha \in [0, 1]$, the number of routes having a non-zero flow is either 2 or 4. We can therefore not comply with Condition 3, and the formula (28) fails, even though the gradient exists.

The problems regarding the applicability of the formula (28) associated with the rank Condition 3 was first observed and commented on by Bell and Iida (1997, p. 97); our example however seems to be the first that has been worked out in detail.

### 6.3.3 A case of non-differentiability where the formula (28) may provide a result

Consider the network shown in Figure 8.

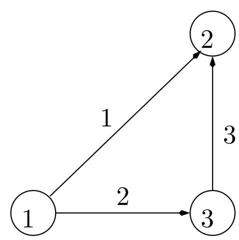


Figure 8: A traffic network concerning non-differentiability.

In this example, there are three OD pairs, with the fixed demands $d_{12} := 1$, $d_{13} := 1$, and $d_{32} := 1$. The link cost functions are

$$t_1(v_1, \rho) := 2v_1 + \rho; \qquad t_2(v_2) := v_2; \qquad t_3(v_3) := v_3.$$

(We thereby comply with Condition 1.) With $\rho^* = 0$, the unique equilibrium link volume is $v^* = (1, 1, 1)^{\mathrm{T}}$. In this case, the route flow is unique: the flow on route $\{(1, 2)\}$ is 1; the flow on route $\{(1, 3), (3, 2)\}$ is 0; the flow on route $\{(1, 3)\}$ is 1; and the flow on route $\{(3, 2)\}$ is 1 as well.

This solution is non-strictly complementary by our definition, since the route $\{(1, 3), (3, 2)\}$ is of least cost but it cannot be used. It is however strictly complementary according to Condition 2, which we thereby satisfy.

We also see that a small negative perturbation in $\rho$ would not affect the equilibrium solution, since the link $\{(1, 2)\}$ (that is, the first route in the first OD pair) is already utilized to send all the demand in the first OD pair. But if the perturbation is positive, we see that the flow in route $\{(1, 2)\}$ would decrease, and the flow on the route $\{(1, 3), (3, 2)\}$ would increase. This is then a case where the directional derivative (which of course exists) is not linear, so $\rho^* = 0$ is a point of non-differentiability.

What happens if we wish to apply the formula (28)? We here have that

$$[\Lambda^{\mathrm{T}} \mid \Gamma] = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix},$$

which has rank 4. Since the flow on the route $\{(1, 3), (3, 2)\}$ is restricted to zero, in all fairness, the formula then breaks down. But it does not really do so for the right reason; there is every possibility of believing that the formula might still work if we either still include the route, or if we were to delete it. In both cases, the formula (28) does produce a result, which in none of the two cases can be interpreted as the value of the gradient at $\rho^*$.

## 6.4 The gradient formula of Cho, Smith, and Friesz

The sensitivity analysis of Cho, Smith, and Friesz (2000) is somewhat related to that in Tobin and Friesz (1988). It replaces all the three conditions 1–3 mentioned in the previous section with weaker ones, and further provides an analysis entirely in link flows. We briefly discuss this analysis below.

(Condition 1'—strict monotonicity) $t(\rho, \cdot)$ is strictly monotone in a neighbourhood of $\rho^*$. Further, the Jacobian matrix $\nabla_v t(\rho^*, v^*)$ is positive definite.

(Condition 2'—strict complementarity) There exists a strictly complementary route flow $h^* \in H^*(\rho^*)$.

Notice that these two conditions together imply differentiability, but that they are stronger than necessary; the latter utilizes the classic definition (27) of strict complementarity, as we do in this paper.

We are next asked to consider, as in Tobin and Friesz (1988), the graph $\mathcal{G}_+$, which only includes links $l \in \mathcal{L}$ with $v_l^* > 0$ at $\rho^*$. In order to state the differences between the analysis in Cho, Smith, and Friesz (2000) and Tobin and Friesz (1988) more clearly, we do not introduce the + notation here, and assume, from now on, that $\mathcal{G} = \mathcal{G}_+$.

Next, suppose that we, for each $(p, q) \in \mathcal{C}$, remove the routes $r \in \mathcal{R}_{pq}$ whose cost is higher than $\pi_{pq}^*$. Thus, we reach a network which we may denote by $\mathcal{G}_0$, in which the set $\mathcal{R}$ is replaced by the subset $\mathcal{R}_0$ of least-cost routes at $v^*$. By Condition 2', they must also be the routes with positive flow at $h^*$.

The sensitivity analysis proceeds with a further reduction:

(Condition 3'—linear independence) Select a subset of the rows of $\Lambda_0$, such that the resulting matrix $[(\Lambda_0')^{\mathrm{T}} \mid \Gamma_0]$ has full (column) rank.

Note that there is no requirement on the rank itself, and therefore this condition is milder than the Condition 3 in Tobin and Friesz (1988).

The sensitivity formula is similar to that in (28), but provides the sensitivity in the link flow space directly, and therefore does not require the selection of a particular equilibrium route flow solution. It is however much more complicated in the sense that the translation between the spaces in $h$ and $v$ implies that several submatrices of, for example, $\Lambda_0'$ must be constructed, collected, inverted and multiplied:

$$\begin{pmatrix} \nabla_\rho v \\ \nabla_\rho \pi \end{pmatrix} = \begin{pmatrix} \nabla_v t(\rho^*, v^*) & -[\Lambda_0'' N_1, -I]^{\mathrm{T}} \\ [\Lambda_0'' N_1, -I] & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\nabla_\rho t(\rho^*, v^*) \\ -\Lambda_0'' N_2 \nabla_\rho g(\rho^*) \end{pmatrix}, \tag{29a}$$

where $\Lambda_0''$ contains the rows of $\Lambda_0$ which are not present in $\Lambda_0'$, and

$$\begin{aligned} N_1 :=& (\Lambda_0')^{\mathrm{T}}[\Lambda_0'(\Lambda_0')^{\mathrm{T}} - \Lambda_0'\Gamma_0(\Gamma_0^{\mathrm{T}}\Gamma_0)^{-1}\Gamma_0^{\mathrm{T}}(\Lambda_0')^{\mathrm{T}}]^{-1} \\ & - \Gamma_0[\Gamma_0^{\mathrm{T}}\Gamma_0 - \Gamma_0^{\mathrm{T}}(\Lambda_0')^{\mathrm{T}}[\Lambda_0'(\Lambda_0')^{\mathrm{T}}]^{-1}\Lambda_0'\Gamma_0]^{-1}\Gamma_0^{\mathrm{T}}(\Lambda_0')^{\mathrm{T}}[\Lambda_0'(\Lambda_0')^{\mathrm{T}}]^{-1}, \end{aligned} \tag{29b}$$

$$\begin{aligned} N_2 :=& -(\Lambda_0')^{\mathrm{T}}[\Lambda_0'(\Lambda_0')^{\mathrm{T}}]^{-1}\Lambda_0'\Gamma_0[\Gamma_0^{\mathrm{T}}\Gamma_0 - \Gamma_0^{\mathrm{T}}(\Lambda_0')^{\mathrm{T}}[\Lambda_0'(\Lambda_0')^{\mathrm{T}}]^{-1}\Lambda_0'\Gamma_0]^{-1} \\ & + \Gamma_0[\Gamma_0^{\mathrm{T}}\Gamma_0 - \Gamma_0^{\mathrm{T}}(\Lambda_0')^{\mathrm{T}}[\Lambda_0'(\Lambda_0')^{\mathrm{T}}]^{-1}\Lambda_0'\Gamma_0]^{-1}. \end{aligned} \tag{29c}$$

While being applicable to a larger set of situations than the analysis in Tobin and Friesz (1988), the analysis in Cho, Smith, and Friesz (2000) is still not valid for problems with a non-strictly complementary equilibrium solution, since it relies on the Implicit Function Theorem. Its main drawback is however its complexity; the formula (29) is rather non-intuitive and computationally burdensome to use. It is also clear from the papers that have been written and referred to during the past few years that the analysis in Tobin and Friesz (1988) is the one favoured, despite the fact that it is less generally applicable.

## 6.5 Conclusion

Interestingly, shortly after Tobin and Friesz published their paper, Qiu and Magnanti (1989) published the first paper which develops a sensitivity analysis of traffic equilibria based on Robinson's (1980) strong regularity condition (albeit under slightly stronger assumptions than necessary; cf. Patriksson, 2004); it is based on a linearized traffic equilibrium model which is similar to (18) in the case of a separable problem. Their paper did however not get much attention from the transportation science community. (See Patriksson, 2004, for an account of the history of the sensitivity analysis of traffic equilibria.) One of the few who has utilized their results is Denault (1994), who applied it in the context of OD matrix estimation. He compared numerically the Qiu/Magnanti sensitivity analysis formulas to that of

Tobin/Friesz, and found that the former was significantly more robust and efficient to use. Our above findings clearly is supportive of that claim.

At the request of one of the anonymous referees, we list some of the papers relying on (or providing analyses similar to) the analysis of Tobin and Friesz (1988), in areas such as network design, transit network optimization, traffic and signal control, OD estimation, car ownership studies, bicriterion traffic equilibrium, network reliability, and congestion pricing: Friesz et al. (1990), Kim and Suh (1990), Yang et al. (1992, 1994), Yang and Yagar (1994, 1995), Yang (1995a, 1995b, 1997), Miyagi and Suzuki (1995), Yang and Lam (1996), Wong and Yang (1997), Leurent (1998), Yang and Bell (1998), Chiou (1999, 2003), Clark and Watling (2000, 2002), Tam and Lam (2000, 2004), Yang, Bell, and Meng (2000), Yang and Meng (2000), Yang and Woo (2000), Wong, Yang, and Lo (2001), Chen and Lam (2002), Chen et al. (2002), Gao and Song (2002), Wong and Wong (2002), Yin and Ieda (2002), Wong et al. (2003), and Gao, Sun, and Shan (2004).

We finally remark here that not only is simplicity important—and the one developed in this paper is certainly simpler, from our point of view, than those in Tobin and Friesz (1988) and Cho, Smith, and Friesz (2000)—, but its conclusive advantage is that it is applicable in more general circumstances.

# 7 A sensitivity analysis tool

## 7.1 The traffic equilibrium solver

The disaggregate simplicial decomposition (DSD) method of Larsson and Patriksson (1992) for the fixed demand problem (22) was taken as the building block of our sensitivity analysis tool. (In the case of elastic demands, one can still solve a fixed demand problem, by first utilizing the fixed demand transformation of Gartner, 1980.) It has the advantage of utilizing route flow information, and therefore the close resemblance between the original problem (22) and the sensitivity problem (18) can be utilized fully.

The DSD algorithm was recoded in Matlab for experimentational purposes, fully aware of the fact that the CPU time will be perhaps two orders of magnitude larger than a final C or Fortran implementation. We refer to Larsson and Patriksson (1992) for the basics of the DSD algorithm, but remind the reader that the most central points of the algorithm are the following: at some iteration point $(h^\tau, v^\tau)$ of consistent route and link flows, the current travel costs are used to solve a shortest route problem for each origin node. The routes that then are generated are compared to sets $\hat{\mathcal{R}}_{pq} \subset \mathcal{R}_{pq}$ that have been generated and stored previously, and those sets are augmented with any routes that were not stored already. (This step is the column, or route, generation one.) With those subsets of routes at hand, the restricted master problem (RMP)—which is the original problem (22) except that $\hat{\mathcal{R}}_{pq}$ replaces $\mathcal{R}_{pq}$ for each $(p,q) \in \mathcal{C}$—is solved using one of several methodologies implemented. The highly structured RMP is either solved by using a gradient projection method or a diagonalized Newton method. (In practice, it appears that the former is the best for smaller networks, but that the Newton method wins for large enough cases.)

## 7.2 The sensitivity solver

The similarity of the sensitivity problem to the equilibrium problem meant that much of the code from the DSD algorithm implementation could be reused. Of course, in the sensitivity context, flow and cost derivatives take the place of flows and costs themselves. For simplicity, these derivatives can be considered "virtual" costs and flows. The RMP solver code then only had to be altered slightly to allow the subset of the routes that were used at equilibrium to take on negative "flow" values.

In order to set up the sensitivity problem, except for the flows and costs, the main part concerns the routes to be included. Remember that only least-cost routes are valid, some of which have a non-negativity requirement. In order to construct this set of routes, we first included only those routes that were used in the equilibrium solution. In order to compensate for the possibility that the equilibrium solution might not perfectly identify these routes, a "fuzz"-factor was used. In other words, to determine whether a route was used, the route flow was compared not to zero but to a very small positive number, obtained by multiplying the OD-pair demand by a tiny factor. In other words, a sign restriction may be included for a route that has a very small amount of flow at the terminal flow of the DSD algorithm. Any remaining set of routes that are potentially interesting for the sensitivity problem could then be included based on a final shortest route calculation and a graph search, together with a "fuzz"-factor similar to the above, allowing for near-shortest routes to be included as well. (Since non-equilibrium routes are strictly more costly than equilibrium ones, we will not include any non-equilibrium routes or discard any equilibrium ones in the sensitivity analysis, if the equilibrium solution is accurate enough and the fuzz factor is small enough.)

This set of routes then is the one that defines the sensitivity problem; no further route generation is necessary, and so the only problem left is a convex quadratic RMP with some variables being free and some being sign restricted. Virtually the same algorithm as in the RMP for the original problem can be used; the only special case stems from the sign restrictions. Further details on the implementation of this and the algorithm in the next section can be found in the first author's master's thesis (Josefsson, 2003).

# 8 An application to equilibrium network design

Network design refers to the optimization of some objective function of the traffic in a network with respect to one or several *design parameters*, which are a collection of attributes of the network that can be altered. Typically, the design parameters are link capacity augmentations, and the optimization is the minimization of some total cost function. The problem is sometimes also referred to as the *capacity expansion problem*. Normally, the total cost function is the sum of the total travel cost in the network and some investment cost function; the latter is, as the name implies, meant to model the costs of rebuilding or expanding parts of the network.

## 8.1 Bilevel programming

The objective functions used in network design problems are functions of both the design parameters (link capacities) and the traffic in the network, but only the link capacities can be controlled directly. Hence, we have a situation where we want to optimize a quantity with respect to the output of another optimization process on our variables. This is an example of the class of problems known as *bilevel optimization* problems. (See Luo, Pang, and Ralph, 1996; and Outrata, Kočvara, and Zowe, 1998, for overviews of the field.)

Conceptually, a bilevel optimization situation can be thought of in game-theoretic terms as a game between a *leader* and a *follower*, each trying to minimize some respective *cost function*. The leader chooses the values of his/her variables, which determines the position of the follower. The follower chooses the values of his/her own variables so as to minimize his/her cost function. These variables in turn influence the cost function of the leader in the sense that he/she anticipates their values when selecting his/her own variable values, based on the knowledge of the follower's cost function. This priority in the variables makes the overall problem hierarchical, and as a result in general very difficult.

In network design, the leader is the network designer, and the follower is defined by the set of users of the network. The cost function of the follower is simply the objective function of the traffic equilibrium problem, parameterized by the leader's choices of parameter values, while the cost function of the leader is the above-mentioned total cost function. We write the problem as that to

$$\begin{aligned}
&\underset{\rho}{\text{minimize}} \;\; \Upsilon(\rho, v) := \sum_{l \in \mathcal{L}} t_l(\rho_l, v_l) v_l + \chi_l(\rho_l), && \text{(30a)}\\
&\text{subject to} \;\; \rho_l \in [\rho_l^{\min}, \rho_l^{\max}], \qquad l \in \mathcal{L}, && \text{(30b)}\\
&\qquad\qquad v \text{ solves (22) parameterized by } \rho, && \text{(30c)}
\end{aligned}$$

where the investment functions $\chi_l : \mathbb{R}_+ \to \mathbb{R}_+$ are differentiable and increasing. We disallow investments in certain links by letting $\rho_l^{\min} = \rho_l^{\max} = 0$ for them.

## 8.2 Solving the network design problem

The network design problem is generally challenging to solve, because the variable $v$ is implicitly given only, and moreover the function $\Upsilon$ is in general both non-convex and non-differentiable because of this nested dependence. In the example algorithm below, we however ignore the possible non-existence of gradients at the iterates. In each coordinate direction the directional derivative exists; the heuristic is however based on the presumption that these derivatives equal the actual derivatives along the coordinate directions, and that assembling the matrix of these derivatives yields the Jacobian matrix $\nabla_\rho v^*$.

Our approach to solving the problem is to apply a gradient projection algorithm for its solution. At $\rho \in \mathbb{R}^{|\mathcal{L}|}$, the "gradient" $\nabla_\rho v^*$ is calculated thus: For each link $l$ where we have a design parameter allowed to become non-zero, we solve the sensitivity problem, where $\rho_l' = 1$ and $\rho_s' = 0$ for $s \in \mathcal{L} \setminus \{l\}$.

In this way, we provide vectors of directional derivatives of $\Upsilon$ with respect to the design parameters. As stated above, we will use these vectors as if they constituted the Jacobian $\nabla_\rho v^*$.

The line search along the projected negative gradient path is performed inexactly, utilizing the Armijo step length rule (see, e.g., Bertsekas, 1999, Section 2.3, for a description of the overall algorithm applied to a differentiable optimization problem over a convex set). One important consideration for the line search is to give it sufficiently loose stopping criteria, the reason being that each calculation of $\Upsilon(\rho)$ is computationally intensive since it requires reoptimizing a traffic equilibrium problem. Furthermore, it will be necessary to solve each such problem quite accurately.

To the fact that we obtain better solutions than in previous heuristics (see below) contribute the facts that our sensitivity analysis is accurate, and that our equilibrium computations are much more accurate. A more sophisticated algorithm uses the directional derivative information in a subgradient algorithm, such as a bundle method, cf. Patriksson (2004).

## 8.3 Numerical experiments

### 8.3.1 Braess' network

Our first test case covers Braess' network defined in Figure 3. It has been used in previous work by several researchers, for various choices of link cost functions, investment cost functions, and demands. The variation we shall look at is the one used by Bell and Iida (1997), since this problem demonstrates the problem of non-uniqueness of local optima very well. Since this network is very simple, it will not be a particularly important test of our algorithm in terms of performance. However, the symmetry and simplicity of the problem will allow us to look at some aspects of it analytically, which will be helpful in understanding the problems involved in performing network design on more complicated networks.

Let the link cost functions for the network be

$$t_j(\rho_j, v_j) = \begin{cases} 1 + 2\left(\dfrac{v_j}{3.2+\rho_j}\right)^2, & j = 1,4,5, \\ 10 + \left(\dfrac{v_j}{3.2+\rho_j}\right)^2, & j = 2,3, \end{cases}$$

and the investment cost function and bounds be defined by

$$\chi_l(\rho_l) := 3\rho_l; \quad \rho_l^{\min} = 0,\ \rho_l^{\max} = 25, \qquad l = 1,2,3,4,5. \tag{31}$$

First note that there is a symmetry in the network (that is, in the network topology, the travel cost functions, and the investment cost functions) between links 1 and 4, and similarly, between links 2 and 3. This means that any local optimum in the subset defined by $\rho_1 = \rho_4$ and $\rho_2 = \rho_3$ will also be a local optimum in the entire space. The symmetry itself does not tell us that the *global* optimum will necessarily be in this hyperplane, and we shall not prove this. However, this seems reasonable, considering that both the travel cost functions and the investment cost functions are convex. Now, to simplify the problem further, we shall assume that $\rho_2$ (and hence also $\rho_3$) is zero. This may seem a very arbitrary assumption, but it is based on the fact that a significant part of the OD-pair flow will always be on the route through links 1, 5, and 4, and hence, increasing $\rho_2$ (and $\rho_3$) from zero will have a positive cost derivative regardless of $\rho_1$ $(= \rho_4)$. The graph of $\Upsilon$ on the plane specified by $\rho_1 = \rho_4$, $\rho_2 = \rho_3 = 0$ is shown in Figure 9.

We see in this figure that there appear to be two regions in the plane, separated by a curve of points where the problem is non-differentiable. Each region has a single optimum. In fact, these two regions represent the sets of capacity expansions that lead to all routes being used [the region containing $(\rho_1, \rho_4) = (0,0)$] and only the route through links 1, 5, and 4 being used [the second region, containing, for example, $(\rho_1, \rho_4) = (10,10)$]. We can analytically calculate the equation of the curve of non-differentiable points separating the regions. Recall that non-differentiability may arise from the equilibrium solution being non-complimentary. The curve between the two regions can be calculated by assuming that the route through links 1, 5, and 4 is the only one used, but that the other routes (i.e. through the links 1 and 3, and 2 and 5, respectively) have the same cost.

The optimum in the second region seems, from the figure, to be the best one. Calculating this solution analytically yields that $\rho_1 = \rho_5 = \sqrt{\frac{4000}{3}} - 3.2 \approx 7.806424162982$. How does it relate to the results obtained using our algorithm? The fact is, that whenever starting at $(\rho_1, \rho_4) = (0,0)$, the algorithm finds the optimum in the first region (the worse one). The reason is that the descent direction for low values

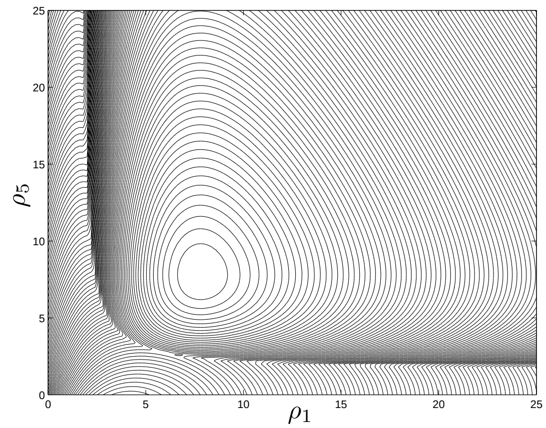


Figure 9: Graph of the design objective function in the plane defined by $\rho_1 = \rho_4$, $\rho_2 = \rho_3 = 0$.

of $\rho_1$ always has a negative $\rho_1$-component (this can be seen in Figure 9). Hence, $\rho_1$ (and $\rho_4$) will never obtain a positive value. The only way to find the better optimum is to use a different starting point. Even so, it can be difficult to predict which local optimum the algorithm will find, when the ridges separating the local optima are unknown.

The better local optimum was found, for example, when the starting point $\rho = (10, 0, 0, 10, 10)^{\mathrm{T}}$ was used. The location of the optimum was calculated to $(7.80642420, 0, 0, 7.80642420, 7.80642420)^{\mathrm{T}}$, which is approximately the same as the exact solution. The design objective function value was found to be 149.7867262002582 (exactly the same as the theoretical value to the precision given by Matlab). It should be noted that also Bell and Iida (1997) found both of these local optima. Bell and Iida used two algorithms that they referred to as the *iterative assignment algorithm* and the *iterative sensitivity algorithm*. The former was based on assuming fixed link flows during the upper level optimization (so as to make it a single level optimization), then re-solving the equilibrium problem and repeating; this is akin to solving the bilevel optimization problem as if it were not a hierarchical game problem but rather a Nash game. (In general it will not lead to the optimal solution to the hierarchical problem.) The iterative sensitivity algorithm is similar to the iterative assignment algorithm, except that also some simple sensitivity information was used at the upper level. Their value for the location of the optimum was $\rho = (7.8064, 0, 0, 7.8064, 7.8064)^{\mathrm{T}}$.

### 8.3.2 The Harker and Friesz network

The network defined by Harker and Friesz (1984) has 6 nodes, 16 links, and 2 OD pairs. It has been used as a network design test network in many previous publications. However, several slightly different network demand scenarios have been used. We shall here consider two scenarios in order to compare our results with as many sources as possible. Capacity improvements are allowed for all 16 links in the network. Network details and the investment cost functions used can be found, for example, in Suwansirikul et al. (1987) or Friesz et al. (1992).

The algorithms with which we compare ours are as follows: Suwansirikul et al. (1987) test both demand scenarios using two algorithms known as the Hooke–Jeeves (H–J) method and the Equilibrium Decomposed Optimization (EDO) algorithms. The Hooke–Jeeves algorithm is a derivative-free method in which one design parameter is optimized at a time while the other variables are kept fixed. (It is similar to the Gauss–Seidel algorithm but has a simpler line search.) EDO is similar except that, in the upper level problem, all design parameters can be modified (this is achieved by a decomposition of the upper level problem into one subproblem for each link).

Friesz et al. (1992) also test both scenarios, using a method known as Simulated Annealing (SA). This is a method inspired by the natural cooling processes that can be described by statistical mechanics. Its advantage is that it (in theory, but only in theory) will always converge to the global minimum. Basically, the algorithm is an iterative method in which there is always a certain probability that a worse solution will be accepted over a previous better one, but where the probability tends towards zero. This is also a derivative-free algorithm.

Suh and Kim (1992) test the first demand scenario using an algorithm denoted the Bilevel Descent

Algorithm (BDA). This algorithm is similar to ours except that Suh and Kim use a simpler sensitivity analysis, similar to that of Tobin and Friesz (1988), and that they utilize the Frank–Wolfe algorithm to solve the traffic equilibrium problems. Huang and Bell (1998), finally, test the second scenario using all of the methods H–J, EDO, and SA.

We will compare all these results with the ones we have obtained using our algorithm, which we denote SBD (as in sensitivity-based descent). In cases where a researcher has published several results for the same algorithm, using different parameters or starting values, the best of these results have been included in the comparison. Tables 2 and 3 below show the comparison of the results, where links with near-zero designs are not shown, and where the interpretation of the reference short-hand should be obvious.

In order to compare the upper-level objective values accurately, each such value $\Upsilon(\rho)$ was recalculated from the corresponding terminal vector $\rho$ using the DSD algorithm to a very strict tolerance level. It was found that many of the previously published solutions really had objective function values that differed quite significantly from what had been stated. Both the originally reported objective function values and those recalculated are therefore presented in the tables.

| Algorithm | SBD | [SFT87]: H–J | [SFT87]: EDO | [Fri+92]: SA | [SuK92]: BDA |
|---|---|---|---|---|---|
| $\rho^0$ | 0 | 0 | — | — | — |
| $\rho^{\max}$ | 20 | — | 20 | 20 | — |
| $\rho_3$ | | 1.2 | 0.13 | | |
| $\rho_6$ | 5.1944681 | 3.0 | 6.26 | 3.1639 | 5.11 |
| $\rho_{15}$ | | 3.0 | 0.13 | | |
| $\rho_{16}$ | 7.5964176 | 2.8 | 6.26 | 6.7240 | 5.71 |
| $\Upsilon$, reported | 199.62532 | 215.08 | 201.84 | 198.10378 | 202.18 |
| $\Upsilon$, recalc. | 199.62526 | 218.20 | 201.20 | 201.33577 | 200.66 |

Table 2: Results for the Harker and Friesz network. Demand scenario I: $d_{16} = 5, d_{61} = 10$.

We note that while the $\Upsilon$-values may seem relatively close to each other, this is due to a large "constant" in $\Upsilon$ because of the zero-congestion link costs; the solutions in $\rho$ are quite different, as is apparent from the table.

It may seem that in the first scenario (Table 2), the H–J solution from Suwansirikul et al. (1987) is quite different from the others and could possibly be in the vicinity of a different local optimum. However, this is not the case, which can be seen by graphing $\Upsilon$ along a straight line through this solution and the solution produced by the SBD algorithm. Apparently, $\Upsilon$ is differentiable and even convex in this one-dimensional subset between the two solutions. Hence, the solution found in Suwansirikul et al. (1987) is in the same region as the others; the objective function is also differentiable at this solution.

The other solutions are all similar in that links 6 and 16 are the only ones with positive capacity expansions. In order to investigate whether there were several local optima in the plane defined by $\rho_l = 0, l \notin \{6, 16\}$, $\Upsilon$ was graphed on this plane; there is however only one local optimum in this plane.

For the second scenario (Table 3), we clearly have solutions in the vicinity of different local optima. An $\Upsilon$-graph shows a non-convexity and even non-differentiability between the solutions obtained using the sensitivity analysis-based algorithm with an upper bound of 20 and 50, respectively. We could have guessed this already when looking at Table 3 since the SBD-execution with $\rho^{\max} = 20$ terminated with a solution not at the upper bound.

A similar graph between the SBD solution (with upper bound 50) and the solution from Friesz et al. (1992) (the SA method) shows that the latter solution is in the vicinity of yet another local optimum.

### 8.3.3 The Sioux Falls network

The 76-link, 24-node, and 728-OD pair network of Sioux Falls, ND, is a classic in algorithm testing in transportation science. The network that has been used for algorithm testing in network design has, however, all OD-pair demands multiplied by 1.1 compared to the standard data. Network design in this network was first attempted by Suwansirikul et al. (1987), who used this modified demand matrix to model peak-hour conditions. They used the Hooke–Jeeves (H–J) and Equilibrium Decomposed Optimization (EDO) algorithms. Capacity improvements were allowed only on a subset of 10 links. The investment cost functions used for these links are described in Suwansirikul et al. (1987).

Apart from Suwansirikul et al, several other researchers have tested network design algorithms on this problem. Friesz et al. (1992) used Simulated Annealing (SA); Huang and Bell (1998) tested all of H–J, EDO, and SA. Lim (2002) uses the Penalty Interior Point Algorithm (PIPA). This is an iterative method

| Algorithm | SBD | SBD | [SFT87]: H–J | [SFT87]: EDO |
|---|---|---|---|---|
| $\rho^0$ | 0 | 0 | 0 | — |
| $\rho^{\max}$ | 20 | 50 | — | 20 |
| $\rho_2$ | 4.6144255 | 4.6144255 | 5.40 | 4.88 |
| $\rho_3$ | 9.8091762 | 9.9102327 | 8.18 | 8.59 |
| $\rho_6$ | 7.7989278 | 7.3744793 | 8.10 | 7.48 |
| $\rho_8$ | 0.5922391 | 0.5922367 | 0.90 | 0.85 |
| $\rho_{14}$ | 1.3152551 | 1.3152551 | 3.90 | 1.54 |
| $\rho_{15}$ | 19.0864543 | | 8.10 | 0.26 |
| $\rho_{16}$ | 0.8823583 | 20.7661321 | 8.40 | 12.52 |
| $\Upsilon$, reported | 557.14050943 | 522.58256876 | 557.22 | 540.74 |
| $\Upsilon$, recalc. | 557.14050943 | 522.58256876 | 561.49 | 540.20 |
| | | | | |
| Algorithm | [Fri+92]: SA | [HuB98]: H–J | [HuB98]: EDO | [HuB98]: SA |
| $\rho^0$ | — | 0 | 0 | 0 |
| $\rho^{\max}$ | 20 | — | 20 | 20 |
| $\rho_2$ | | 4.613 | 4.616 | 1.230 |
| $\rho_3$ | 10.1740 | 9.872 | 12.341 | 10.432 |
| $\rho_6$ | 5.7769 | 5.859 | 7.659 | 7.332 |
| $\rho_8$ | | | 0.593 | 0.207 |
| $\rho_{14}$ | | 0.816 | 1.314 | 2.238 |
| $\rho_{15}$ | | 0.591 | 0.005 | 0.022 |
| $\rho_{16}$ | 17.2786 | 17.288 | 19.995 | 19.472 |
| $\Upsilon$, reported | 528.497 | 526.495 | 525.973 | 531.931 |
| $\Upsilon$, recalc. | 533.329 | 526.513 | 525.871 | 532.327 |

Table 3: Results for the Harker and Friesz network. Demand scenario II: $d_{16} = 10, d_{61} = 20$.

for bilevel optimization problems based on the removal of some feasibility constraints for the traffic equilibrium problem in favour of a penalty term in the objective function, according to how infeasible a solution is with respect to the relaxed constraints. This is the only method tested here that does not rely on repeatedly solving traffic equilibrium problems. (See Luo, Pang, and Ralph, 1996, for a convergence analysis of PIPA.)

Our method was initiated at three different values of $\rho^0$ in order to compare its performance to other methods, which have been applied from different starting points. Table 4 collects the numerical results.

Clearly, the SBD algorithm achieves much better solutions than those previously reported for other algorithms. The poor performance of these algorithms may, at least in part, be due to the low accuracy used in solving the equilibrium problems in the other algorithms—the large discrepancy between the reported and the accurately recalculated objective values of these results certainly suggests this.

It seems likely that all four solutions produced by the SBD algorithm are in the vicinity of the same local optimum. As for the other solutions, they are not accurate enough (that is, near enough to a local optimum) to make a qualified guess. The fact that all upper-level objective values are quite similar does not in itself indicate that all solutions are near the same optimum, particularly as the design objective function for this problem is very flat.

# Final comments

It is certainly true that with probability one the equilibrium link flow solution is differentiable under the assumptions stated; in other words, by randomly picking a vector $\rho^*$ we will not very likely encounter a non-differentiable point. The reality of optimization is however not kind to optimistic thoughts of this nature: a locally optimal solution is, by its nature, extremal, and the likelihood of such a point to be non-differentiable is not zero. Moreover, we have illustrated that if we start our bilevel optimization search in the wrong neighbourhood then we must pass through regions where the equilibrium is non-differentiable, and therefore we must also use tools that are able to detect them. Such a tool is offered in this paper, together with the investigations of subgradients in Patriksson (2004).

Heuristic attempts to avoid differentiable points will most likely fail when the optimal solution is non-differentiable, by preventing the algorithm to ever reaching it. Also, if we ignore non-differentiability, we will probably encounter serious numerical problems when calculating derivatives near a non-differentiable point, quite in the same manner as when inverting a near-singular matrix. The network approach provided in this paper is more likely to give stable results. The results of this paper can be used to construct reliable methods for bilevel optimization together with subgradient techniques, and which can, and should, replace

| Algorithm | SBD | SBD | SBD | SBD | [SFT87]: H–J | [SFT87]: EDO |
|---|---|---|---|---|---|---|
| $\rho^0$ | 0 | 2 | 3 | 6.25 | 2 | — |
| $\rho^{\max}$ | 25 | 25 | 25 | 25 | — | 25 |
| $\rho_{16}$ | 5.3027 | 5.1492 | 5.3457 | 5.2773 | 4.8 | 4.59 |
| $\rho_{17}$ | 2.0560 | 2.0214 | 1.9786 | 2.0533 | 1.2 | 1.52 |
| $\rho_{19}$ | 5.3430 | 5.1679 | 5.3741 | 5.3002 | 4.8 | 5.45 |
| $\rho_{20}$ | 1.9901 | 2.0012 | 1.9460 | 2.0369 | 0.8 | 2.33 |
| $\rho_{25}$ | 2.5216 | 2.4945 | 2.7856 | 2.7670 | 2.0 | 1.27 |
| $\rho_{26}$ | 2.5548 | 2.5447 | 2.8245 | 2.8222 | 2.6 | 2.33 |
| $\rho_{29}$ | 2.9883 | 2.9535 | 2.9257 | 3.0124 | 4.8 | 0.41 |
| $\rho_{39}$ | 4.8559 | 4.8330 | 4.7528 | 4.7348 | 4.4 | 4.59 |
| $\rho_{48}$ | 3.0026 | 2.9798 | 2.9732 | 2.9746 | 4.8 | 2.71 |
| $\rho_{74}$ | 4.8496 | 4.8212 | 4.7347 | 4.7511 | 4.4 | 2.71 |
| $\Upsilon$, reported | 79.9969 | 79.9968 | 79.9987 | 80.0026 | 80.78 | 83.08 |
| $\Upsilon$, recalculated | 79.9961 | 79.9971 | 79.9990 | 80.0043 | 80.67 | 82.34 |
| | | | | | | |
| Algorithm | [Fri+92]: SA | [HuB98]: H–J | [HuB98]: EDO | [HuB98]: SA | [Lim02]: PIPA | |
| $\rho^0$ | 6.25 | 3 | 0 | 6.25 | — | |
| $\rho^{\max}$ | 25 | — | 25 | 25 | — | |
| $\rho_{16}$ | 5.38 | 4.507 | 4.276 | 5.322 | 5.4680 | |
| $\rho_{17}$ | 2.26 | 4.509 | 2.288 | 2.596 | 2.0039 | |
| $\rho_{19}$ | 5.50 | 4.520 | 4.080 | 5.664 | 5.4471 | |
| $\rho_{20}$ | 2.01 | 4.052 | 1.618 | 1.309 | 1.9395 | |
| $\rho_{25}$ | 2.64 | 4.299 | 1.654 | 2.498 | 2.9448 | |
| $\rho_{26}$ | 2.47 | 2.949 | 1.130 | 2.732 | 2.8191 | |
| $\rho_{29}$ | 4.54 | 3.000 | 3.219 | 4.123 | 3.4039 | |
| $\rho_{39}$ | 4.45 | 3.601 | 3.326 | 4.508 | 4.8061 | |
| $\rho_{48}$ | 4.21 | 3.006 | 1.981 | 3.736 | 3.2364 | |
| $\rho_{74}$ | 4.67 | 3.200 | 3.190 | 3.903 | 4.7779 | |
| $\Upsilon$, reported | 80.87 | 83.316 | 83.703 | 81.983 | 80.8669 | |
| $\Upsilon$, recalculated | 80.42 | 81.185 | 81.345 | 80.304 | 80.0528 | |

Table 4: Results for the Sioux Falls network

the heuristics that have been presented so far (including the one illustrated in this paper!).

We finally comment on a popular means to avoid non-differentiabilities in bilevel optimization. "Smoothing" refers to a methodology by which a constraint in the MPEC model implying a non-differentiability, such as the complementarity condition (3a) in the context of the problem (30), is transformed into a sequence of smooth constraints that eventually tend to the original, non-smooth, one. In the context of the traffic equilibrium model we have access to a possible such smoothing instrument, namely by replacing the UE condition by a SUE condition. Suppose, for example, that the logit SUE condition is introduced; then, the equilibrium solution is differentiable with respect to $\rho$, cf. Patriksson (2004), and algorithms for differentiable optimization can be applied to the resulting MPEC problem. As the parameter $\theta$ tends to infinity we approach the UE condition, and hence we tend to attack a problem that, although differentiable, resembles more and more the original problem. (We can also replace the logit SUE with another SUE model, such as one based on the probit SUE model, with the same result.) While the advantage is that we work with a sequence of differentiable MPEC problems instead of a non-differentiable one, we also require the enumeration of every route in the network. Such algorithms could, however, be a viable alternative for small-scale networks.

**Acknowledgments** The second author wishes to thank an anonymous referee for several helpful comments on the analysis, its consequences, and its presentation. He also acknowledges discussions with professor Terry Friesz, and with professors David Boyce and David Watling, especially on the applicability and interpretations of the sensitivity analysis of this paper and in Tobin and Friesz (1988).

## References

M. J. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*, Yale University Press, New Haven, CT, 1956.

M. G. H. Bell and Y. Iida, *Transportation Network Analysis*, John Wiley & Sons, Chichester, UK, 1997.

D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Bellmont, MA, second ed., 1999.

D. Braess, "Über ein Paradox der Verkehrsplannung," *Unternehmensforchung*, 12 (1968), pp. 258–268.

J. V. Burke and M. C. Ferris, “Characterization of solution sets of convex programs,” *Operations Research Letters*, 10 (1991), pp. 57–60.

A. Chen, H. Yang, H. K. Lo, and W. H. Tang, “Capacity reliability of a road network: an assessment methodology and numerical results,” *Transportation Research*, 36B (2002), pp. 225–252.

K. S. Chen and W. H. K. Lam, “Optimal speed detector density for the network with travel time information,” *Transportation Research*, 36B (2002), pp. 203–223.

S. W. Chiou, “Optimization of area traffic control for equilibrium network flows,” *Transportation Science*, 33 (1999), pp. 279–289.

S. W. Chiou, “TRANSYT derivatives for area traffic control optimisation with network equilibrium flows,” *Transportation Research*, 37 (2003), pp. 263–290.

H.-J. Cho, T. E. Smith and T. L. Friesz, “A reduction method for local sensitivity analyses of network equilibrium arc flows,” *Transportation Research*, 34B (2000), pp. 31–51.

S. D. Clark and D. P. Watling, “Probit-based sensitivity analysis for general traffic networks,” *Transportation Research Record*, 1733 (2000), pp. 88–95.

S. D. Clark and D. P. Watling, “Sensitivity analysis of the probit-based stochastic user equilibrium assignment model,” *Transportation Research*, 36B (2002), pp. 617–635.

L. Denault, *Étude de deux méthods d’adjustement de matrices origina–destination à partir des flots des véhicules observés*, PhD thesis, Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1994.

A. L. Dontchev and R. T. Rockafellar, “Ample parameterization of variational inclusions,” *SIAM Journal on Optimization*, 12 (2001), pp. 170–187.

T. L. Friesz, R. L. Tobin, H.-J. Cho, and N. J. Metha, “Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints,” *Mathematical Programming*, 48 (1990), pp. 265–284.

T. L. Friesz, H.-J. Cho, N. J. Metha, R. L. Tobin, and G. Anandalingam, “A simulated annealing approach to the network design problem with variational inequality constraints,” *Transportation Science*, 26 (1992), pp. 18–26.

Z. Y. Gao and Y. F. Song, “A reserve capacity model of optimal signal control with user-equilibrium route choice,” *Transportation Research*, 36B (2002), pp. 313–323.

Z. Y. Gao, H. J. Sun, and L. L. Shan, “A continuous equilibrium network design model and algorithm for transit systems,” *Transportation Research*, 38B (2004), pp. 235–250.

N. H. Gartner, “Optimal traffic assignment with elastic demands: A review. Part II: Algorithmic approaches,” *Transportation Science*, 14 (1980), pp. 192–208.

P. T. Harker and T. L. Friesz, “Bounding the solution of the continuous equilibrium network design problem,” in *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, VNU Science Press, Delft, The Netherlands, 1984, pp. 233–252.

H.-J. Huang and M. G. H. Bell, “Continuous equilibrium network design problem with elastic demand: Derivative-free solution methods,” in *Transportation Networks: Recent Methodological Advances*, M. G. H. Bell (ed.), Pergamon, Amsterdam, The Netherlands, 1998, pp. 175–193.

M. Josefsson, “Sensitivity analysis of traffic equilibria,” Master’s thesis, Department of mathematics, Chalmers University of Technology, Gothenburg, 2003.

T. J. Kim and S. Suh, *Advanced Transport and Spatial Systems Models*, Springer-Verlag, New York, NY, 1990.

J. Kyparisis, “Perturbed solution of variational inequality problems over polyhedral sets,” *Journal of Optimization Theory and Applications*, 66 (1988), pp. 121–135.

J. Kyparisis, “Solution differentiability for variational inequalities,” *Mathematical Programming*, 48 (1990), pp. 285–301.

T. Larsson and M. Patriksson, “Simplicial decomposition with disaggregated representation for the traffic assignment problem,” *Transportation Science*, 26 (1992), pp. 4–17.

F. Leurent, “Sensitivity and error analysis of the dual criteria traffic assignment model,” *Transportation Research*, 32B (1998), pp. 189–204.

A. C. Lim, “Transportation network design problems: An MPEC approach,” Ph.D. thesis, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD, 2002.

Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge

University Press, Cambridge, UK, 1996.

T. Miyagi and T. Suzuki, "A Ramsey price equilibrium model for urban transit systems: A bi-level programming approach with transportation network equilibrium constraints." Conference paper, presented at the 6th World Conference on Transportation Research, Sydney, Australia, 17–20 July, 1995.

J. V. Outrata, "On a special class of mathematical programs with equilibrium constraints," in *Recent Advances in Optimization*, Proceedings of the 8th French–German Conference on Optimization, Trier, July 21–26, 1996, P. Gritzmann, R. Horst, E. Sachs, and R. Tichatschke, eds., Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1997, pp. 246–260.

J. V. Outrata, M. Kočvara, and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, no. 28 in Nonconvex Optimization and Its Applications, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

M. Patriksson, *The Traffic Assignment Problem—Models and Methods*, Topics in Transportation, VSP BV, Utrecht, The Netherlands, 1994.

M. Patriksson, "Sensitivity analysis of traffic equilibria," *Transportation Science*, 37 (2004), pp. 258–281.

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

M. Patriksson and R.T. Rockafellar, "Sensitivity analysis of variational inequalities over aggregated polyhedra, with application to traffic equilibria," *Transportation Science*, 37 (2003), pp. 56–68.

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

S. M. Robinson, "Strongly regular generalized equations," *Mathematics of Operations Research*, 5 (1980), pp. 43–62.

S. Suh and T.J. Kim, "Solving nonlinear bilevel programming models of the equilibrium network design problem: A comparative review," in *Hierarchical Optimization*, G. Anandalingam and T. L. Friesz, eds., vol. 34 of Annals of Operations Research, J.C. Baltzer AG, Basel, Switzerland, 1992, pp. 203–218.

C. Suwansirikul, T. L. Friesz, and R. L. Tobin, "Equilibrium decomposed optimization: A heuristic for the continuous equilibrium network design problem," *Transportation Science*, 21 (1987), pp. 254–260.

M. L. Tam and W. H. K. Lam, "Maximum car ownership under constraints of road capacity and parking space," *Transportation Research*, 34A (2000), pp. 145–170.

M. L. Tam and W. H. K. Lam, "Balance of car ownership under user demand and road network supply conditions: Case study in Hong Kong," *Journal of Urban Planning and Development/ASCE*, 130 (2004), pp. 24–36.

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

C. K. Wong and S. C. Wong, "Lane-based optimization of traffic equilibrium settings for area traffic control," *Journal of Advanced Transportation*, 36 (2002), pp. 349–386.

S. C. Wong, Y. C. Du, H. W. Ho, and L. J. Sun, "Simultaneous optimization formulation of a discrete-continuous transportation system," *Transportation Research Record*, 1857 (2003), pp. 11–20.

S. C. Wong and H. Yang, "Reserve capacity of a signal-controlled road network," *Transportation Research*, 31B (1997), pp. 397–402.

S. C. Wong, H. Yang, and H. K. Lo, "A path-based traffic assignment algorithm based on the TRANSYT traffic model," *Transportation Research*, 35B (2001), pp. 163–181.

H. Yang, "Heuristic algorithms for the bilevel origin–destination matrix estimation problems," *Transportation Research*, 29B (1995), pp. 231–242.

H. Yang, "Sensitivity analysis for queuing equilibrium network flow and its application to traffic control," *Mathematical and Computer Modelling*, 22 (1995), pp. 247–258.

H. Yang, "Sensitivity analysis for the elastic-demand network equilibrium problem with applications," *Transportation Research*, 31B (1997), pp. 55–70.

H. Yang and M. G. H. Bell, "Models and algorithms for road network design: A review and some new developments," *Transportation Reviews*, 18 (1998), pp. 257–278.

H. Yang, M. G. H. Bell, and Q. Meng, "Modeling the capacity and level of service of urban transportation networks," *Transportation Research*, 34B (2000), pp. 255–275.

H. Yang and W. H. K. Lam, "Optimal road tolls under conditions of queueing and congestion," *Transportation*

*Research*, 30A (1996), pp. 319–332.

H. Yang and Q. Meng, "Highway pricing and capacity choice in a road network under a build-operate-transfer scheme," *Transportation Research*, 34A (2000), pp. 207–222.

H. Yang, T. Sasaki, Y. Iida, and Y. Asakura, "Estimation of origin–destination matrices from link traffic counts on congested networks," *Transportation Research*, 26B (1992), pp. 417–434.

H. Yang and K. K. Woo, "Competition and equilibria of private toll roads in a traffic network," *Transportation Research Record*, 1733 (2000), pp. 15–22.

H. Yang and S. Yagar, "Traffic assignment and traffic control in general freeway-arterial corridor systems," *Transportation Research*, 28B (1994), pp. 463–486.

H. Yang and S. Yagar, "Traffic assignment and signal control in saturated road networks," *Transportation Research*, 29A (1995), pp. 125–139.

H. Yang, S. Yagar, Y. Iida, and Y. Asakura, "An algorithm for the inflow control problem on urban freeway networks with user-optimal flows," *Transportation Research*, 28B (1994), pp. 123–139.

N. D. Yen, "Lipschitz continuity of solutions of variational inequalities with a parametric polyhedral constraint," *Mathematics of Operations Research*, 20 (1995), pp. 695–708.

Y. F. Yin and H. Ieda, "Optimal improvement scheme for network reliability," *Transportation Research Record*, 1783 (2002), pp. 1–6.