Available online at www.sciencedirect.com

SCIENCE DIRECT®

Transportation Research Part B 39 (2005) 769–795

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

Sensitivity analysis of stochastic user equilibrium flows in a bi-modal network with application to optimal pricing

Jiang Qian Ying [a,*], Hai Yang [b]

[a] Faculty of Regional Studies, Gifu University, Yanagido, Gifu 501, Japan
[b] Department of Civil Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

Received 16 May 2002; received in revised form 3 May 2003; accepted 16 September 2003

Abstract

Sensitivity analysis methods for transport systems having an automobile road network and a physically separate transit network are studied. For the case that both the automobile and transit networks are congested in the sense that link cost functions increase with the flow, a general computational method is presented for sensitivity analysis. For another case where the automobile network is congested but the transit network simply consists of independent lines connecting Origin–Destination pairs and may have economies of scale with the increase of the passengers, conditions under which sensitivity analysis can be properly conducted are investigated. The sensitivity analysis algorithm is applied to the optimal pricing problem in a combined network with transit lines exhibiting economies of scale as well as congestion diseconomies.

Keywords: Transportation networks; Transit; Sensitivity analysis; Congestion; Scale economies; Pricing

* Corresponding author. Tel.: +81 58 2933306; fax: +81 58 2933118.
E-mail address: ying@cc.gifu-u.ac.jp (J.Q. Ying).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2003.09.003

1. Introduction

Sensitivity analysis is an important method for solving various optimization problems in transportation systems (see, e.g., Luo et al., 1996; Miyagi and Suzuki, 1996; Bell and Iida, 1997; Yang, 1997; Yang et al., 2001). Efficient computational algorithms for sensitivity analysis of the multinominal logit stochastic user equilibrium model (Sheffi, 1985) have been proposed by Davis (1994), Ying and Miyagi (2001), Ying et al. (2001), Huang et al. (2001) for traffic networks with fixed and elastic travel demands. Recently, a formulation of sensitivity analysis method has been provided by Clark and Watling (2002) which can deal with more general SUE models including the probit model.

The purpose of this work is to develop effective sensitivity analysis methods for solving various optimization problems in a combined transport system consisting of an automobile road network along with a physically separate public transit network, e.g., a railway or subway network. The explicit treatment of transit network is important in transportation system optimization problems. In particular, due to the complex cost structure of transit system, there is a need to explicitly formulate its cost functions in order to set road tolls and transit fares that is optimal from the social welfare point of view. In this paper, we assume that the supplier of the transit service is a zero-profit firm. We also assume that the subsidies for part of the fixed cost for the construction of the transit and an operational cost from the public sector are given in lump sum, the rest of the costs are shared by the passengers so that the firm breaks even. Even in this simplified framework, there are many factors that make up the complexity of the structure of the cost incurred by transit users. On the one hand, in a metropolitan area, congestion is usually a dominating factor that causes external diseconomies for the transit users. On the other hand, in rural regions where the passenger flows are far less than the potential capacity of the transit system, the increase of passengers brings about extra profit to the supply side of the transit service, which may be re-invested for accomplishing a higher frequency operation schedule, thus shortening the waiting time in using transit and decrease the general cost of the passengers. Such a case involves external economies or economies of scale. Another well-studied factor accounting for the economies of scale is that the fixed cost for construction of a transit system, even after being partially subsidized, is still very large. If the transit service is offered under a break-even constraint, as is assumed in this work, increased passengers will lead to a lower fare because they share the fixed and variable service cost.

In this paper we provide sensitivity analysis methods for two kinds of combined transport systems. The first is the case where a transit network with general topology exists along with the road network and both networks are congested, that is, the cost incurred by passenger is a function increasing absolutely with the passenger flow on each link in each network. An efficient link-based sensitivity analysis method is proposed for computing the derivatives of road link vehicle flows and transit link passenger flows with respect to the link parameters. Under the congestion assumptions, the equilibrium traffic flow assignment problem has a unique solution, the proposed sensitivity analysis method always works for such an equilibrium. These results are presented in Section 2. The second case is a system with a road network and transit lines directly connecting Origin–Destination (O–D) pairs. The performance function of each transit line may have external diseconomies caused by congestion, as well as economies of scale. In such a case with economies of scale on the transit, the modal split and traffic assignment problem involves the difficulty of singularity. For this case a sensitivity analysis algorithm is also provided and conditions are examined under which the sensitivity analysis can be applied. These results are presented in Section 3. An example is given in Section 4 for demonstrating the effectiveness of the proposed sensitivity analysis algorithm for solving the optimal pricing problem in a transport system with a road network and O–D-specific transit lines which may have economies of scale. A social utility for the combined transportation system is introduced based on random utility theory. The example shows that besides the well-known marginal cost pricing scheme, there are other pricing schemes that optimize the social utility. Some related problems are described in Section 5.

2. Sensitivity analysis for congested road and transit networks

2.1. Stochastic user equilibrium on the road traffic network

Notations for road network:

- $N = \{i, j, \ldots\}$: set of nodes
- $A = \{ij, \ldots\}$: set of links
- $W = \{rs, \ldots\}$: set of O–D pairs
- S_{rs}: the expected minimum cost for O–D rs on the road network
- q_{rs}: elastic O–D demand in the road network
- $\boldsymbol{q} = (q_{rs})_{rs\in W}$ denotes the column vector of O–D demands
- $R_{rs} = \{k, p, \ldots\}$: set of paths connecting rs
- h_k^{rs}: flow on path k with origin r and destination s
- P_k^{rs}: probability that a traveler from r to s chooses path k
- P_{ij}^{rs}: probability that a traveler from r to s traces link ij
- x_{ij}: link flow, for $ij \in A$
- $t_{ij}(x_{ij}, \epsilon_{ij})$: differentiable cost function of link ij with respect to flow x_{ij}, and parameter ϵ_{ij}. It is assumed that t_{ij} is strongly monotonically increasing with respect to x_{ij}. For given ϵ_{ij}, the inverse of the cost function is denoted as $x_{ij}(t_{ij}, \epsilon_{ij})$, which is also strongly monotonically increasing in t_{ij}
- $(x_{ij})_{t_{ij}}$: partial derivative of x_{ij} with respect to t_{ij}
- $(x_{ij})_{\epsilon_{ij}}$: partial derivative of x_{ij} with respect to ϵ_{ij}
- $\boldsymbol{x} = (x_{ij})_{ij\in A}$, $\boldsymbol{t} = (t_{ij})_{ij\in A}$ and $\boldsymbol{\epsilon} = (\epsilon_{ij})_{ij\in A}$ denote the column vectors of all link flows, link costs and uncertainty parameters, respectively
- $\delta_{ij,k}^{rs} = \begin{cases} 1 & \text{if } ij \text{ is a link on path } k \\ 0 & \text{otherwise} \end{cases}$
- $\delta_{ij,gh}^{rs} = \begin{cases} 1 & \text{if } ij = gh \in A \\ 0 & \text{otherwise} \end{cases}$
- $c_k^{rs} = \sum_{ij\in A} t_{ij}\delta_{ij,k}^{rs}$: the total cost of traveling on a path $k \in R_{rs}$
- θ: a dispersion parameter in SUE

- For simplicity, summation notations $\sum_{k\in R_{rs}}$, $\sum_{p\in R_{rs}}$, $\sum_{rs\in W}$ will be abbreviated as $\sum_k$, $\sum_p$, $\sum_{rs}$, respectively

We assume that the costs t_{ij} and c_k^{rs} are a kind of generalized costs containing prices, time, and other factors, measured by monetary term. The road network is said to be "congested" in the sense that t_{ij} is strongly monotonically increasing with respect to x_{ij}, as is assumed in this paper. Note that it is implicitly assumed that there is no interference between link flows.

In a multinominal logit-based stochastic user equilibrium (SUE), for a traveler on O–D pair $rs \in W$, the probability P_k^{rs} at which the path k is chosen is given by (see, e.g., Sheffi, 1985)

$$P_k^{rs} = \frac{\exp(-\theta c_k^{rs})}{\sum_p \exp(-\theta c_p^{rs})}, \quad k \in R_{rs}. \tag{1}$$

At stochastic user equilibrium, the path flows are

$$h_k^{rs} = q_{rs}\frac{\exp(-\theta c_k^{rs})}{\sum_p \exp(-\theta c_p^{rs})}, \quad k \in R_{rs}. \tag{2}$$

The link flows are

$$x_{ij} = \sum_{rs}\sum_k h_k^{rs}\delta_{ij,k}^{rs} = \sum_{rs} q_{rs}\frac{\sum_k \exp(-\theta c_k^{rs})\delta_{ij,k}^{rs}}{\sum_p \exp(-\theta c_p^{rs})}, \quad ij \in A. \tag{3}$$

In traffic assignment literature, θ is usually understood as a dispersion parameter indicating how precisely a driver can correctly choose the shortest routes; the larger the θ, the higher the probability that a driver chooses the shortest routes (Sheffi, 1985). However, in compatible with the economic theory of discrete choice, $1/\theta$ is better understood as a measure of the individual's preference for variety of choice. The expected disutility S_{rs} for using the road network is given by

$$S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})) = -\frac{1}{\theta}\ln\sum_k \exp(-\theta c_k^{rs}).$$

$-S_{rs}$ is the *expected utility* of using the road network. This quantity has been justified theoretically sound for measuring benefit of the network. See Williams (1977) and Section 4.1 of this paper for more details.

2.2. Stochastic user equilibrium on the transit network

Notations for transit network:

- $\widehat{N} = \{i, j, \ldots\}$: set of stops
- $\hat{A} = \{ij, \ldots\}$: set of links
- $\widehat{S}_{rs}$: the expected minimum cost for O–D rs on the transit network
- $\hat{q}_{rs}$: demand for O–D pair rs
- $\hat{\boldsymbol{q}} = (\hat{q}_{rs})_{rs\in W}$ denotes the vector of all O–D demands
- $\widehat{R}_{rs} = \{k, p, \ldots\}$: set of paths connecting rs

- $\hat{h}_k^{rs}$: flow on path k with origin r and destination s
- $\widehat{P}_k^{rs}$: probability that a traveler from r to s chooses path k
- $\widehat{P}_{ij}^{rs}$: probability that a traveler from r to s traces link ij
- $\hat{x}_{ij}$: link flow, for $ij \in \hat{A}$
- $\hat{t}_{ij}(\hat{x}_{ij}, \hat{\epsilon}_{ij})$: *average cost* function of transit link ij, differentiable with respect to flow $\hat{x}_{ij}$, and parameter $\hat{\epsilon}_{ij}$
  It is assumed that $\hat{t}_{ij}$ is strongly monotonically increasing with respect to $\hat{x}_{ij}$. For given $\hat{\epsilon}_{ij}$, the inverse of the cost function is denoted as $\hat{x}_{ij}(\hat{t}_{ij}, \hat{\epsilon}_{ij})$, which is also strongly monotonically increasing in $\hat{t}_{ij}$
- $(\hat{x}_{ij})_{\hat{t}_{ij}}$: partial derivative of $\hat{x}_{ij}$ with respect to $\hat{t}_{ij}$
- $(\hat{x}_{ij})_{\hat{\epsilon}_{ij}}$: partial derivative of $\hat{x}_{ij}$ with respect to $\hat{\epsilon}_{ij}$
- $\hat{\boldsymbol{x}} = (\hat{x}_{ij})_{ij\in\hat{A}}$, $\hat{\boldsymbol{t}} = (\hat{t}_{ij})_{ij\in\hat{A}}$ and $\hat{\boldsymbol{\epsilon}} = (\hat{\epsilon}_{ij})_{ij\in\hat{A}}$ denote the vectors of all link flows, link costs and uncertainty parameters, respectively
- $\hat{c}_k^{rs} = \sum_{ij\in\hat{A}} \hat{t}_{ij}\delta_{ij,k}^{rs}$: the total cost of traveling on a path $k \in \widehat{R}_{rs}$
- $\hat{\theta}$: a dispersion parameter in SUE

As in the case of road network, we assume that the costs $\hat{t}_{ij}$ and $\hat{c}_k^{rs}$ are a kind of generalized costs containing fare, time, and other factors, measured by monetary term. The transit network is said to be "congested" in the sense that $\hat{t}_{ij}$ is strongly monotonically increasing with respect to $\hat{x}_{ij}$, as is assumed in this section.

Note that the transit network has no interference with the road network. In practice, the set of nodes of public transit network is usually a subset of that of road network, covering all the origins and destinations for travel demands. See Abrahamsson and Lundqvist (1999) for a model for Stockholm.

For a traveler on O–D pair $rs \in W$, the probability $\widehat{P}_k^{rs}$ at which the path k is chosen is given by

$$\widehat{P}_k^{rs} = \frac{\exp(-\theta\hat{c}_k^{rs})}{\sum_p \exp\left(-\hat{\theta}\hat{c}_p^{rs}\right)}, \quad k \in \widehat{R}_{rs}. \tag{4}$$

At stochastic user equilibrium, the path flows are

$$\hat{h}_k^{rs} = \hat{q}_{rs}\frac{\exp(-\hat{\theta}\hat{c}_k^{rs})}{\sum_p \exp\left(-\hat{\theta}\hat{c}_p^{rs}\right)}, \quad k \in \widehat{R}_{rs}. \tag{5}$$

The link flows are

$$\hat{x}_{ij} = \sum_{rs}\sum_k \hat{h}_k^{rs}\delta_{ij,k}^{rs} = \sum_{rs}\hat{q}_{rs}\frac{\sum_k \exp\left(-\hat{\theta}\hat{c}_k^{rs}\right)\delta_{ij,k}^{rs}}{\sum_p \exp\left(-\hat{\theta}\hat{c}_p^{rs}\right)}, \quad ij \in \hat{A}. \tag{6}$$

The disutility $\widehat{S}_{rs}$ for using the transit network is

$$\widehat{S}_{rs}(\hat{\boldsymbol{c}}^{rs}(\hat{\boldsymbol{t}})) = -\frac{1}{\hat{\theta}}\ln\sum_k \exp\left(-\hat{\theta}\hat{c}_k^{rs}\right).$$

2.3. Travel mode choice

Notations for demands:

- $\bar{q}_{rs}$: total demand for O–D pair rs
- $q_{rs} = D_{rs}(S_{rs}, \widehat{S}_{rs})$: elastic O–D demand in the transit network; D_{rs} is a strictly decreasing function with S_{rs} and increasing with $\widehat{S}_{rs}$
- $\hat{q}_{rs} = \widehat{D}_{rs}(\widehat{S}_{rs}, S_{rs}) = \bar{q}_{rs} - q_{rs}$, the demand on the transit network

In a logit model, the road network travel demand function takes the form

$$D_{rs}\left(S_{rs}, \widehat{S}_{rs}\right) = \bar{q}_{rs} \frac{1}{1 + \exp\left(-\alpha\left(\widehat{S}_{rs} - S_{rs}\right)\right)}, \tag{7}$$

where α is a parameter reflecting the characteristics of the traveler's behavior regarding travel mode choice. The demand for transit use from r to s is

$$\widehat{D}_{rs}\left(\widehat{S}_{rs}, S_{rs}\right) = \bar{q}_{rs} - q_{rs} = \bar{q}_{rs} \frac{1}{1 + \exp\left(-\alpha\left(S_{rs} - \widehat{S}_{rs}\right)\right)}. \tag{8}$$

The expected disutility at the modal split level is measured by

$$\overline{S}_{rs} = -\frac{1}{\alpha} \ln\left(\exp(-\alpha S_{rs}) + \exp\left(-\alpha \widehat{S}_{rs}\right)\right).$$

The economic meaning of this measure will be discussed in Section 4. In the following we reformulate the SUE conditions on the road and transit network and establish the positive definiteness of certain matrix that plays an important role in the sensitivity analysis.

2.4. SUE conditions and properties

2.4.1. Reformulation of SUE conditions

It is easy to derive

$$\frac{\partial S_{rs}}{\partial t_{ij}} = \frac{\sum_k \exp(-\theta c_k^{rs}) \delta_{ij,k}^{rs}}{\sum_p \exp(-\theta c_p^{rs})} = P_k^{rs},$$

thus the equilibrium conditions are characterized by the following equations:

$$F_{ij}(\boldsymbol{t}, \hat{\boldsymbol{t}}; \boldsymbol{\epsilon}) = x_{ij}(t_{ij}, \epsilon_{ij}) - \sum_{rs} q_{rs} \frac{\partial S_{rs}}{\partial t_{ij}} = 0, \quad ij \in A, \tag{9}$$

or in a vector form

$$\boldsymbol{F}(\boldsymbol{t}, \hat{\boldsymbol{t}}, \boldsymbol{\epsilon}) = \boldsymbol{x} - \sum_{rs} q_{rs} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})))^{\mathrm{T}} = \boldsymbol{0}, \tag{10}$$

where ∇ is the gradient operator and T denotes transposed matrix. In the equations $q_{rs} = D_{rs}(S_{rs}, \widehat{S}_{rs})$ are road network travel demands. Note that the vector of variables $\hat{\boldsymbol{t}}$ enters into F_{ij} through q_{rs}. Similarly, given $\boldsymbol{t}$ fixed, the SUE on the transit network is defined by

$$\widehat{\boldsymbol{F}}(\hat{\boldsymbol{t}}, \boldsymbol{t}; \boldsymbol{\epsilon}) = \hat{\boldsymbol{x}} - \sum_{rs} \hat{q}_{rs} \left(\nabla_{\boldsymbol{t}} \widehat{S}_{rs}(\hat{\boldsymbol{c}}^{rs}(\hat{\boldsymbol{t}})) \right)^{\mathrm{T}} = \boldsymbol{0}. \tag{11}$$

Eqs. (10) with (11) describe the complete equilibrium condition in the combined transportation system and define the link costs $\boldsymbol{t}$ and $\hat{\boldsymbol{t}}$ as implicit functions in $\boldsymbol{\epsilon}$ and $\hat{\boldsymbol{\epsilon}}$. We will show that the following Jacobian matrix:

$$\begin{bmatrix} \nabla_{\boldsymbol{t}} \boldsymbol{F} & \nabla_{\hat{\boldsymbol{t}}} \boldsymbol{F} \\ \nabla_{\boldsymbol{t}} \widehat{\boldsymbol{F}} & \nabla_{\hat{\boldsymbol{t}}} \widehat{\boldsymbol{F}} \end{bmatrix} \tag{12}$$

is a positive definite matrix and thus Eqs. (10) and (11) have a unique solution for $\boldsymbol{t}$ and $\hat{\boldsymbol{t}}$, and hence for $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, $\boldsymbol{q}$ and $\hat{\boldsymbol{q}}$.

2.4.2. *Positive definiteness of* $\begin{bmatrix} \nabla_{\boldsymbol{t}} \boldsymbol{F} & \nabla_{\hat{\boldsymbol{t}}} \boldsymbol{F} \\ \nabla_{\boldsymbol{t}} \widehat{\boldsymbol{F}} & \nabla_{\hat{\boldsymbol{t}}} \widehat{\boldsymbol{F}} \end{bmatrix}$ *and uniqueness of SUE flows*

From (9) we have

$$\begin{aligned} \frac{\partial F_{ij}}{\partial t_{gh}} &= (x_{ij})_{t_{ij}} \delta_{ij,gh} \\ &\quad - \sum_{rs} q_{rs} \left[\frac{-\theta \sum_k \exp(-\theta c_k^{rs}) \delta_{ij,k}^{rs} \delta_{gh,k}^{rs}}{\sum_p \exp(-\theta c_p^{rs})} - \frac{-\theta \left(\sum_k \exp(-\theta c_k^{rs}) \delta_{ij,k}^{rs} \right) \left(\sum_l \exp(-\theta c_l^{rs}) \delta_{gh,l}^{rs} \right)}{\left(\sum_p \exp(-\theta c_p^{rs}) \right)^2} \right] \\ &\quad - \sum_{rs} \frac{\partial D_{rs}}{\partial S_{rs}} \frac{\left(\sum_k \exp(-\theta c_k^{rs}) \delta_{ij,k}^{rs} \right) \left(\sum_l \exp(-\theta c_l^{rs}) \delta_{gh,l}^{rs} \right)}{\left(\sum_p \exp(-\theta c_p^{rs}) \right)^2}, \quad ij, gh \in A, \end{aligned} \tag{13}$$

or in a compact form

$$\nabla_{\boldsymbol{t}} \boldsymbol{F} = \mathrm{diag}((x_{ij})_{t_{ij}})_{ij} - \sum_{rs} q_{rs} \nabla_{\boldsymbol{t}}^2 S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})) - \sum_{rs} \frac{\partial D_{rs}}{\partial S_{rs}} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})))^{\mathrm{T}} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t}))), \tag{14}$$

and

$$\nabla_{\hat{\boldsymbol{t}}} \boldsymbol{F} = -\sum_{rs} \frac{\partial D_{rs}}{\partial \widehat{S}_{rs}} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})))^{\mathrm{T}} \left(\nabla_{\hat{\boldsymbol{t}}} \widehat{S}_{rs}(\hat{\boldsymbol{c}}^{rs}(\hat{\boldsymbol{t}})) \right). \tag{15}$$

For function $\widehat{\boldsymbol{F}}$, we have similarly

$$\nabla_{\hat{\boldsymbol{t}}} \widehat{\boldsymbol{F}} = \mathrm{diag}((\hat{x}_{i'j'})_{\hat{t}_{i'j'}})_{i'j'} - \sum_{rs} \hat{q}_{rs} \nabla_{\hat{\boldsymbol{t}}}^2 \widehat{S}_{rs}(\hat{\boldsymbol{c}}^{rs}(\hat{\boldsymbol{t}})) - \sum_{rs} \frac{\partial \widehat{D}_{rs}}{\partial \widehat{S}_{rs}} \left(\nabla_{\hat{\boldsymbol{t}}} \widehat{S}_{rs}(\hat{\boldsymbol{c}}^{rs}(\hat{\boldsymbol{t}})) \right)^{\mathrm{T}} \left(\nabla_{\hat{\boldsymbol{t}}} \widehat{S}_{rs}(\hat{\boldsymbol{c}}^{rs}(\hat{\boldsymbol{t}})) \right), \tag{16}$$

and

$$\nabla_{\boldsymbol{t}} \widehat{\boldsymbol{F}} = -\sum_{rs} \frac{\partial \widehat{D}_{rs}}{\partial S_{rs}} \left(\nabla_{\hat{\boldsymbol{t}}} \widehat{S}_{rs}(\hat{\boldsymbol{c}}^{rs}(\hat{\boldsymbol{t}})) \right)^{\mathrm{T}} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t}))). \tag{17}$$

It is trivial to show that

$$\frac{\partial D_{rs}}{\partial S_{rs}} = -\frac{\partial D_{rs}}{\partial \widehat{S}_{rs}} = \frac{\partial \widehat{D}_{rs}}{\partial S_{rs}} = -\frac{\partial \widehat{D}_{rs}}{\partial \widehat{S}_{rs}} < 0.$$

Thus we have

$$\begin{bmatrix} \nabla_{\boldsymbol{t}}\boldsymbol{F} & \nabla_{\hat{\boldsymbol{t}}}\boldsymbol{F} \\ \nabla_{\boldsymbol{t}}\widehat{\boldsymbol{F}} & \nabla_{\hat{\boldsymbol{t}}}\widehat{\boldsymbol{F}} \end{bmatrix} = \begin{bmatrix} \mathrm{diag}((x_{ij})_{t_{ij}})_{ij} & 0 \\ 0 & \mathrm{diag}\left((\hat{x}_{i'j'})_{\hat{t}_{i'j'}}\right)_{i'j'} \end{bmatrix} + \begin{bmatrix} -\sum_{rs} q_{rs}\nabla^2_{\boldsymbol{t}} S_{rs} & 0 \\ 0 & -\sum_{rs}\hat{q}_{rs}\nabla^2_{\hat{\boldsymbol{t}}}\widehat{S}_{rs} \end{bmatrix}$$
$$- \sum_{rs} \frac{\partial D_{rs}}{\partial S_{rs}} \begin{bmatrix} (\nabla_{\boldsymbol{t}} S_{rs})^{\mathrm{T}}(\nabla_{\boldsymbol{t}} S_{rs}) & -(\nabla_{\boldsymbol{t}} S_{rs})^{\mathrm{T}}(\nabla_{\hat{\boldsymbol{t}}}\widehat{S}_{rs}) \\ -(\nabla_{\hat{\boldsymbol{t}}}\widehat{S}_{rs})^{\mathrm{T}}(\nabla_{\boldsymbol{t}} S_{rs}) & (\nabla_{\hat{\boldsymbol{t}}}\widehat{S}_{rs})^{\mathrm{T}}(\nabla_{\hat{\boldsymbol{t}}}\widehat{S}_{rs}) \end{bmatrix}. \tag{18}$$

The diagonal matrix $\mathrm{diag}((x_{ij})_{t_{ij}})_{ij}$ is positive definite, since each diagonal entry is positive from the assumption that x_{ij} is strongly monotonically increasing in t_{ij}. It is well known (see, e.g., Sheffi, 1985, p. 278) that $S_{rs}(\boldsymbol{c}^{rs})$ is concave with respect to $\boldsymbol{c}^{rs}$. As $\boldsymbol{c}^{rs}$ is a vector with components which are linear combinations of t_{ij}, it is thus shown that the function $-S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t}))$ is convex with respect to t_{ij}, and the matrix $-\sum_{rs} q_{rs}\nabla^2_{\boldsymbol{t}} S_{rs}$ is positive semidefinite. The same statements also hold for $\mathrm{diag}((\hat{x}_{i'j'})_{\hat{t}_{i'j'}})_{i'j'}$ and $-\sum_{rs}\hat{q}_{rs}\nabla^2_{\hat{\boldsymbol{t}}}\widehat{S}_{rs}$. We have thus shown that the first and second terms in the right hand side of (18) are positive definite and positive semidefinite, respectively. The third term can be rewritten as

$$-\sum_{rs} \frac{\partial D_{rs}}{\partial S_{rs}} \begin{bmatrix} (\nabla_{\boldsymbol{t}} S_{rs})^{\mathrm{T}} \\ -(\nabla_{\hat{\boldsymbol{t}}}\widehat{S}_{rs})^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} (\nabla_{\boldsymbol{t}} S_{rs}) & -(\nabla_{\hat{\boldsymbol{t}}}\widehat{S}_{rs}) \end{bmatrix}.$$

Invoking the fact that $\frac{\partial D_{rs}}{\partial S_{rs}} < 0$, this term is also easily shown to be a positive semidefinite matrix. Thus the left side is a positive definite matrix. Therefore when ϵ and $\hat{\epsilon}$ are given, the solution $\boldsymbol{t}$ and $\hat{\boldsymbol{t}}$ for Eqs. (10) and (11) are unique, so are the link flows and mode-specific demands.

2.5. Sensitivity analysis method

In Davis (1994), Ying and Miyagi (2001), it has been shown that

(i) $$\frac{\sum_k \exp(-\theta c_k^{rs})\delta_{ij,k}^{rs}}{\sum_k \exp(-\theta c_k^{rs})}$$

and

(ii) $$\frac{\sum_k \exp(-\theta c_k^{rs})\delta_{ij,k}^{rs}\delta_{gh,k}^{rs}}{\sum_k \exp(-\theta c_k^{rs})}$$

can be computed by using Dial's algorithm in such a way that the paths do not have to be enumerated. Therefore $\nabla_{\boldsymbol{t}}\boldsymbol{F}$, $\nabla_{\hat{\boldsymbol{t}}}\boldsymbol{F}$, $\nabla_{\boldsymbol{t}}\widehat{\boldsymbol{F}}$, and $\nabla_{\hat{\boldsymbol{t}}}\widehat{\boldsymbol{F}}$, can be efficiently computed in a link-based manner.

In the following we establish the formulae for computing the derivatives of $\boldsymbol{t}$, $\hat{\boldsymbol{t}}$, $\boldsymbol{x}$, $\hat{\boldsymbol{x}}$, $\boldsymbol{q}$ and $\hat{\boldsymbol{q}}$ with respect to ϵ and $\hat{\epsilon}$. From the equations

$$\begin{cases} \boldsymbol{F}(\boldsymbol{t}, \hat{\boldsymbol{t}}; \boldsymbol{\epsilon}) = 0, \\ \widehat{\boldsymbol{F}}(\hat{\boldsymbol{t}}, \boldsymbol{t}; \hat{\boldsymbol{\epsilon}}) = 0, \end{cases} \tag{19}$$

the following formulae are derived:

$$\begin{bmatrix} \nabla_{\boldsymbol{t}}\boldsymbol{F} & \nabla_{\hat{\boldsymbol{t}}}\boldsymbol{F} \\ \nabla_{\boldsymbol{t}}\widehat{\boldsymbol{F}} & \nabla_{\hat{\boldsymbol{t}}}\widehat{\boldsymbol{F}} \end{bmatrix} \begin{bmatrix} \frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\epsilon}} & \frac{\partial \boldsymbol{t}}{\partial \hat{\boldsymbol{\epsilon}}} \\ \frac{\partial \hat{\boldsymbol{t}}}{\partial \boldsymbol{\epsilon}} & \frac{\partial \hat{\boldsymbol{t}}}{\partial \hat{\boldsymbol{\epsilon}}} \end{bmatrix} + \begin{bmatrix} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\epsilon}} & 0 \\ 0 & \frac{\partial \widehat{\boldsymbol{F}}}{\partial \hat{\boldsymbol{\epsilon}}} \end{bmatrix} = 0, \tag{20}$$

$$\begin{bmatrix} \frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\epsilon}} & \frac{\partial \boldsymbol{t}}{\partial \hat{\boldsymbol{\epsilon}}} \\ \frac{\partial \hat{\boldsymbol{t}}}{\partial \boldsymbol{\epsilon}} & \frac{\partial \hat{\boldsymbol{t}}}{\partial \hat{\boldsymbol{\epsilon}}} \end{bmatrix} = -\begin{bmatrix} \nabla_{\boldsymbol{t}}\boldsymbol{F} & \nabla_{\hat{\boldsymbol{t}}}\boldsymbol{F} \\ \nabla_{\boldsymbol{t}}\widehat{\boldsymbol{F}} & \nabla_{\hat{\boldsymbol{t}}}\widehat{\boldsymbol{F}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\epsilon}} & 0 \\ 0 & \frac{\partial \widehat{\boldsymbol{F}}}{\partial \hat{\boldsymbol{\epsilon}}} \end{bmatrix}, \tag{21}$$

where

$$\begin{bmatrix} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\epsilon}} & 0 \\ 0 & \frac{\partial \widehat{\boldsymbol{F}}}{\partial \hat{\boldsymbol{\epsilon}}} \end{bmatrix} = \begin{bmatrix} \mathrm{diag}(((x_{ij})_{\epsilon_{ij}})_{ij} & 0 \\ 0 & \mathrm{diag}((\hat{x}_{i'j'})_{\hat{\epsilon}_{i'j'}})_{i'j'} \end{bmatrix}$$

can be directly computed from the link functions. Subsequently we have

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\epsilon}} = (\boldsymbol{x})_{\boldsymbol{\epsilon}} + (\boldsymbol{x})_{\boldsymbol{t}} \frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\epsilon}}, \tag{22}$$

or in an expanded version

$$\left(\frac{\partial x_{ij}}{\partial \epsilon_{ij}}\right)_{ij,gh} = \left((x_{ij})_{\epsilon_{ij}} \delta_{ij,gh} + (x_{ij})_{t_{ij}} \frac{\partial t_{ij}}{\partial \epsilon_{gh}}\right)_{ij,gh};$$

and

$$\frac{\partial \hat{\boldsymbol{x}}}{\partial \boldsymbol{\epsilon}} = (\hat{\boldsymbol{x}})_{\hat{\boldsymbol{t}}} \frac{\partial \hat{\boldsymbol{t}}}{\partial \boldsymbol{\epsilon}}, \tag{23}$$

or in an expanded version

$$\left(\frac{\partial \hat{x}_{i'j'}}{\partial \epsilon_{gh}}\right)_{i'j',gh} = \left((\hat{x}_{i'j'})_{\hat{t}_{i'j'}} \frac{\partial \hat{t}_{i'j'}}{\partial \epsilon_{gh}}\right)_{i'j',gh}.$$

The terms $(x_{ij})_{t_{ij}}$, $(x_{ij})_{\epsilon_{ij}}$, $(\hat{x}_{i'j'})_{\hat{t}_{i'j'}}$, $(\hat{x}_{i'j'})_{\hat{\epsilon}_{i'j'}}$ are directly computable from the explicit link cost functions. Therefore all these derivatives can be easily obtained, once $\nabla_{\boldsymbol{t}}\boldsymbol{F}$, $\nabla_{\hat{\boldsymbol{t}}}\boldsymbol{F}$, $\nabla_{\boldsymbol{t}}\widehat{\boldsymbol{F}}$, and $\nabla_{\hat{\boldsymbol{t}}}\widehat{\boldsymbol{F}}$ have been computed, by the algorithm developed in Ying and Miyagi (2001). The derivatives of q_{rs} with respect to ϵ_{ij} can be computed as follows:

$$\frac{\partial q_{rs}}{\partial \epsilon_{ij}} = \frac{\partial D_{rs}}{\partial S_{rs}} \sum_{gh} \frac{\partial S_{rs}}{\partial t_{gh}} \frac{\partial t_{gh}}{\partial \epsilon_{ij}} + \frac{\partial D_{rs}}{\partial \widehat{S}_{rs}} \sum_{g'h'} \frac{\partial \widehat{S}_{rs}}{\partial \hat{t}_{g'h'}} \frac{\partial \hat{t}_{g'h'}}{\partial \epsilon_{ij}}. \tag{24}$$

As

$$\frac{\partial S_{rs}}{\partial t_{gh}} = \frac{\sum_k \exp(-\theta c_k^{rs})\delta_{gh,k}^{rs}}{\sum_k \exp(-\theta c_k^{rs})}$$

and

$$\frac{\partial \widehat{S}_{rs}}{\partial \hat{t}_{g'h'}} = \frac{\sum_k \exp(-\hat{\theta}\hat{c}_k^{rs})\delta_{g'h',k}^{rs}}{\sum_k \exp(-\hat{\theta}\hat{c}_k^{rs})}$$

can be efficiently computed, and in the logit model

$$\frac{\partial D_{rs}}{\partial S_{rs}} = \bar{q}_{rs}\frac{\alpha\exp(\alpha(S_{rs}-\widehat{S}_{rs}))}{[1+\exp(\alpha(S_{rs}-\widehat{S}_{rs}))]^2},$$

so $\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{\epsilon}} = \left(\frac{\partial q_{rs}}{\partial \epsilon_{ij}}\right)_{rs,ij}$ and $\frac{\partial \hat{\boldsymbol{q}}}{\partial \boldsymbol{\epsilon}} = -\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{\epsilon}}$ can be efficiently computed. The computational formulae for $\frac{\partial \boldsymbol{x}}{\partial \hat{\boldsymbol{\epsilon}}}, \frac{\partial \hat{\boldsymbol{x}}}{\partial \hat{\boldsymbol{\epsilon}}}, \frac{\partial \boldsymbol{q}}{\partial \hat{\boldsymbol{\epsilon}}}, \frac{\partial \hat{\boldsymbol{q}}}{\partial \hat{\boldsymbol{\epsilon}}}$ are similar and are omitted here.

3. System with complicated transit performance

In last section it was assumed that for both the road and transit networks the costs are increasing with flows. Such an assumption is not realistic particularly for some kind of transit networks. In regions where the transit passenger flow is below the potential transport ability, an increase in transit passenger population may some times decrease the individual cost for each passenger, as discussed in the Introduction section of this paper. In other words, transit line cost function $\hat{t}_{ij}$ may be a decreasing function in the flow $\hat{x}_{ij}$. For such a case exhibiting economies of scale, the sensitivity analysis method presented in last section may not apply because the positive definiteness of the matrix $\begin{bmatrix} \nabla_{\boldsymbol{t}}\boldsymbol{F} & \nabla_{\hat{\boldsymbol{t}}}\boldsymbol{F} \\ \nabla_{\boldsymbol{t}}\widehat{\boldsymbol{F}} & \nabla_{\hat{\boldsymbol{t}}}\widehat{\boldsymbol{F}} \end{bmatrix}$ may not be satisfied. (Note that the positive definiteness for the congested case was established on the congestion condition that the transit line cost function $\hat{t}_{ij}$ is a strongly monotonically increasing function in the flow $\hat{x}_{ij}$.)

In this section we investigate the case where the transit consists of independent lines connecting O–D pairs, each line may have economies of scale as well as congestion diseconomies. Instead of general networks, simple transit lines with complex cost structure are considered because they result in a combined transportation system sophisticated enough to reveal the complicated phenomena when there exhibit economies of scale. Also for practical applications, though the assumption of independence of transit lines is not realistic in metropolitan areas, where, however, congestion diseconomies are usually observed, the algorithms presented in Section 2 can be applied. Therefore the work to be presented in this section is significant especially for suburban and rural areas with simple transit line systems.

3.1. Equilibrium formulation

Let p_{rs} denote some relevant parameter characterizing the performance of the transit line from origin r to destination s. Typical p_{rs} can be a tax or subsidy for the use of transit. The disutility of using transit $\widehat{S}_{rs}$ is a function in p_{rs} and the transit demand $\hat{q}_{rs} = \bar{q}_{rs} - q_{rs}$:

$$\widehat{S}_{rs} = \widehat{S}_{rs}(p_{rs}, \bar{q}_{rs} - q_{rs}).$$

At equilibrium, the demand on the road network satisfies the following equation:

$$q_{rs} = D_{rs}(S_{rs}, \widehat{S}_{rs}(p_{rs}, \bar{q}_{rs} - q_{rs})).$$

In a logit model we have

$$q_{rs} = \bar{q}_{rs} \frac{1}{1 + \exp(\alpha(S_{rs} - \widehat{S}_{rs}(p_{rs}, \bar{q}_{rs} - q_{rs})))}. \tag{25}$$

Solving this equation for q_{rs}, we can obtain the automobile travel demand in the form

$$q_{rs} = q_{rs}(S_{rs}, p_{rs}).$$

For simplicity of expression, in the rest of this paper we use the notation

$$q_{rs} = D_{rs}(S_{rs}, p_{rs}),$$

where D_{rs} differs from the one used previously in that it is here an explicit function (multivalued in general) in S_{rs} and p_{rs}, while it was previously an explicit function in S_{rs} and $\widehat{S}_{rs}$. The equilibrium conditions are characterized by the following equations:

$$F_{ij}(\boldsymbol{t}, \epsilon, \boldsymbol{p}) = x_{ij}(t_{ij}, \epsilon_{ij}) - \sum_{rs} q_{rs} \frac{\partial S_{rs}}{\partial t_{ij}} = x_{ij}(t_{ij}, \epsilon_{ij}) - \sum_{rs} q_{rs} \frac{\sum_k \exp(-\theta c_k^{rs}) \delta_{ij,k}^{rs}}{\sum_p \exp(-\theta c_p^{rs})} = 0, \quad ij \in A, \tag{26}$$

where $q_{rs} = D_{rs}(S_{rs}, p_{rs})$; or in a vector form

$$\boldsymbol{F}(\boldsymbol{t}, \epsilon, \boldsymbol{p}) = \boldsymbol{x} - \sum_{rs} q_{rs} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})))^{\mathrm{T}} = \boldsymbol{0}. \tag{27}$$

Remark. In the transport system treated in Section 2, there are two partial equilibrium equations, one for the road network and the other for the transit, and an equilibrium for modal choice, which have to be simultaneously satisfied. For the system with direct O–D specific transit lines, by using the modified demand function which is obtained from the equilibrium for transport mode choice, Eq. (27) fully characterizes the equilibrium on the road network, and subsequently determines the flow on the transit lines.

3.2. Properties of $\nabla_{\boldsymbol{t}} \boldsymbol{F}$

From (27) is derived the following equation:

$$\nabla_{\boldsymbol{t}} \boldsymbol{F} = \mathrm{diag}((x_{ij})_{t_{ij}})_{ij} - \sum_{rs} q_{rs} \nabla_{\boldsymbol{t}}^2 S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})) - \sum_{rs} \frac{\partial q_{rs}}{\partial S_{rs}} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t})))^{\mathrm{T}} (\nabla_{\boldsymbol{t}} S_{rs}(\boldsymbol{c}^{rs}(\boldsymbol{t}))). \tag{28}$$

By the same argument as in Section 2.4.2, it can be shown that the first term is positive definite, the second is positive semidefinite. If

$$\frac{\partial D_{rs}}{\partial S_{rs}} < 0,$$

then $\nabla_t \boldsymbol{F}$ is a positive definite matrix. If the transit lines are congested, $\frac{\partial \widehat{S}_{rs}}{\partial \hat{q}_{rs}} > 0$, it can be shown $\frac{\partial q_{rs}}{\partial S_{rs}} < 0$. Thus it is seen that congestion on the transit lines is a sufficient condition for $\nabla_t \boldsymbol{F}$ to be positive definite.

However, $\frac{\partial q_{rs}}{\partial S_{rs}} < 0$ may fail to hold when the economies of scale prevail. To put this in a precise way let us consider the demand function defined by (25). From (25) we have

$$\frac{\partial q_{rs}}{\partial S_{rs}} = -\left[1 - \bar{q}_{rs}\frac{\alpha\exp(\alpha(S_{rs}-\widehat{S}_{rs}))}{[1+\exp(\alpha(S_{rs}-\widehat{S}_{rs}))]^2}\frac{\partial \widehat{S}_{rs}}{\partial q_{rs}}\right]^{-1}\cdot \bar{q}_{rs}\frac{\alpha\exp(\alpha(S_{rs}-\widehat{S}_{rs}))}{[1+\exp(\alpha(S_{rs}-\widehat{S}_{rs}))]^2}.$$

$\frac{\partial q_{rs}}{\partial S_{rs}} < 0$ holds if and only if

$$1 - \bar{q}_{rs}\frac{\alpha\exp(\alpha(S_{rs}-\widehat{S}_{rs}))}{[1+\exp(\alpha(S_{rs}-\widehat{S}_{rs}))]^2}\frac{\partial \widehat{S}_{rs}}{\partial q_{rs}} > 0. \tag{29}$$

In the case that economies of scale prevail in the transit line, we have

$$\frac{\partial \widehat{S}_{rs}}{\partial \hat{q}_{rs}} < 0 \quad \text{or} \quad \frac{\partial \widehat{S}_{rs}}{\partial q_{rs}} > 0.$$

Inequality (29) holds only when the quantity $\frac{\partial \widehat{S}_{rs}}{\partial q_{rs}}$ is not too large.

In general, there seems to be no meaningful economic or behavioral explanation of (29). However, for some special class of transit performance functions, implication of (29) can be further explored.

Suppose that F is the fixed cost of the transit line, $a_1\hat{q}_{rs}^2 + c_1\hat{q}_{rs}$ the variable operating cost, both to be shared by the transit users. Then the fare charged to each passenger so that the transit firm breaks even is

$$F/\hat{q}_{rs} + a_1\hat{q}_{rs} + c_1.$$

Besides the fare, suppose that the cost in monetary term accounting for other factors including travel time incurred by a passenger is given by

$$a_2\hat{q}_{rs} + c_2.$$

Let $a = a_1 + a_2$, $c = c_1 + c_2$, the aggregate cost incurred by a passenger is

$$\widehat{C}_{rs} = F/\hat{q}_{rs} + a_1\hat{q}_{rs} + c_1 + a_2\hat{q}_{rs} + c_2 = F/\hat{q}_{rs} + a\hat{q}_{rs} + c.$$

F is a large positive number denoting the fixed cost, thus the first term reflects the scale economies of the transit system; c is a positive constant cost incurred by individual user; a is also assumed to be a positive coefficient reflecting a congestion effect when q_{rs} grows large.

As it is assumed that the transit has only one line connecting an O–D pair rs, the disutility perceived by the transit user is equal to the cost

$$\widehat{S}_{rs} = \widehat{C}_{rs} = F/\hat{q}_{rs} + a\hat{q}_{rs} + c,$$

inequality (29) is reduced to

$$\frac{\hat{q}_{rs}}{q_{rs}} + a\alpha\frac{\hat{q}_{rs}^2}{\bar{q}_{rs}} > \frac{\alpha F}{\bar{q}_{rs}}.$$

This inequality is satisfied if

$$\frac{\hat{q}_{rs}}{q_{rs}} > \frac{\alpha F}{\bar{q}_{rs}},$$

that is, if the ratio of the transit users to the car drivers is greater than α times the ratio of the fixed transit cost to the number of total travelers.

When $\boldsymbol{\epsilon}$ and $\boldsymbol{p}$ are given, if $\nabla_{\boldsymbol{t}}\boldsymbol{F}$ is always positive definite, then the solution $\boldsymbol{t}$ for the equation $\boldsymbol{F}(\boldsymbol{t}, \boldsymbol{\epsilon}, \boldsymbol{p}) = \boldsymbol{0}$ is unique, i.e., there is a unique equilibrium.

If inequality (29) is satisfied, then $\nabla_{\boldsymbol{t}}\boldsymbol{F}$ is positive definite. And when $\boldsymbol{\epsilon}$ and $\boldsymbol{p}$ are given, the solution $\boldsymbol{t}$ for the equations

$$F_{ij}(\boldsymbol{t}, \boldsymbol{\epsilon}, \boldsymbol{p}) = 0, \quad ij \in A$$

is unique, so are the link flows and path flows.

In the case that $\nabla_{\boldsymbol{t}}\boldsymbol{F}$ is regular but is not positive definite, there may be multiple solutions for the above equation. In this case, suppose that a solution is specified that describes an equilibrium state, local unique solution still exists around that state, sensitivity analysis can be undertaken by the method developed in Section 3.3.

There are two cases for which $\nabla_{\boldsymbol{t}}\boldsymbol{F}$ is not regular. The first is the case

$$1 - \bar{q}_{rs}\frac{\alpha\exp\left(\alpha\left(S_{rs} - \widehat{S}_{rs}\right)\right)}{\left[1 + \exp\left(\alpha\left(S_{rs} - \widehat{S}_{rs}\right)\right)\right]^2}\frac{\partial\widehat{S}_{rs}}{\partial q_{rs}} = 0.$$

In this case,

$$\frac{\partial q_{rs}(S_{rs}, p_{rs})}{\partial S_{rs}} = \infty,$$

$\nabla_{\boldsymbol{t}}\boldsymbol{F}$ is not well-defined. $\frac{\partial q_{rs}}{\partial S_{rs}} = \infty$ occurs at the point in the S_{rs}–q_{rs} plane where the tangent vector of the curve of the multivalued function $q_{rs}(S_{rs}, p_{rs})$ (p_{rs} being fixed) turns downward. See Fig. 1 for an illustration.

The second case is that $\nabla_{\boldsymbol{t}}\boldsymbol{F}$ is well defined but is singular. For both cases, before sensitivity analysis could be discussed, the properties of the equilibrium equations have to be investigated in depth, which will be left for future research.

3.3. Sensitivity analysis method

In the following, with the assumption that the inverse matrix $(\nabla_{\boldsymbol{t}}\boldsymbol{F})^{-1}$ exists, we establish the formulae for computing the derivatives of $\boldsymbol{t}$, $\boldsymbol{x}$ and $\boldsymbol{q}$ with respect to $\boldsymbol{\epsilon}$ and $\boldsymbol{p}$.

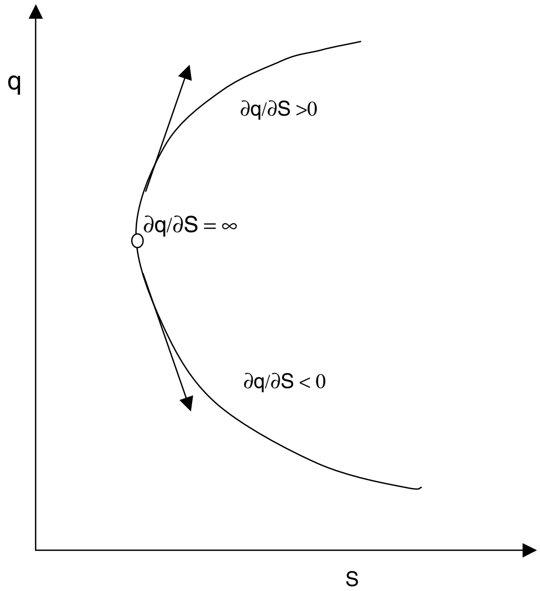

Fig. 1. When q is large, the transit has a small passenger volume $\hat{q}_{rs}$ and has large scale economies, $\partial D_{rs}/\partial S_{rs} > 0$; when $\hat{q}_{rs}$ grows large and the transit becomes congested, $\partial D_{rs}/\partial S_{rs}$ turns downward ($\partial D_{rs}/\partial S_{rs} < 0$).

3.3.1. Computing $\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\epsilon}}, \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\epsilon}}$ and $\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{\epsilon}}$

In the equilibrium equation

$$\boldsymbol{F}(\boldsymbol{t}, \epsilon, \boldsymbol{p}) = \boldsymbol{0}, \tag{30}$$

looking $\boldsymbol{x}$ as intermediate (apparent) variables in $\boldsymbol{F}$, $\boldsymbol{p}$ as constants, $\boldsymbol{\epsilon}$as free variables and $\boldsymbol{t}$ as functions of $\boldsymbol{\epsilon}$, we have

$$(\boldsymbol{F})_{\boldsymbol{x}}(\boldsymbol{x})_{\epsilon} + \nabla_{\boldsymbol{t}}\boldsymbol{F}\left(\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\epsilon}}\right) = \boldsymbol{0}.$$

Since the terms $(x_{ij})_{t_{ij}}$, $(x_{ij})_{\epsilon_{ij}}$ are directly computable from the explicit cost functions $t_{ij}(x_{ij}, \epsilon_{ij})$, once $(\nabla_{\boldsymbol{t}}\boldsymbol{F})^{-1}$ has been computed, all the derivatives can be immediately computed as follows:

$$\left(\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{\epsilon}}\right) = -(\nabla_{\boldsymbol{t}}\boldsymbol{F})^{-1}(\boldsymbol{F})_{\boldsymbol{x}}(\boldsymbol{x})_{\epsilon} = -(\nabla_{\boldsymbol{t}}\boldsymbol{F})^{-1}\mathrm{diag}((x_{ij})_{\epsilon_{ij}})_{ij}, \tag{31}$$

$$\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\epsilon}}\right) = \left((x_{ij})_{\epsilon_{ij}}\delta_{ij,gh} + (x_{ij})_{t_{ij}}\frac{\partial t_{ij}}{\partial \epsilon_{gh}}\right)_{ij,gh}, \tag{32}$$

and

$$\left(\frac{\partial \boldsymbol{q}_{rs}}{\partial \boldsymbol{\epsilon}_{ij}}\right) = \left(\frac{\partial q_{rs}}{\partial S_{rs}} sum_{gh}\frac{\partial S_{rs}}{\partial t_{gh}}\frac{\partial t_{gh}}{\partial \epsilon_{ij}}\right)_{rs,ij}. \tag{33}$$

3.3.2. Computing $\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{p}}$, $\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}}$ and $\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}$

Looking $\boldsymbol{q}$ as intermediate variables in $\boldsymbol{F}$, ϵ as constants, $\boldsymbol{p}$ as free variables and $\boldsymbol{t}$ as functions of $\boldsymbol{p}$, the following formulae are derived:

$$(\boldsymbol{F})_{\boldsymbol{q}}(\boldsymbol{q})_{\boldsymbol{p}} + \nabla_{\boldsymbol{t}}\boldsymbol{F}\left(\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{p}}\right) = \boldsymbol{0},$$

$$\left(\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{p}}\right) = -(\nabla_{\boldsymbol{t}}\boldsymbol{F})^{-1}(\boldsymbol{F})_{\boldsymbol{q}}(\boldsymbol{q})_{\boldsymbol{p}}, \tag{34}$$

where

$$(\boldsymbol{F})_{\boldsymbol{q}} = ((F_{ij})_{q_{rs}})_{ij,rs} = \left(-\frac{\sum_k \exp(-\theta c_k^{rs})\delta_{ij,k}^{rs}}{\sum_p \exp(-\theta c_p^{rs})}\right)_{ij,rs}$$

and

$$(\boldsymbol{q})_{\boldsymbol{p}} = \operatorname{diag}\left(\frac{\partial D_{rs}}{\partial p_{rs}}\right)_{rs}.$$

Once $\frac{\partial \boldsymbol{t}}{\partial \boldsymbol{p}}$ is known, we can immediately compute

$$\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{p}}\right) = \left((x_{ij})_{t_{ij}}\frac{\partial t_{ij}}{\partial p_{rs}}\right)_{ij,rs}, \tag{35}$$

and

$$\left(\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{p}}\right) = \left(\frac{\partial D_{rs}}{\partial p_{uv}} + \frac{\partial D_{rs}}{\partial S_{rs}}\sum_{gh}\frac{\partial S_{rs}}{\partial t_{gh}}\frac{\partial t_{gh}}{\partial p_{uv}}\right)_{rs,uv}. \tag{36}$$

The derivatives can be computed in a link based fashion (Ying and Miyagi, 2001).

Note. It is implicitly assumed that if $uv \neq rs$, D_{rs} does not depend on parameter p_{uv}, thus $\partial D_{rs}/\partial p_{uv} = 0$. However, this assumption is not necessary for our computational formulas to hold.

Example. In the logit model for a transportation system consisting of a road network and transit network as described earlier, we have

$$\frac{\partial D_{rs}}{\partial p_{rs}} = \left[1 - \bar{q}_{rs}\frac{\alpha\exp(\alpha(S_{rs} - \widehat{S}_{rs}))}{[1 + \exp(\alpha(S_{rs} - \widehat{S}_{rs}))]^2}\frac{\partial \widehat{S}_{rs}}{\partial q_{rs}}\right]^{-1} \cdot \bar{q}_{rs}\frac{\alpha\exp(\alpha(S_{rs} - \widehat{S}_{rs}))}{[1 + \exp(\alpha(S_{rs} - \widehat{S}_{rs}))]^2}\frac{\partial \widehat{S}_{rs}}{\partial p_{rs}}.$$

Remark. In reality, if not being used, automobile road may also lose its attractiveness and finally rust out. This phenomenon reveals the potential existence of economies of scale on road network. For such a general case, the property of the equilibrium and the applicability of the sensitivity analysis algorithm developed above depend solely on the property of the matrix $\nabla_{\boldsymbol{t}}\boldsymbol{F}$. Therefore the above analysis also applies to this general case.

For a bi-modal network, instead of $\nabla_t \boldsymbol{F}$, what plays an essential role in equilibrium analysis and sensitivity analysis is the matrix $\begin{bmatrix} \nabla_t \boldsymbol{F} & \nabla_{\hat{t}} \boldsymbol{F} \\ \nabla_t \widehat{\boldsymbol{F}} & \nabla_{\hat{t}} \widehat{\boldsymbol{F}} \end{bmatrix}$, as has been seen for congested networks described in Section 2. For the case of a general bi-modal network with congestion as well as scale economies, this matrix is no longer positive definite in general. However, equilibrium analysis and sensitivity analysis can still be conducted within the framework provided in this section.

4. Application to optimal pricing

The aim of this section is to use a numerical example to demonstrate the effectiveness of the sensitivity analysis method developed in Section 3, as it is theoretically less complete compared to the method developed in Section 2 for a combined transport with all modes congested. The optimal pricing problem on a transport system consisting of a road network and a transit line is examined based on the sensitivity analysis method. The individual user cost of the transit line is assumed to have the same structure as described in Section 3.2. Scale economies may prevail when the flow of passengers is small but may be congested when the flow grows large.

4.1. Social utility and its optimization

For simplicity of expression, it is assumed that there is only one O–D pair, so that the suffix "*rs*" can be omitted in the O–D-specific notations. For example, travel demand $\bar{q}_{rs}$ will be abbreviated as $\bar{q}$. As stated earlier, the expected minimum cost or disutility S for using the road network is

$$S(\boldsymbol{c}(\boldsymbol{t})) = -\frac{1}{\theta} \ln \sum_k \exp(-\theta c_k).$$

The negative of this measure is the traveler's utility for using the road network with probabilistic route choice behavior. $1/\theta$ can be explained as an indicator revealing the degree of the passenger's preference for variety of choice. In the absence of income effects, as will be assumed throughout this paper, $-S$ is an economically sound utility measure in the sense that its variation due to price or cost change is equivalent to that of a generalized Marshallian surplus (see Williams, 1977). The expected minimum cost $\widehat{S}$ for using the transit network

$$\widehat{S}(\hat{\boldsymbol{c}}(\hat{\boldsymbol{t}})) = -\frac{1}{\hat{\theta}} \ln \sum_k \exp(-\hat{\theta} \hat{c}_k)$$

has a similar economic meaning. If there is only one route with cost $\widehat{C}$, then $\widehat{S} = \widehat{C}$.

The disutility of using in a random manner the combined transportation system consisting of two modes is given by

$$\overline{S} = -\frac{1}{\alpha} \ln(\exp(-\alpha S) + \exp(-\alpha \widehat{S})).$$

$1/\alpha$ reflects the degree of traveler's preference for variety of modal choice. $-\overline{S}$ is an economically sound utility measure in that its variation is equivalent to that of a generalized Marshallian surplus (see Williams, 1977).

4.1.1. Pricing scheme

It is assumed that there is a governmental transportation department that collects road tolls from drivers and transit taxes from transit users, transit tax may have negative value that means the transit users are actually subsidized. The transit is assumed to be managed by a zero-profit firm.

Let T_{ij} denote the toll imposed on link ij in the road network. The road link cost is

$$c_{ij} = t_{ij} + T_{ij}.$$

The disutility of using the road network is

$$S = -\frac{1}{\theta} \ln \left(\sum_k \exp \left(-\theta \sum_{ij} c_{ij} \delta_{k,ij} \right) \right).$$

For the transit system, as in Section 3.2, it is assumed that the average cost charged to each passenger so that the transit firm breaks even is

$$\widehat{C}_1(\hat{q}).$$

In addition, the cost in monetary term accounting for other factors including travel time incurred by a passenger is given by

$$\widehat{C}_2(\hat{q}).$$

The aggregate cost incurred by an individual passenger is

$$\widehat{C} = \widehat{C}_1 + \widehat{C}_2.$$

Let 'Tax' denote the tax collected from the transit user. Tax < 0 means the transit users are actually subsidized. Thus the transit user incurs the following disutility:

$$\widehat{S} = \widehat{C} + \mathrm{Tax} = \widehat{C}_1 + \mathrm{Tax} + \widehat{C}_2.$$

In such a scheme, the fare paid by each passenger is $\widehat{C}_1 + \mathrm{Tax}$, of which only the average transit cost $\widehat{C}_1$ goes to the transit firm, 'Tax' goes to the government. If Tax < 0, then a user is subsidized with an amount $-$Tax. A transit user in general does not know the amount of the tax (or subsidy), but just measures its disutility by the fare and $\widehat{C}_2$. In any case, it is assumed that the transit firm takes properly the amount $\widehat{C}_1$ to break even. Under this assumption, a pricing scheme consists of the following components:

- Transit tax: Tax;
- Road tolls: T_{ij}.

What is directly relevant in determining the road tolls and transit tax are the road cost functions and the individual transit user cost function $\widehat{C}$. In the following no particular concern will be paid for the average transit cost $\widehat{C}_1$ and the fare $\widehat{C}_1 + \mathrm{Tax}$, which are only parts of disutility perceived by the transit user and are assumed to be set properly by the transit firm. In the following a pricing scheme with the road tolls and transit tax set to be zero is referred to as a "0-pricing" policy.

4.1.2. Social utility

Suppose that the road tolls and transit taxes collected are returned to the traveling community by certain means, a social utility is obtained as

$$\mathrm{SU} = -\bar{q} \cdot \overline{S} + \sum_{ij} x_{ij} \cdot T_{ij} + \hat{q} \cdot \mathrm{Tax}. \tag{37}$$

It is easily shown that as $\theta \to \infty$ and $\alpha \to \infty$,

$$\bar{q} \cdot \overline{S} \to \sum_{ij} x_{ij} \cdot (t_{ij} + T_{ij}) + \hat{q} \cdot \left(\widehat{C} + \mathrm{Tax}\right) = \left(\sum_{ij} x_{ij} \cdot t_{ij} + \hat{q} \cdot \widehat{C}\right) + \left(\sum_{ij} x_{ij} \cdot T_{ij} + \hat{q} \cdot \mathrm{Tax}\right),$$

hence

$$-\mathrm{SU} \to \left(\sum_{ij} x_{ij} \cdot t_{ij} + \hat{q} \cdot \widehat{C}\right) = \text{total travel cost.}$$

This means that when the travelers choose exactly the least cost mode and route, the social utility is equal to the negative of the total transportation cost.

4.1.3. Sensitivity analysis based optimization

For consistence of notations with those used previously, let

$$\epsilon_{ij} = T_{ij}, \quad p = \mathrm{Tax}.$$

If no constraints are imposed on the range of tolls and tax, then possible optimal solutions that maximize SU can be obtained as the points where the gradient

$$\nabla_{\boldsymbol{T}} \mathrm{SU} = \left(\frac{\partial \mathrm{SU}}{\partial p}, \ldots, \frac{\partial \mathrm{SU}}{\partial \epsilon_{ij}}, \ldots\right) = \mathbf{0},$$

where $\boldsymbol{T} = (p, \ldots, \epsilon_{ij}, \ldots)$ denotes the vector of transit tax and automobile road link tolls. We have

$$\frac{\partial \mathrm{SU}}{\partial p} = -q\frac{\partial S}{\partial p} - \hat{q}\frac{\partial \widehat{S}}{\partial p} + \sum_{ij} \frac{\partial x_{ij}}{\partial p}\epsilon_{ij} + \hat{q} + \frac{\partial \hat{q}}{\partial p}p,$$

where

$$\frac{\partial S}{\partial p} = \sum_{ij} \frac{\partial S}{\partial t_{ij}} \frac{\partial t_{ij}}{\partial p},$$

$$\frac{\partial \widehat{S}}{\partial p} = 1 + \frac{\partial \widehat{C}}{\partial q} \frac{\partial q}{\partial p},$$

and

$$\frac{\partial \hat{q}}{\partial p} = -\frac{\partial q}{\partial p}.$$

The terms $\frac{\partial S}{\partial t_{ij}}, \frac{\partial t_{ij}}{\partial p}, \frac{\partial q}{\partial p}, \frac{\partial x_{ij}}{\partial p}$ can be computed by the sensitivity analysis method provided in Section 3, and $\frac{\partial \widehat{C}}{\partial q}$ can be directly computed from the transit cost function, thus $\frac{\partial \mathrm{SU}}{\partial p}$ can be efficiently computed. $\frac{\partial \mathrm{SU}}{\partial \epsilon_{ij}}$, $ij \in A$, can be computed in a similar manner. By using the gradient, conventional optimization methods, e.g., the steepest decent method (see, e.g., Luenberger, 1984), can then be applied to find a pricing scheme that optimizes the social utility.

4.1.4. Marginal cost pricing

In this paper we are interested in the welfare of the traveler's community, and the marginal cost is defined as the additional cost imposed to all the users by the new entry of a unit of passenger volume

$$\mathrm{MC} = \frac{\mathrm{d}(\hat{q}\widehat{C})}{\mathrm{d}\hat{q}} = \frac{\mathrm{d}(\hat{q}(\widehat{C}_1 + \widehat{C}_2))}{\mathrm{d}\hat{q}}.$$

A marginal cost pricing (MCP) scheme is to set the link performance functions as the marginal cost functions in each mode. Or equivalently, the road tolls and transit tax are set as follows:

$$T_{ij} = \frac{\mathrm{d}(x_{ij}t_{ij})}{\mathrm{d}x_{ij}} - t_{ij} = x_{ij}\frac{\mathrm{d}(t_{ij})}{\mathrm{d}x_{ij}},$$

$$\mathrm{Tax} = \mathrm{MC} - \widehat{C} = \frac{\mathrm{d}(\hat{q}\widehat{C})}{\mathrm{d}\hat{q}} - \widehat{C} = \hat{q}\frac{\mathrm{d}(\widehat{C})}{\mathrm{d}\hat{q}}.$$

Note. Various cost structures for transit lines and their marginal cost pricing implications are discussed in the literature. See, e.g., Chapters 7 and 12 of Mohring (1976), Section 3.2 of Small (1992), and Kraus (1991). However, in these researches modal shift between transit and automobile is not considered.

It is known that for congested road networks (Yang, 1999) and combined networks with all modes congested (Bellei et al., 2002), a marginal cost pricing scheme is optimal in the sense that the social utility SU is maximized. However, the optimality of MCP in a transport system consisting of both congested mode and mode where economies of scale prevail has not been investigated yet. While a sound theoretic study of optimal pricing in such a system is out of the scope of this paper, some numerical results will be given which suggest that MCP may still yield an optimal solution. In the following example we will use sensitivity analysis based numerical method to examine some properties of MCP for the case when the transit may have economies of scale. In the example it will be shown that the gradient of the social utility with respect to the variation of the road tolls and transit tax are 0 when a MCP scheme is imposed. And, starting from 0 tolls

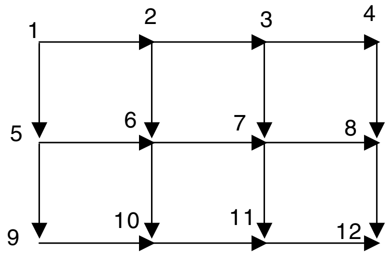

Fig. 2. Mobile network. The numbers denote nodes.

and 0 tax, by applying our sensitivity algorithm and the steepest gradient method, an optimal solution (pricing pattern) could be obtained which is different from the MCP scheme but yields the same transit and network flows and the same optimal social utility.

4.2. Numerical example

The road network, as shown in Fig. 2, consists of 17 links, each link has a BPR performance function

$$t(x) = C_0\left(1 + b\left(\frac{x}{\mathrm{Cap}}\right)^4\right),$$

with parameters described in Table 1.

Table 1
Mobile network link coefficients

Link no. [head->tail]	C_0	Cap	b
1 [1->2]	20	1000	0.15
2 [2->3]	23	500	0.15
3 [3->4]	17	500	0.15
4 [1->5]	18	1500	0.15
5 [2->6]	19	500	0.15
6 [3->7]	16	500	0.15
7 [4->8]	22	500	0.15
8 [5->6]	14	1000	0.15
9 [6->7]	17	1000	0.15
10 [7->8]	13	1000	0.15
11 [5->9]	24	800	0.15
12 [6->10]	20	500	0.15
13 [7->11]	26	500	0.15
14 [8->12]	19	1000	0.15
15 [9->10]	7	800	0.15
16 [10->11]	18	800	0.15
17 [11->12]	17	800	0.15

C_0: 0-flow travel cost; Cap: link capacity; b: parameter for the congestion term; 0-pricing network flows, costs.

We assume that the only travel demand is from node 1 to 12,

$$q + \hat{q} = \bar{q} = \bar{q}_{1\to 12} = 3750.$$

In the example there is a transit line for the single O–D. The individual transit user cost is assumed to be

$$\widehat{C} = \frac{F}{\hat{q}} + a\hat{q} + c = \frac{24151}{\hat{q}} + 0.01\hat{q} + 100.$$

In this example we assume

$$\theta = 0.5, \quad \alpha = 0.1.$$

Given link performance functions be specified either by a pricing pattern or by the marginal cost functions, the equilibria of the combined transport system are computed by a modified MSA (method of successive average; see Sheffi, 1985). For the case that no tolls are collected from drivers and that the transit is neither taxed nor subsidized, here two equilibrium solutions are found. One is a solution where the number of transit users is zero and thus the individual transit user cost is infinity. This solution is trivial in the sense that if $\hat{q}$ is set to be 0, then $\widehat{S} = \infty$, the infinitely large average transit cost prevents traveler to choose the transit mode, thus $\hat{q} = 0$ with a unique assignment of travelers on road network is a solution for the bi-modal network equilibrium problem. (Such a trivial solution is sometimes the only solution in reality: in the prevailing trend of public transit deregulation which actually forces the firm to charge the average cost to the passengers, some local rail transit lines with small volume of passengers have to be discarded.)

The other solution is one with a non-zero number of transit users. For these solutions the social utilities are

$$\mathrm{SU} = -682050 \quad \text{and} \quad \mathrm{SU} = -463527,$$

respectively. For the second non-trivial solution, we have verified that the gradient $\nabla_{\boldsymbol{T}}\mathrm{SU} \neq \mathbf{0}$ (see Table 4).

If the marginal cost pricing is imposed, that is,

$$T_{ij} = x_{ij}\frac{\mathrm{d}t_{ij}}{\mathrm{d}t_{ij}},$$

$$\mathrm{Tax} = \mathrm{MC} - \widehat{C} = 0.01\hat{q} - \frac{24151}{\hat{q}},$$

the traffic flows, tolls and tax are as shown in Table 2, and

$$\mathrm{SU} = -386630.$$

Note that in this case Tax = 2.34, which implies that the transit line also becomes congested by a large increase of passengers, and a congestion tax is imposed. Under the MCP scheme, the gradient $\nabla_{\boldsymbol{T}}\mathrm{SU}$ is verified numerically to be $\mathbf{0}$.

By using the steepest decent method, with initial tolls and tax set to be 0, a toll and tax pattern that maximizes SU is obtained as shown in Table 2. This pricing pattern is different from the one obtained by the marginal cost pricing scheme, but yields exactly the same maximum SU and the

Table 2
Network flows and tolls under MCP/SA-optimized pricing schemes

Link no.	Link flow	Link cost	MCP toll	SA-optimized toll
1	738.2	20.9	3.6	4.8
2	404.9	24.5	5.9	2.3
3	155.1	17	0.1	0.3
4	1336.5	19.7	6.8	7.4
5	333.3	19.6	2.3	2.4
6	249.8	16.1	0.6	2.1
7	155.1	22	0.1	0.3
8	787.5	14.8	3.2	4.1
9	896.2	18.6	6.6	4.3
10	977.1	14.8	7.1	6
11	549	24.8	3.2	3.3
12	224.6	20.1	0.5	2.2
13	168.9	26.1	0.2	0.3
14	1132.2	23.7	18.7	6.3
15	549	7.2	0.9	3.3
16	773.6	20.4	9.4	5.6
17	942.5	21.9	19.6	5.9
			Tax	Tax
			2.34	−12.17

Table 3
Comparison of various pricing schemes

	SU	Total toll	Total tax	q	q_{rs}
0-Pricing	−682050	0	0	3750	0
0-Pricing	−463527	0	0	3000	750
MCP	−386630	79889	3917	2074.7	1675.3
SA-optimal	−386630	49793	−20386	2074.7	1675.3
	S	S_{rs}	Tax	AC	MC
0-Pricing	181.88	∞	0	∞	$-\infty$
0-Pricing	125.84	139.7	0	139.7	115
MCP	131.37	133.51	2.34	131.17	133.51
SA-optimal	116.86	119	−12.17	131.17	133.51

same traffic flows. Note that in this case the transit tax is −12.17, it means that the transit users are actually subsidized. By the "minimum-revenue" criterion (see, e.g., Dial, 1971, 1999), this pricing pattern is preferable to the MCP scheme. See also Table 3, where is given a comparison of social utility, total collected toll and tax, automobile and transit users, automobile and transit user disutilities, transit tax, individual and marginal transit costs, for the equilibria under the 0-pricing (both the road tolls and transit tax are set 0) policy, the marginal cost pricing policy, and the sensitivity analysis based optimized pricing policy.

The sensitivity analysis based optimization process converges with about 10 iterations, as shown in Fig. 3. The gradients at the equilibria in the initial state, the 1st, 10th and 100th itera-

Table 4
Convergence of gradient

$\partial SU/\partial T_i$	Initial	1st iteration	10th iteration	100th iteration
$i = 1$	2348.64	−114.53	3.81	0.04
2	1394.44	−42.95	1.44	0.05
3	1033.86	−77.22	0.25	0.03
4	3148.91	103.92	−4.17	0.01
5	954.2	−71.58	2.38	−0.01
6	360.58	34.17	1.19	0.02
7	1033.86	−77.22	0.25	0.03
8	1886.36	−92.54	−0.41	−0.01
9	2297.76	−257.12	0.54	−0.02
10	1936.37	−12.69	0.34	−0.02
11	1262.55	196.46	−3.76	0.02
12	542.8	93	1.42	0
13	721.97	−210.26	1.38	0.01
14	2970.23	−89.81	0.6	0.01
15	1262.55	196.46	−3.76	0.02
16	1805.34	289.46	−2.34	0.03
17	2527.31	79.2	−0.95	0.04
$\partial SU/\partial Tax$	−5497.54	10.61	0.36	−0.05

The initial state is an equilibrium with 0 tolls and tax.

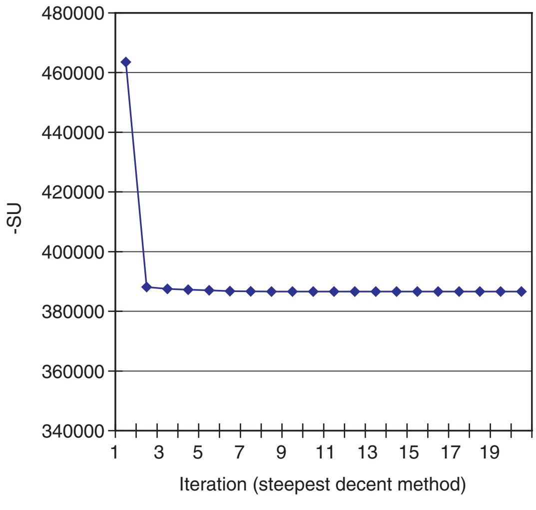

Fig. 3. Convergence of −SU.

tions are shown in Table 4. For the initial state with zero toll and tax, the transit users are few and the economies of scale prevail, we had $\frac{\partial \widehat{S}}{\partial q} = 0.033 > 0$, $\frac{\partial D}{\partial S} = 53.14 > 0$, and negative eigenvalues were observed in the Jacobian matrix $\nabla_{\boldsymbol{t}} \boldsymbol{F}$, as shown in Table 5. But from the 1st iteration, we had $\frac{\partial D}{\partial S} < 0$, and all the eigenvalues were verified to be positive. If the pricing pattern is adjusted

Table 5
Eigenvalues of the Jacobian at the initial state, 1st and 100th iteration

Initial	1st iteration	100th iteration
2142.9	2473814	1773.5
1222	1911585	1591.2
868.4	12021	1336.5
665.4	1698.9	948.2
559.4	940.4	925.4
449.1	593.3	682.8
432.1	451.6	524.9
174.5	380.5	510.5
154.9	288.7	339
130	248.3	289.6
81.4	190	267.7
64.2	178.7	200.7
55.5	167.1	178.9
53	125.9	142
−49.8	111.6	116.2
42.4	84	105.5
25.3	63.4	68.1

The initial state is an equilibrium with 0 tolls and tax.

continuously from 0 to the one yielded by the first iteration, then there is a point where $\nabla_{\boldsymbol{t}}\boldsymbol{F}$ has a 0 eigenvalue thus is singular. This point is jumped over by the first iteration. In general numerical applications, care should be taken to treat such kind of singular points.

Remark. (i) Consider the following ad hoc dynamic flow adjustment process:

$$\dot{\boldsymbol{x}} = -\gamma \boldsymbol{F}, \quad \gamma > 0. \tag{38}$$

This process is compatible with the MSA algorithm in the following sense. Let $\boldsymbol{x}_n$ denote the link flow vector at the nth MSA iteration. In the MSA algorithm,

$$\boldsymbol{x}_n = \boldsymbol{x}_{n-1} + \gamma_n \left(\left(\sum_{rs} q_{rs}(\boldsymbol{t}(\boldsymbol{x}_{n-1})) P_{ij}^{rs}(\boldsymbol{t}(\boldsymbol{x}_{n-1})) \right)_{ij} - \boldsymbol{x}_{n-1} \right),$$

where γ_n is a step size (see, e.g., Sheffi, 1985). A local continuous version of this MSA flow adjustment process is expressed exactly by the above differential equation. Since $\nabla_{\boldsymbol{x}}\boldsymbol{F} = \nabla_{\boldsymbol{t}}\boldsymbol{F} \cdot \mathrm{diag}(((t_{ij})_{x_{ij}})_{ij}$, and $(t_{ij})_{x_{ij}} > 0$, the eigenvalues of $\nabla_{\boldsymbol{x}}\boldsymbol{F}$ have the same signs as those of $\nabla_{\boldsymbol{t}}\boldsymbol{F}$. Therefore $\nabla_{\boldsymbol{x}}\boldsymbol{F}$ has a negative eigenvalue at the initial 0-pricing equilibrium, which is an unstable equilibrium point for the ad hoc dynamic system; even though, the MSA algorithm converged to this unstable equilibrium point. A plausible explanation for this fact is that the domain where the flow vector moves away from the unstable equilibrium is so small that the MSA algorithm actually generates a sequence of points outside of that domain which converges approximately to the true equilibrium point.

(ii) The 0-gradient condition is sufficient only for local optima, because the optimization problem is of a non-convex nature. See the following Conclusions section for a discussion regarding the non-convexities.

5. Conclusions

Pricing problem is a typical *network design problem* (see, Chapter 9, Bell and Iida, 1997). For a general design problem, there are some *design parameters* or *control variables* (e.g., tolls, road capacities, transit fares) to be determined in order to optimize a certain objective function. The objective function usually depend on some *state variables* (e.g., link costs, link flows), as well as the control variables. The state variables are endogeneously determined under equilibrium constraints, therefore the objective function finally depends only on the control variables. This dependency is in general *non-convex* in nature (see e.g., Chapter 9, Bell and Iida, 1997), even if the equilibrium problem is itself convex for given control variables. As is clarified in Section 2 of this paper, the convexity of the equilibrium problem is generally ensured if all the links in each mode are "congested" in the sense defined there. Due to this convexity property, the sensitivity analysis of the equilibrium solution is simple.

The equilibrium problem itself becomes a non-convex one if there exhibit economies of scale on some links and transit lines. The main complex features inherent in such a non-convex problem are: multiple equilibria, instability and singularity problems encountered in sensitivity analysis algorithms. We have investigated all these features under the SUE model setting for a transportation system consisting of a congested road network and separate transit lines; the transit lines are assumed to exhibit economies of scale by imposing a break-even constraints. Our theoretical observations and analysis have been demonstrated by a numerical example. Although our work is far from complete, it does provide a general methodological framework for dealing with the two-fold non-convexity in optimal design and management of transportation systems.

Some related themes worth further studying are listed as follows:

(i) A comprehensive theoretical study on the properties of equilibria in a general multimodal transport system where each mode consists of a general network which may have economies of scale and external diseconomies caused by congestion; Development of related computational techniques.
(ii) Studies on theoretical and practical aspects of sensitivity analysis methods for such a general combined system; Design of global optimization algorithms for finding global optimal solution.
(iii) A study on the dynamic evolution property of such a system. We note that this problem has been addressed by Wilson as early as 1976 for a bi-modal system, and recently by Watling (1999) and others for a single mode network but in greater depth.
(iv) Integration of the frequency-based SUE modeling for the transit network. In Lam et al. (1999) and Lam et al. (2002) such a model is adequately developed for modeling a detailed transit network. Sensitivity analysis for such kind of transit equilibrium models has to be conducted first, and then combined with the well established method for road network.
(v) Consideration of the physical interaction of public transit mode with the automobile network. This is particularly important when buses are main public transit tools.

Acknowledgments

We would like to express our thanks to H.P. Lu of Tsinghua University and T. Miyagi of Gifu University for their valuable discussions. We are also grateful to the anonymous referees and M.G.H. Bell of Imperial College London for their comments that have improved the presentation of the paper and stimulated a few important issues for further research as well. The second author acknowledges the financial support by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKUST6033/02E).

References

Abrahamsson, T., Lundqvist, L., 1999. Formulation and estimation of combined network models with applications to Stockholm. Transportation Science 33, 80–100.

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

Bellei, G., Gentile, G., Papola, N., 2002. Network pricing optimization in multi-user and multimodal context with elastic demand. Transportation Research B 36, 779–798.

Clark, S.D., Watling, D.P., 2002. Sensitivity analysis of the probit-based stochastic user equilibrium assignment model. Transportation Research B 36, 617–635.

Davis, G., 1994. Exact solution of the continuous network design problem via stochastic user equilibrium assignment. Transportation Research B 28, 61–75.

Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research 5, 83–111.

Dial, R.B., 1999. Minimal-revenue congestion pricing part I: a fast algorithm for the single-origin case. Transportation Research B 33, 189–202.

Huang, H., Wang, S., Bell, M., 2001. A bi-level formulation and quasi-Newton algorithm for stochastic equilibrium network design problem with elastic demand. Journal of Systems Science and Complexity 14, 40–53.

Kraus, M., 1991. Discomfort externalities and marginal transit fares. Journal of Urban Economics 29, 249–259.

Lam, W.H.K., Gao, Z.Y., Chan, K.S., Yang, H., 1999. A stochastic user equilibrium assignment model for congested transit networks. Transportation Research B 33, 1–18.

Lam, W.H.K., Zhou, J., Sheng, Z.-H., 2002. A capacity restraint transit assignment with elastic line frequency. Transportation Research B 36, 919–938.

Luenberger, D.G., 1984. Linear and Nonlinear Programming, second ed. Addison-Wesley, Reading, MA.

Luo, Z.-Q., Pang, J.-S., Ralph, D., 1996. Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge, UK.

Miyagi, T., Suzuki, T., 1996. A Ramsey price equilibrium model for urban transit systems: a bilevel programming approach with transportation network equilibrium constraints. In: Proceedings of the 7th World Conference on Transport Research, vol. 2, pp. 65–78.

Mohring, H., 1976. Transportation Economics. Balinger, Cambridge, MA.

Sheffi, Y., 1985. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall, Englewood Cliffs, NJ.

Small, K.A., 1992. Urban Transportation Economics. Harwood Academic Publishers GmbH, Lausanne, Switzerland.

Watling, D.P., 1999. Stability of the stochastic equilibrium assignment problem: a dynamic systems approach. Transportation Research B 33, 281–312.

Williams, H.C.W.L., 1977. On the formation of travel demand models and economics measures of user benefit. Environment and Planning A 9, 285–344.

Wilson, A.G., 1976. Catastrophe theory and urban modeling: an application to modal choice. Environment and Planning A 8, 351–356.

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

Yang, H., 1999. System optimum, stochastic user equilibrium and optimal link tolls. Transportation Science 33, 354–360.

Yang, H., Meng, Q., Bell, M.G.H., 2001. Simultaneous estimation of the origin–destination matrices and travel-cost coefficient for congested networks in a stochastic equilibrium. Transportation Science 35, 107–123.

Ying, J.Q., Miyagi, T., 2001. Sensitivity analysis for stochastic user equilibrium network flows—a dual approach. Transportation Science 35, 124–133.

Ying, J.Q., Liu, B.Y., Miyagi, T., 2001. Sensitivity analysis based method for optimal road network pricing. In: Proceedings of the 5th International Conference on Optimization: Techniques and Applications, December, Hong Kong, China, vol. 1, pp. 63–71.