Available online at www.sciencedirect.com

SCIENCE DIRECT®

Transportation Research Part B xxx (2004) xxx–xxx

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

A first best toll pricing framework for variable demand traffic assignment problems

Mehmet Bayram Yildirim [a,*], Donald W. Hearn [b]

[a] *Industrial and Manufacturing Engineering Department, Wichita State University, Wichita, KS 67260, United States*
[b] *Center for Applied Optimization, Industrial and Systems Engineering Department, University of Florida, United States*

Received 21 June 2002; received in revised form 15 March 2004; accepted 10 August 2004

Abstract

In this paper, we present a toll pricing framework for a general variable demand traffic assignment problem with side constraints, where the demand between an origin destination pair is a function of the least total travel cost for making the trip. This general demand model unifies earlier toll pricing treatments of the variable demand models including elastic demand traffic assignment problems and combined distribution assignment problems. All of these models have the constant toll revenue property. Given that users experience the side constraints, we show that when they are charged by a toll vector in the first best toll set, the system optimal flows and demands are achieved. We then present a toll pricing framework by which a traffic planner might find the most appropriate toll vector given certain restrictions and objectives on the network. Finally, we derive the toll sets and illustrate the toll pricing framework for specific instances of the general variable demand models.

Keywords: Congestion toll pricing; Combined distribution assignment models; Elastic demand traffic assignment problem

* Corresponding author. Tel.: +1 316 978 3426.
E-mail addresses: bayram.yildirim@wichita.edu (M.B. Yildirim), hearn@ise.uf1.edu (D.W. Hearn).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2004.08.001

1. Introduction

Congestion is becoming an inevitable part of everyday life in most metropolitan areas all over the world. Increasing population and wealth result in more automobiles than current transportation networks can handle. Due to limited expansion possibilities of the transportation network, congestion has increased drastically over the last decade. Arnott and Small (1994) estimate that the annual cost of driving in congested areas in 39 metropolitan areas of the US is around $48 billion or $640 per driver. Japan's international co-operation agency calculated that Bangkok loses one-third of its potential output due to congestion (The Economist, 1998). In 1995, 75% of San Francisco's, 66.5% of Los Angeles's, 63.8% of San Bernardino and Riverside's and 60% of San Jose's rush hour traffic was under congested conditions (Schiller, 1998).

Traffic planners often charge users in order to utilize the system resources more efficiently and restrain the number of travellers on the transportation network, based on the time, distance and congestion level. Economic theory argues that to achieve economic efficiency in a market, the price of a good or a service should be at its full cost to society, but this is not generally the case in transportation networks. Pigou (1920), Armstrong-Wright (1986), Beckmann et al. (1956), Elliot (1986), Johnson (1964), Luk and Chung (1997) and Arnott and Small (1994) recommend the marginal social cost pricing (MSCP) tolls that are equal to the negative externalities imposed on other users (such as cost of congestion, travel delays, air pollution, and accidents) in order to have an efficient utilization of the transportation system. The MSCP tolls are easy to compute by a formula which prices the time value of an additional user on the system. This makes MSCP tolls one of the most popular tools for road pricing applications. Economists define the MSCP tolls as the "first best" tolls since they achieve the optimal utilization of the transportation system by changing the user behavior to system optimal behavior. We extend the definition for the first best pricing by also including all toll vectors which achieve the most efficient utilization of the transportation system. We define the set of all such vectors as the *First Best Toll Set*. Bergendorff (1995), Bergendorff et al. (1997) and Hearn and Ramana (1998) show that there exist toll vectors other than the MSCP toll vector in the first best toll set for the fixed demand traffic assignment problems. A similar result for the elastic demand traffic assignment problems is given by Hearn and Yildirim (2002). Hearn and Ramana (1998) define the procedure for finding alternative toll vectors as the *Toll Pricing Framework*. For the fixed demand traffic assignment problems, Bergendorff (1995), Hearn and Ramana (1998) and Hearn et al. (2001) show that cheaper and more implementable toll vectors compared to MSCP tolls can be found among such solutions. Hearn and Yildirim (2002) and Larsson and Patriksson (1998) show that elastic demand traffic assignment problems have a constant toll revenue property.

This paper mainly focuses on traffic equilibrium models which have the constant toll revenue property. The system problem for these models usually has the network balance constraints and some side constraints. Side constraints can be used to describe the transportation authority's goals, control policies and physical constraints. Hearn (1980), Ferrari (1995), Larsson and Patriksson (1994, 1995a,b, 1998, 1999) and Yang and Bell (1997) analyze the side constrained traffic equilibrium models in detail to show that an unconstrained tolled user problem will have the same equilibrium flows as the side constrained one. Furthermore, Larsson and Patriksson (1998) present a toll pricing model based on Lagrange multipliers and show that the constant toll revenue property holds for elastic demand problems with side constraints.

In this paper, our primary goal is to generalize the first best elastic demand toll pricing framework to variable demand models. Variable demand traffic assignment problems model the situation where the demand between origin–destination (OD) pairs might change substantially based on the "service level," which is usually a function of the travel time on the transportation network. When service level varies, users might decide to take the trip, or not to take the trip at all. We propose a general variable demand (GVD) model which can encompass several traffic assignment problems such as the elastic demand (ED) traffic assignment problem, elastic demand with capacities on links (ED-C) and combined distribution assignment models (CDAM). We show that the toll set for each of these models is an instance of the toll set of the GVD model and all of these models have the constant toll revenue property.

2. General variable demand TA models

Let $\mathscr{G}$ denote the network model of a transportation system which consists of streets, $\mathscr{A}$, and intersections, $\mathscr{N}$. Mathematically, this is a network, $\mathscr{G} = (\mathscr{N}, \mathscr{A})$ where $\mathscr{N}$ is the node set and $\mathscr{A}$ is the link set. Let A be the node-arc incidence matrix of $\mathscr{G}$.We define a commodity by an origin, p, and a destination, q. Let $\mathscr{K}$ represent the set indexing all such origin–destination pairs $k = (p, q)$. The kth *commodity flow* vector is denoted by x^k and the sum of all the commodity flow vectors is the *aggregate flow vector* v. We assume that a continuously differentiable **cost map** $s : \mathscr{A} \to \mathscr{A}$ is given. ∇s denotes the Jacobian of s. When the aggregate flow on the network is v, the travel time for a user on arc a is given by $s_a(v)$. Let c_k be the generalized cost of travel for commodity k, and t_k denote a nonnegative invertible function of c_k. The demand for travel from some origin p to destination q is expressed as $t_k(c_k)$. It is generally assumed that $t_k(c_k)$ are monotonically decreasing and bounded from above. Let $w_k(t_k)$ indicate the inverse of t_k. The vector t has components t_k and the vector function $w(t)$ has $w_k(t_k)$ as components. Then the set of inequalities which define all possible feasible flows and demands can be stated as

$$
\begin{array}{llll}
v & = \sum_{k \in \mathscr{K}} x^k & & : \mu \\
Ax^k & = E_k t_k & \forall k \in \mathscr{K} & : \rho^k \\
g_m(v) & \leqslant 0 & \forall m \in M & : \gamma_m \\
h_n(t) & = 0 & \forall n \in N & : \theta_n \\
-x^k & \leqslant 0 & \forall k \in \mathscr{K} & : \tau^k \\
-t_k & \leqslant 0 & \forall k \in \mathscr{K} & : \phi_k
\end{array}
$$

where $E_k = e_p - e_q$, a column incidence vector for commodity k, and e_p and e_q are unit vectors. The first constraint is the **aggregate flow** constraint and the second constraint is the **network balance** constraint. $g_m(v) \leqslant 0$, $m \in M$, are constraints on the aggregate flows, and $h_n(t) = 0$, $n \in N$, are constraints on the demand, and $(\mu, \rho, \gamma, \theta, \tau, \phi)$ are multipliers related to the constraints. Some of the constraints on flows and demands can be physical constraints on the network such as road capacity and environmental regulations. Others might be traffic management constraints usually imposed by regulators. We define the set of all feasible flows and demands, Ω, as

Table 1
$W(t)$, $g(v)$ and $h(t)$ for the special cases of the GVD model

Model	$W(t)$	$g(v)$	$h(t)$
ED	$-\sum_{k\in\kappa}\int_0^{t_k} w_k(z)\,\mathrm{d}z$		
ED-C	$-\sum_{k\in\kappa}\int_0^{t_k} w_k(z)\,\mathrm{d}z$	$v_a - C_a \leqslant 0$	
CDAM	$\frac{1}{\zeta}\sum_{k\in K} t_k \ln t_k$		$\sum_q t_k = O_p, \sum_p t_k = D_q, k = (p,q)$

$$\Omega = \{(v,t) : (v,t) \text{ satisfies the constraints above}\}.$$

The system and user objectives that we consider throughout this paper are of the form $S(v) - W(t)$. We use $S(v) = \sum_{a\in\mathscr{A}} s_a(v)v_a$ for the system problems and $S(v) = \sum_{a\in\mathscr{A}}\int_0^{v_a} s_a(z)\,\mathrm{d}z$ for the user problems. We assume that $S(v)$ and $-W(t)$ are convex. $W(t)$, $g(v)$ and $h(t)$ for the models that we consider are listed in Table 1. We assume that $g(v)$ and $h(t)$ are linear, and $s_a(v)$ and $w_k(t_k)$ are separable. The resulting system and user problems have a convex nonlinear objective and linear constraints. Thus a constraint qualification automatically holds and multipliers will exist at the local optima. [1] Note also that the results in this paper can be extended, under assumptions such as in Larsson and Patriksson (1998), when g and h are nonlinear, provided that some constraint qualification holds.

2.1. System problem

In the general variable demand model, when the system objective is to minimize the total system cost (or maximize the total social welfare), the goal is to utilize the network resources in the most efficient way. The system problem (SOPT-GVD) for the GVD model can be defined as:

$$\begin{aligned} \min \quad & s(v)^{\mathrm{T}} v - \sum_{k\in\mathscr{K}} \int_0^{t_k} w_k(z)\,\mathrm{d}z \\ \text{subject to} \quad & (v,t) \in \Omega. \end{aligned}$$

The system optimal flows and demands are characterized by the following lemma:

Lemma 1. *Assume that $(\bar{v},\bar{t}) \in \Omega$ is an optimal solution to the SOPT-GVD problem. Then there exists $(\mu,\rho,\gamma,\theta,\tau,\phi)$ such that the following conditions hold:*

$$\begin{aligned} s_a(\bar{v}) + \frac{\partial s_a(\bar{v})}{\partial v_a}\bar{v}_a + \sum_{m\in M}\gamma_m \frac{\partial g_m(\bar{v})}{\partial v_a} &= \mu_a && \forall a \in \mathscr{A} \\ \mu_a + (\rho_i^k - \rho_j^k) &= \tau_a^k && \forall k \in \mathscr{K},\ \forall a = (i,j) \in \mathscr{A} \\ -w_k(\bar{t}_k) - \sum_{n\in N}\theta_n \frac{\partial h_n(\bar{t})}{\partial t_k} + (\rho_q^k - \rho_p^k) &= \phi_k && \forall k = (p,q) \in \mathscr{K} \\ \gamma_m g_m(\bar{v}) &= 0 && \forall m \in M \\ \sum_{k\in K}\left(w_k(\bar{t}_k) + \sum_{n\in N}\theta_n \frac{\partial h_n(\bar{t})}{\partial t_k}\right)\bar{t}_k &= \sum_{a\in\mathscr{A}} \mu_a \bar{v}_a \\ zpc\tau, \phi, \gamma &\geqslant 0 \end{aligned}$$

[1] See Bazaraa et al. (1993, Chapter 5).

Proof. Since $(\bar{v}, \bar{t})$ is a KKT point for the SOPT-GVD problem, there exists $(\mu, \rho, \gamma, \theta, \tau, \phi)$ such that the following conditions hold:

$$\frac{\partial L}{\partial v_a} = s_a(\bar{v}) + \frac{\partial s_a(\bar{v})}{\partial v_a}\bar{v}_a - \mu_a + \sum_{m\in M}\gamma_m \frac{\partial g_m(\bar{v})}{\partial v_a} = 0 \quad \forall a \in \mathscr{A}$$

$$\frac{\partial L}{\partial x_a^k} = \mu_a + (\rho_i^k - \rho_j^k) - \tau_a^k = 0 \quad \forall k \in \mathscr{K}, \ \forall a \in \mathscr{A}$$

$$\frac{\partial L}{\partial t_k} = -w_k(\bar{t}_k) + (\rho_q^k - \rho_p^k) - \sum_{n\in N}\theta_n \frac{\partial h_n(\bar{t})}{\partial t_k} - \phi_k = 0 \quad \forall k = (p, q) \in \mathscr{K}$$

$$\gamma_m g_m(\bar{v}) = 0 \quad \forall m \in M$$

$$\tau_a^k \bar{x}_a^k = 0 \quad \forall k \in \mathscr{K}, \ \forall a =\in \mathscr{A}$$

$$\phi_k \bar{t}_k = 0 \quad \forall k \in \mathscr{K}$$

$$\tau, \phi, \gamma \geqslant 0$$

where $L(x, v, t, \mu, \rho, \gamma, \theta, \tau, \phi)$ is the Lagrangian of the system problem. We can aggregate the complementarity conditions, $\tau_a^k \bar{x}_a^k = 0$ and $\phi_k \bar{t}_k = 0$ to

$$\sum_{k\in\mathscr{K}}\sum_{a\in\mathscr{A}}\tau_a^k \bar{x}_a^k = \sum_{k\in\mathscr{K}}\sum_{a\in\mathscr{A}}\mu_a \bar{x}_a^k - \sum_{k\in\mathscr{K}}(x^{-k})^{\mathrm{T}} A^{\mathrm{T}} \rho^k = 0$$

and

$$\sum_{k\in\mathscr{K}}\phi_k \bar{t}_k = -\sum_{k\in\mathscr{K}}\left(w_k(\bar{t}_k) + \sum_{n\in N}\theta_n \frac{\partial h_n(\bar{t})}{\partial t_k}\right)\bar{t}_k + \sum_{k\in\mathscr{K}}\bar{t}_k E_k^{\mathrm{T}}\rho^k$$
$$= -\sum_{k\in\mathscr{K}}\left(w_k(\bar{t}_k) + \sum_{n\in N}\theta_n \frac{\partial h_n(\bar{t})}{\partial t_k}\right)\bar{t}_k + \sum_{k\in\mathscr{K}}(A\bar{x}^k)^{\mathrm{T}}\rho^k = 0$$

Summing up these equations we get,

$$\sum_{k\in\mathscr{K}}\sum_{a\in\mathscr{A}}\tau_a^k \bar{x}_a^k + \sum_{k\in\mathscr{K}}\phi_k \bar{t}_k = 0$$

$$\sum_{a\in\mathscr{A}}\mu_a \bar{v}_a - \sum_{k\in\mathscr{K}}(\bar{x}^k)^{\mathrm{T}} A^{\mathrm{T}}\rho_k - \sum_{k\in\mathscr{K}}\left(w_k(\bar{t}_k) + \sum_{n\in N}\theta_n \frac{\partial h_n(\bar{t})}{\partial t_k}\right)\bar{t}_k + \sum_{k\in\mathscr{K}}(A\bar{x}^k)^{\mathrm{T}}\rho_k = 0$$

$$\sum_{a\in\mathscr{A}}\mu_a \bar{v}_a - \sum_{k\in\mathscr{K}}\left(w_k(\bar{t}_k) + \sum_{n\in N}\theta_n \frac{\partial h_n(\bar{t})}{\partial t_k}\right)\bar{t}_k = 0. \qquad \square$$

Note that the multipliers $(\mu, \rho, \gamma, \theta, \tau, \phi)$ exist, since the system problem is a convex nonlinear programming problem with linear constraints (thus, a constraint qualification automatically holds).

2.2. User problem

As all deterministic traffic assignment models do, the user problem assumes that each traveller has complete and precise information about all routes available. The user problem is defined by Wardrop's first principle (1952). The underlying assumption is that all users taking a trip between an OD pair have the same travel time which is less than or equal to the travel time on any unutilized path.

In the system problem, there are side constraints which are used either to model the physical constraints or the management goals. However, it is usually assumed that the users perceive only conditions on the network such as the travel times, $s(v)$ (Hearn, 1980). In addition, they do not have any information on constraints imposed on the network, neither the physical ones nor the management goals. Although this might not be the case when there are physical link capacity constraints, as with bottlenecks, we will analyze the user problem without side constraints to model the user behavior.

Mathematically, the user problem can be formulated as a variational inequality. A given aggregate flow and demand vector $(\bar{v}, \bar{t})$ is a user equilibrium flow if and only if

$$s(\bar{v})^{\mathrm{T}}(v - \bar{v}) - w(\bar{t})^{\mathrm{T}}(t - \bar{t}) \geqslant 0 \quad \forall (v, t) \in V,$$

where V is the constraint set defined by the aggregate flow and network balance constraints.

Lemma 2 characterizes the user equilibrium (UOPT-GVD) flows and demands as follows:

Lemma 2. *A feasible point* $(\bar{v}, \bar{t})$ *is a user equilibrium flow if and only if there exists* $(\mu, \rho, \gamma, \theta, \tau, \phi)$ *such that the following holds*:

$$\begin{aligned} s_a(\bar{v}) + (\rho_i^k - \rho_j^k) &= \tau_a^k && \forall k \in \mathscr{K}, \ \forall a = (i, j) \in \mathscr{A} \\ -w_k(\bar{t}_k) + (\rho_q^k - \rho_p^k) &= \phi_k && \forall k \in \mathscr{K} \\ \sum_{k \in \mathscr{K}} w_k(\bar{t}_k)\bar{t}_k &= \sum_{a \in \mathscr{A}} \mu_a \bar{v}_a && \\ \tau, \phi &\geqslant 0 && \end{aligned}$$

Proof. Let $(\bar{v}, \bar{t})$ be the equilibrium flows and demand, then

$$s(\bar{v})^{\mathrm{T}} v - w(\bar{t})^{\mathrm{T}} t \geqslant s(\bar{v})^{\mathrm{T}} \bar{v} - w(\bar{t})^{\mathrm{T}} \bar{t} \quad \forall (v, t) \in V$$

holds. This implies

$$\min_{(v,t) \in V} s(\bar{v})^{\mathrm{T}} v - w(\bar{t})^{\mathrm{T}} t \geqslant s(\bar{v})^{\mathrm{T}} \bar{v} - w(\bar{t})^{\mathrm{T}} \bar{t}.$$

Consider now

$$\begin{aligned} \min \quad & s(\bar{v})^{\mathrm{T}} v - w(\bar{t})^{\mathrm{T}} t \\ \text{subject to} \quad & (v, t) \in V \end{aligned}$$

From the inequality above it is clear that $(\bar{v}, \bar{t})$ solves this linear program. Therefore, by linear programming duality, there exists $(\mu, \rho, \gamma, \theta, \tau, \phi)$ such that the following conditions hold:

$$\begin{aligned}
s_a(\bar{v}) - \mu_a &= 0 && \forall a \in \mathscr{A} \\
\mu_a + (\rho_i^k - \rho_j^k) - \tau_a^k &= 0 && \forall k \in \mathscr{K},\ \forall a \in \mathscr{A} \\
-w_k(\bar{t}_k) + (\rho_q^k - \rho_p^k) - \phi^k &= 0 && \forall k \in \mathscr{K} \\
\tau_a^k \bar{x}_a^k &= 0 && \forall k \in \mathscr{K},\ \forall a \in \mathscr{A} \\
\phi_k \bar{t}_k &= 0 && \forall k \in \mathscr{K} \\
\tau, \phi &\geqslant 0
\end{aligned}$$

where $\tau, \phi, \gamma \geqslant 0$. The last three equations are the complementary slackness conditions. Further, $\tau_a^k \bar{x}_a^k = 0$ and $\phi_k \bar{t}_k = 0$ can be aggregated in a manner similar to Lemma 1 to obtain

$$\sum_{k \in \mathscr{K}} w_k(\bar{t}_k)\bar{t}_k = \sum_{a \in \mathscr{A}} \mu_a \bar{v}_a$$

Then the lemma follows. □

3. First best toll pricing for the GVD model

In this section, we extend the toll set idea in Bergendorff (1995), Hearn and Ramana (1998) and Hearn and Yildirim (2002) to the GVD model. We will assume that the transportation authority aims to exactly replicate the system optimal flows and demands, and the effect of side constraints. We will first describe this framework for UOPT-GVD and then discuss the relation with other user problems.

3.1. Toll pricing theory for the GVD model

Suppose that $(\bar{\gamma}, \bar{\theta})$ are the multipliers for the side constraints in the SOPT-GVD problem. Assume that on each link users are charged by an amount equal to

$$\lambda_a = \beta_a + \sum_{m \in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a}$$

where β is a toll vector and $\sum_{m \in \mathscr{M}} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a}$ is the constraint cost. Furthermore, each user is rewarded (or penalized) for an amount equal to

$$\xi_k = \sum_{n \in N} \bar{\theta}_m \frac{\partial h_n(\bar{t})}{\partial t_k}$$

for making a trip between OD pair $k = (p, q)$.

Suppose $s_\lambda(v) = s(v) + \lambda$ is the perturbed cost map. Further, let $w_\xi(t) = w(t) + \xi$ be the perturbed inverse demand function. Then the perturbed user problem can be stated as a variational inequality:

$(\bar{v}, \bar{t})$ is a tolled user equilibrium flow if and only if

$$(s(\bar{v}) + \lambda)^{\mathrm{T}}(v - \bar{v}) - (w(\bar{t}) + \xi)^{\mathrm{T}}(t - \bar{t}) \geqslant 0 \quad \forall (v, t) \in V$$

Let U^*_β be the set of all perturbed equilibrium solutions, i.e., those $(\bar{v},\bar{t}) \in V$ that solve the above variational inequality, and let S^* be the optimal solution set for SOPT-GVD. We would like to identify all toll vectors such that the resulting perturbed user equilibrium problem has a system optimal solution, i.e.,

$$\emptyset \neq U^*_\beta \subseteq S^*$$

Any such β is defined to be a valid toll vector. Let the toll set denoted by $\mathscr{T}$ be the set of all such vectors, $\mathscr{T} := \{\beta \mid \emptyset \neq U^*_\beta \subseteq S^*\}$. Now, for a given vector $(\bar{v},\bar{t}) \in V$, define

$$W_{\text{GVD}}(\bar{v},\bar{t},\bar{\gamma},\bar{\theta}) = \{\beta \mid (\bar{v},\bar{t}) \in U^*_\beta\}$$

as the set of all tolls which ensures that $(\bar{v},\bar{t})$ is a solution of UOPT$_{(\lambda,\xi)}$-GVD. In fact, we can define $W_{\text{GVD}}(\bar{v},\bar{t},\bar{\gamma},\bar{\theta})$ by the following result.

Lemma 3. *Given that* $(\bar{v},\bar{t}) \in V$ *and* $(\bar{\gamma},\bar{\theta})$ *are the system optimal multipliers,* $W_{\text{GVD}}(\bar{v},\bar{t},\bar{\gamma},\bar{\theta})$ *is the polyhedron given by the* β *part of the linear system defined in* (β,ρ):

$$\left(s_a(\bar{v}) + \beta_a + \sum_{m\in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a}\right) + (\rho_i^k - \rho_j^k) \geqslant 0 \quad \forall k \in \mathscr{K},\ \forall a = (i,j) \in \mathscr{A}$$

$$\left(w_k(\bar{t}_k) + \sum_{n\in N} \bar{\theta}_n \frac{\partial h_n(\bar{t})}{\partial t_k}\right) - (\rho_q^k - \rho_p^k) \leqslant 0 \quad \forall k \in \mathscr{K}$$

$$\sum_{a\in\mathscr{A}} \left(s_a(\bar{v}) + \beta_a + \sum_{m\in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a}\right)\bar{v}_a = \sum_{k\in\mathscr{K}} \left(w_k(\bar{t}_k) + \sum_{n\in N} \bar{\theta}_n \frac{\partial hn(\bar{t})}{\partial t_k}\right)\bar{t}_k$$

Proof. Using Lemma 2 and the perturbed cost vector $s_a(v_a) + \beta_a$ and the perturbed inverse demand function $w_\xi(t) = w(t) + \xi$, we observe that the solution to the perturbed user equilibrium satisfies the following:

$$\left(s_a(\bar{v}) + \beta_a + \sum_{m\in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a}\right) = \mu_a = \rho_j^k - \rho_i^k + \tau_a^k \quad \forall k \in \mathscr{K},\ \forall a \in \mathscr{A}$$

and

$$\rho_q^k - \rho_p^k - \left(w_k(\bar{t}_k) + \sum_{n\in N} \bar{\theta}_n \frac{\partial h_n(\bar{t})}{\partial t_k}\right) = \phi^k \quad \forall k \in \mathscr{K}$$

hold. Combining the above with $\tau_a^k \geqslant 0$ and $\phi^k \geqslant 0$ yields the first two inequalities in the lemma. The lemma follows. □

When the optimality conditions of the system and tolled user problems are compared, it is clear that they differ in the definition of μ and the complementarity conditions, $\bar{\gamma}_m g_m(\bar{v}) = 0 \quad \forall m \in M$ and $\bar{\gamma} \geqslant 0$. The difference between μ for the system and tolled user problems is $\nabla_s(\bar{v})\bar{v}$, the MSCP toll vector. The complementarity conditions also hold for the tolled user problem since $(\bar{\gamma},\bar{\theta})$ are the multipliers from the system problem. We can conclude from these observations that if the individuals were charged with the MSCP tolls and constraint costs, the perturbed user equilibrium flows and demands would be the same as the system optimal flows and demands.

The toll set is defined by variables (β, ρ) and by parameters $(\bar{v}, \bar{t}, \bar{\gamma}, \bar{\theta})$. The interpretation of what these variables mean is as follows: β is the toll vector. ρ_i^k is the total travel cost from origin p to node i for the commodity $k = (p, q)$ on the transportation network. Furthermore, $\bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a}$ is the amount that users will pay if there were no $g_m(v) \leqslant 0$ constraint and similarly, $\bar{\theta}_n \frac{\partial h_n(\bar{t})}{\partial t_k}$ is the amount that users will pay in the absence of experiencing the $h_n(t) = 0$ constraint.

Note that if the user and system problems (thus the tolled user problem) have unique solutions, then $\mathscr{T} = W_{\mathrm{GVD}}(\bar{v}, \bar{t}, \bar{\gamma}, \bar{\theta})$. A complete description of characterization of the toll sets is given in Hearn and Ramana (1998). In addition, the toll set where $(\bar{\gamma}, \bar{\theta})$ are fixed has similar characteristics as the elastic demand toll set (Hearn and Yildirim, 2002; Larsson and Patriksson, 1998). The following corollary shows that the total toll revenue for the GVD models is constant as it is in the elastic demand traffic assignment problems:

Corollary 1. *The toll revenue for the variable demand traffic assignment models is constant when the toll set is defined by* $W_{\mathrm{GVD}}(\bar{v}, \bar{t}, \bar{\gamma}, \bar{\theta})$.

Proof. This is obvious from the definition of $W_{\mathrm{GVD}}(\bar{v}, \bar{t})$. The total toll revenue is

$$\beta^{\mathrm{T}}\bar{v} - \sum_{k\in\mathscr{K}} \sum_{n\in N} \bar{\theta}_n \frac{\partial h_n(\bar{t})}{\partial t_k} \bar{t}_k + \sum_{a\in\mathscr{A}} \sum_{m\in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a} \bar{v}_a = \sum_{k\in\mathscr{K}} w_k(\bar{t}_k)\bar{t}_k - \sum_{a\in\mathscr{A}} s_a(\bar{v})\bar{v}_a. \qquad \square$$

However, note that this is not the case in traffic assignment problems with fixed demand (Bergendorff et al., 1997; Hearn and Ramana, 1998; Larsson and Patriksson, 1998).

Wardrop's First Principle implies that at equilibrium the utilized paths for an origin–destination pair have the same travel time (travel costs) and the unused ones do not have lower travel times (travel costs). The next corollary verifies that the generalized cost version of this principle holds (a similar result for the fixed demand case is given by Larsson and Patriksson (1999). Let r_k be a path for commodity k, χ_{ar_k} be 1 if link a is on path r_k and zero otherwise,

$$s_a(\bar{v}_a) + \beta_a + \sum_{m\in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a}$$

be the "generalized cost" of travelling on arc a and

$$w_k(\bar{t}_k) + \sum_{n\in N} \bar{\theta}_n \frac{\partial h_n(\bar{t})}{\partial t_k}$$

be the "generalized benefit" for commodity k.

Corollary 2. *At the UOPT$_\beta$-GVD solution $(\bar{v}, \bar{t})$ the total generalized cost on any path is greater than or equal to the generalized benefit for any commodity k, i.e.,*

$$\sum_{a\in\mathscr{A}} \chi_{ar_k} \left(s_a(\bar{v}_a) + \beta_a + \sum_{m\in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a} \right) \geqslant w_k(\bar{t}_k) + \sum_{n\in N} \bar{\theta}_m \frac{\partial h_n(\bar{t})}{\partial t_k}$$

For any path with positive flow, the inequality holds as an equality. Therefore the costs on the utilized paths are constant for any commodity.

Proof. For $k = (p,q)$ and $a = (i,j) \in r_k$,

$$s_a(\bar{v}_a) + \beta_a + \sum_{m\in M} \bar{\gamma}_m \frac{\partial g_m(\bar{v})}{\partial v_a} \geqslant \rho_j^k - \rho_i^k$$

holds as an equality when $\bar{x}_a^k > 0$ and similarly,

$$w_k(\bar{t}_k) + \sum_{n\in N} \bar{\theta}_n \frac{\partial h_n(\bar{t})}{\partial t_k} \leqslant \rho_j^k - \rho_i^k$$

holds as an equality when $\bar{t}_k > 0$. That is, the complementarity conditions force the inequality to hold as an equality when there is flow on that path. Thus, when they are summed over the arcs on the utilized paths, the inequality in the corollary holds as an equality. The conclusion of the corollary then follows. □

In other words, this corollary can be interpreted by saying that every commodity reaches an equilibrium exactly when the generalized path costs, including tolls, equals the generalized benefit at the final demand level.

Note that the Lagrangean multipliers related to side constraints are usually not unique (Larsson and Patriksson, 1998). Detailed discussions on traffic management through link tolls where the Lagrangean multipliers are allowed to vary can be found in Larsson and Patriksson (1998) and Yang and Bell (1997). In this case, β can absorb any contribution of the Lagrange multipliers for the side constraints. This leads to some simplification in notation. However, we have not used the conditions in Larsson and Patriksson (1998) to be compatible with Hearn (1980) and Yang and Huang (1998).

4. First best toll pricing framework for the GVD models

Assume that SOPT-GVD and UOPT-GVD models have unique solutions. Furthermore, assume that the traffic planner is interested in obtaining the system optimal flows and demands and the constraint costs $(\bar{\gamma}, \bar{\theta})$. Then, the first best toll pricing framework for GVD models can be summarized as follows:

Step 1: Solve the SOPT-GVD to obtain the system optimal solution $(\bar{v}, \bar{t})$ and $(\bar{\gamma}, \bar{\theta})$.
Step 2: Define a toll set which is the β part of $W_{\text{GVD}}(\bar{v}, \bar{t}, \bar{\gamma}, \bar{\theta})$.
Step 3: Define and optimize an objective function over the toll set, possibly intersected with other constraints (see Table 2 for examples).

As in the fixed demand case, various objectives in Step 3 lead to linear or mixed integer programs. For the variable demand models, the natural choices can be D-MINSYS, D-MINREV, MINTB and MINMAX objectives (see Table 2). The first of these aims to minimize the total tolls collected while constraining the toll vector to be nonnegative. D-MINREV is similar to D-MINSYS, but tolls are free to be negative as well as positive. Note that the marginal social cost pricing tolls, $\beta_{\text{MSCP-GVD}} = \nabla s(\bar{v})\bar{v}$, are optimal for the GVD model when the objective is minimization of the total toll revenue. This is because of the constant toll revenue property. MINTB minimizes the

Table 2
Alternative optimization formulations

Toll	Objective ($\overline{Z}$)	Extra constraints ($\widehat{T}$)
D-MINREV	$\beta^T \bar{v}$	
D-MINSYS	$\beta^T \bar{v}$	$\beta \geqslant 0$
MINMAX	z	$z \geqslant \beta_a + \bar{\gamma}_a \ \forall a \in \mathscr{A}, \beta \geqslant 0$
MINTB	$\sum_{a \in \mathscr{A}} y_a$	$\beta_a + \bar{\gamma}_a \leqslant M y_a \ \forall a \in \mathscr{A}, y_a \in \{0,1\}, \beta \geqslant 0$
MINTB/MINREV	$\sum_{a \in \mathscr{A}} y_a$	$\beta_a + \bar{\gamma} \leqslant M y_a \ \forall a \in \mathscr{A}, y_a \in \{0,1\}$

number of toll booths, and the MINMAX minimizes the maximum toll on the transportation network. Note that any valid toll vector including MSCP toll vector produces the same toll revenue. As a result, D-MINSYS and D-MINREV formulations are not very interesting for the GVD networks, however, having some alternative toll vectors might help the traffic planner to propose different toll pricing schemes for various scenarios. In all of these toll pricing problems, the transportation authority collects both β and $\bar{\gamma}$ as the total toll charge on a link. In this paper, we use GAMS optimization modeling package (1995) and linear and nonlinear solvers (CPLEX, 2001 and MINOS, 1983) to implement the GVD Toll Pricing Framework to obtain alternative toll vectors. The GAMS code is available upon request.

Actually, the toll pricing framework has interesting implications: The traffic managers do not need to put physical constraints (such as closing a lane, etc.) to restrict users on the transportation network. In fact, users can be charged by the "correct" toll amount for not only sustaining the management goals but also maintain the flows to be consistent with the constraints on the transportation system.

5. GVD models in the literature

In this section, we present special cases of the GVD model and define the toll set for each model. Furthermore, we illustrate the toll pricing framework for each model on the Nine Node network (Fig. 1) which has nine nodes and 18 links, and all of the links have cost functions with the same structure:

$$s_a(v) = s_a(v_a) = T_a(1 + 0.15(v_a/C_a)^4)$$

where T_a is a measure of travel time when there is zero flow and C_a is the practical capacity of link a. In fact $s(v)$ is strictly convex and separable. There are four OD pairs. The particular choices of $s(v)$ and $w(t)$ results in having unique solutions to the system and user problems. Thus, the toll set for each model is $\mathscr{T} = W_{\text{GVD}}(\bar{v}, \bar{t}, \bar{\gamma}, \bar{\theta})$. In this paper, tolls are expressed in time units (Arnott and Small (1994) discuss how conversion to dollars can be made based on studies in the US).

5.1. Elastic demand TA models

The system model with elastic demand assumes that the goal of transportation planners is to maximize the net economic benefit (Hearn and Yildirim, 2002; Yang and Huang, 1998) which is the difference between the total network user benefit, $\sum_{k\in\mathscr{K}}\int_0^{t_k} w_k(z)\,\mathrm{d}z$, and the system cost, $s(v)^T v$. The system problem (SOPT-ED) can be stated as

$$\begin{aligned} &\max \quad && \sum_{k\in\mathscr{K}}\int_0^{t_k} w_k(z)\,\mathrm{d}z - s(v)^{\mathrm{T}} v \\ &\text{subject to} \quad && (v,t)\in V. \end{aligned}$$

where V denotes the set of all feasible flows and demands, which is defined by the aggregate flow, network balance and nonnegativity constraints, i.e., there are no g and h constraints.

The elastic demand user equilibrium problem (UOPT-ED) models Wardrop's first principle (1952). Mathematically, the user problem can be stated as a variational inequality:

Find $(\bar{v},\bar{t})\in V$ such that

$$s(\bar{v})^{\mathrm{T}}(v-\bar{v}) - w(\bar{t})^{\mathrm{T}}(t-\bar{t}) \geqslant 0 \quad \forall (v,t)\in V.$$

Using Lemma 3 of the GVD model, we can extend the notion of toll pricing to the elastic demand traffic assignment models. [2] We will give a description of the toll set in the next section.

5.2. Elastic demand TA models with link capacities

Traditionally, traffic planners handle capacities using special social cost functions like

$$s_a(v_a) = T_a\left(1 + 0.15\left(\frac{v_a}{C_a}\right)^4\right).$$

However, using these functions will not guarantee that the attained flows do not exceed the capacity on each link. For example, in Fig. 1, the capacity of link $(5,7)$ is 11, but as it can be seen from Table 4, the uncapacitated user problem allows 26.44 users and the system problem has 17.98 users on link $(5,7)$, both of which are far above the capacity. Thus it might be important to use the capacity constraints explicitly in some cases.

[2] Refer to Hearn and Yildirim (2002) for a detailed description of first best toll pricing theory for elastic demand traffic assignment problems.