



A modern congestion pricing policy for urban traffic: subsidy plus toll

Saeed Asadi Bagloee¹  · Majid Sarvi¹

Received: 22 November 2016/Revised: 27 March 2017/Accepted: 29 March 2017/Published online: 22 April 2017
© The Author(s) 2017. This article is an open access publication

Abstract Congestion pricing is seen as an effective policy to address traffic congestion. In such policies where money, people and authorities are involved, the success generally hinges upon two factors: equity (being fair) and acceptability (to both people and authorities). The primary concern is the equity, for which “tradable credit scheme (TCS)” has been introduced and extensively studied in the literature. Nevertheless, due to the complexity of the trading schemes, the TCS has yet to find any foot in the real world. To this end, a novel idea of rewarding has substituted the trading component to be known as toll-and-subsidy scheme (TSS). The idea is to charge the drivers on some roads (toll) while rewarding them to use other alternative—and perhaps underutilized—roads (subsidy). The research of the TSS is in its infancy stage. The problem to be tackled in this study is as follows: Given a set of roads constituting a cordon line around the central business district (CBD) or across a screen line, how much toll or subsidy should be assigned to each road? The problem is first transformed into a capacitated traffic assignment problem. We employ a solution method based on augmenting the travel time of roads up to the level at which the traffic volumes do not exceed some target rates. A real dataset from the city of Winnipeg, Canada, is used as a pilot study. We then discuss policy-related applications of the TSS. It is proved in the literature that one can obtain

optimal TSSs for various objectives and considerations. To this end, the non-negativity of the toll values is relaxed which results in a valid toll set. Nevertheless, the computational time is found to be of highest significance. Our method differs in the fact that the traffic volumes are bounded from the above and it is quite affordable. The main contribution is first to highlight the concept of subsidy along with traditional thought of merely toll. Second is to interpret the Lagrangian values of the capacity constraints as the values of the toll/subsidy.

Keywords Congestion pricing · Toll and subsidy · Tradable credit scheme · Lagrangian values

1 Introduction

Traffic congestion has a hefty toll on the society and economy. In one estimate, the cost of congestion in 2010 in the urban areas of the USA amounted to USD 101 billion [1, 2]. A recent study shows that the avoidable cost of congestion for the Australian capital cities is estimated to be around AUD16.5 billion for the 2015 financial year, which is a significant increase from about AUD12.8 billion for 2010 [3]. It seems there is no ceiling cap for such costs. For instance, traffic congestion could cost the UK economy more than £300 billion over the next 16 years which represents a 63% increase in annual costs by 2030 [4]. Building more roads should not be the put forward as the first solution to the congestion for the following reasons: (1) More roads encourages more trips; something is known as induced demand [5]. (2) More roads may adversely cause more congestion; the phenomenon is known as Braess’ paradox [6, 7]. (3) Roads’ constructions are capital intensive and time-consuming [8, 9].

✉ Saeed Asadi Bagloee
saeed.bagloee@unimelb.edu.au

Majid Sarvi
majid.sarvi@unimelb.edu.au

¹ Department of Infrastructure Engineering, Melbourne School of Engineering, The University of Melbourne, Parkville, VIC 3010, Australia

By casting the transport in a supply–demand context, the congestion occurs when the demand outweighs the supply. Hence, the solution to the congestion can be attained either by adding to the supply or by lowering the demand. In the supply side (more roads), “Building more roads to prevent congestion is like a fat man loosening his belt to prevent obesity” (Lewis Mumford 1955). In the demand side, congestion pricing, originally proposed in the economic literature [10], has emerged as a significant leverage to bring balance back to the supply–demand equation. Economists, who believe in the power of a free market, seek solutions in the cost and pricing mechanism when they see any supply–demand imbalance [11, 12].

A traditional approach is to flag some (highly congested) roads as “toll roads” aiming to redistribute the traffic load evenly across the network including less congested roads. Labeling road with a price tag does not fall delicately in the mind of the people. Furthermore, equity (in terms of who has to pay and how much) is the main challenge in such scheme. To this end, some scholars introduced some sort of tradable credits which will be discussed in depth in the next section.

In the same spirit, we propose an unorthodox scheme in which some roads are tolled and some other (underutilized) roads are designated as subsidy roads. The aim is to encourage commuters to avoid highly congested roads and see alternative roads in the periphery area by means of credit incentive (subsidy).

The remainder of the article is organized as follows. The relevant studies in the literature are reviewed in Sect. 2. The concepts and underlining mathematical features of the methodology are elaborated in Sect. 3. Numerical results are presented in Sect. 4 followed by conclusion remarks in Sect. 5.

2 Literature review

This section presents a comprehensive review through which we cover a variety of subjects pertaining to the toll pricing, credit-based schemes, system optimal, etc.

2.1 Actual cost of congestion and system optimal

What we pay for transport is fuel cost, fare/tolls and value of travel time while we ignore other hidden costs imposed upon the society such as environmental degradation—namely externality cost. Hence, if we pay the actual cost of what we are consuming (the product here is the mobility), the market will correct itself and no such imbalance would occur. As the result, the question now is what is the actual cost of using a road, so that if it is charged as a toll the

congestion would disappear? The pioneers on this idea are [13–15], and the following excerpt summarizes their take on the matter.

Although many traffic flow patterns could exist, it is widely believed that people seek to minimize their individual travel cost/time which results in a user equilibrium (UE) traffic pattern. Unfortunately, the UE is not the best pattern as far as the total costs of the transport system (the entire road network) are concerned. The “best” traffic pattern is the one at which the costs of the entire transport system are minimized, referred to as the system optimal (SO). The traffic congestion can be defined and formulated as “minimization of the cost of the system” that is to forcing the traffic pattern from UE to SO. To do so, congestion pricing is a plausible approach. It is mathematically proved [16] that the price tags of the roads are their respective marginal costs. Accordingly, the UE turns to the SO. The marginal cost can represent some externality costs such as pollutions and inequity [17]. From a mathematical point of view, it is easy to identify the marginal costs of the roads [16].

2.2 First-best congestions pricing versus second-best and equity concerns

In reality, it is not possible to communicate the marginal costs of all roads of the network with the drivers (i.e., tolling all roads) and expect to see a full SO traffic flow (known as first-best congestion pricing [18]). Rather, a few roads, presumably the highly congested ones, are selected as tolled roads and some fairly consistent toll rates are assigned to them (known as second-best congestion pricing [19]). Examples are the imposition of tolls on the roads forming cordon lines encompassing a central business district (CBD) or on the roads forming some screen lines which bisect the road network. One of the main challenges in the pricing schemes is the equity [20]. Equity concern is associated with the distribution of costs and benefits among the members of the society. A policy can be called equitable if it meets a normative standard of fairness (so equity means fairness). Following is one of the typical complaints: “I am living just one block after the cordon line in the CBD and my office is in the outer CBD. Beside the fact that you put my house inside the cordon line, I have no business in the CBD, so it is totally unfair to tax me.” In the USA, many congestion pricing proposals have been rejected based on worries that they are inequitable [21].

The equity concerns arise when people feel what they are paying is more than what they are getting. The inequity makes the concept of congestion pricing hard sell to both public and traffic authorities which is also known as acceptability issue, [2, 22–24]).

2.3 Credit-based scheme, an answer to equity concerns

One answer to the equity issue raised in the above typical complaint can be set out as follows: The authorities would issue a limited number of cordon-passing credits, at a reasonable price, to the people working or living in the CBD. Commuters who receive the credits can then sell them in a free market to other frequent commuters. Such a scheme is called “tradable credit scheme (TCS).” Although the TCS has been extensively studied in the past 20 years, there is no real implementation in the real world [25]. The TCS is hampered by three crucial questions: who are eligible to receive the initial credit?; what should be the initial price?; and, what would be the trading mechanism?

Given the above-mentioned complexities involved with the TCS, recently, the inclusion of “subsidy” in the congestion pricing has been introduced; that is, instead of traversing the main (tolled) roads, commuters can choose some other roads (un-tolled) slightly away from the main (tolled) roads. In addition, by traveling on these un-tolled roads commuters are given credits; so they are called subsidized roads.¹ The concept is called “toll-and-subsidy scheme (TSS).” In contrast to the TCS, there is no provision for the credit trading, nor does it need to carefully pick up the eligible drivers which both are advantageous.² There are also a number of other advantages associated with such schemes: (1) The equity and acceptability features are highly upheld. (2) It would strongly rally the public behind the scheme since it cannot be considered as a different form of taxation. (3) The scheme can be used to promote the public transport mode. (4) Districts in the vicinity of subsidized roads are expected to receive added values which can be then exploited by zoning authorities in the land-use planning initiatives (a practice called value capture [26]).

2.4 TSS and capacitated traffic assignment

Although finding the locations of the tolled roads can also be cast in a mathematical programming framework, past experiences have shown that it is usually a decision to be made by the authorities considering many technical, societal as well as political factors. One practical, intuitive and less computationally intensive approach is via traffic

¹ In such schemes, these subsidized roads are largely underutilized; hence—by adding subsidy—they would take some traffic loads from already congested main (tolled) roads. Commuters can then spend the collected credit either to pass the main road, parking fee, public transport, metro, car’s annual registration fee, traffic fine, etc.

² Given the current available technology even in the developing countries, implementation of such schemes is not a big challenge.

control policies, which are discussed in the following exposition. In principle, congestion pricing aims to keep the traffic volumes under some certain levels. These levels could be simply set to prevent traffic jam on some specific part of the cities, such as downtowns or the CBD. Traffic authorities have a fairly good understanding of the acceptable level of traffic loads in major roads or total incoming traffic in the CBD. Or in a more insightful way, the acceptable level of the incoming traffic to the CBD can be found from the macroscopic fundamental diagram [27]. Therefore, the location of the roads to be tolled and accepted level of traffic loads on them are assumed known and exogenously given in the TSS.

The only remaining issue is the pricing regime. The TSS’ pricing problem is articulated as follows: Given a set of roads with capped traffic volumes to be either tolled or subsidized, which one must be tolled, which one must be subsidized and how much?

The past studies have mathematically proved that this problem always has a solution [28]. The main focus of the current research is to present a methodology to reach a TSS tailored to large-sized networks. Since the congestion pricing is subject to a set of capped traffic volumes, we first transform the problem to a conventional “capacitated traffic assignment problem (CTAP).” We then interpret the values of the Lagrange multipliers of the capacity constraints as the toll/subsidy values.³ To solve the resulting CTAP, we relax the capacity constraints by adding their Lagrange multiplier values to the travel time of the respective roads, viz. inflated travel time [29]. By doing so, the CTAP is transformed into a simple (un-capacitated) traffic assignment problem (TAP) which is easier to solve. The solution algorithm is designed as an iterative process: The travel times are increased to the level at which the traffic volumes do not exceed the roads’ capacities.

We present the TSS in the context of congestion pricing and then the tradable schemes. The interested reader for a comprehensive review of the former subject can consult with [18, 30, 31] and for the latter with [1, 20, 24, 32–35].

2.5 Literature on TSS

This section is dedicated to a comprehensive review on the subject of the TSS which sometimes referred to as toll-cum-rebate or price-rewarding schemes as well.

³ The primary aim of this research is to bring down traffic volume below a certain target values for a certain roads (it is a regular practice in traffic control in which—based on historical traffic county—traffic authorities know what are desirable traffic loads specially during peak hours heading to or off the down towns). Therefore, subjecting traffic assignment problem to the above-mentioned target values is equivalent to solving a capacitated traffic assignment problem.

Bernstein [36] proposed a time-varying pricing scheme that includes a toll and a subsidy on the same routes. Commuters are charged if they arrive at the peak hour and are subsidized if they choose off-peak times. Over a simple network, it was demonstrated that the scheme could reduce the equilibrium cost. Adler and Cetin [37] discussed a redistribution scheme in which, revenue collected from a desirable route is transferred to drivers on a less desirable route. In contrast to the SO traffic pattern, they showed the redistribution scheme renders almost identical results. Guo and Yang [38] investigated the existence of Pareto-improving⁴ TSS considering multiple users. In the presence of the restricting Pareto-improving condition, they show that the TSS problem always has a solution if a significant gap exists (The gap is the room for improvement and it refers to the difference in total travel time between the UE and SO traffic patterns). Even in a more restrictive scheme where no subsidy exists, the Pareto-improving solution exists [39, 40]. Furthermore, in a fashion reminiscent of the TSS, the revenue from tolling highway roads is used to subsidize the fare adjustments on transit lines. In a controlled environment, Ben-Elia and Ettema [41] investigated the impact of a reward scheme on the behavior of commuters in the Netherlands. The results suggest that the reward scheme can be certainly effective in the short run, while it has yet to be investigated in the long run. It is worth noting that, recently, the idea of using rewards to change commuters' behavior has been implemented in the Netherlands—though in limited scale—as part of a program called Spitsmijden [41]. Using a bilevel Stackelberg game approach, Maillé and Stier-Moses [42] investigated a mechanism based on a pure subsidy (without any toll) that aims at reducing congestion in urban networks. They concluded that the subsidies can significantly lower the social cost and part of the budget for rebates may come from the savings that arise from the most efficient use of the system.

In conjunction with the TSS, some scholars seek the revenue-neutral property; that is, the total toll revenue collected is set equal to the total amount of subsidies paid (essentially, the average commuter pays nothing [43]). As the result, the authorities can attain an optimized system by merely redistributing the revenue without engaging in financial transfer to or from the commuters. The scheme is also called Robin Hood toll scheme [44], which can be characterized by the following simple equation: toll–subsidy = 0.

By relaxing the non-negativity of the toll values [44] introduced the idea of the valid toll set. Based on their

work, by establishing a number of linear programs one can obtain optimal TSSs for various objectives and considerations. Nevertheless, the computational time is found to be of highest significance. Our method differs in the fact that the traffic volumes are bounded from the above and the computational expenses is of no concern.

Compared to the tradable schemes (TCS), the Robin Hood method tries to obviate the “trade” part, which comes at a restrictive cost compared to what our method (TSS) offers. In other words, in our approach, we relax the equality sign in the equation; that is, the sum of the toll minus subsidy could be positive or negative (while in Robin Hood, this must equate to zero). This is aligned with what we are after (i.e., certain roads associated with certain target volumes to be achieved either by toll or subsidy); if we enforce Robin Hood method (adding additional constraints), the problem may become infeasible.

Regardless of the challenging task of the toll redistribution, it is highly appealing to merge the two ideas of revenue-neutral and Pareto-improving TSS, where everyone is a winner. For a single origin–destination (O–D) pair, Eliasson [45] showed that compared to the do-nothing scenario, in a revenue-neutral TSS, where the total travel time is reduced, no commuter is found to be worse off. Liu, Guo and Yang [46] and Nie and Liu [47] examined the existence of Pareto-improving and revenue-neutral pricing scheme in a simple bimodal network consisting of road and a parallel transit line. Results provided by Liu, Guo and Yang [48] suggests that the condition for every user being better off is weaker than the condition for total system travel time reduction. In other words, any revenue-neutral pricing scheme is Pareto-improving as long as it reduces the total system travel time. The results provided by Nie and Liu [47] imply that the revenue-neutral property is a restricting condition which may compromise the existence of a solution. In other words, external subsidies may be required to make every commuter happy (Pareto-improving). Xiao and Zhang [48] showed that on a one-origin or one-destination network, a Pareto-improving, system-optimal and revenue-neutral TSS always exists. It turns out that such a scheme may not always exist for a multiorigin network; hence, they sought the maximum possible revenue collected by the TSS problem subject to the Pareto-improving constrains. Given the results of the above studies, it is safe to say that maintaining both the Pareto-improving and revenue-neutral properties is difficult; possibly the best bet is to drop the revenue-neutral in order to secure a solution. Chen and Yang [28] proved that in the context of nonnegative toll schemes and/or toll-cum-rebate schemes for Pareto-efficient control, the TSS always has a solution. Considering that tolls and subsidies are represented by positive and negative values, then, a negative cycle is a distinct closed path along which the sum of the road travel costs is negative. If a negative cycle exists, commuters can

⁴ In a naïve congestion pricing, although the entire transport system are sought to be better off, some commuters might become worse off. To this end, the Pareto-improving seek a pricing scheme based on which no one becomes worse off.

make money merely by traversing along the negative cycle. Chen and Yang [28] proved that the TSS—as we set out here—has always a solution, if, loosely speaking, the amount of toll is higher than that of subsidy.

3 The mathematical model

As noted early, the problem to be undertaken is as follows: Given a set of roads associated with capped volumes (let us call it capacity) to be tolled or subsidized, what is a pricing scheme. In other words, what is the price tag (toll or subsidy) of each road? In this section, we first articulate the mathematical expression of the problem and then propose the solution algorithm.

3.1 Formulation of the problem

The problem as set out above with the capped traffic volume essentially is a capacitated traffic assignment problem (CTAP). Consider $G(N, A)$ a traffic network as a graph which consists of N nodes and A roads on which $R \subset N \times N$ is a set of origin–destination pairs. Set of roads nominated as toll/subsidy is denoted by $\bar{A} \subset A$. The CTAP can be formulated as a nonlinear programming problem as follows (throughout the paper, all terms are nonnegative unless otherwise stated):

[CTAP]:

$$\min z(x) = \sum_{a \in A} \int_0^{x_a} t_a(x) dx, \tag{1}$$

s.t.:

$$\sum_p f_{p,r} = q_r \quad r \in R \rightarrow \text{Lagrange multiplier } (w_r), \tag{2}$$

$$f_{p,r} \geq 0 \quad p \in P_r, r \in R, \tag{3}$$

$$x_a = \sum_r \sum_p f_{p,r} \delta_{a,p}^r \quad a \in A, p \in P_r, r \in R, \tag{4}$$

$$x_a \leq C_a \quad a \in \bar{A} \rightarrow \text{Lagrange multiplier } (\beta_a), \tag{5}$$

where z is the Beckmann objective function to be minimized; x_a is the traffic flow on road a ; q_r is the travel demand pertaining to OD pair $r \in R$; $f_{p,r}$ is the traffic flow on path p connecting OD pair r ; P_r is the set of all paths connecting OD pair r ; $\delta_{a,p}^r$ is the road–path incidence (1 if road a belongs to path p connecting OD pair r , and 0 otherwise); C_a is the capped traffic volume of road a .⁵

⁵ Note that the target values are literally considered as physical capacities of the respective roads. The formulation can in fact be extended to include physical capacities of all the roads. Nevertheless,

3.2 Karush–Kuhn–Tucker conditions and Lagrange multiplier

It is proved that the CTAP subject to linear constraints is a strictly convex problem which renders a unique global optimal solution of road flows [49–54]. Let us consider w_r and β_a as Lagrange multipliers for travel demand and capacity constraints, respectively; hence, the Karush–Kuhn–Tucker (KKT) conditions are established as

$$f_{p,r}(u_{p,r} + \sum_{a \in A} \delta_{a,p}^r \beta_a - w_r) = 0 \quad p \in P_r, r \in R, \tag{6}$$

$$\beta_a(C_a - x_a) = 0 \quad a \in \bar{A}, \tag{7}$$

$$u_{p,r} + \sum_{a \in A} \delta_{a,p}^r \beta_a - w_r \geq 0 \quad p \in P_r, r \in R, \tag{8}$$

$$C_a - x_a \geq 0 \quad a \in \bar{A}, \tag{9}$$

$$f_{p,r} \geq 0 \quad p \in P_r, r \in R, \tag{10}$$

$$\beta_a \geq 0 \quad a \in \bar{A}, \tag{11}$$

$$\sum_p f_{p,r} = q_r \quad r \in R, \tag{12}$$

where $u_{p,r} = \sum_{a \in A} \delta_{a,p}^r t_a$ is the total travel time perceived by commuters on a path p connecting OD pair r . Let us define $\hat{u}_{p,r}$, an “inflated” travel time of the respective path as

$$\hat{u}_{p,r} = \sum_{a \in A} \delta_{a,p}^r (t_a + \beta_a) \quad a \in \bar{A} \tag{13}$$

Introduction of (13) into (6) and (8) results in

$$f_{p,r}(\hat{u}_{p,r} - w_r) = 0 \quad p \in P_r, r \in R, \tag{14}$$

$$\hat{u}_{p,r} - w_r \geq 0 \quad p \in P_r, r \in R. \tag{15}$$

With respect to Eqs. (14) and (15), it can be proved that w_r is the travel time of the shortest path connecting OD pair r ; hence, if a path takes some traffic volume ($f_{p,r} > 0$), it is certainly the shortest path. In other words, the first principle of Wardrop holds and the global optimum solution of the CTAP is, in fact, a UE traffic flow. According to Eqs. (7), (9) and (11), if the capacity constraint is binding, which means saturation ($x_a = C_a$), then the corresponding lagrange multipliers are nonzero ($\beta_a > 0$); otherwise, it is zero. There are two terms contributing to the augmented travel time [Eq. (13)]: normal or cruise travel time (t_a) and beta (β_a). The beta can be interpreted as toll/subsidy value or alternatively the additional delay to prevent the respective road become oversaturated (i.e., traffic volume above the capacity). In the literature, the beta is also interpreted as the waiting time

Footnote 5 continued
the target traffic loads of the tolled/subsidized roads of the case studies are set up below their actual respective capacities.

caused by the queue buildup in the oversaturated roads [49–52, 55–58]. As shown, the beta which represents the value of toll or subsidy is a positive number, whereas the subsidy is supposed to be a negative value. We will later handle this issue when an especial provision is devised to prevent any negative cycle.

3.3 Augmented travel time

Let us first assume that there exists only toll and no subsidy. After we establish the formulation for the tolls, we will then extend the discussion to the subsidy. Let us consider the delay function as $t_a = t_a^0(1 + f(x_a))$, where t_a^0 is the free flow travel time perceived by commuters on the road a and $f(x_a) \geq 0$ is a non-decreasing and convex function of x_a such that $f(x_a) = 0 | x_a = 0$. The aforementioned function can accommodate a variety of known delay functions including the widely used function proposed by the US Bureau of Public Roads (the BPR delay function) [59]. Therefore, we can intuitively argue that, as long as a road is empty or uncongested, its travel time is the free flow travel time ($x_a \approx 0 \Rightarrow t_a \approx t_a^0$); as the traffic volume builds up, t_a , the travel time, increases—certainly higher than the free flow travel time (this increase can be considered as a factor of t_a^0 which is greater than 1, i.e., $((1 + f(x_a)) \geq 1)$). According to Eq. (13), \hat{t}_a the inflated travel time of “saturated” road⁶ a can be formulated as

$$\hat{t}_a = (t_a^0 + b_a)(1 + f(x_a)), \quad (16)$$

$$\beta_a = \hat{t}_a - t_a = b_a(1 + f(x_a)), \quad (17)$$

where b_a is an additional penalty in the free flow time in Eq. (16). In other words, b_a is the value of beta at $x_a = 0$; hence, we referred to it as the “initial-beta.”

Note that we have proved that the additional penalty term in the travel time is identical to the Lagrangian value of the respective capacity constraint [see β_a in Eq. (13)]. For more information, interested reader can also consult with [60]).

As a result, the travel time in the CTAP is replaced with the augmented travel time ($t_a \leftarrow \hat{t}_a$) and the capacity constraint is also dropped. It is due to the fact, the beta, the Lagrange multiplier of the corresponding capacity constraint, now contributes to the travel time. Therefore, the CTAP is transformed to an un-capacitated TAP. If we had already known the global optimum value of the beta (in the CTAP), we would have just needed to solve the TAP based on the inflated travel times using any known algorithm or software. Of course, this is not the case; hence, the values of the (initial) betas are updated iteratively in the course of

solving the TAP as explained in the following section. Moreover, the value of the beta is zero unless the corresponding road is oversaturated. In other words, Eqs. (16) and (17) apply only to the oversaturated roads to bring them down at their capacity levels (to become saturated). Hence, nonzero values are assigned to the initial-beta of the (over)saturated roads while they are set to be updated in the next iterations.

The proposed formulation differs from the second-best method which is in fact a bilevel programming problem. In the second-best methods, the aim is to minimize the total travel time spent in the network by enforcing a pricing regime, which is extremely hard to solve due to the computational complexity (also known as NP hard problem). In our proposed method, the target traffic loads of a number of selected roads (to be either tolled or subsidized) are already known and the aim is to set up the amount of tolls and subsidies to reach the target values. The main contribution of this study therefore can be attributed to avoiding a bilevel programming formulation, instead to articulate it as a CTAP which is much easier to solve even for large-sized networks. The important point to note is that, although a CTAP is basically solved, the toll/subsidy values are derived from the Lagrangian values of the respective capacity constraints.

3.4 A heuristic method to update the initial-beta

A variety of algorithms has been proposed to solve the CTAP for which Bagloee and Sarvi [29] have recently presented a comprehensive review. The inherent mathematical complexities associated with the capacity constraints have resulted in solution algorithms laden with a number of parameters to be calibrated, which is a prohibitive factor. In addition, arriving at an initial feasible solution to launch the algorithm is also a challenge. Alternatively, in this study, the mathematical complexity of the CTAP is overcome by adopting an intuitive interpretation of capacity; that is, the Lagrange multipliers of the capacity constraints are interpreted as additional delay imposed on the oversaturated roads to make them saturated. Such an interpretation has also been employed in the congestion pricing, namely the “trial-and-error” methods [18, 61–64].

The concept embedded in Eqs. (16) and (17) is to lift the delay function of the oversaturated roads until the traffic volume stabilizes at capacity level. The value of the initial-beta in the delay functions is updated iteratively, for which the main challenge is to progress toward convergence and user equilibrium. There is a strong correlation between the capacitated traffic assignment and the general theory of the congestion pricing [18] which is centered on charging commuters based on the marginal cost. Inspired by such an

⁶ In this manuscript, roads at and above capacity (i.e., $x_a = C_a$ and $x_a > C_a$, respectively) are referred to as saturated and oversaturated.

observation, the amount of additional delay to be added to obtain the inflated travel time is derived from the concept of marginal cost as follows [16, 49, 65]:

$$\tilde{t}_a(x_a) = t_a(x_a) + x_a \cdot \partial t_a(x_a) / \partial x_a, \tag{18}$$

where $\tilde{t}_a(x_a)$ is the marginal cost or travel time experienced by an additional commuter added to x_a who are already traversing road a , and $\partial t_a(x_a) / \partial x_a$ is the additional travel time experienced by each driver among x_a . The marginal cost enforces system-optimal flow across the network to extend the underutilized roads will absorb the traffic discharged off the oversaturated roads. Hence, $x_a \cdot \partial t_a(x_a) / \partial x_a$ —the additional delay imposed on the respective road—is considered as a template to update the initial-betas as follows:

$$\hat{t}_{a,C}^{(i)} = (t_a^0 + b_a^{(i)}) (1 + f(C_a)), \tag{19}$$

$$\nabla b_a^{(i)} = (x_a^{(i)} - C_a) \frac{\hat{t}_a^{(i)} - \hat{t}_{a,C}^{(i)}}{C_a}, \tag{20}$$

$$b_a^{(i+1)} = b_a^{(i)} + \nabla b_a^{(i)}, \tag{21}$$

where superscript i and a denote the current iteration and respective (over)saturated road; $\hat{t}_{a,C}^{(i)}$ represents travel time at capacity ($x_a = C_a$) of the inflated delay function; $b_a^{(i)}$ is the initial-beta or an additional penalty to free flow travel time ($b_a^{(i)}$); $\nabla b_a^{(i)}$ is the pace of the initial-beta computed at the current iteration, while $b_a^{(i+1)}$ is the updated initial-beta computed for the next iteration, initialized by zero ($b_a^{(1)} = 0$).

The formulation provided in Eq. (20) follows the spirit of marginal cost (i.e., $x_a \cdot \partial t_a(x_a) / \partial x_a$) to update the pace values. The “ x ” is replaced by the excess traffic ($x_a^{(i)} - C_a$) and the slope of the delay function (dt/dx) is replaced by the slope of the inflated travel time minus the pace value normalized by the value of the capacity $(\hat{t}_a^{(i)} - \hat{t}_{a,C}^{(i)}) / C_a$. In Fig. 1, Exhibit 1, the above formulations [Eqs. (19)–(21)] for three iterations on the (inflated) delay function are graphically shown. In the first iteration when there is no initial-beta ($b_a^{(1)} = 0$), the volume stands at $x_a^{(1)} > C_a$, and the pace $\nabla b_a^{(1)} > 0$ is computed as shown graphically, which lifts the delay function for the next iteration ($b_a^{(2)} = 0 + \nabla b_a^{(1)}$). In the second iteration, even though it is an inflated travel time ($\nabla b_a^{(1)}$), the traffic volume still stands above the capacity ($x_a^{(2)} > C_a$). Hence, the value of the pace ($\nabla b_a^{(2)} > 0$) is uplifted again in the third iteration: ($b_a^{(3)} = b_a^{(2)} + \nabla b_a^{(2)}$). The third iteration is executed and the volume stands at capacity ($x_a^{(3)} = C_a$). The three key-words or components of the proposed algorithm, beta (β_a),

initial-beta (b_a) and the pace (∇b_a) are shown in Fig. 1. During this progressive approach, if an (over)saturated road is found unsaturated—at an intermediate iteration—its corresponding penalty is nullified ($p_a^{(i+1)} = 0 | x_a^{(i)} < C_a$) as shown in Fig. 1, Exhibit 2.

The above conventions for computing the initial-beta can be summarized by two rules:

$$b_a^{(i+1)} = \begin{cases} b_a^{(i)} + \nabla b_a^{(i)} & \text{from (20)} & x_a^{(i)} \geq C_a \\ 0 & & x_a^{(i)} < C_a. \end{cases} \tag{22}$$

The above setup complies with the KKT conditions at the stationary point of the capacitated assignment problem, in which the beta of the non-saturated roads must be set zero [18].

For ease of reference, the proposed methodology is referred to as the “inflated travel time” (ITT) method, similar to the terminology used in the literature. In the context of the other methods, the numerical results showed promising convergence behavior.

3.5 Subsidy

We can now include the subsidy in the formulation. As noted early, a negative cycle may spoil the whole pricing process. To avoid such a case, we ensure the roads’ travel times to take always positive values. To this end, we first zero out the free flow travel times of all the toll/subsidy roads. We then initialize the initial-beta to the respective free flow travel time. Effectively, nothing changes and the delay functions remain intact⁷:

$$\begin{cases} b_a^{(0)} = t_a^0 \\ t_a^0 = 0 \end{cases} \quad a \in \bar{A}. \tag{23}$$

This process is graphically shown in Exhibit 3 of Fig. 1. Let us have a close look at Exhibit 3: In the end of the computations (the last iteration), due to the value of the beta (or the last initial-beta), the resulting delay function can either stand at A or B (above and below the free flow time axis). The beta minus the free flow time ($\beta_a - t_a^0$) can result in either a positive value (delay function A) or a negative value (delay function B) which represents toll or subsidy, respectively. Moreover, the absolute value of the subsidy cannot exceed the free flow time. In other words, no road will take a negative travel time; hence, no negative cycle will appear.

Therefore, given the outputs at the end of the last iteration (iteration n), the toll/subsidy values ($-t_a^0 \leq s_a \leq +\infty$) can be calculated as follows:

⁷ In other words, though the toll values can theoretically be as high as infinity (i.e., $0 \leq \text{toll} \leq +\infty$), the absolute value of subsidy cannot be higher than the free flow travel time or $-t_a^0 \leq \text{subsidy} \leq 0$.

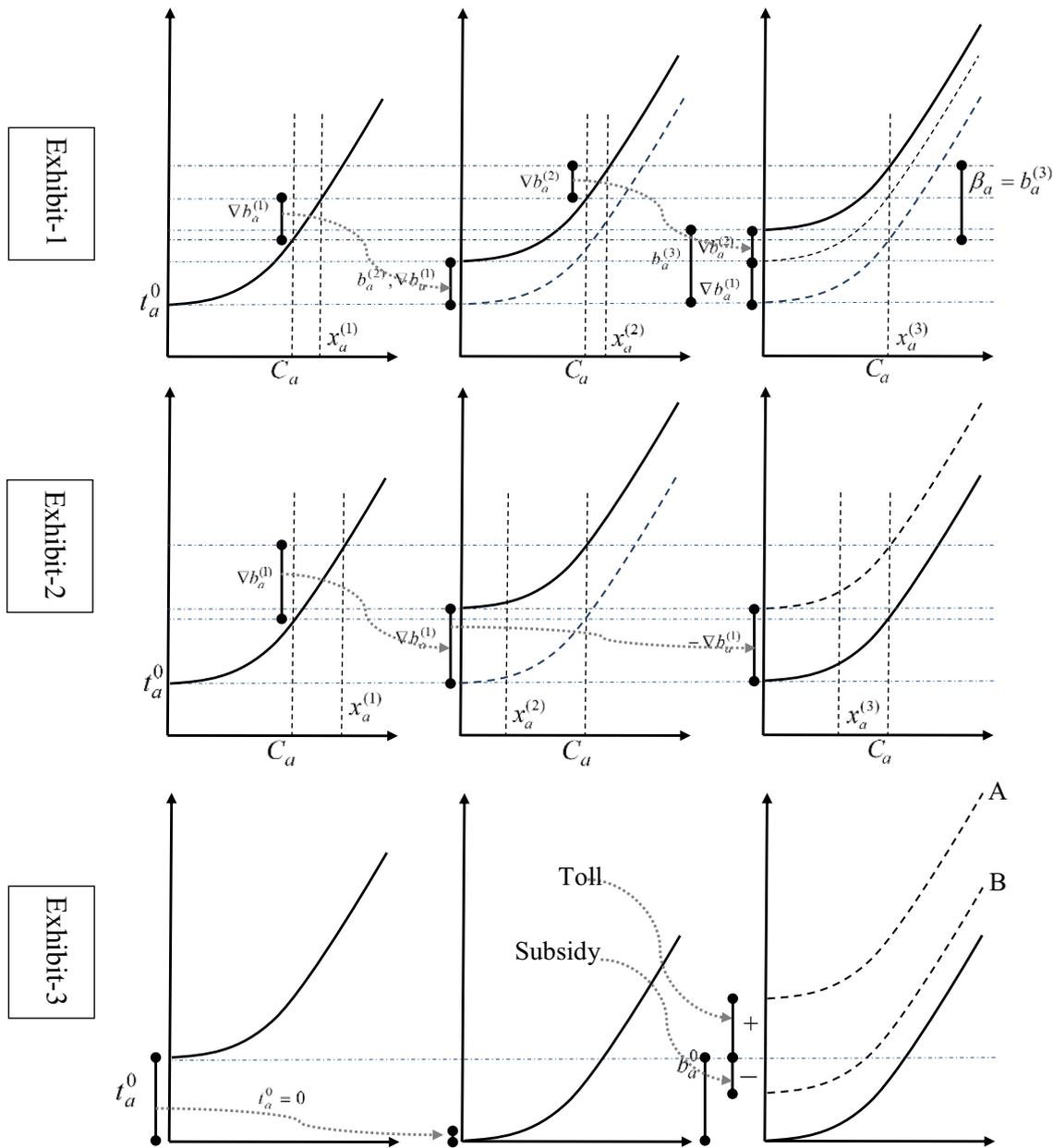


Fig. 1 Conceptual representation of the proposed methodology for the road delay functions

$$s_a = \hat{t}_a^{(n)} - t_a^0(1 + f(\bar{C}_a)) \quad a \in \bar{A}. \tag{24}$$

3.6 Capacity feasibility

In case the travel demand is higher than the capacity of the network, the problem becomes infeasible. It is important for any algorithm to have some mechanism to detect and address the infeasibility cases. To this end, one can introduce a dummy node connected with all zones via un-capacitated roads associated with high travel time. Therefore, the problem always remains feasible.

The dummy roads not only obviate any feasibility concern, they are also devised to replicate reality. Consider a single OD connected via a single road with a capacity of 10 vehicles. Faced with 15 vehicles demand, the algorithm allows 10 vehicles to drive through while 5 vehicles have no chance to do so; instead, they are assigned to the dummy roads as the leftover traffic volume. In reality, these excess 5 vehicles have no choice except to change their departure times which is also studied separately in the literature. In other words, the residual demand remains off the network until the next available traffic assignment

interval, which is discussed in the literature on dynamic traffic assignment [66].

3.7 Termination conditions

The proposed algorithm can be easily integrated into a conventional solution algorithm for the TAP such as Frank–Wolfe. Hence, given fixed rates of the initial-betas, the solution algorithm itself needs to converge and meet its own termination criterion which is mainly driven by a relative gap. Boyce, Ralevic-Dekic and Bar-Gera [67] and Dial [68] recommend a relative gap of 0.0001 to ensure a perfect convergence to road flow stability. This criterion is adopted in our numerical evaluations. It is also expected that the initial-betas show convergence behavior over the successive iterations. Given the gradual buildup of the initial-beta, a convergence criterion can be defined as pace values falling below a trivial value (denoted by ε):

$$\max_a |\nabla b_a^{(i)} / \hat{t}_a| \leq \varepsilon. \tag{25}$$

The numerical results showed that $\varepsilon = 1\%$ is sufficient to obtain reliable results. Consequently, the algorithm does not terminate unless both criteria the relative gap and the relative pace values are met.

The attempt to find the Lagrange multiplier values of the capacity constraints iteratively is, in fact, equivalent to solving a partial dual transformation of the original CTAP [62]. It can be shown that the gradient of the partial dual problem leads to a vector of excessive traffic volume for which the sub-gradient method is deemed to be a competent solution algorithm [69, 70]. The ITT algorithm is essentially a sub-gradient method because the dual values are updated iteratively based on the excessive traffic volume.

A key component of the sub-gradient methods is the step size which is uniformly applied to all roads and may vary across iterations. In order to ensure the convergence, the values of the step size must comply with a set of rules [71]. Hence, finding an appropriate scheme for the step size is a challenge. Instead, we proposed an intuitive step-size mechanism (analogous to marginal cost), exclusive to each road and iteration.

Note that the optimality conditions as well as the convergence aspects of the ITT method have been discussed in the literature ([29]). The novelty of our method in this research is to translate the Lagrangian values as toll/subsidy values. To this end, free flow travel time is first zeroed out; when the CTAP terminates, the difference between the ultimate travel time and the original travel time function specifies the toll/subsidy values (this process is also graphically illustrated in Fig. 1). We are yet to prove that

the step size devised for the ITT abides by the aforementioned rules.

In the sequel, we report on the numerical results. As a recap, Fig. 2 shows a flowchart of the proposed methodology for the toll/subsidy pricing scheme.

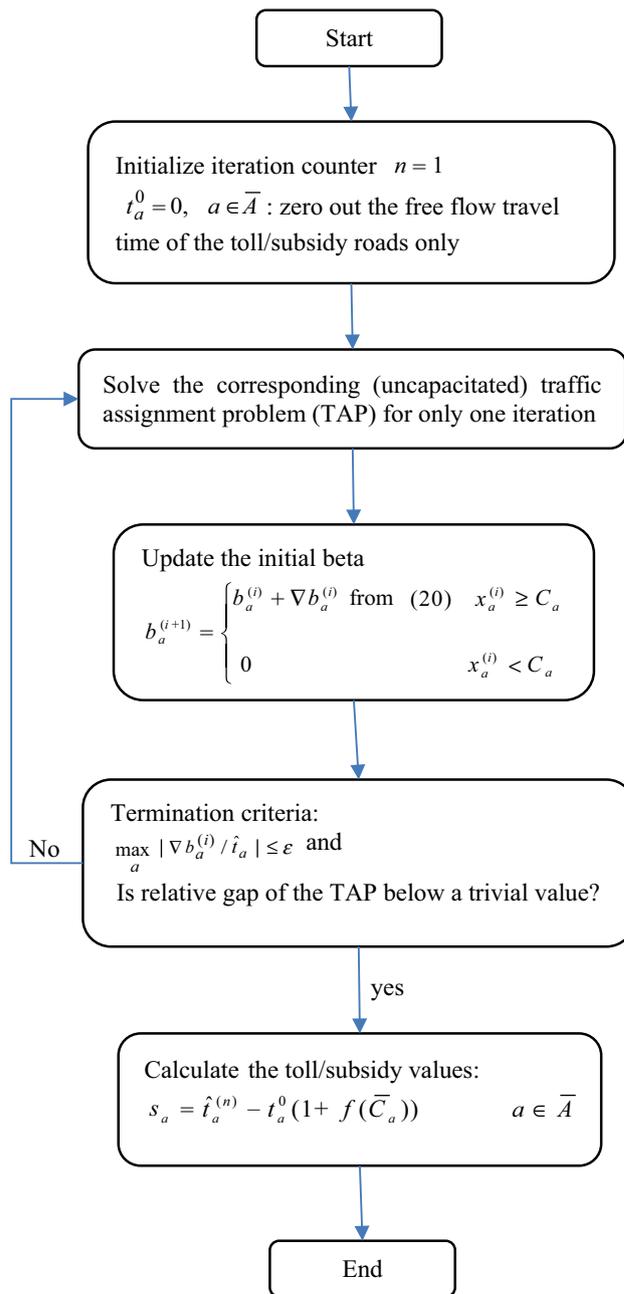


Fig. 2 Flowchart of the proposed methodology for the toll/subsidy pricing. Note that the only loop forwarded to the second block in the flowchart (where it says “Solve the corresponding (un-capacitated) traffic assignment problem (TAP) for only one iteration”) is repeated until the corresponding TAP converges to a trivial relative gap. In other words, that does not mean the TAP is solved only for one iteration. In fact, it means a lot more iterations while being paused to provide an opportunity to update the initial-beta in the subsequent block

4 Numerical results

The methodology is coded in a “macro,” which is the programming language of the EMME 3 [72]—a transportation planning software, in which the un-capacitated TAP is solved using the Frank–Wolfe algorithm. A desktop PC with a 3.60 GHz CPU and 16 GB of RAM is employed. In this section, we first report of numerical analysis of a pedagogical example to illustrate the concept of toll-and-subsidy. Second, a large-sized road network is undertaken to show applicability of the proposed methodology to real-life examples.

4.1 A Pedagogical example based on the Braess’ network

Figure 3 depicts the famous Braess network consisting of 5 roads connecting an origin to a destination as well as roads’ delay functions in which the travel demand is 6 [16]. Figure 3a also shows traffic volumes of the existing condition with no toll/subsidy such that the total travel time becomes 552 (see also Table 1). It is proved that road #5 is a Braess tainted road; that is, if it is closed, the traffic circulation improves (or in other words, the total travel time in the network decreases to 498, see Fig. 3b and Table 1). To this end, a traffic authority would like to retain the road #5, instead to toll it to decrease the total traffic load from current 2 to 0.5. Figure 3c shows the results in

which amount of toll imposed on the road #5 became 9.75 and the total travel time decreased to 506.63. As can be seen, the traffic volume of road #1 has reached 2.75. In the last try, the traffic authority would like to retain the traffic volume of the road #5 at 0.5 and increase the traffic volume of the road #1 to 3.5. It is intuitively conceivable that to increase more traffic passing through the road #1 there must be some incentive or subsidy, whereas the road #5 still has to be tolled. Figure 3d shows the results in which the target traffic loads have been achieved at the cost of toll price of 1.5 (for road #5) and subsidy of −16.5 (for road #1), while the total travel time stands at 519, which is still lower than that in the existing network scenario (Fig. 3a). All the calculations are reported in Table 1.

4.2 Winnipeg case study

The ITT algorithm is applied to a large-sized network of the city of Winnipeg, Canada, which is widely used in the literature [73] (it is also provided by [72] as part of the EMME 3). The case study consisted of 154 zones, 943 nodes and 3075 directional roads with hourly travel demand of 56,219 vehicles. The delay functions are assumed to comply with the general format of the BPR (Bureau of Public Roads) functions.

The city is bisected by two rivers which merge in the middle of the city. According to traffic surveys and field

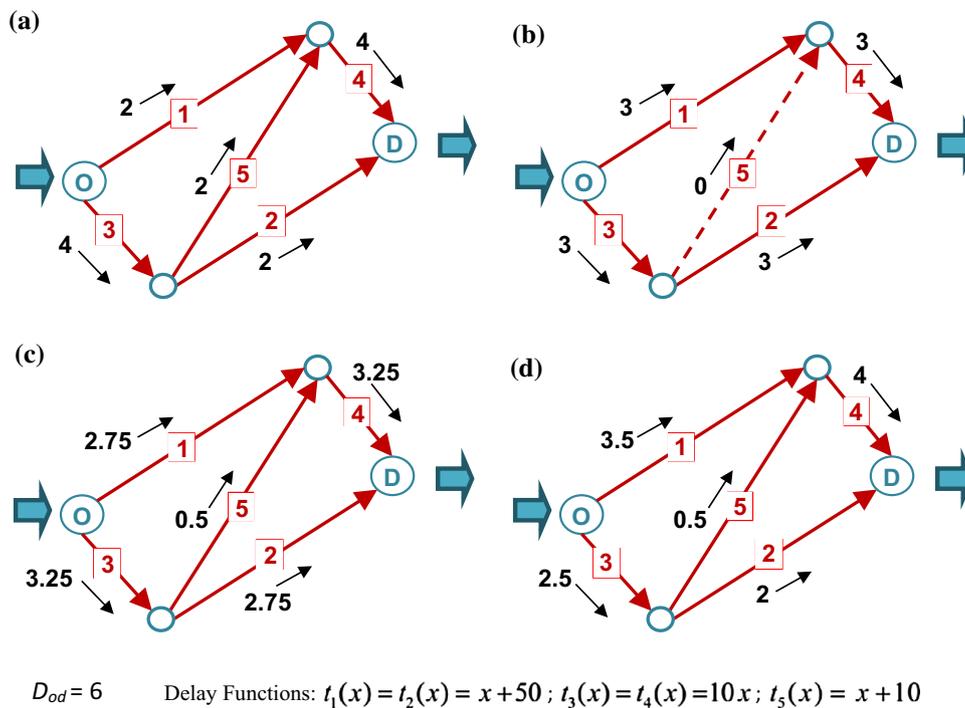


Fig. 3 Braess example, different pricing regimes

Table 1 Braess example, different toll/subsidy scenarios

Road ID	Delay function	Figure 3a: No toll/subsidy			Figure 3b: No toll/subsidy, road 5 is closed			Figure 3c: Road 5 is tolled for target value of 0.5			Figure 3d: Road 1 is subsidized for target value of 3.5 and road 5 is tolled for target value of 0.5		
		x	t	xt	x	t	xt	x	t	xt	x	t	xt
1	$x + 50$	2	52	104	3	53	159	2.75	52.75	145.06	3.5**	53.5	187.25
2	$x + 50$	2	52	104	3	53	159	2.75	52.75	145.06	2	52	104
3	$10x$	4	40	160	3	30	90	3.25	32.5	105.63	2.5	25	62.5
4	$10x$	4	40	160	3	30	90	3.25	32.5	105.63	4	40	160
5	$x + 10$	2	12	24	0	10	0	0.5*	10.5	5.25	0.5***	10.5	5.25
Sum				552			498			506.63			519

Note x denotes traffic volume, and t denotes traffic time. Data in bold pertains to the links subjected to subsidy/toll

* Toll is +9.75

** Subsidy is -16.5

*** Toll is +1.5



Fig. 4 Winnipeg case study, locations of the river crossing nominated for the TSS

observations, traffic authorities have come up with a desirable traffic distribution over the west-north river crossings to avoid any major gridlock in the AM peak hour. Figure 4 depicts 11 bridges which are subject to the traffic control.

Note that it is postulated that the target traffic values of the toll/subsidy roads are set in such a way the total system travel time improves. This can be derived from macroscopic fundamental diagrams [27] or also known as perimeter control which is an active area of research

Table 2 Winnipeg case study, input and output values

Road ID	Direction	Road direction ID	t_a^0 : Free flow travel time	\bar{C}_a : Physical capacity	C_a : Capped traffic volume	Existing traffic volume	x_a : Traffic volume of the TSS	x_a/C_a (%)	\hat{t}_a : Inflated travel time (min)	s_a : Net toll/subsidy	Monetary value of the net toll/subsidy (\$)
1	$j - j^*$	1-1	0.77	2500	1800	828.93	1796.92	99.83	2.66	1.86	0.52**
	$j - i$	1-2	0.77	2500	1800	183.76	185	10.28	0.03	-0.74	-0.21
2	$j - j$	2-1	0.84	750	750	722.32	727.05	96.94	0.03	-0.92	-0.26
	$j - i$	2-2	0.84	750	750	762.35	750.24	100.03	7.08	6.11	1.7
3	$j - j$	3-1	0.25	2500	1800	600.48	348.27	19.35	0.01	-0.24	-0.07
	$j - i$	3-2	0.25	2500	1800	2759.71	1801.39	100.08	8.63	8.37	2.33
4	$j - j$	4-1	0.35	2500	1800	1205.92	1800.78	100.04	7.2	6.84	1.9
	$j - i$	4-2	0.35	2500	1800	84.03	108.25	6.01	0.01	-0.33	-0.09
5	$j - j$	5-1	0.44	1750	1750	1641.31	1749.8	99.99	9.52	9.02	2.51
	$j - i$	5-2	0.44	1750	1750	469.82	403.75	23.07	0.02	-0.42	-0.12
6	$j - j$	6-1	0.23	2625	1800	2446.11	1802.87	100.16	12.18	11.95	3.32
	$j - i$	6-2	0.23	1750	1750	785.81	592.48	33.86	0.01	-0.22	-0.06
7	$j - j$	7-1	0.16	2500	1800	2313.34	1796.73	99.82	12.85	12.68	3.53
	$j - i$	7-2	0.16	2500	1800	656.64	389.2	21.62	0.01	-0.15	-0.04
8	$j - j$	8-1	0.23	1750	1750	2035.5	1751.26	100.07	11.73	11.47	3.19
	$j - i$	8-2	0.23	2625	1800	783	752.7	41.82	0.01	-0.22	-0.06
9	$j - j$	9-1	0.57	2625	1800	1683.56	1795.59	99.76	12.75	12.16	3.38
	$j - i$	9-2	0.47	2625	1800	903.52	824.2	45.79	0.02	-0.45	-0.12
10	$j - j$	10-1	0.30	2550	1800	1547.86	1092.88	60.72	0.01	-0.29	-0.08
	$j - i$	10-2	0.27	3750	1800	3261.11	1802.74	100.15	15.13	14.86	4.13
11	$j - j$	11-1	0.59	2500	1800	189.06	629.38	34.97	0.02	-0.57	-0.16
	$j - i$	11-2	0.59	2500	1800	289.45	1505.82	83.66	0.02	-0.58	-0.16

* “ i ” and “ j ” are the node numbers such that $i < j$

** The value of time is assumed \$16.69 per hour (Canadian dollar): $0.52 = (1.86/60) \times 16.69$. The positive values are toll and the negative values are subsidy

[74, 75].⁸ In our study, the location and target traffic values of the toll/subsidy are known and are exogenously given. To this end, in Table 2, characteristics of the crossings including the exiting traffic volume, free flow travel times t_a^0 , physical capacities \bar{C}_a as well as the capped traffic volume C_a are indicated. In this section, we first report on results of above-mentioned network and demand scenario. As a sensitivity analysis effort, given the same network, we then project the travel demand to 5% and report results of the congestion pricing.

⁸ From mathematical point of view, the objective function of the TSS is not tied to the total system travel time. In other words, the total system travel time can either increase or decrease due to the implementation of the TSS.

4.2.1 Case study of existing network and existing travel demand

The algorithm runs for 237 iterations until the relative gap and the relative pace value falls below 0.0001 and 0.01, respectively. The computation time lasts only 15 s. The results including the traffic volumes and the inflated travel time as well as the toll/subsidy values are also shown in Table 2. The restrictions on the traffic volume are upheld, and the slight violations—as shown under x_a/C_a column—are less than 1%, which can be attributed to the fact that the ITT is a heuristic algorithm. Figure 5 demonstrates variations of traffic volume and the inflated travel times over successive iterations pertaining to the roads 1, 2 and 3 (note that we chose the first three roads; otherwise, the figure would be inundated with myriad of data becoming hard to read). The results indicate that the variation is volatile in

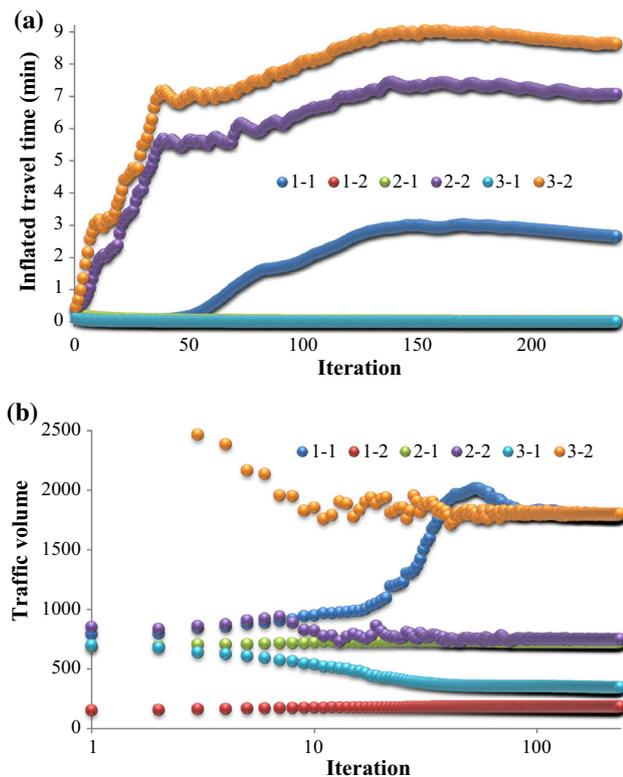


Fig. 5 Winnipeg case study, the first three bridges over successive iterations: **a** variation of the traffic volumes; **b** variation of the inflated travel time

early iterations, but it then stabilizes as the results head toward the convergence. Similar erratic variations in solving the CTAP have also been reported in the literature [55].

Figure 6a illustrates traffic volumes before and after implementation of the TSS. It is obvious that the traffic volumes are squeezed below the capped values. Variations of the pace values over successive iterations are depicted in Fig. 6b. In this figure, the maximum of the pace values (in absolute value) across the toll/subsidy roads are plotted. As can be seen, the pace values rapidly converge to zero. The overall convergence of the algorithm is depicted in Fig. 6c in which the relative gaps (in percentage) as well as the Beckmann values are shown. The last column in Table 2 presents the toll/subsidy value—in Canadian dollar unit—in which the value of time is assumed \$16.69 per hour [76]. The maximum toll and subsidy found to attain the target crossing volume are \$4.13 and \$0.26 (twenty-six cents). Given the rapid expansion of communication technology, devices such as GPS are now indispensable gadget. Therefore, communicating a set of diverse rates of toll/subsidy—as is the case in the output—should not be a problem.⁹

⁹ GPS devices nowadays can offer minimum route as well as toll-free routes. So it seems highly plausible to get the GPS devices programed to advice on a variety of routes based on the users budget and toll prices.

4.2.2 Case study of existing network and projected travel demand of 5%

The travel demand is projected to 5% which stands at 59,029.94 vehicle trips. Growth of 5% can be considered as annual or biannual growth of urban traffic, so that one can gain some insight of the annual changes in the toll/subsidy amounts. Similar to Table 3, comparative results of the projected demand scenarios are shown in Table 3. As seen, the total traffic flow passing through the toll/subsidy road increases from 24,407.3 to 25,609.1, which is equivalent to 4.9% growth. The amounts of subsidy and toll also have seen moderate changes which is an endorsement to the stability of the final results.

5 Conclusion

We studied the idea of the TSS for congestion pricing. The element of subsidy is an appealing factor that can greatly compensate some shortcomings related to equity and acceptability in the previous schemes. A fully fledged TSS problem ought to be postulated as a bilevel programming problem which is highly intractable. Nonetheless, in practice, the problem is usually as simple as follows: The location of the tolled (or subsidized) roads is known; it could be a cordon line surrounding a CBD or screen lines bisecting the urban network. Even—based on the traffic survey or historical data—the target traffic volume on the tolled road is also given. Therefore, the question of interest is: What is the best pricing regime, or, which road should be tolled and which ones should be subsidized and how much? To answer this question, given the capped values of the target roads, we first transformed the problem to a conventional CTAP. Second, the Lagrangian values of the respective capacity constraints are interpreted as the toll/subsidy. Third, we relaxed the capacity constraints by moving their Lagrangian values to the respective roads' delay function. In other words, the delay functions receive an additional term to be iteratively updated in such a way the traffic volumes remain below the capped values. We referred to the process as “inflated travel time.” The algorithm was then applied to the real dataset of the Winnipeg, Canada, in which 11 river crossings bridges of the city of Winnipeg were subject to the TSS. The maximum values of toll and subsidy were found to be 4 and 0.2 Canadian dollars.

The subsidy can also be well played as a leverage to address land-use-related problems. The area in the vicinity of subsidized roads can be a magnet for business and job opportunity. The immediate role of the subsidy can be attributed to the fact that it makes the pricing product more

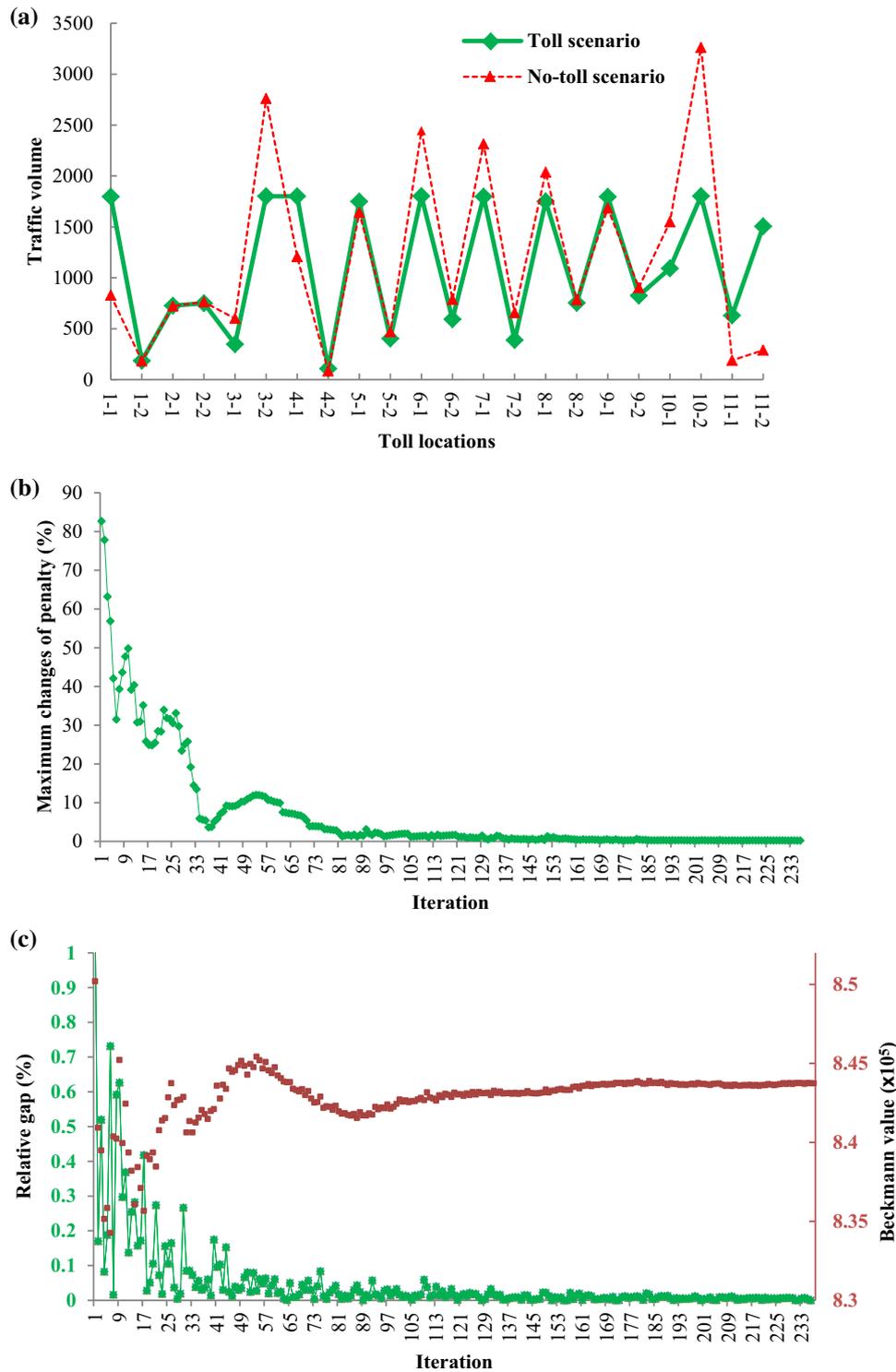


Fig. 6 Winnipeg case study: **a** traffic volume on the screen line before and after TSS; **b** maximum of changes in penalty; **c** convergence behavior

sellable to the public since it brushes off the stain of being another form of taxation.

The concept of the subsidy does not need any payment mechanism. The collected subsidy can be later reimbursed

in many different ways such as toll payment, parking fee, public transport ticket, traffic fine, annual car registration to name a few. It is also easy to protect the system from any mall practices, for instance, total daily value of collected

Table 3 Winnipeg case studies, comparisons of two scenarios: existing demand and projected demand of 5%

Road ID	Direction	Capped traffic volume	Existing network and demand				Existing network and projected demand of 5%			
			Existing traffic volume	Traffic volume of the toll–subsidy	Volume per capped volume (%)	Monetary value of the net toll/subsidy (\$)	Existing traffic volume	Traffic volume of the toll–subsidy	Volume per capped volume (%)	Monetary value of the net toll/subsidy (\$)
1	$i - j$	1800	828.93	1796.92	99.83	0.52	884.15	1897.14	105.4	1.65
	$j - i$	1800	183.76	185	10.28	-0.21	171.15	198.69	11.04	-0.2
2	$i - j$	750	722.32	727.05	96.94	-0.26	693.08	753.73	100.5	-0.22
	$j - i$	750	762.35	750.24	100.03	1.7	738.28	785.66	104.75	3.09
3	$i - j$	1800	600.48	348.27	19.35	-0.07	712.76	362.41	20.13	-0.06
	$j - i$	1800	2759.71	1801.39	100.08	2.33	2894.1	1872.34	104.02	3.55
4	$i - j$	1800	1205.92	1800.78	100.04	1.9	1344.33	1870.29	103.91	3.08
	$j - i$	1800	84.03	108.25	6.01	-0.09	52.04	115.48	6.42	-0.09
5	$i - j$	1750	1641.31	1749.8	99.99	2.51	1583.63	1817.77	103.87	3.95
	$j - i$	1750	469.82	403.75	23.07	-0.12	407.18	421.54	24.09	-0.11
6	$i - j$	1800	2446.11	1802.87	100.16	3.32	2656.02	1832.61	101.81	4.47
	$j - i$	1750	785.81	592.48	33.86	-0.06	937.85	586.84	33.53	-0.06
7	$i - j$	1800	2313.34	1796.73	99.82	3.53	2449.24	1849.18	102.73	4.69
	$j - i$	1800	656.64	389.2	21.62	-0.04	636.55	408.02	22.67	-0.04
8	$i - j$	1750	2035.5	1751.26	100.07	3.19	2104.15	1784.64	101.98	4.71
	$j - i$	1800	783	752.7	41.82	-0.06	826.34	773.17	42.95	-0.06
9	$i - j$	1800	1683.56	1795.59	99.76	3.38	1610.9	1841.02	102.28	4.5
	$j - i$	1800	903.52	824.2	45.79	-0.12	754.71	888.55	49.36	-0.12
10	$i - j$	1800	1547.86	1092.88	60.72	-0.08	1678.71	1110.07	61.67	-0.08
	$j - i$	1800	3261.11	1802.74	100.15	4.13	3444.08	1876.04	104.22	5.02
11	$i - j$	1800	189.06	629.38	34.97	-0.16	228.85	728	40.44	-0.15
	$j - i$	1800	289.45	1505.82	83.66	-0.16	306.38	1835.87	101.99	-0.16
	Sum		26,153.6	24,407.3			27,114.5	25,609.1		
	Min					-0.26				-0.22
	Max					4.13				5.02

subsidies can be limited, and hence, no one can make money by simply driving in the city (a condition to ensure that the problem always has a solution [28]).

In the same spirit, Morosan and Florian [77] have recently solved a distance-based pricing with minimum and maximum toll rates based on which commuters are charged based on vehicle kilometers traveled (VKT). Therefore, integration of the VKT-based TSS can be a worthwhile research.

On other threads for future studies, the followings are worth mentioning. In this study, we assumed that the travel demand is fixed and all drivers have an identical perception about the toll/subsidy. It is, therefore, a worthy attempt to extend the research to multiclass travel demand in which the value of time varies across different sectors of the society. The elasticity of the travel demand to the changes in the pricing regime by itself deserves a further investigation. With the same spirit, consideration of the mutual impact of transit and private modes in the presence of congestion

pricing needs to be further studied. As noted before, in order to subject different classes of cars (e.g., heavy truck vs. passenger cars), different amount of toll/subsidy, the multi-class modeling is inevitable. Furthermore, replacing the link-based traffic assignment with path-based method is a significant boon to the model, such that one can get more insight of the origin destination and the trajectory of the drivers. This information can then be used for planning purpose such as distance-based congestion pricing which is an interested line of research. Investigating the congestion pricing via driving simulator and its applications in the intelligent transportation system to promote the freight modes are some worthy stream of research [78, 79]

Acknowledgements The authors would like to thank Prof Zhao the editor in chief and reviewers for their constructive comments which contributed to improving the quality and the clarity of the final version. The authors would like to express their sincere appreciation to Prof Michael Florian for his contribution to the early version of this research [80].

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Grant-Muller S, Xu M (2014) The role of tradable credit schemes in road traffic congestion management. *Transp Rev* 34(2):128–149
- Hensher DA, Bliemer MC (2014) What type of road pricing scheme might appeal to politicians? Viewpoints on the challenge in gaining the citizen and public servant vote by staging reform. *Transp Res Part A Policy Pract* 61:227–237
- BITRE (2015) Traffic and congestion cost trends for Australian capital cities. Bureau of Infrastructure, Transport and Regional Economics, Information Sheet 74, Canberra
- Alonze S (2015) “Traffic congestion to cost the UK economy more than £300 billion over the next 16 years,” 20/12/2015. <http://inrix.com/press/traffic-congestion-to-cost-the-uk-economy-more-than-300-billion-over-the-next-16-years>
- Lee D Jr, Klein L, Camus G (1999) Induced traffic and induced demand. *Transp Res Rec J Transp Res Board* 1659:68–75
- Bagloee SA, Ceder A, Tavana M, Bozic C (2014) A heuristic methodology to tackle the Braess Paradox detecting problem tailored for real road networks. *Transportmet A Transp Sci* 10(5):437–456
- Braess D, Nagurney A, Wakolbinger T (2005) On a paradox of traffic planning. *Transp sci* 39(4):446–450
- Bagloee SA, Asadi M (2015) Prioritizing road extension projects with interdependent benefits under time constraint. *Transp Res Part A Policy Pract* 75:196–216
- Bagloee SA, Tavana M (2012) An efficient hybrid heuristic method for prioritising large transportation projects with interdependent activities. *Int J Logist Syst Manag* 11(1):114–142
- Pigou AC (1920) *The economics of welfare*. Palgrave Macmillan, London
- Bagloee SA, Asadi M (2015) Heterogeneous parking market subject to parking rationing. *Math Probl Eng* 2015:1–13. doi:10.1155/2015/847324
- Bagloee SA, Asadi M, Richardson L (2012) Methodology for parking modeling and pricing in traffic impact studies. *Transp Res Rec J Transp Res Board* 2319(1):1–12
- Walters AA (1961) The theory and measurement of private and social cost of highway congestion. *Econom J Econom Soc* 29(4):676–699
- Beckmann M (1965) On optimal tolls for highways, tunnels and bridges. In: Edie LC, Herman R, Rothery R (eds) *Vehicular traffic science*. Elsevier, New York, pp 331–341
- Vickrey WS (1969) Congestion theory and transport investment. *Am Econ Rev* 59:251–260
- Sheffi Y (1985) *Urban transportation networks: equilibrium analysis with mathematical programming methods*. Prentice-Hall Inc, Englewood Cliffs
- Bigazzi AY, Figliozzi MA (2013) Marginal costs of freeway traffic congestion with on-road pollution exposure externality. *Transp Res Part A Policy Pract* 57:12–24
- Yang H, Huang H-J (2005) *Mathematical and economic theory of road pricing*. Elsevier Science, Oxford, UK. <http://www.emeraldinsight.com/doi/book/10.1108/9780080456713>
- Verhoef ET (2002) Second-best congestion pricing in general networks. Heuristic algorithms for finding second-best optimal toll levels and toll points. *Transp Res Part B Methodol* 36(8):707–729
- Zhu DL, Yang H, Li CM, et al (2014) Properties of the multiclass traffic network equilibria under a tradable credit scheme. *Transp Sci* 49:519–534
- Ecola L, Light T (2009) Equity and congestion pricing. *Rand Corp* pp 1–45. http://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR680.pdf
- Daganzo CF, Lehe LJ (2015) Distance-dependent congestion pricing for downtown zones. *Transp Res Part B Methodol* 75:89–99
- Liu B, Huang C (2014) Alternative tradable credit scheme on general transportation network. *Wuhan Univ J Nat Sci* 19(3):193–198
- Wu D, Yin Y, Lawphongpanich S, Yang H (2012) Design of more equitable congestion pricing and tradable credit schemes for multimodal transportation networks. *Transp Res Part B Methodol* 46(9):1273–1287
- Chu Z, Cheng L, Chen H, et al (2014) Tradable credit schemes in managing network mobility: brief review and new developments. In: *CICTP 2014: Safe, smart, and sustainable multimodal transportation systems*, Changsha, China, 4–7 July 2014, pp 3388–3400. doi:10.1061/9780784413623.325
- Bagloee SA, Heshmati M, Tavana M, et al (2017) A logit-based model for measuring the effects of transportation infrastructure on land value. *Transp Plan Tech* 40(2):143–166
- Geroliminis N, Daganzo CF (2008) Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transp Res Part B Methodol* 42(9):759–770
- Chen L, Yang H (2012) Managing congestion and emissions in road networks with tolls and rebates. *Transp Res Part B Methodol* 46(8):933–948
- Bagloee SA, Sarvi M (2015) Heuristic approach to capacitated traffic assignment problem for large-scale transport networks. *Transp Res Rec J Transp Res Board* 2498:1–11
- Tsekeris T, Voß S (2009) Design and evaluation of road pricing: state-of-the-art and methodological advances. *Netnomics* 10(1):5–52
- de Palma A, Lindsey R (2011) Traffic congestion pricing methodologies and technologies. *Transp Res Part C Emerg Technol* 19(6):1377–1399
- Yang H, Wang X (2011) Managing network mobility with tradable credits. *Transp Res Part B Methodol* 45(3):580–594
- Wang X, Yang H, Zhu D, Li C (2012) Tradable travel credits for congestion management with heterogeneous users. *Transp Res Part E Logist Transp Rev* 48(2):426–437
- Nie YM (2012) Transaction costs and tradable mobility credits. *Trans Res Part B Methodol* 46(1):189–203
- Zhang X, Yang H, Huang H-J (2011) Improving travel efficiency by parking permits distribution and trading. *Trans Res Part B Methodol* 45(7):1018–1034
- Bernstein D (1993) Congestion pricing with tolls and subsidies. In: *Pacific rim trans tech conference—volume II: international ties, management systems, propulsion technology, strategic highway research program*, Seattle, Washington, USA, July 25–28, pp 145–151
- Adler JL, Cetin M (2001) A direct redistribution model of congestion pricing. *Trans Res Part B Methodol* 35(5):447–460
- Guo X, Yang H (2010) Pareto-improving congestion pricing and revenue refunding with multiple user classes. *Transp Res Part B* 44(8):972–982. <http://scholar.uwindsor.ca/odettepub/50>
- Lawphongpanich S, Yin Y (2010) Solving the Pareto-improving toll problem via manifold suboptimization. *Transp Res Part C Emerg Technol* 18(2):234–246

40. Wu D, Yin Y, Lawphongpanich S (2011) Pareto-improving congestion pricing on multimodal transportation networks. *Eur J Op Res* 210(3):660–669
41. Ben-Elia E, Ettema D (2011) Rewarding rush-hour avoidance: a study of commuters' travel behavior. *Transp Res Part A Policy Pract* 45(7):567–582
42. Maillé P, Stier-Moses NE (2009) Eliciting coordination with rebates. *Transp Sci* 43(4):473–492
43. Kockelman KM, Kalmanje S (2005) Credit-based congestion pricing: a policy proposal and the public's response. *Transp Res Part A Policy Pract* 39(7):671–690
44. Hearn DW, Ramana MV (1998) Solving congestion toll pricing models. In: Marcotte P, Nguyen S (eds) *Equilibrium and advanced transportation modeling*. Kluwer Academic Publishers, Springer, Boston, pp 109–124
45. Eliasson J (2001) Road pricing with limited information and heterogeneous users: a successful case. *Ann Reg Sci* 35(4):595–604
46. Liu Y, Guo X, Yang H (2009) Pareto-improving and revenue-neutral congestion pricing schemes in two-mode traffic networks. *Netnomics Econ Res Electron Netw* 10(1):123–140
47. Nie Y, Liu Y (2010) Existence of self-financing and Pareto-improving congestion pricing: impact of value of time distribution. *Transp Res Part A Policy Pract* 44(1):39–51
48. Xiao F, Zhang H (2014) Pareto-improving toll and subsidy scheme on transportation networks. *EJTIR* 14(1):46–65
49. Patriksson P (1994) *The traffic assignment problem: models and methods*. VSP BV, The Netherlands. Facsimile reproduction published in 2014 by Dover Publications, Inc., Mineola, New York
50. Marcotte P, Patriksson M (2007) Traffic equilibrium. *Handb Op Res Manag Sci* 14:623–713
51. Larsson T, Patriksson M (1995) An augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems. *Trans Res Part B Methodol* 29(6):433–455
52. Nie Y, Zhang H, Lee D-H (2004) Models and algorithms for the traffic assignment problem with link capacity constraints. *Trans Res Part B Methodol* 38(4):285–312
53. Inouye H (1986) Traffic equilibria and its solution in congested road networks. In: *The 5th IFACIFIPFORS conference on control in transportation systems*, Vienna, Austria.
54. Hearn D (1980) Bounding flows in traffic assignment models. In: *Technical report, vol. Research Report 80-4, no. Department of Industrial and Systems Engineering*, University of Florida, Gainesville
55. Shahpar AH, Aashtiani HZ, Babazadeh A (2008) Dynamic penalty function method for the side constrained traffic assignment problem. *Appl Math Comput* 206(1):332–345
56. Marcotte P, Nguyen S, Schoeb A (2004) A strategic flow model of traffic assignment in static capacitated networks. *Op Res* 52(2):191–212
57. Yang H, Yagar S (1994) Traffic assignment and traffic control in general freeway-arterial corridor systems. *Trans Res Part B Methodol* 28(6):463–486
58. Yang H, Yagar S (1995) Traffic assignment and signal control in saturated road networks. *Transp Res Part A Policy Pract* 29(2):125–139
59. BPR, Bureau of Public Roads (BPR) (1964) *Traffic assignment manual*. Department of Commerce, Urban Planning Division, Washington D.C, p 1964
60. Larsson T, Patriksson M (1999) Side constrained traffic equilibrium models—analysis, computation and applications. *Trans Res Part B Methodol* 33(4):233–264
61. Zhou B, Bliemer M, Yang H, He J (2015) A trial-and-error congestion pricing scheme for networks with elastic demand and link capacity constraints. *Trans Res Part B Methodol* 72:77–92
62. Meng Q, Xu W, Yang H (2005) Trial-and-error procedure for implementing a road-pricing scheme. *Transp Res Rec J Transp Res Board* 1923(1):103–109
63. Yang H, Meng Q, Lee D-H (2004) Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions. *Trans Res Part B Methodol* 38(6):477–493
64. Yang H, Xu W, He B-S, Meng Q (2010) Road pricing for congestion control with unknown demand and cost functions. *Transp Res Part C Emerg Technol* 18(2):157–175
65. Beckmann M, McGuire C, Winsten CB (1956) *Studies in the economics of transportation*. Yale University Press, New Haven
66. Zhong RX, Sumalee A, Friesz TL, Lam WHK (2011) Dynamic user equilibrium with side constraints for a traffic network: theoretical development and numerical solution algorithm. *Trans Res Part B Methodol* 45(7):1035–1061
67. Boyce D, Ralevic-Dekic B, Bar-Gera H (2004) Convergence of traffic assignments: how much is enough? *J Transp Eng* 130(1):49–55
68. Dial RB (2006) A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Trans Res Part B Methodol* 40(10):917–936
69. Larsson T, Patriksson M, Strömberg A-B (1996) Conditional subgradient optimization—theory and applications. *Eur J Op Res* 88(2):382–403
70. Gustavsson E, Patriksson M, Strömberg A-B (2015) Primal convergence from dual subgradient methods for convex optimization. *Math Program* 150(2):365–390
71. Anstreicher KM, Wolsey LA (2009) Two “well-known” properties of subgradient optimization. *Math Program* 120(1):213–220
72. INRO (2009) “EMME3 v 3.2”, EMME3 user's guide. Montreal, Quebec, Canada
73. Bar-Gera H (2016) Transportation network test problems. <http://www.bgu.ac.il/~bargera/tntp/>
74. Ortigosa J, Menendez M, Tapia H (2014) Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. *EURO J Transp Logist* 3(3–4):245–266
75. Zheng N, Waraich RA, Axhausen KW, Geroliminis N (2012) A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transp Res Part A Policy Pract* 46(8):1291–1303
76. HDR (2015) *Current and projected costs of congestion in metro vancouver: final report*, TransLink Canada
77. Morosan CD, Florian M (2015) A network model for capped link-based tolls. *EURO J Transp Logist*. doi:10.1007/s13676-015-0078-4
78. Sarvi M, Kuwahara M (2008) Using ITS to improve the capacity of freeway merging sections by transferring freight vehicles. *IEEE Trans Intell Transp Sys* 9(4):580–588
79. Sarvi M, Kuwahara M, Ceder A (2004) Freeway ramp merging phenomena in congested traffic using simulation combined with a driving simulator. *Comput Aided Civil Infrastruct Eng* 19(5):351–363
80. Bagloee SA, Florian M, Sarvi M (2016) A new policy in congestion pricing: why only toll? why not subsidy? Interuniversity research centre on enterprise networks, logistics and transportation (CIRRELT). Montreal, Canada. <https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2016-19.pdf>