Trial-and-error operation schemes for bimodal transport systems

Ren-Yong Guo[a], W.Y. Szeto[b,*], Jiancheng Long[c]

[a] *School of Economics and Management, Beihang University, Beijing 100191, China; and Key Lab of Complex System Analysis and Management Decision, Ministry of Education, China*

[b] *Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong*

[c] *School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei 230009, China*

Abstract. We concern the modal choice of commuters in a transport system comprising a highway, which is only used by autos, in parallel to a transit line, which is only used by buses. In the transport system, the in-vehicle congestion of passengers in bus carriages is treated as a negative externality cost of affecting the modal choice of commuters and commuters choose their travel modes according to the perceived travel costs of transport modes. We propose two trial-and-error operation schemes for the transport system without resorting to both the function of in-vehicle congestion costs and the distribution of perceived travel cost errors. In the first operation scheme, the manager (or the government) determines the transit fare charged from (or financial subsidy to) bus users from period to period so as to minimize the system time cost of the transport system. The second operation scheme is established from the viewpoint of a private firm that operates the public transit line. The operator determines the transit fare and bus run frequency from period to period in order to maximize its operating profit. Moreover, we demonstrate the effectiveness of the two operation schemes for optimizing the system time cost and the operating profit by both theoretical analyses and numerical examples.

Keywords: Bimodal transport; trial-and-error operation scheme; system time optimization; operating profit optimization

1. Introduction

In most cities around the world, multiple transport modes exist to provide substitutable

---

[*] Corresponding author. Tel.: +852 28578552.

E-mail addresses: buaa_guorenyong@126.com (R.-Y. Guo), ceszeto@hku.hk (W.Y. Szeto), jianchenglong@hfut.edu.cn (J. Long).

transportation services for people, e.g., people can choose either private cars or public transit to travel from their origins to their destinations. Some of these cities, such as London, Singapore, Milan, and Stockholm, further implement toll charging to manage travel demand.

So far, a number of studies have concerned travelers' modal choices and toll charge regimes in multi-modal transport systems, which comprise highways (used by private cars) and transit lines (operated by public transit companies). These existing studies mainly focus on the planning and operation of multi-modal transport systems and address two problems. The first problem is how the manager of a multi-modal transport system sets the decision variables in the system (e.g., transit fares, transit capacities, auto toll charges, and bus run frequencies) so as to cut down or minimize the total social cost of the system (e.g., Huang, 2000, 2002; Kraus, 2003; Gonzales and Daganzo, 2012; Li et al., 2012; David and Foucart, 2014; Wu and Huang, 2014; Gonzales, 2015; Liu et al., 2016) or maximize consumer surplus or government surplus (e.g., Arnott and Yan, 2000; Pels and Verhoef, 2007; Ahn, 2009; van den Berg and Verhoef, 2014). The second problem is, when the public transit lines in a multi-modal transport system are operated by one or more private firms, how the private firms determine bus fares or bus run frequencies in order to improve or maximize their operating profits (e.g., Pels and Verhoef, 2007; Li et al., 2012; van der Weijde et al., 2013; Wu and Huang, 2014; Zhang et al., 2014; Li and Yang, 2016). Some of the existing studies also examine the responses of commuters or service operators to various policies implemented in multi-modal transport systems, e.g., the first-best pricing for system optimum (Huang, 2002), the second-best pricing (Arnott and Yan, 2000), auto mode underpricing (Kraus, 2003; Ahn, 2009), the location-based or distance-based pricing (Li et al., 2012), the flat toll charge (Wu and Huang, 2014), parking reservation (Liu et al., 2016), and the provision of traffic forecasts by smart transport information providers/agencies (Liu et al., 2017). For a comprehensive review of transport pricing in multi-modal transport systems, the readers may refer to the study of Tirachini and Hensher (2012).

When public transit modes are involved in a transport system, the in-vehicle congestion in transit carriages should be considered (de Palma et al., 2017). On one hand, mass transit may bring passengers discomfort generated by the in-vehicle congestion in transit carriages if passengers are crowded in carriages. On the other hand, the in-vehicle congestion also leads to the loss of independence and privacy of passengers (Huang, 2000, 2002). Therefore, the in-vehicle congestion in carriages can be treated as a class of negative externality costs that significantly affect the modal choices of travelers between transit services and other transport modes.

The public transport crowding was modeled in some theoretical and case studies. In the studies of Huang (2000) and van den Berg and Verhoef (2014), the in-vehicle congestion cost of passengers is assumed to be a linear function of the number of travelers, who use the bus mode, to generate some analytical results. In the studies of Huang (2002), Huang et al. (2007), Parry and Small (2009), Li et al. (2012), and Wu and Huang (2014), the in-vehicle congestion cost of passengers in transit carriages is assumed to be an increasing function of the number of travelers selecting the public transit mode. Parry and Small (2009) considered in-vehicle crowding costs as a dimension of a bi-modal problem in their theoretical framework analyzing the optimal level of public transport subsidies. Pel et al. (2014) introduced the effect of in-vehicle crowding into the train passenger assignment model of the Dutch National Model System and captured travelers' behavioral response to in-vehicle crowding.

Recently, public transport crowding was also studied from an empirical viewpoint. For instance, Li and Hensher (2011) identified three measures used to value crowding and associated ways of representing crowding in stated preference experiments, by using a number of primary studies conducted in the UK, the USA, Australia, and Israel. Wardman and Whelan (2011) reviewed evidence from the British experience of the valuation of rail crowding obtained over 20 years from 17 studies. Using the original survey data from Paris, Haywood and Koning (2015) assessed the distribution of comfort costs of congestion in public transport and applied their results to the cost-benefit analysis of a Parisian public transport project.

To more realistically formulate the modal choice of commuters in a multi-modal transport system, the stochastic user equilibrium (SUE) model is generally adopted to prescribe that commuters choose their travel modes according to the perceived travel costs of all travel modes rather than the actual travel costs. The perceived errors of commuters for the travel cost of a mode are explained as the differences among travel costs perceived by different commuters and result from the coincidence of plenty of factors that affect commuters' modal choice, e.g., commuters' values of time (VOT), commuters' tolerance degrees for in-vehicle congestion in carriages, commuters' inertia for choosing a travel mode, the amount of travel cost information known by commuters, the possession of private cars, and so on. A commuter with a smaller VOT may think that the perceived travel cost of one mode with a higher travel time and a lower toll charge is less than the perceived travel cost of another mode with a lower travel time and a higher toll charge, and hence he/she may choose to use the former travel mode. A commuter, who has no information regarding the travel costs of all modes, may choose his/her travel mode randomly. A commuter, who has no private car, has to choose to travel by public transit. A commuter, who has a private car, may always choose to travel by the private car mode due to inertia or aversion to in-vehicle congestion in bus carriages.

The perceived travel cost errors can be modeled by different distributions. By assuming that the perceived travel cost errors follow independently and identically distributed (i.i.d.) Gumbel distributions, the modal choice of commuters at a stationary state can be formulated as a logit-based discrete choice model (Huang, 2002; Huang et al., 2007). Under the assumption that the perceived travel cost errors follow normal distributions, the modal choice of commuters at a stationary state can be formulated as a probit-based discrete choice model (Sheffi, 1985). The logit model has been commonly adopted due to its ease of use. Through a binary logit model, Cantarella et al. (2015) formulated the joint adjustment of modal choice and transit operation from day to day in a bi-modal transport system, in which the frequency of bus runs is prefixed to meet the demand with all the buses available or is daily updated to meet the demand with the minimum number of buses needed to avoid oversaturation. By assuming that commuters choose their travel modes according to the perceived travel costs of travel modes and that the perceived travel cost errors follow i.i.d. Gumbel distributions, Liu and Geroliminis (2017) modeled and controlled a multi-region and multi-modal transportation system, in which both the day-to-day and within-day traffic dynamics are involved, and developed an adaptive mechanism to update parking pricing from period to period so as to improve the system's efficiency.

Some studies of multi-modal transport systems adopted a general distribution of the perceived travel cost errors to account for various realistic situations where the logit and probit assumptions are not satisfied. For instance, David and Foucart (2014) studied the choice of transportation modes within a city, where commuters have heterogeneous preferences for the car mode, and examined two policy tools, i.e., taxation and traffic separation. Li and Yang (2016) investigated travelers' day-to-day modal choice in a bi-modal transportation system with responsive transit services under various economic objectives. Guo and Szeto (2018) designed a control strategy to control the day-to-day modal choice of commuters in a bi-modal transportation system so as to simultaneously reduce the daily total travel cost of the transportation system and achieve a Pareto improvement or zero-sum revenue target.

Some of the aforementioned studies of multi-modal transport systems involved the generalized travel cost functions of public transit passengers that commonly consist of travel time cost, waiting time cost, in-vehicle congestion cost, and so on. To obtain these generalized travel cost functions, not only the in-vehicle congestion cost function but also the coefficient of the in-vehicle congestion cost need to be known first. However, *it is generally difficult to obtain the in-vehicle congestion cost function and the coefficient*. Moreover, *it is also difficult to obtain the distribution function of perceived travel cost errors and the calibrated parameters in the distribution function*. Thus, we bring up a question, i.e., whether an operating scheme for a multi-modal transport system without resorting to the function of in-vehicle congestion costs and the distribution of perceived travel cost errors can be developed to optimize the system time cost or operating profit of the transport system. To the best of our knowledge, the question has not been addressed in these existing studies with respect to travelers' modal choice and toll charge regimes in multi-modal transport systems.

The methodology for solving our investigated problem relates to the trial-and-error methods proposed by a series of papers for optimizing the system costs of transport networks or restricting the flows on links below desirable upper bounds with unknown demand functions or travel cost functions. For example, Li (2002) proposed a bisection iterative procedure in deriving the congestion toll for system optimum for a single road in the absence of the demand function. Wang and Yang (2012) demonstrated the non-convergence of the bisection iterative procedure by Li (2002) and modified the iterative procedure to guarantee its convergence. Yang et al. (2004) and Han and Yang (2009) developed a class of trial-and-error implementation schemes of marginal-cost pricing on a general road network when the demand functions are unknown. Yang et al. (2010) presented a convergent trial-and-error implementation method for finding a set of effective link toll patterns to reduce link flows to below a desirable target level when both the link travel time functions and demand functions are unknown. Zhou et al. (2015) proposed a trial-and-error congestion pricing scheme that not only considers the minimization of the total system cost but also addresses the capacity constraints. Wang et al. (2018) proposed a trial-and-error fare design scheme to alleviate the boarding/alighting congestion of commuters at train stations to a certain level. However, *these trial-and-error methods, mentioned above, cannot be directly applied to solve the proposed operating problem in a multi-modal transport system with unknown in-vehicle congestion cost functions and perceived travel cost error distributions*. It is thus essential to design and develop new trial-and-error procedures to solve the proposed operating problem.

In this paper, we propose two trial-and-error operation schemes for a bimodal transport system, which comprises a highway, which is only used by autos, in parallel to a transit line, which is only used by buses. In the first operation scheme, the manager of the transport system sets the transit fare charged from (or financial subsidy to) bus users during each period according to the reaction of commuters to the bus fare (or subsidy) in the previous period so as to minimize the system time cost of the transport system. In the second operation scheme, the operator of the public transit line sets the bus fare and bus run frequency from period to period in order to optimize its operating profit. In both the operation schemes, both the function of in-vehicle congestion costs and the distribution of perceived travel cost errors are not needed. Moreover, we theoretically demonstrate that the two operation schemes are effective for optimizing the system time cost and the operating profit. The contributions of this paper are as follows: (1) This paper introduces a new operating problem in a bimodal transport system with unknown in-vehicle congestion cost functions and perceived travel cost error distributions, and a variant with a different objective. (2) This paper proposes a novel convergent trial-and-error operation scheme for each of the two operating problems.

The remainder of this paper is organized as follows. In the next section, the trial-and-error operation scheme for minimizing the system time cost and its several properties are presented. In Section 3, we propose the trial-and-error operation scheme for maximizing the profit of operating a public transit line. Several numerical examples are given to show the effectiveness of the two operation schemes in Section 4. Section 5 concludes this paper.

2. Trial-and-error implementation for system time optimization

2.1. System description

We consider a bi-modal transport network, which comprises an origin-destination (OD) pair connected by a transit line in parallel to a highway. In every morning, a fixed number d (>0) of commuters travel from the origin to the destination. The two transport modes are separated. Commuters can choose to travel by bus running on the transit line or auto running on the highway, i.e., they have two discrete choices. The number of bus users on the transit line is denoted by x_b (≥ 0) and the number of auto users on the highway is then $d-x_b$ (≥ 0).

Let y (≥ 0) be the frequency of bus runs on the transit line. $w(x_b, y)$ (≥ 0) denotes the average waiting time cost of bus users at the bus stop on the transit line. The function w is twice-continuously differentiable with respect to (x_b, y). Meanwhile, it is assumed that $\partial w(x_b, y)/\partial x_b > 0$, $\partial w(x_b, y)/\partial y < 0$, and $\partial^2 w(x_b, y)/\partial x_b^2 \geq 0$. That is to say, the waiting time cost of bus users is positively (negatively) proportion to the number of bus users (the frequency of bus runs), and the function w is increasing and convex (decreasing) with respect to x_b (y). Let t_b (>0) stand for the average in-vehicle travel time cost of bus users on the transit line. The notation $g(x_b, y)$ (≥ 0) denotes the average in-vehicle congestion cost of passengers in bus carriages and it reflects the discomfort generated by in-vehicle congestion, which has a significant effect on the choices of commuters between transit services and other transport modes (Huang, 2000, 2002; Huang et al., 2007; Parry and Small, 2009; Li et al., 2012; Pel et al., 2014; van den Berg and Verhoef, 2014; Wu and Huang, 2014). The function g is continuously differentiable with respect to (x_b, y). It is assumed that $\partial g(x_b, y)/\partial x_b > 0$ and $\partial g(x_b, y)/\partial y < 0$, i.e., the in-vehicle congestion cost increases (decreases) as the number of bus users (the frequency of bus runs) increases.

For simplicity, the occupancy of each auto is assumed to be one. $t_a(x_b)$ (>0) represents the average travel time cost of auto users on the highway, including both the free flow travel cost and the road congestion cost. The function t_a is twice-continuously differentiable with respect to x_b. Moreover, it is supposed that $\mathrm{d}t_a(x_b)/\mathrm{d}x_b < 0$ and $\mathrm{d}^2 t_a(x_b)/\mathrm{d}x_b^2 \geq 0$. This implies that the function t_a is increasing and convex with respect to the number of auto users $d - x_b$. After all, a higher number of auto users generate more congestion for auto users because there are more autos on the highway.

The notation p_b stands for the transit fare charged from (or financial subsidy to) each bus user on the transit line. A positive (negative) p_b-value represents a toll charge (financial subsidy). p_a (≥ 0) is the toll charge from each auto user on the highway. There are two travel modes and also the total travel demand is fixed in the transport system. Thus, to achieve the system time optimization objective, it is enough to adjust the transit fare (or financial subsidy) p_b for bus users while setting the toll charge p_a for auto users as a fixed value. Thus, it is assumed that the toll charge p_a is constant. All those costs and prices, mentioned above, are measured in monetary units.

Commuters choose their travel modes according to the perceived generalized travel costs of the two modes. The perceived generalized travel costs $c_b(x_b, y, p_b)$ and $c_a(x_b)$ of commuters using buses on the transit line and using autos on the highway are respectively formulated as

$$c_b(x_b, y, p_b) = w(x_b, y) + t_b + g(x_b, y) + p_b + \xi_b \quad \text{and} \tag{1}$$

$$c_a(x_b) = t_a(x_b) + p_a + \xi_a, \tag{2}$$

where the terms $w(x_b, y) + t_b + g(x_b, y) + p_b$ and $t_a(x_b) + p_a$ are the actual generalized travel costs of bus and auto users, respectively, and ξ_b and ξ_a are two random error terms. The random error terms prescribe the differences between travel costs perceived by different commuters. Let ε_b and ε_a be two realizations of the two random variables ξ_b and ξ_a, respectively. Let $\xi = \xi_b - \xi_a$ be the difference between ξ_b and ξ_a and $\varepsilon = \varepsilon_b - \varepsilon_a$ be a realization of the random variable ξ. The random variable ξ follows a distribution with the probability density function $\varepsilon \mapsto f(\varepsilon)$ and the distribution function $\varepsilon \mapsto F(\varepsilon)$ over the support $(-\infty, +\infty)$. The function f is continuous with respect to ε, and hence F is related to f by $F'(\varepsilon) = f(\varepsilon)$. In addition, it is assumed that $f(\varepsilon) > 0$ holds for any ε.

At a stochastic user equilibrium (SUE) state, no commuter can reduce his/her perceived travel cost by unilaterally altering his/her travel mode (Sheffi, 1985). Mathematically, the SUE condition can be formulated as

$$x_b = d \Pr\left(c_b(x_b, y, p_b) < c_a(x_b)\right) = d \Pr\left(\xi < \Delta(x_b, y, p_b)\right) = dF\left(\Delta(x_b, y, p_b)\right), \tag{3}$$

where

$$\Delta(x_b, y, p_b) = t_a(x_b) + p_a - w(x_b, y) - t_b - g(x_b, y) - p_b. \tag{4}$$

2.2. Optimization of total system time cost

Given a frequency of bus runs y on the transit line, the optimization problem of minimizing the total system time cost of the transport system is formulated as

$$\min_{0 \le x_b \le d} V = x_b\left(w(x_b, y) + t_b\right) + (d - x_b)t_a(x_b). \tag{5}$$

The total system time cost V is the sum of the total waiting and travel time cost $x_b\left(w(x_b, y) + t_b\right)$ of all bus passengers and the total travel time cost $(d - x_b)t_a(x_b)$ of all auto users.

The feasible set $[0, d]$ of the decision variable x_b is nonempty, compact, and convex. The system time cost V is continuous with respect to x_b. Furthermore, the second derivative of the system time cost V satisfies

$$\frac{\mathrm{d}^2 V}{\mathrm{d}x_b^2} = 2\frac{\partial w(x_b, y)}{\partial x_b} + x_b \frac{\partial^2 w(x_b, y)}{\partial x_b^2} - 2\frac{dt_a(x_b)}{dx_b} + (d - x_b)\frac{\mathrm{d}^2 t_a(x_b)}{\mathrm{d}x_b^2} > 0,$$

due to $\partial w(x_b, y)/\partial x_b > 0$, $\partial^2 w(x_b, y)/\partial x_b^2 \ge 0$, $\mathrm{d}t_a(x_b)/\mathrm{d}x_b < 0$, and $\mathrm{d}^2 t_a(x_b)/\mathrm{d}x_b^2 \ge 0$. Thus, V is strictly convex with respect to x_b. It follows that the globally optimal solution to the optimization problem (5) exists and also is unique. Meanwhile, x_b^* is the globally optimal solution (or globally minimum point) to the optimization problem (5) if and only if it satisfies the following Karush–Kuhn–Tucker (KKT) conditions:

$$w(x_b, y) + t_b + x_b \frac{\partial w(x_b, y)}{\partial x_b} - t_a(x_b) + (d - x_b)\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} = \mu_1 - \mu_2, \tag{6}$$

$$\mu_1 \ge 0,\ x_b \ge 0,\ \mu_1 x_b = 0,\ \mu_2 \ge 0,\ d - x_b \ge 0, \text{ and } \mu_2(d - x_b) = 0, \tag{7}$$

where μ_1 and μ_2 are the Lagrange multipliers associated with the constraints $x_b \geq 0$ and $d - x_b \geq 0$, respectively.

It is assumed that the globally optimal solution x_b^* is in the interior of the feasible set $[0, d]$, i.e., both $x_b^* > 0$ and $d - x_b^* > 0$ hold. Then, the KKT conditions (6) and (7) can be further reduced as

$$w(x_b, y) + t_b + x_b \frac{\partial w(x_b, y)}{\partial x_b} - t_a(x_b) + (d - x_b)\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} = 0. \tag{8}$$

2.3. Trial-and-error procedure

In the transport system, the in-vehicle congestion in bus carriages is considered to be a negative externality cost of affecting the modal choice of commuters. Meanwhile, commuters choose their travel modes according to the perceived travel costs of travel modes and the modal split at a stationary state satisfies the SUE condition (3). How does the manager of the transport system set the transit fare (or financial subsidy) for bus users, without resorting to the function of in-vehicle congestion costs and the distribution of perceived travel cost errors, so as to achieve a system time optimization objective? To answer this question, we propose a trial-and-error implementation of the bus fare (or subsidy) for the optimization objective.

It is worth mentioning that the trial-and-error methods in existing studies, e.g., Li (2002), Yang et al. (2004), Han and Yang (2009), Yang et al. (2010), Wang and Yang (2012), Zhou et al. (2015), and Wang et al. (2018), cannot be directly applied to solve the operating problem in our study because the implementation objectives and preconditions in those existing studies and in our study are different. In those existing studies, the system cost of a transport network is minimized under the precondition that the demand functions of travelers are unknown or the traffic flows on a set of links are restricted below desirable upper bounds under the precondition that the demand functions and travel cost functions of travelers are unknown in general. However, in our study, the system time cost of a transport system is minimized and the operating profit of a transit operator is maximized without resorting to both the function of in-vehicle congestion costs and the distribution of perceived travel cost errors (in addition, the social/system cost, including not only waiting and travel time cost but also in-vehicle congestion cost and transit operating cost, is optimized without resorting to the function of in-vehicle congestion costs in the appendix of this paper).

Assuming a fixed frequency of bus runs y, we summarize the iterative trial-and-error procedure for determining the transit fare (or subsidy) as follows.

Step 1. Let $p_b^{(0)}$ be the initial transit fare (or subsidy) for bus users, set an initial iterative step size α_0 (>0) and a convergence tolerance $\tilde{\varepsilon}$ (>0), and set $n=0$.

Step 2. Observe the revealed number of bus users $x_b^{(n)}$ at the equilibrium state after the imposition of the transit fare (or subsidy) $p_b^{(n)}$.

Step 3. Set the transit fare (or subsidy) $\overline{p}_b^{(n)} = p_b^{(n)} + \Delta p$ and then observe the revealed number of bus users $\overline{x}_b^{(n)}$ at the equilibrium state.

Step 4. Compute the iterative direction $G_p^{(n)}$ of the transit fare (or subsidy) according to

$$G_p^{(n)} = \left(w\left(x_b^{(n)}, y\right) + t_b + x_b^{(n)} \frac{\partial w\left(x_b^{(n)}, y\right)}{\partial x_b^{(n)}} - t_a\left(x_b^{(n)}\right) + \left(d - x_b^{(n)}\right) \frac{\mathrm{d}t_a\left(x_b^{(n)}\right)}{\mathrm{d}x_b^{(n)}} \right) \frac{x_b^{(n)} - \overline{x}_b^{(n)}}{\Delta p}. \quad (9)$$

Step 5. If $n < n_0$, then set $\alpha^{(n)} = \alpha_0$; otherwise, set $\alpha^{(n)} = \alpha^{(n-1)}/2$.

Step 6. Update the transit fare (or subsidy) by the following formula:

$$p_b^{(n+1)} = \begin{cases} p_b^{(n)} - \alpha^{(n)}, & \text{if } G_p^{(n)} < 0, \\ p_b^{(n)} + \alpha^{(n)}, & \text{otherwise.} \end{cases} \quad (10)$$

Step 7. If $\left| p_b^{(n+1)} - p_b^{(n)} \right| < \tilde{\varepsilon}$, then stop; otherwise, set $n = n+1$ and go to Step 2.

We call the above trial-and-error implementation the trial-and-error system time optimization (TESTO) procedure. In the TESTO procedure, the difference parameter Δp is positive and sufficiently small. The convergence tolerance $\tilde{\varepsilon}$ is sufficiently small. Step 2 (or Step 3) means that, under the implementation of the transit fare (or subsidy) $p_b^{(n)}$ (or $\overline{p}_b^{(n)}$), the number of bus users $x_b^{(n)}$ (or $\overline{x}_b^{(n)}$), which satisfies the SUE condition (3), is observed. Obviously, in the TESTO procedure, the function of in-vehicle congestion costs and the distribution of perceived travel cost errors are not involved, i.e., the manager need not know the function of congestion costs and the distribution of perceived travel cost errors to implement the TESTO procedure.

To show how the TESTO procedure takes effect, we first prove that x_b is a continuously differentiable function of p_b and y in the SUE condition (3).

Proposition 1. In the SUE condition (3), x_b is a continuously differentiable function of p_b and y and the partial derivatives of x_b with respect to p_b and y can be governed by

$$\frac{\partial x_b}{\partial p_b} = \frac{-df\left(\Delta(x_b, y, p_b)\right)}{1 + df\left(\Delta(x_b, y, p_b)\right)\left(-\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} + \frac{\partial w(x_b, y)}{\partial x_b} + \frac{\partial g(x_b, y)}{\partial x_b}\right)} \quad \text{and} \quad (11)$$

$$\frac{\partial x_b}{\partial y} = \frac{-df\left(\Delta(x_b, y, p_b)\right)\left(\frac{\partial w(x_b, y)}{\partial y} + \frac{\partial g(x_b, y)}{\partial y}\right)}{1 + df\left(\Delta(x_b, y, p_b)\right)\left(-\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} + \frac{\partial w(x_b, y)}{\partial x_b} + \frac{\partial g(x_b, y)}{\partial x_b}\right)}. \quad (12)$$

Proof. We first prove that x_b is a function of p_b and y. Let a function H be formulated as

$$H(x_b, p_b, y) = x_b - dF\left(\Delta(x_b, y, p_b)\right). \tag{13}$$

Due to $f(\varepsilon) > 0$ for any $\varepsilon \in (-\infty, +\infty)$, it is concluded that

$$H(0, p_b, y) = 0 - dF\left(\Delta(0, y, p_b)\right) < 0 \text{ and } H(d, p_b, y) = d - dF\left(\Delta(d, y, p_b)\right) > 0,$$

for any (p_b, y). On the other hand, the partial derivative of the function H with respect to x_b satisfies

$$\frac{\partial H(x_b, p_b, y)}{\partial x_b} = 1 + df\left(\Delta(x_b, y, p_b)\right)\left(-\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} + \frac{\partial w(x_b, y)}{\partial x_b} + \frac{\partial g(x_b, y)}{\partial x_b}\right) > 0, \tag{14}$$

owing to $f\left(\Delta(x_b, y, p_b)\right) > 0$, $\mathrm{d}t_a(x_b)/\mathrm{d}x_b < 0$, $\partial w(x_b, y)/\partial x_b > 0$, and $\partial g(x_b, y)/\partial x_b > 0$. That is to say, the function H is increasing with respect to x_b. In addition, the function H is continuous with respect to x_b. Thus, given any (p_b, y), there is a unique x_b, which satisfies $H(x_b, p_b, y) = 0$, or equivalently, x_b is a function of p_b and y.

We then prove that the function $(p_b, y) \mapsto x_b$ determined by the SUE condition is continuously differentiable. On one hand, the partial derivatives of the function H with respect to p_b and y are formulated as

$$\frac{\partial H(x_b, p_b, y)}{\partial p_b} = df\left(\Delta(x_b, y, p_b)\right) \text{ and} \tag{15}$$

$$\frac{\partial H(x_b, p_b, y)}{\partial y} = df\left(\Delta(x_b, y, p_b)\right)\left(\frac{\partial w(x_b, y)}{\partial y} + \frac{\partial g(x_b, y)}{\partial y}\right). \tag{16}$$

In formulae (14) to (16), the function f is continuous and the functions t_a, w, and g are continuously differentiable. Thus, the function H is continuously differentiable with respect to (x_b, p_b, y). On the other hand, $\partial H(x_b, p_b, y)/\partial x_b \neq 0$ holds. Therefore, by Theorem 6.4.1 by Trench (2003), x_b is continuously differentiable with respect to (p_b, y). Moreover, the partial derivatives of x_b with respect to p_b and y are expressed as

$$\frac{\partial x_b}{\partial p_b} = -\frac{\dfrac{\partial H(x_b, p_b, y)}{\partial p_b}}{\dfrac{\partial H(x_b, p_b, y)}{\partial x_b}} = \frac{-df\left(\Delta(x_b, y, p_b)\right)}{1 + df\left(\Delta(x_b, y, p_b)\right)\left(-\dfrac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} + \dfrac{\partial w(x_b, y)}{\partial x_b} + \dfrac{\partial g(x_b, y)}{\partial x_b}\right)} \text{ and}$$

$$\frac{\partial x_b}{\partial y}=-\frac{\dfrac{\partial H(x_b,p_b,y)}{\partial y}}{\dfrac{\partial H(x_b,p_b,y)}{\partial x_b}}=\frac{-df\left(\Delta(x_b,y,p_b)\right)\left(\dfrac{\partial w(x_b,y)}{\partial y}+\dfrac{\partial g(x_b,y)}{\partial y}\right)}{1+df\left(\Delta(x_b,y,p_b)\right)\left(-\dfrac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b}+\dfrac{\partial w(x_b,y)}{\partial x_b}+\dfrac{\partial g(x_b,y)}{\partial x_b}\right)}.$$

This completes the proof. ■

The system time cost V in the optimization problem (5) is determined by the variable x_b. Given a fixed y-value, the variable x_b is a function of p_b and hence the system time cost V is finally determined by p_b. The derivative of V with respect to p_b can be governed by

$$\frac{\mathrm{d}V}{\mathrm{d}p_b}=\left(w(x_b,y)+t_b+x_b\frac{\partial w(x_b,y)}{\partial x_b}-t_a(x_b)+(d-x_b)\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b}\right)\frac{\partial x_b}{\partial p_b}. \tag{17}$$

To minimize the system time cost V by using a gradient-based method, the decision variable p_b could be updated in the negative gradient direction $-\mathrm{d}V/\mathrm{d}p_b$ in each iteration. However, the in-vehicle congestion cost function and the perceived travel cost error distribution (i.e., the functions g and f) are unknown, and hence the negative gradient direction $-\mathrm{d}V/\mathrm{d}p_b$ cannot be computed by formula (17) and cannot be directly applied to solve the optimization problem (5).

In the TESTO procedure, Steps 3 and 4 are employed to estimate the derivative of x_b with respect to p_b without resorting to the functions g and f. As the difference parameter Δp is positive and sufficiently small, it is concluded that

$$\frac{\overline{x}_b^{(n)}-x_b^{(n)}}{\overline{p}_b^{(n)}-p_b^{(n)}}=\frac{\overline{x}_b^{(n)}-x_b^{(n)}}{\Delta p}\approx\frac{\partial x_b^{(n)}}{\partial p_b^{(n)}}.$$

It immediately follows that

$$\left(w\left(x_b^{(n)},y\right)+t_b+x_b^{(n)}\frac{\partial w\left(x_b^{(n)},y\right)}{\partial x_b^{(n)}}-t_a\left(x_b^{(n)}\right)+\left(d-x_b^{(n)}\right)\frac{\mathrm{d}t_a\left(x_b^{(n)}\right)}{\mathrm{d}x_b^{(n)}}\right)\frac{x_b^{(n)}-\overline{x}_b^{(n)}}{\Delta p}\approx-\frac{\mathrm{d}V}{\mathrm{d}p_b^{(n)}},$$

i.e., in Step 4 of the TESTO procedure, the decision variable p_b is updated in an approximate negative gradient direction of V (i.e., a descent direction) despite that the functions g and f are not used.

The following proposition shows that the optimal bus fare (or subsidy) for minimizing the system time cost V exists and is unique. Moreover, V is decreasing with respect to p_b on the left-hand side of the optimal bus fare (or subsidy) and is increasing with respect to p_b on the right-hand side of the optimal bus fare (or subsidy).

Proposition 2. There exists a unique bus fare (or subsidy) p_b^* so that the system time cost V is minimum at the point p_b^*. Furthermore, V is decreasing with respect to $p_b \in (-\infty, p_b^*)$ and increasing with respect to $p_b \in (p_b^*, +\infty)$.

Proof. By the SUE condition (3), it is obtained that when the bus fare (or subsidy) $p_b \to -\infty$, the number of bus users $x_b \to d$; when $p_b \to +\infty$, $x_b \to 0$. In addition, in virtue of $f\left(\Delta(x_b, y, p_b)\right) > 0$, $\mathrm{d}t_a(x_b)/\mathrm{d}x_b < 0$, $\partial w(x_b, y)/\partial x_b > 0$, and $\partial g(x_b, y)/\partial x_b$, the partial derivative formulated by expression (11) is always negative, i.e., $\partial x_b/\partial p_b < 0$. That is to say, x_b is decreasing with respect to $p_b \in (-\infty, +\infty)$. Thus, the increase in p_b from $-\infty$ to $+\infty$ corresponds to the decrease in x_b from d to 0.

The system time cost V in the optimization problem (5) is strictly convex with respect to $x_b \in [0, d]$. Meanwhile, the globally optimal solution x_b^* to the optimization problem (5) is in the interior of the feasible set $[0, d]$. Therefore, V is decreasing with respect to $x_b \in [0, x_b^*)$ and increasing with respect to $x_b \in (x_b^*, d]$.

Associating the above two facts results in that there exists a unique optimal bus fare (or subsidy) p_b^*, at which the system time cost V is minimum, and also V is decreasing with respect to $p_b \in (-\infty, p_b^*)$ and increasing with respect to $p_b \in (p_b^*, +\infty)$. This completes the proof. ■

Figure 1 shows the relationship between p_b, x_b, and V, which can help to understand the proof of proposition 2. In the figure, p_b^1, p_b^*, and p_b^2 satisfy $p_b^1 < p_b^* < p_b^2$. Through the function relationship between p_b, x_b, and V, p_b^1, p_b^*, and p_b^2 correspond to V^1, V^*, and V^2, respectively. It immediately follows that $V^1 > V^*$ holds if $p_b^1 < p_b^*$, and $V^2 > V^*$ holds if $p_b^2 > p_b^*$. This indicates that the system time cost V takes the minimum value at the point p_b^*, and V is decreasing with respect to $p_b \in (-\infty, p_b^*)$ and increasing with respect to $p_b \in (p_b^*, +\infty)$.

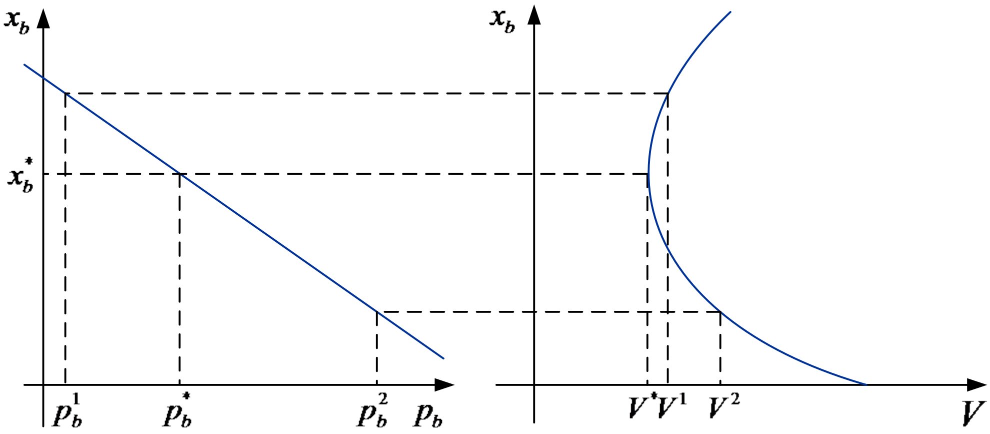

Figure 1. The relationship between p_b, x_b, and V.

The number n_0 in Step 5 is used to record the iteration, in which the iterative direction $G_p^{(n_0)}$ is different from the previous iterative directions $G_p^{(n_0-1)}$, $G_p^{(n_0-2)}$, $\cdots$, $G_p^{(1)}$, and $G_p^{(0)}$. That is to say, before iteration n_0, the iterative directions remain unchanged; in iteration n_0, the iterative direction changes for the first time. In each of the first $n_0 - 1$ iterations, after the iterative direction $G_p^{(n)}$ is determined in Step 4, the iterative step size $\alpha^{(n)}$ takes the fixed value α_0 in Step 5 to search for the interval that contains the optimal solution. The interval of containing the optimal solution has a property, i.e., the derivative of V with respect to p_b, $\mathrm{d}V/\mathrm{d}p_b$, at the left boundary point of the interval is negative and the derivative $\mathrm{d}V/\mathrm{d}p_b$ at the right boundary point of the interval is positive. Thus, if the iterative directions in two successive iterations are different for the first time, then the interval of containing the optimal solution is known. After the interval containing the optimal solution is determined, the bisection iterative method is used to gradually reduce the interval so that the optimal solution is found. Figure 2 gives an example, which shows how the TESTO procedure takes effect for optimizing the total system time cost V.

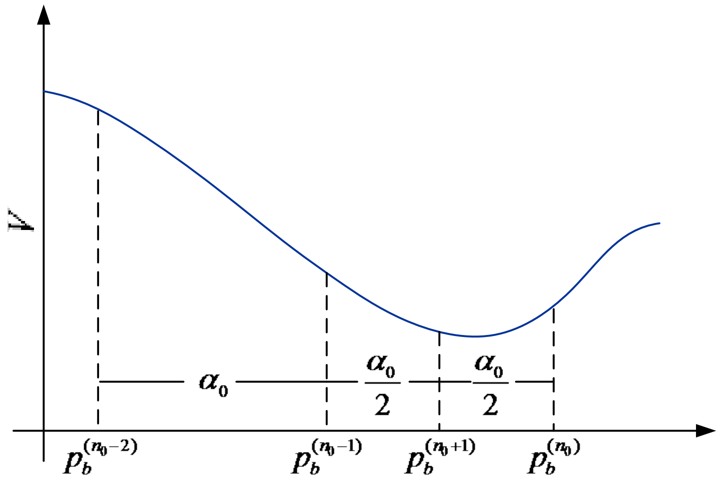

Figure 2. An iterative procedure for determining p_b for optimizing the total system time cost V.

Based on the above analyses, we obtain the following proposition about the convergence of the TESTO procedure to the optimum solution.

Proposition 3. For an initial point $p_b^{(0)} \in (-\infty, +\infty)$, the TESTO procedure is convergent, i.e., the sequence $\left\{x_b^{(n)}, n = 0,1,2,\cdots\right\}$ is convergent to the optimum solution.

The TESTO procedure can be extended to handle cases with multiple bus lines with different fares and the flow interactions between buses and cars in a general network. This extension is described in Appendix A.

3. Trial-and-error implementation for transit profit maximization

3.1. Optimization of transit's operating profit

When the transit line is operated by a private transit operator, the operator would like to deliberately set the transit fare and bus run frequency so as to maximize its operating profit. Let $k(y)$ be the operating cost of the transit line. The function k is continuously differentiable with respect to y. It is supposed that $\mathrm{d}k(y)/\mathrm{d}y > 0$, which this means that the function k is increasing with respect to y. The total operating profit of the transit operator is governed by

$$U = x_b p_b - k(y), \tag{18}$$

i.e., the total profit is equal to the difference between the revenue $x_b p_b$ from the transit fare and the operating cost $k(y)$.

The operator maximizes the total profit at the SUE state by determining the transit fare p_b and the bus run frequency y. The optimal transit fare and bus run frequency are obtained by solving the following optimization problem

$$\max_{(p_b,y)} U = x_b p_b - k(y), \tag{19}$$

where the decision variable (p_b, y) is subject to

$$x_b = dF\left(\Delta(x_b, y, p_b)\right), \quad p_b \geq 0, \text{ and } \quad y \geq 0. \tag{20}$$

The first constraint is the SUE condition formulated by expression (3) and the last two constraints are the non-negativity constraints for the transit fare and the bus run frequency. It is assumed that the optimal solution to the optimization problem (19) exists.

3.2. Trial-and-error procedure

We now propose an iterative trial-and-error procedure for determining the transit fare and bus run frequency to globally or locally maximize the operating profit of the operator without resorting to the function of in-vehicle congestion costs and the distribution of perceived travel cost errors. The iterative trial-and-error procedure is summarized as follows.

Step 1. Let the initial transit fare and bus run frequency be $p_b^{(0)}$ (≥ 0) and $y^{(0)}$ (≥ 0), set an initial iterative step size of transit fare β_{p0} (>0) and an initial iterative step size of bus run frequency β_{y0} (>0), and set $n=0$.

Step 2.1. Let $\tilde{p}_b^{(0)} = p_b^{(n)}$ and set $m=0$.

Step 2.2. Observe the revealed number of bus users $\tilde{x}_b^{(m)}$ at the SUE state under the implementation of the transit fare $\tilde{p}_b^{(m)}$ and the bus run frequency $y^{(n)}$.

Step 2.3. Set the transit fare $\overline{p}_b^{(m)} = \tilde{p}_b^{(m)} + \Delta p$, the bus run frequency $y^{(n)}$ remains unchanged, and then observe the revealed number of bus users $\overline{x}_b^{(m)}$ at the SUE state.

Step 2.4. Compute the iterative direction $G_p^{(m)}$ of the transit fare according to

$$G_p^{(m)} = \tilde{p}_b^{(m)} \frac{\overline{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\Delta p} + \tilde{x}_b^{(m)}. \tag{21}$$

Step 2.5. If $m < m_{p0}$, then $\beta_p^{(m)} = \beta_{p0}$; otherwise, set $\overline{\beta} = \left|\tilde{p}_b^{(m)} - \tilde{p}_b^{(m-1)}\right|$ and $\beta_p^{(m)} = \overline{\beta}/2$.

Step 2.6. Update the transit fare by the following formula

$$\tilde{p}_b^{(m+1)} = \begin{cases} \max\left\{\tilde{p}_b^{(m)} - \beta_p^{(m)}, 0\right\}, & \text{if } G_p^{(m)} < 0, \\ \tilde{p}_b^{(m)} + \beta_p^{(m)}, & \text{otherwise.} \end{cases} \tag{22}$$

Step 2.7. If $\left|\tilde{p}_b^{(m+1)} - \tilde{p}_b^{(m)}\right| < \overline{\varepsilon}_p$, then set $p_b^{(n+1)} = \tilde{p}_b^{(m+1)}$ and go to Step 3.1; otherwise, set $m = m+1$ and go to Step 2.2.

Step 3.1. Let $\tilde{y}^{(0)} = y^{(n)}$ and set $m=0$.

Step 3.2. Observe the revealed number of bus users $\tilde{x}_b^{(m)}$ at the SUE state under the implementation of the bus run frequency $\tilde{y}^{(m)}$ and the transit fare $p_b^{(n+1)}$.

Step 3.3. Set the bus run frequency $\bar{y}^{(m)} = \tilde{y}^{(m)} + \Delta y$, the transit fare $p_b^{(n+1)}$ remains unchanged, and then observe the revealed number of bus users $\bar{x}_b^{(m)}$ at the SUE state.

Step 3.4. Compute the iterative direction $G_y^{(m)}$ of the bus run frequency according to

$$G_y^{(m)} = p_b^{(n+1)} \frac{\bar{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\Delta y} - \frac{\mathrm{d}k\left(\tilde{y}^{(m)}\right)}{\mathrm{d}\tilde{y}^{(m)}}. \tag{23}$$

Step 3.5. If $m < m_{y0}$, then $\beta_y^{(m)} = \beta_{y0}$; otherwise, set $\tilde{\beta} = \left|\tilde{y}^{(m)} - \tilde{y}^{(m-1)}\right|$ and $\beta_y^{(m)} = \tilde{\beta}/2$.

Step 3.6. Update the bus run frequency by the following formula

$$\tilde{y}^{(m+1)} = \begin{cases} \max\left\{\tilde{y}^{(m)} - \beta_y^{(m)}, 0\right\}, & \text{if } G_y^{(m)} < 0, \\ \tilde{y}^{(m)} + \beta_y^{(m)}, & \text{otherwise.} \end{cases} \tag{24}$$

Step 3.7. If $\left|\tilde{y}^{(m+1)} - \tilde{y}^{(m)}\right| < \bar{\varepsilon}_y$, then set $y^{(n+1)} = \tilde{y}^{(m+1)}$ and go to Step 4; otherwise, set $m = m+1$ and go to Step 3.2.

Step 4. If $\left\|\left(p_b^{(n+1)}, y^{(n+1)}\right) - \left(p_b^{(n)}, y^{(n)}\right)\right\| < \bar{\varepsilon}$, then stop; otherwise, set $n = n+1$ and go to Step 2.1.

We call the above iterative trial-and-error procedure the trial-and-error profit maximization (TEPM) procedure. In the TEPM procedure, the difference parameters Δp and Δy are positive and also sufficiently small. The operator $\max\{\cdot, 0\}$ is used to guarantee the feasibility of p_b and y after their update in formulae (22) and (24). $\bar{\varepsilon}_p$, $\bar{\varepsilon}_y$, and $\bar{\varepsilon}$ are three convergence tolerances and they are positive and sufficiently small. $\|\cdot\|$ in Step 4 denotes the Euclidean norm, e.g., for a row vector $\mathbf{v}$, $\|\mathbf{v}\| = \sqrt{\mathbf{v}\mathbf{v}^{\mathrm{T}}}$. One can see that the operator need not know the function of in-vehicle congestion costs and the distribution of perceived travel cost errors (i.e., the functions g and f) to implement the TEPM procedure.

The operating profit U in the optimization problem (19) is determined by the variables x_b, y, and p_b. By Proposition 1, the variable x_b is a function of y and p_b, and hence the operating profit U is finally determined by y and p_b. The partial derivatives of U with respect to p_b and y can be formulated as

$$\frac{\partial U}{\partial p_b} = p_b \frac{\partial x_b}{\partial p_b} + x_b \text{ and } \frac{\partial U}{\partial y} = p_b \frac{\partial x_b}{\partial y} - \frac{\mathrm{d}k(y)}{\mathrm{d}y}, \tag{25}$$

where $\partial x_b / \partial p_b$ and $\partial x_b / \partial y$ are given by expressions (11) and (12), respectively. To optimize the operating profit U by using a gradient-based algorithm, the decision variable (p_b, y) could be updated in the gradient direction $\left(\partial U/\partial p_b, \partial U/\partial y\right)$ in each iteration.

However, the in-vehicle congestion costs of commuters in bus carriages and the perceived errors of commuters for travel costs (i.e., the functions g and f) are unknown, and hence the gradient direction $\left(\partial U/\partial p_b, \partial U/\partial y\right)$ cannot be computed by formula (25) to solve the optimization problem (19).

In the TEPM procedure, Steps 2.3 and 2.4 are used to estimate the partial derivative of x_b with respect to p_b and Steps 3.3 and 3.4 are used to estimate the partial derivative of x_b with respect to y without resorting to the functions g and f. As the difference parameters Δp and Δy are positive and also sufficiently small, we conclude

$$\frac{\overline{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\overline{p}_b^{(m)} - \tilde{p}_b^{(m)}} = \frac{\overline{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\Delta p} \approx \frac{\partial \tilde{x}_b^{(m)}}{\partial \tilde{p}_b^{(m)}} \text{ and } \frac{\overline{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\overline{y}^{(m)} - \tilde{y}^{(m)}} = \frac{\overline{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\Delta y} \approx \frac{\partial \tilde{x}_b^{(m)}}{\partial \tilde{y}^{(m)}}.$$

It immediately follows that

$$\tilde{p}_b^{(m)} \frac{\overline{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\Delta p} + \tilde{x}_b^{(m)} \approx \frac{\partial U}{\partial \tilde{p}_b^{(m)}} \text{ and } p_b^{(n+1)} \frac{\overline{x}_b^{(m)} - \tilde{x}_b^{(m)}}{\Delta y} - \frac{\mathrm{d}k\left(\tilde{y}^{(m)}\right)}{\mathrm{d}\tilde{y}^{(m)}} \approx \frac{\partial U}{\partial \tilde{y}^{(m)}}.$$

In Steps 2.1 to 2.7 of the TEPM procedure, the bus run frequency $y^{(n)}$ remains unchanged and the bus fare p_b is updated in an approximate ascent direction $G_p^{(m)}$ of U with respect to p_b until a maximum point $p_b^{(n+1)}$ is generated. After the iterative direction $G_p^{(m)}$ of the transit fare is determined in Step 2.4, the iterative step size $\beta_p^{(m)}$ is generated in Step 2.5 in a similar way as the iterative step size $\alpha^{(n)}$ in the TESTO procedure in Section 2.3. The number m_{p0} in Step 2.5 records the iteration, in which the iterative direction changes for the first time. In each of the first $m_{p0}-1$ iterations, the iterative step size $\beta_p^{(m)}$ takes the fixed value β_{p0} to obtain the interval in which the optimal solution lies.

In iteration m_{p0} and subsequent iterations, a bisection iterative method is used to find the optimal solution in the interval. However, the bisection iterative method in the TEPM procedure is different from that in the TESTO procedure due to the non-negative constraint on the bus fare p_b. In fact, in the TEPM procedure, when the bus fare p_b is updated in the left direction, the width of the obtained interval of containing the optimal solution may be less than the step size β_{p0} because the bus fare p_b is non-negative. For example, in the iterative procedure shown in Figure 3, $\tilde{p}_b^{(1)} - \beta_{p0} < 0$ holds. It follows that $p_b^{(2)} = \max\left\{\tilde{p}_b^{(1)} - \beta_{p0}, 0\right\} = 0$ and the width of the interval containing the optimal solution $\left|\tilde{p}_b^{(2)} - \tilde{p}_b^{(1)}\right| < \beta_{p0}$. Thus, in the bisection iterative method, half of $\overline{\beta} = \left|\tilde{p}_b^{(m)} - \tilde{p}_b^{(m-1)}\right|$ rather than half of β_{p0} is used in each iteration.

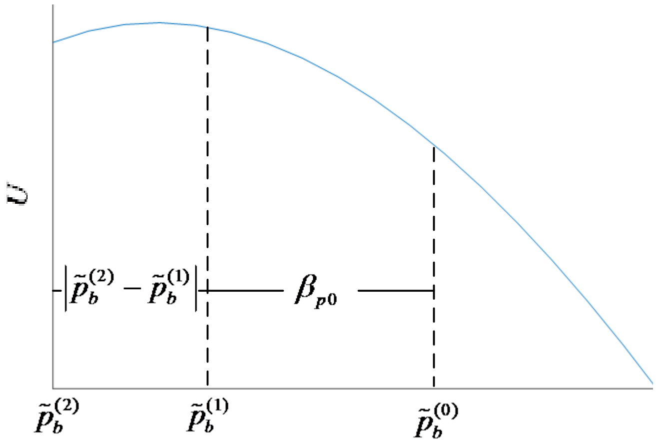

Figure 3. An iterative procedure for determining p_b for optimizing the operating profit U .

In Steps 3.1 to 3.7 of the TEPM procedure, given the transit fare $p_b^{(n+1)}$, the bus run frequency y is updated in an approximate ascent direction $G_y^{(m)}$ of U with respect to y until a maximum point $y^{(n+1)}$ is obtained. The update rules of Steps 3.1 to 3.7 are identical to the update rules of Steps 2.1 to 2.7. The number m_{y0} in Step 3.5 records the iteration, in which the iterative direction changes for the first time. In each of the first $m_{y0}-1$ iterations, the iterative step size $\beta_y^{(m)}$ takes the fixed value β_{y0} to obtain the interval of containing the optimal solution. In iteration m_{y0} and subsequent iterations, a bisection iterative method is used to find the optimal solution in the interval.

In this way, in each iteration of the TEPM procedure, the operating profit U increases by a certain value until the iterative trajectory gets to a globally or locally maximum point. Based on the above analyses, we conclude the following convergence of the TEPM procedure.

Proposition 4. For an initial point $\left(p_b^{(0)}, y^{(0)}\right)$ that satisfies $p_b^{(0)} \geq 0$ and $y^{(0)} \geq 0$, the TEPM procedure is convergent, i.e., the sequence $\left\{\left(p_b^{(n)}, y^{(n)}\right), n=0,1,2,\cdots\right\}$ is convergent to a locally or globally optimal solution.

The problem (19) is an optimization problem with equilibrium constraints, and hence the optimal solution to the problem may not be unique. When the problem has multiple optimal solutions, the TEPM procedure can be used together with other methods, e.g., a segmentation technology of the feasible set or the implementation of the TEPM from different initial points, to find the globally optimal solution.

Very often a bus operator cannot adjust its price and bus run frequency freely. For example, the government always sets a bound for the bus fare or defines the minimum bus frequency for the profit-maximizing case. When the bus fare has an upper bound or the bus run frequency has a lower bound, the trial-and-error procedure can be designed and developed in a similar way. In this case, the transit fare and the bus run frequency are required to be iterated and updated in their feasible sets with interval constraints.

In the profit maximization objective (18), the operating cost can be related to the flow of bus users. The proposed trial-and-error method works for an operation problem, in which the in-vehicle congestion cost and the perceived error terms do not appear in the manager or operator's optimization objective. Thus, when an operating cost that relates to the flow of bus users is considered in the profit maximization objective, the trial-and-error procedure can also be developed similarly.

The TEPM procedure can be extended to handle cases with more bus lines in a general network and the flow interactions between cars and buses. This extension is introduced in Appendix B.

4. Numerical examples

In this section, we give a set of numerical examples to show the effectiveness of the two classes of trial-and-error implementations. In the trial-and-error procedures, the dynamic modal choice of commuters can be formulated as (Li and Yang, 2016; Guo and Szeto, 2018)

$$x_b^{(i+1)} = (1-\delta)x_b^{(i)} + \delta dF\left(\Delta\left(x_b^{(i)}, y^{(i)}, p_b^{(i+1)}\right)\right), \tag{26}$$

for $i = 0,1,2,\cdots$. The superscript (i) refers to the ith day, e.g., $x_b^{(i)}$ represents the number of bus users on the transit line on day i. The adjustment parameter $\delta \in (0,1]$. Formula (26) states that on day $i+1$, a portion of commuters does not change their travel modes chosen on the previous day i due to inertia. The other portion of commuters does reconsider their travel modes based on the waiting time, travel time, and congestion costs $w\left(x_b^{(i)}, y^{(i)}\right) + t_b + g\left(x_b^{(i)}, y^{(i)}\right)$ and $t_a\left(x_b^{(i)}\right)$ on the previous day i and the transit fare (or financial subsidy) and the auto toll $p_b^{(i+1)}$ and p_a on that day $i+1$. These commuters with the perceived generalized travel cost of buses $c_b\left(x_b^{(i)}, y^{(i)}, p_b^{(i+1)}\right)$ less than (more than) the perceived generalized travel cost of autos $c_a\left(x_b^{(i)}\right)$ choose to use the bus mode (auto mode). Given the bus run frequency y, the transit fare (or subsidy) p_b, and the auto toll p_a, the stationary state of the dynamical system (26) is equivalent to the SUE state that satisfies condition (3) and also the trajectory of the dynamical system is convergent to the SUE state (Li and Yang, 2016).

The total number of commuters is $d = 2\times10^4$. The toll charge for auto users is $p_a = 0$.

The waiting time cost of bus users at the bus stop on the transit line is governed by

$$w(x_b, y) = 5\times10^{-4}\times\left(\frac{x_b}{y+10^{-5}}\right)^2.$$

The in-vehicle travel time cost of bus users on the transit line is $t_b = 10$. The in-vehicle congestion cost of passengers on the transit line is expressed as

$$g(x_b, y) = 3\times10^{-4}\times\left(\frac{x_b}{y+10^{-5}}\right)^2.$$

The travel time cost of auto users on the highway is formulated as

$$t_a(x_b) = 8\times10^{-3}\times\left(\frac{d-x_b}{2400}\right)^4 + 8.$$

The random variable ξ follows a bimodal distribution with the probability density function

$$f(\varepsilon) = \frac{1}{2\sqrt{2\pi}\sigma}\left(\exp\left(-\frac{(\varepsilon-\bar{\mu}_1)^2}{2\sigma^2}\right)+\exp\left(-\frac{(\varepsilon-\bar{\mu}_2)^2}{2\sigma^2}\right)\right),\quad \varepsilon\in(-\infty,+\infty),$$

where the parameters $\bar{\mu}_1 = -3$, $\bar{\mu}_2 = 4$, and $\sigma = 2$. The bimodal distribution is a mixture of two different unimodal distributions and it prescribes that two groups of commuters have an obvious preference to the two modes, respectively. For example, some commuters have no private cars, and hence they have to choose to travel by public transit. Some commuters, who have private cars, may always choose to travel by private auto mode due to inertia or aversion to in-vehicle congestion in bus carriages. Moreover, the two parameters $\bar{\mu}_1$ and $\bar{\mu}_2$ satisfy $|\bar{\mu}_1| < \bar{\mu}_2$. This means that commuters prefer to use private transport.

First, we show the effectiveness of the TESTO procedure in Section 2.3. The frequency of bus runs is $y = 200$. The adjustment parameter δ in formula (26) takes 0.01. The initial iterative step size is set as $\alpha_0 = 5$, the difference parameter $\Delta p = 0.1$, and the convergence tolerance $\tilde{\varepsilon} = 10^{-10}$. Figure 4 shows the relation of the system time cost V of the transport system and the transit fare (or subsidy) p_b for bus users. One can see that the V-value is minimum and is equal to 212652.71 at the point of $p_b = -5.05$. On the left (right) hand side of the minimum point, V is decreasing (increasing) with respect to p_b.

Figure 5 displays the convergence of p_b when the TESTO procedure is applied to the transport system and the initial transit fare (or subsidy) $p_b^{(0)}$ varied from -20 to 50 with an interval of 5. It can be seen that the p_b-values for all these initial points converge to the same value of -5.05. Figure 6 depicts the corresponding changes in the total system time cost V as the number of iterations increases. One can see that all of the V-values finally reduce to the same value of 212652.71. Therefore, the TESTO procedure is effective for optimizing the total system time cost of the transport system, regardless of initial points.

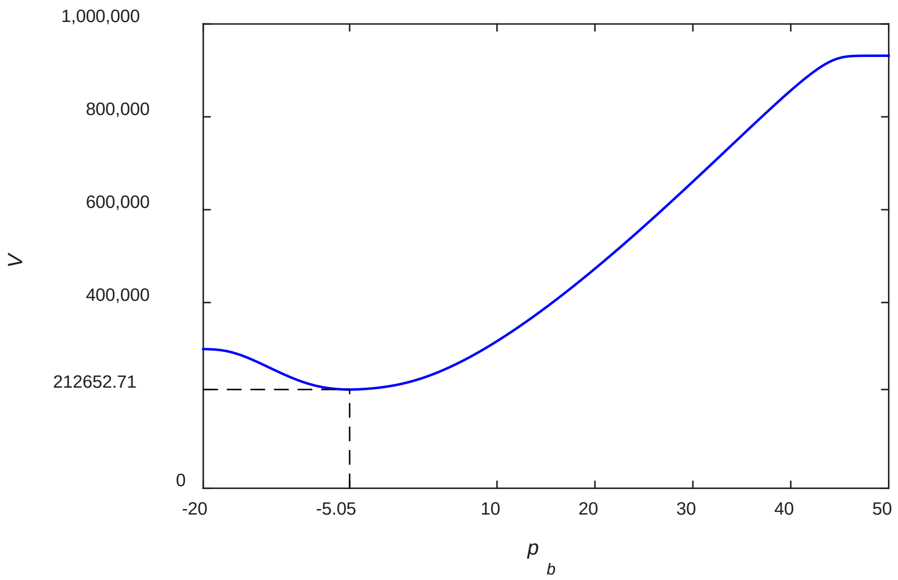

Figure 4. The total system time cost V of the transport system against the transit fare (or subsidy) p_b for bus users.

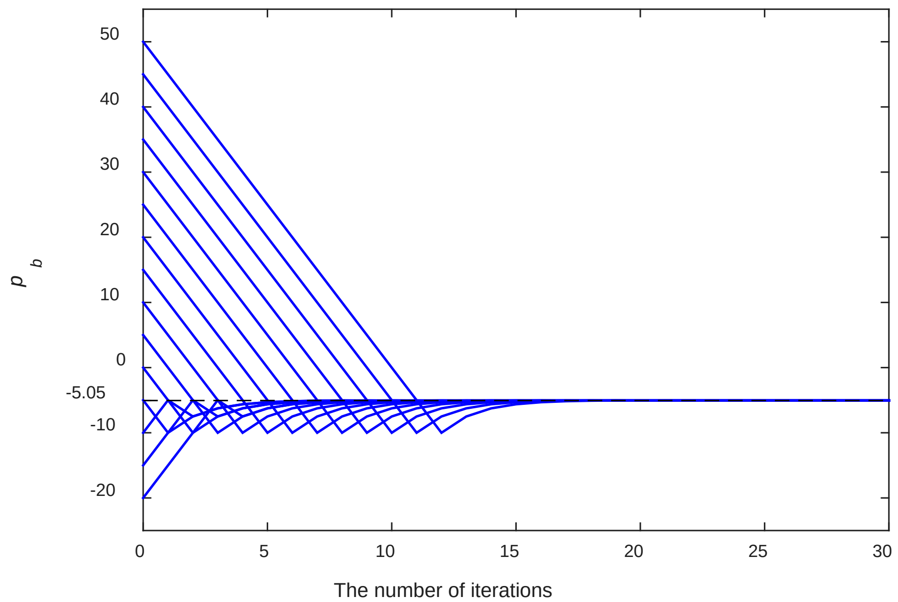

Figure 5. The convergence of p_b when the TESTO procedure was applied to the transport system and the initial transit fare (or subsidy) $p_b^{(0)}$ varied from -20 to 50 with an interval of 5.

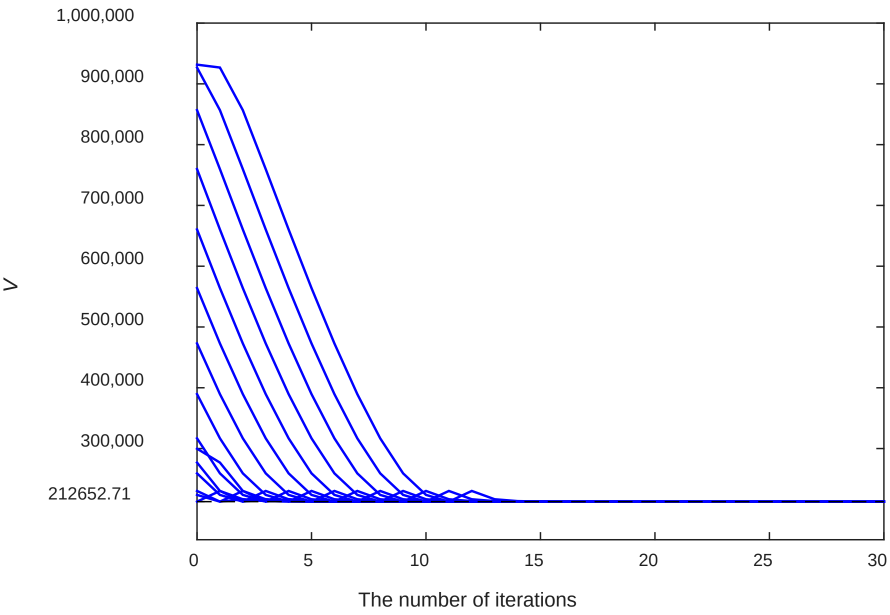

Figure 6. The convergence of the total system time cost V for the initial transit fare (or subsidy) $p_b^{(0)}$ that varies from -20 to 50 with an interval of 5.

We then demonstrate the effectiveness of the TEPM procedure for optimizing transit operating profit in Section 3.2. The operating cost of the transit line is governed by

$$k(y) = 50y + 3\times 10^4 .$$

The adjustment parameter δ in formula (26) takes 0.01. The initial iterative step sizes are set as $\beta_{p0} = 5$ and $\beta_{y0} = 10$, the difference parameters $\Delta p = 0.1$ and $\Delta y = 0.1$, and the convergence tolerances $\overline{\varepsilon} = 10^{-5}$, $\overline{\varepsilon}_p = 10^{-10}$, and $\overline{\varepsilon}_y = 10^{-10}$. Figure 7 shows the relation of the total operating profit U, formulated by expression (18), to the transit fare p_b and the bus run frequency y that belong to the feasible set

$$\Omega = \left\{ (p_b, y) \mid 0 \le p_b \le 50, 1 \le y \le 400 \right\} .$$

One can see that the total operating profit U is a unimodal function with respect to (p_b, y). The total operating profit U takes the maximum value 30420.63 at the point $(p_b, y) = (18.72, 111.99)$. Figure 8 depicts the iterative trajectories of (p_b, y) when the TEPM procedure is applied to the transport system and the initial points $\left(p_b^{(0)}, y^{(0)}\right)$ are on the boundary of the feasible set Ω. It can be seen that all these trajectories of (p_b, y) adjust to the same optimal point $(p_b, y) = (18.72, 111.99)$. Figure 9 exhibits the changes of the total operating profit U for all these initial points as the number of iterations increases. One can see that the total operating profits for all these initial points converge to the same maximum value 30420.63. Thus, the TEPM procedure is effective for optimizing transit operating profit.

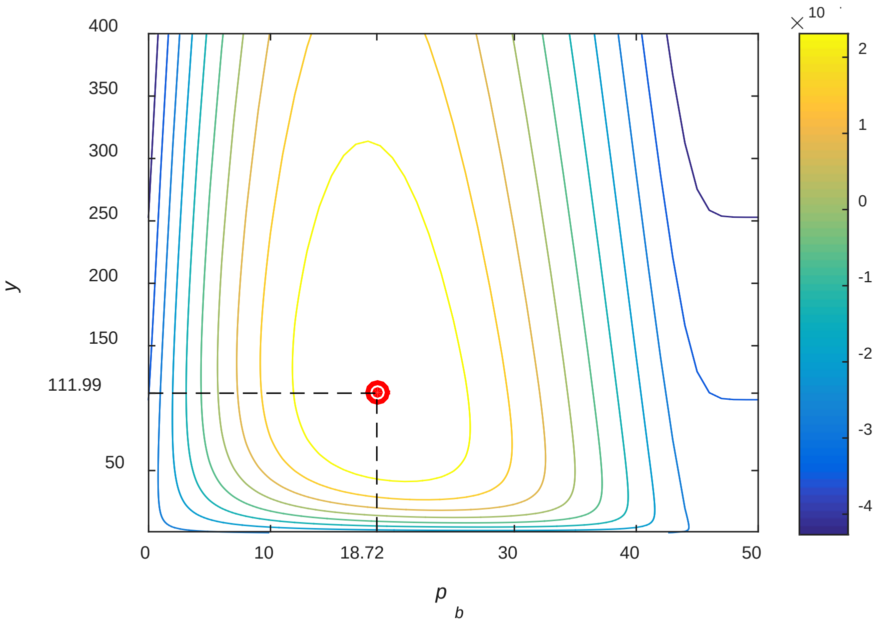

Figure 7. The relation of the total operating profit U, formulated by expression (18), to the transit fare p_b and the bus run frequency y that belong to the feasible set Ω.

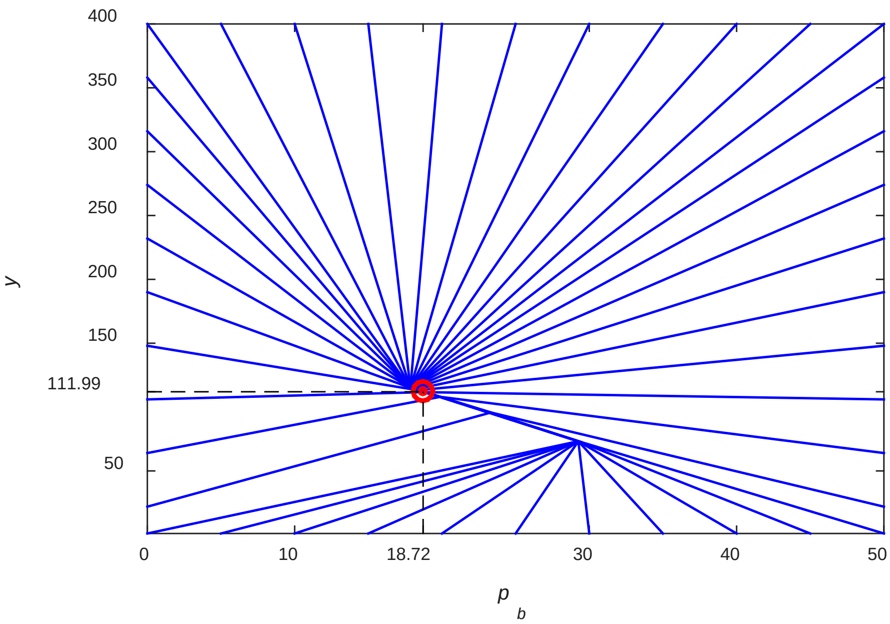

Figure 8. The iterative trajectories of (p_b, y) when the TEPM procedure is applied to the transport system and the initial points $\left(p_b^{(0)}, y^{(0)}\right)$ are on the boundary of the feasible set Ω.

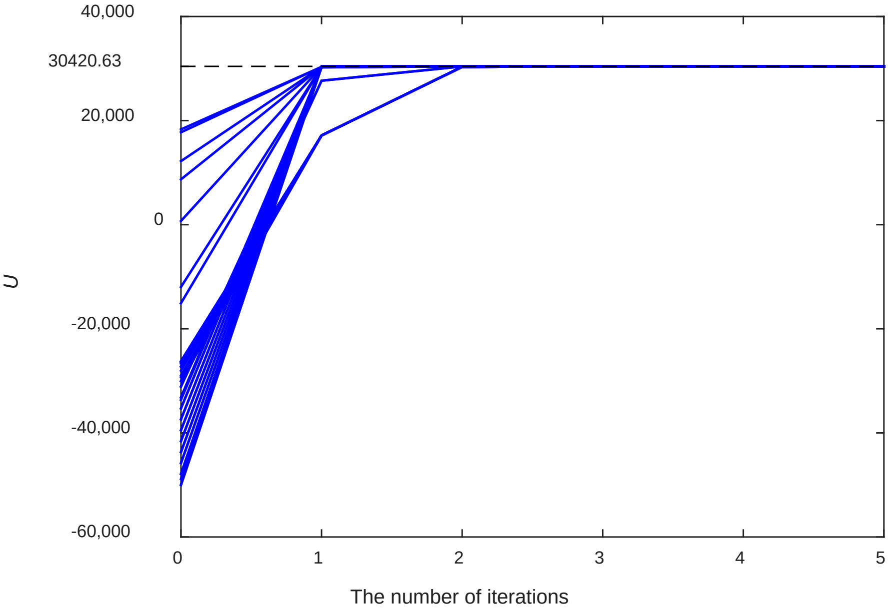

Figure 9. The changes in the total operating profit U for all these initial points on the boundary of the set Ω as the number of iterations increases.

5. Conclusions

In this paper, we concern a bimodal transport system that comprises an origin-destination (OD) pair connected by a transit line in parallel to a highway. The highway is only used by autos and the transit line is only used by buses. The transport system has two characteristics. First, the in-vehicle congestion in bus carriages is regarded as a negative external cost for commuters to choose their travel modes, i.e., commuters choose their travel modes according to not only the waiting time and travel time costs of both modes but also the in-vehicle congestion costs in bus carriages. Second, commuters choose their travel modes according to the perceived travel costs of both modes rather than the actual travel costs and the modal split at a stationary state follows the stochastic user equilibrium (SUE) rather than the deterministic user equilibrium (DUE).

It is difficult to obtain the function of in-vehicle congestion costs and the distribution of perceived travel cost errors. Thus, we propose two trial-and-error procedures for the transport system to achieve two objectives without resorting to both the in-vehicle congestion cost function and the perceived travel cost error distribution. The first objective is to minimize the total system time cost of the transport system from the viewpoint of the government through reasonably determining the transit fare (or financial subsidy) for bus users. The second objective is to maximize the total profit of operating the transit line through setting the transit fare and the bus run frequency when the transit line is operated by a private firm. Moreover, we demonstrate that the two trial-and-error procedures are effective for achieving the two objectives by both theoretical analyses and numerical examples. This study gives new insights and avenues for the practical implementation of both congestion control and public transit operation schemes. The results indicate that it is unnecessary to know the function of in-vehicle congestion costs and the distribution of perceived travel cost errors so as to optimize the system time cost or operating profit of a multi-modal transport system.

In Section 2.3, only the waiting and travel time cost of all commuters is optimized by adjusting the transit fare (or subsidy) for bus users while the frequency of bus runs is fixed. In Appendix C, we extend the trial-and-error method to optimize the social/system cost of the transport system, including not only the waiting and travel time cost of all commuters but also the in-vehicle congestion cost of passengers in bus carriages and the operating cost of bus transit, through adjusting both the bus fare (or subsidy) and bus run frequency. In the extended trial-and-error procedure, the function of in-vehicle congestion costs is not needed; however, the distribution of perceived travel cost errors has to be involved.

In this paper, the travel demand of each individual mode is elastic but the total number of travelers selecting the two modes is fixed. We used this assumption not only to simplify the analysis but also to illustrate the nature of the problem clearly. The extension to the elastic demand case is interesting and is left to future study.

Acknowledgments

The work described in this paper was jointly supported by grants from the National Natural Science Foundation of China (71622005, 71890972/71890970), a grant from the Research Grants Council of the Hong Kong Special Administrative Region of China (HKU17218916), and a grant from the University Research Committee of the University of Hong Kong (201711159034). The authors are grateful to the two reviewers for their constructive comments.

Appendix A. Extension of the TESTO procedure to handle a general network instance

In this appendix, we extend the TESTO procedure for the system time optimization of a general traffic network with both private auto and public transit modes. Let W be the set of OD pairs. Commuters between OD pair $w \in W$ can choose to travel by either auto running on auto routes or bus running on transit routes. The travel demands in the network are assumed to be fixed and denoted by a column vector $\mathbf{d} = \left(d_w, w \in W\right)^{\mathrm{T}}$, where d_w is the travel demand between OD pair $w \in W$. g_w stands for the flow of bus users between OD pair $w \in W$ and $\mathbf{g} = \left(g_w, w \in W\right)^{\mathrm{T}}$ is the vector of flows of bus users. The occupancy of each auto is assumed to be one. Let f_w be the flow of auto users between OD pair $w \in W$ and $\mathbf{f} = \left(f_w, w \in W\right)^{\mathrm{T}}$ be the vector of flows of auto users. Naturally, we have $\mathbf{d} = \mathbf{g} + \mathbf{f}$. The set Ω of feasible flows of bus and auto users is formulated as

$$\Omega \equiv \left\{(\mathbf{g}, \mathbf{f}) \middle| \mathbf{d} = \mathbf{g} + \mathbf{f}, \mathbf{g} \geq \mathbf{0}, \mathbf{f} \geq \mathbf{0}\right\}. \tag{A.1}$$

In the transport system, the total flow of bus users on all transit routes (auto users on all auto routes) between an OD pair is regarded as an entirety and the specific flow of bus users on a transit route (auto users on an auto route) is not considered. The interactions between bus user flows, bus run frequencies, and auto user flows (on either an OD pair or different OD pairs) are formulated by non-separable cost functions. Thus, we actually examine a general case in this appendix.

Let y_w be the frequency of bus runs between OD pair $w \in W$ and $\mathbf{y} = \left(y_w, w \in W\right)^{\mathrm{T}}$ is the vector of bus run frequencies. Let $\hat{c}_w(\mathbf{g}, \mathbf{y})$ denote the waiting time cost of bus users between OD pair $w \in W$ and $\hat{\mathbf{c}}(\mathbf{g}, \mathbf{y}) = \left(\hat{c}_w(\mathbf{g}, \mathbf{y}), w \in W\right)^{\mathrm{T}}$ is the corresponding vector. That is to say, the waiting time cost of bus users between an OD pair depends on not only the bus user flow and bus run frequency between the OD pair but also the bus user flows and bus run frequencies between other OD pairs. Let $\overline{c}_w(\mathbf{f}, \mathbf{y})$ stand for the in-vehicle travel time cost of bus users between OD pair $w \in W$ and $\overline{\mathbf{c}}(\mathbf{f}, \mathbf{y}) = \left(\overline{c}_w(\mathbf{f}, \mathbf{y}), w \in W\right)^{\mathrm{T}}$ is the corresponding vector. This indicates that there are interactions between public transit and private car mode on roads. The notation $\tilde{c}_w(\mathbf{g}, \mathbf{y})$ denotes the in-vehicle congestion cost of passengers in bus carriages between OD pair $w \in W$ and $\tilde{\mathbf{c}}(\mathbf{g}, \mathbf{y}) = \left(\tilde{c}_w(\mathbf{g}, \mathbf{y}), w \in W\right)^{\mathrm{T}}$ is the corresponding vector. This means that the in-vehicle congestion cost of passengers between an OD pair depends on not only the bus user flow and bus run frequency between the OD pair but also the bus user flows and bus run frequencies between other OD pairs.

Let $c_w(\mathbf{f}, \mathbf{y})$ represent the travel time cost of auto users between OD pair $w \in W$ and $\mathbf{c}(\mathbf{f}, \mathbf{y}) = \left(c_w(\mathbf{f}, \mathbf{y}), w \in W\right)^{\mathrm{T}}$ is the corresponding vector. That is to say, there are interactions between private cars and public transit on roads. The notation p_w (≥ 0) stands for the transit fare charged from each bus user between OD pair $w \in W$ and $\mathbf{p} = \left(p_w, w \in W\right)^{\mathrm{T}}$ is the vector of transit fares. All those costs and prices, mentioned above, are measured in monetary units.

Commuters choose their travel modes according to their perceived travel costs. The SUE conditions are governed by

$$g_w = d_w G_w(\mathbf{g}, \mathbf{f}, \mathbf{y}, \mathbf{p}) \text{ and } f_w = d_w F_w(\mathbf{g}, \mathbf{f}, \mathbf{y}, \mathbf{p}), \ \forall w \in W, \tag{A.2}$$

where $G_w(\mathbf{g},\mathbf{f},\mathbf{y},\mathbf{p})$ is the probability that the public transit mode between OD pair $w \in W$ is chosen and $F_w(\mathbf{g},\mathbf{f},\mathbf{y},\mathbf{p})$ is the probability that the private car mode between OD pair $w \in W$ is chosen.

Given a vector $\mathbf{y}$ of bus run frequencies, the optimization problem of minimizing the total system time cost of the transport system is formulated as

$$\min_{(\mathbf{p},\mathbf{g},\mathbf{f})} V = \mathbf{g}^{\mathrm{T}}\left(\hat{\mathbf{c}}(\mathbf{g},\mathbf{y})+\overline{\mathbf{c}}(\mathbf{f},\mathbf{y})\right)+\mathbf{f}^{\mathrm{T}}\mathbf{c}(\mathbf{f},\mathbf{y}), \tag{A.3}$$

where the decision variables $\mathbf{p}$, $\mathbf{g}$, and $\mathbf{f}$ are subject to the SUE condition (A.2), $(\mathbf{g},\mathbf{f}) \in \Omega$, and $\mathbf{p} \geq \mathbf{0}$.

In the trial-and-error procedure, the SUE condition (A.2) is met under the implementation of the transit fares $\mathbf{p}$ in each iteration, that is to say, an equilibrium flow assignment $(\mathbf{g},\mathbf{f})$ is observed once the transit fares $\mathbf{p}$ are implemented. Thus, when the trial-and-error procedure is applied to solve optimization problem (A.3), the system time cost V is finally determined by $\mathbf{p}$. The iterative trial-and-error procedure for determining $\mathbf{p}$ to solve problem (A.3) is summarized as follows.

Step 1. Let $\mathbf{p}^{(0)}$ ($\geq \mathbf{0}$) be the initial vector of transit fares, and set a convergence tolerance $\tilde{\varepsilon}$ (>0) and $n=0$.

Step 2. Observe the revealed vectors $\mathbf{g}^{(n)}$ and $\mathbf{f}^{(n)}$ of bus and auto user flows at the SUE state after the imposition of the transit fares $\mathbf{p}^{(n)}$.

Step 3. For any OD pair $w \in W$, set the transit fare $\overline{p}_w^{(n)} = p_w^{(n)} + \Delta p$ and $\overline{p}_u^{(n)} = p_u^{(n)}$ for any $u \in W \setminus \{w\}$, and then observe the revealed vectors $\overline{\mathbf{g}}^{(n,w)}$ and $\overline{\mathbf{f}}^{(n,w)}$ of bus and auto user flows at the SUE state, respectively.

Step 4. Compute the iterative direction $\mathbf{P}^{(n)} = \left(P_w^{(n)}, w \in W\right)^{\mathrm{T}}$ of transit fares according to

$$\begin{aligned} P_w^{(n)} = {} & \sum_{u \in W}\left(\hat{c}_u\left(\mathbf{g}^{(n)},\mathbf{y}\right)+\overline{c}_u\left(\mathbf{f}^{(n)},\mathbf{y}\right)+\sum_{v \in W} g_v^{(n)}\frac{\partial \hat{c}_v\left(\mathbf{g}^{(n)},\mathbf{y}\right)}{\partial g_u^{(n)}}\right)\frac{g_u^{(n)}-\overline{g}_u^{(n,w)}}{\Delta p} \\ & +\sum_{u \in W}\left(\sum_{v \in W} g_v^{(n)}\frac{\partial \overline{c}_v\left(\mathbf{f}^{(n)},\mathbf{y}\right)}{\partial f_u^{(n)}}+c_u\left(\mathbf{f}^{(n)},\mathbf{y}\right)+\sum_{v \in W} f_v^{(n)}\frac{\partial c_v\left(\mathbf{f}^{(n)},\mathbf{y}\right)}{\partial f_u^{(n)}}\right)\frac{f_u^{(n)}-\overline{f}_u^{(n,w)}}{\Delta p}. \end{aligned} \tag{A.4}$$

Step 5. Update the transit fares by the following formula:

$$\mathbf{p}^{(n+1)} = \max\left\{\mathbf{p}^{(n)}+\alpha^{(n)}\mathbf{P}^{(n)},\mathbf{0}\right\}. \tag{A.5}$$

Step 6. If $\left\|\mathbf{p}^{(n+1)}-\mathbf{p}^{(n)}\right\| < \tilde{\varepsilon}$, then stop; otherwise, set $n=n+1$ and go to Step 2.

In the above procedure, the convergence tolerance $\tilde{\varepsilon}$ is sufficiently small. The difference parameter Δp is positive and sufficiently small. $\alpha^{(n)}$ (>0) is the step size in iteration n. The step sizes $\left\{\alpha^{(n)}, n=0,1,2,\cdots\right\}$ can be a predetermined sequence that satisfies the following conditions:

$\sum_{n=0}^{\infty} \alpha^{(n)} = +\infty$ and $\sum_{n=0}^{\infty} \left(\alpha^{(n)}\right)^2 < +\infty$.

For example, a typical step size is $\alpha^{(n)} = 1/(n+1)$. The operator $\max\{\cdot, \mathbf{0}\}$ is used to guarantee the feasibility of $\mathbf{p}$ after its update. Obviously, in the above procedure, the function of in-vehicle congestion costs and the distribution of perceived travel cost errors are not involved.

Appendix B. Extension of the TEPM procedure to handle a general network instance

We extend the TEPM procedure for transit profit maximization to handle the general network introduced in Appendix A. When all transit lines are operated by a private transit operator, the operator would like to deliberately set the vector $\mathbf{p}$ of transit fares and the vector $\mathbf{y}$ of bus run frequencies so as to maximize its operating profit at the SUE state. Let $k_w(\mathbf{y})$ be the operating cost of the transit route between OD pair $w \in W$. The optimal transit fares and bus run frequencies are obtained by solving the following optimization problem:

$$\max_{(\mathbf{p},\mathbf{y},\mathbf{g},\mathbf{f})} U = \mathbf{g}^{\mathrm{T}}\mathbf{p} - \sum_{w \in W} k_w(\mathbf{y}), \tag{A.6}$$

where the decision variables $\mathbf{p}$, $\mathbf{y}$, $\mathbf{g}$ and $\mathbf{f}$ are subject to the SUE condition (A.2), $\mathbf{p} \geq \mathbf{0}$, and $\mathbf{y} \geq \mathbf{0}$.

An iterative trial-and-error procedure for solving problem (A.6) is summarized as follows.

Step 1. Let $\mathbf{p}^{(0)}$ ($\geq \mathbf{0}$) and $\mathbf{y}^{(0)}$ ($\geq \mathbf{0}$) be the initial vectors of transit fares and bus run frequencies respectively, and set $n = 0$.

Step 2. Observe the revealed vectors $\mathbf{g}^{(n)}$ and $\mathbf{f}^{(n)}$ of bus and auto user flows at the SUE state respectively under the implementation of the transit fares $\mathbf{p}^{(n)}$ and the bus run frequencies $\mathbf{y}^{(n)}$.

Step 3. For any OD pair $w \in W$, set the transit fare $\overline{p}_w^{(n)} = p_w^{(n)} + \Delta p$ and $\overline{p}_u^{(n)} = p_u^{(n)}$ for any $u \in W \setminus \{w\}$ (the bus run frequency vector remains unchanged), and then observe the revealed vectors $\overline{\mathbf{g}}^{(n,w)}$ and $\overline{\mathbf{f}}^{(n,w)}$ of bus and auto user flows at the SUE state; set the bus run frequency $\overline{y}_w^{(n)} = y_w^{(n)} + \Delta y$ and $\overline{y}_u^{(n)} = y_u^{(n)}$ for any $u \in W \setminus \{w\}$ (the transit fare vector remains unchanged), and then observe the revealed vectors $\tilde{\mathbf{g}}^{(n,w)}$ and $\tilde{\mathbf{f}}^{(n,w)}$ of bus and auto user flows at the SUE state.

Step 4. Compute the iterative directions $\mathbf{P}^{(n)} = \left(P_w^{(n)}, w \in W\right)^{\mathrm{T}}$ and $\mathbf{Y}^{(n)} = \left(Y_w^{(n)}, w \in W\right)^{\mathrm{T}}$ of transit fares and bus run frequencies according to

$$P_w^{(n)} = \sum_{u \in W} p_u^{(n)} \frac{\bar{g}_u^{(n,w)} - g_u^{(n)}}{\Delta p} + g_w^{(n)} \text{ and } Y_w^{(n)} = \sum_{u \in W} p_u^{(n)} \frac{\tilde{g}_u^{(n,w)} - g_u^{(n)}}{\Delta y} - \sum_{u \in W} \frac{\partial k_u \left(\mathbf{y}^{(n)}\right)}{\partial y_w^{(n)}}. \quad \text{(A.7)}$$

Step 5. Update the transit fares and bus run frequencies by the following formulae:

$$\mathbf{p}^{(n+1)} = \max\left\{\mathbf{p}^{(n)} + \beta^{(n)}\mathbf{P}^{(n)}, \mathbf{0}\right\} \text{ and } \mathbf{y}^{(n+1)} = \max\left\{\mathbf{y}^{(n)} + \beta^{(n)}\mathbf{Y}^{(n)}, \mathbf{0}\right\}. \quad \text{(A.8)}$$

Step 6. If $\left\|\left(\mathbf{p}^{(n+1)}, \mathbf{y}^{(n+1)}\right) - \left(\mathbf{p}^{(n)}, \mathbf{y}^{(n)}\right)\right\| < \bar{\varepsilon}$, then stop; otherwise, set $n = n+1$ and go to Step 2.

In the above procedure, the difference parameters Δp and Δy are positive and also sufficiently small. $\beta^{(n)}$ (>0) is the step size in iteration n, e.g., a typical step size is $\beta^{(n)} = 1/(n+1)$. $\bar{\varepsilon}$ is a convergence tolerance and it is positive and sufficiently small. One can see that the transit operator need not know the function of in-vehicle congestion costs and the distribution of perceived travel cost errors to implement the above procedure.

Appendix C. Extension to the optimization of social/system cost

In this appendix, we extend the trial-and-error method to optimize the social/system cost of the transport system. The optimization problem is formulated as

$$\min_{(x_b, y)} \bar{V} = x_b \left(w(x_b, y) + t_b + g(x_b, y)\right) + (d - x_b) t_a(x_b) + k(y), \quad \text{(A.9)}$$

where the variable (x_b, y) is subject to $0 \le x_b \le d$ and $y \ge 0$. Compared with the optimization problem (5) for minimizing system time cost, the optimization objective of the optimization problem (A.9) includes not only the waiting and travel time cost $x_b\left(w(x_b, y) + t_b\right)$ of bus passengers and the travel time cost $(d - x_b)t_a(x_b)$ of auto users but also the in-vehicle congestion cost $x_b g(x_b, y)$ of bus passengers and the operating cost $k(y)$ of bus transit. The toll charges or financial subsidies are transferred between commuters and the manager of the transport system. Therefore, the toll charges or financial subsidies are excluded from the objective function.

We develop an iterative trial-and-error procedure for determining the transit fare (or subsidy) p_b and the bus run frequency y to optimize the social/system cost of the transport system without resorting to the function of in-vehicle congestion costs. The iterative trial-and-error procedure is summarized as follows.

Step 1. Let $p_b^{(0)}$ and $y^{(0)}$ (≥ 0) be the initial transit fare (or subsidy) and bus run frequency respectively, set an initial iterative step size of transit fare γ_{p0} (>0) and an initial iterative step size of bus run frequency γ_{y0} (>0), and set $n = 0$.

Step 2.1. Let $\tilde{p}_b^{(0)} = p_b^{(n)}$ and set $m = 0$.

Step 2.2. Observe the revealed number of bus users $\tilde{x}_b^{(m)}$ at the SUE state under the implementation of the transit fare (or subsidy) $\tilde{p}_b^{(m)}$ and the bus run frequency $y^{(n)}$.

Step 2.3. Set the transit fare (or subsidy) $\bar{p}_b^{(m)} = \tilde{p}_b^{(m)} + \Delta p$, the bus run frequency $y^{(n)}$ remains unchanged, and then observe the revealed number of bus users $\bar{x}_b^{(m)}$ at the SUE state.

Step 2.4. Compute the iterative direction $G_p^{(m)}$ of the transit fare (or subsidy) according to

$$G_p^{(m)} = \left(p_a - \tilde{p}_b^{(m)} - F^{-1}\left(\tilde{x}_b^{(m)}/d\right) - \frac{\tilde{x}_b^{(m)}}{df\left(F^{-1}\left(\tilde{x}_b^{(m)}/d\right)\right)} + d\frac{\mathrm{d}t_a\left(\tilde{x}_b^{(m)}\right)}{\mathrm{d}\tilde{x}_b^{(m)}} \right) \frac{\tilde{x}_b^{(m)} - \bar{x}_b^{(m)}}{\Delta p} + \tilde{x}_b^{(m)}. \quad (A.10)$$

Step 2.5. If $m < m_{p0}$, then $\gamma_p^{(m)} = \gamma_{p0}$; otherwise, set $\gamma_p^{(m)} = \gamma_p^{(m-1)}/2$.

Step 2.6. Update the transit fare (or subsidy) by the following formula

$$\tilde{p}_b^{(m+1)} = \begin{cases} \tilde{p}_b^{(m)} - \gamma_p^{(m)}, & \text{if } G_p^{(m)} < 0, \\ \tilde{p}_b^{(m)} + \gamma_p^{(m)}, & \text{otherwise.} \end{cases} \quad (A.11)$$

Step 2.7. If $\left|\tilde{p}_b^{(m+1)} - \tilde{p}_b^{(m)}\right| < \hat{\varepsilon}_p$, then set $p_b^{(n+1)} = \tilde{p}_b^{(m+1)}$ and go to Step 3.1; otherwise, set $m = m+1$ and go to Step 2.2.

Step 3.1. Let $\tilde{y}^{(0)} = y^{(n)}$ and set $m = 0$.

Step 3.2. Observe the revealed number of bus users $\tilde{x}_b^{(m)}$ at the SUE state under the implementation of the bus run frequency $\tilde{y}^{(m)}$ and the transit fare (or subsidy) $p_b^{(n+1)}$.

Step 3.3. Set the bus run frequency $\bar{y}^{(m)} = \tilde{y}^{(m)} + \Delta y$, the transit fare (or subsidy) $p_b^{(n+1)}$ remains unchanged, and then observe the revealed number of bus users $\bar{x}_b^{(m)}$ at the SUE state.

Step 3.4. Compute the iterative direction $G_y^{(m)}$ of the bus run frequency according to

$$G_y^{(m)} = \left(p_a - p_b^{(n+1)} - F^{-1}\left(\tilde{x}_b^{(m)}/d\right) - \frac{\tilde{x}_b^{(m)}}{df\left(F^{-1}\left(\tilde{x}_b^{(m)}/d\right)\right)} + d\frac{\mathrm{d}t_a\left(\tilde{x}_b^{(m)}\right)}{\mathrm{d}\tilde{x}_b^{(m)}} \right)$$

$$\times \frac{\tilde{x}_b^{(m)} - \bar{x}_b^{(m)}}{\Delta y} - \frac{\mathrm{d}k\left(\tilde{y}^{(m)}\right)}{\mathrm{d}\tilde{y}^{(m)}}. \quad (A.12)$$

Step 3.5. If $m < m_{y0}$, then $\gamma_y^{(m)} = \gamma_{y0}$; otherwise, set $\tilde{\gamma} = \left|\tilde{y}^{(m)} - \tilde{y}^{(m-1)}\right|$ and $\gamma_y^{(m)} = \tilde{\gamma}/2$.

Step 3.6. Update the bus run frequency by the following formula

$$\tilde{y}^{(m+1)} = \begin{cases} \max\left\{\tilde{y}^{(m)} - \gamma_y^{(m)}, 0\right\}, & \text{if } G_y^{(m)} < 0, \\ \tilde{y}^{(m)} + \gamma_y^{(m)}, & \text{otherwise.} \end{cases} \quad (A.13)$$

Step 3.7. If $\left|\tilde{y}^{(m+1)} - \tilde{y}^{(m)}\right| < \hat{\varepsilon}_y$, then set $y^{(n+1)} = \tilde{y}^{(m+1)}$ and go to Step 4; otherwise, set $m = m+1$ and go to Step 3.2.

Step 4. If $\left\|\left(p_b^{(n+1)}, y^{(n+1)}\right) - \left(p_b^{(n)}, y^{(n)}\right)\right\| < \hat{\varepsilon}$, then stop; otherwise, set $n = n+1$ and go to Step

2.1.

We call the above trial-and-error implementation the trial-and-error social/system cost optimization (TESCO) procedure. In the TESCO procedure, the difference parameters Δp and Δy are positive and also sufficiently small. The operator $\max\{\cdot, 0\}$ is used to guarantee the feasibility of y after its update in formula (A.13). $\hat{\varepsilon}_p$, $\hat{\varepsilon}_y$, and $\hat{\varepsilon}$ are three convergence tolerances and they are positive and sufficiently small. One can see that the manager need not know the function of in-vehicle congestion costs to implement the above trial-and-error procedure.

The social/system cost $\bar{V}$ in the optimization problem (A.9) is determined by the variables x_b and y. By Proposition 1, the variable x_b is a function of p_b and y, and hence the social/system cost $\bar{V}$ is finally determined by p_b and y. The partial derivatives of $\bar{V}$ with respect to p_b and y can be formulated as

$$\frac{\partial \bar{V}}{\partial p_b} = \left(w(x_b, y) + t_b + g(x_b, y) + x_b \left(\frac{\partial w(x_b, y)}{\partial x_b} + \frac{\partial g(x_b, y)}{\partial x_b} \right) - t_a(x_b) + (d - x_b) \frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} \right) \frac{\partial x_b}{\partial p_b} \quad \text{and} \tag{A.14}$$

$$\frac{\partial \bar{V}}{\partial y} = \left(w(x_b, y) + t_b + g(x_b, y) + x_b \left(\frac{\partial w(x_b, y)}{\partial x_b} + \frac{\partial g(x_b, y)}{\partial x_b} \right) - t_a(x_b) + (d - x_b) \frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} \right) \frac{\partial x_b}{\partial y} + x_b \left(\frac{\partial w(x_b, y)}{\partial y} + \frac{\partial g(x_b, y)}{\partial y} \right) + \frac{\mathrm{d}k(y)}{\mathrm{d}y}, \tag{A.15}$$

where $\partial x_b / \partial p_b$ and $\partial x_b / \partial y$ are given by expressions (11) and (12), respectively. To optimize the social/system cost $\bar{V}$ by using a gradient-based algorithm, the decision variables (p_b, y) could be updated in the negative gradient direction $-\left(\partial \bar{V} / \partial p_b, \partial \bar{V} / \partial y \right)$ in each iteration. However, the in-vehicle congestion costs of commuters in bus carriages are unknown, and hence the negative gradient direction $-\left(\partial \bar{V} / \partial p_b, \partial \bar{V} / \partial y \right)$ cannot be computed by formulae (A.14) and (A.15) to solve the optimization problem (A.9).

It follows from the SUE condition (3), definition (4), and expressions (11) and (12) that

$$g(x_b, y) = t_a(x_b) + p_a - w(x_b, y) - t_b - p_b - F^{-1}\left(x_b / d \right), \tag{A.16}$$

$$\frac{\partial g(x_b, y)}{\partial x_b} = -\frac{1}{\partial x_b / \partial p_b} - \frac{1}{df\left(F^{-1}\left(x_b / d \right) \right)} + \frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} - \frac{\partial w(x_b, y)}{\partial x_b}, \text{ and} \tag{A.17}$$

$$\frac{\partial g(x_b, y)}{\partial y} = \frac{\partial x_b / \partial y}{\partial x_b / \partial p_b} - \frac{\partial w(x_b, y)}{\partial y}. \tag{A.18}$$

Substituting formulae (A.16) to (A.18) into (A.14) and (A.15) generates

$$\frac{\partial \bar{V}}{\partial p_b} = \left(p_a - p_b - F^{-1}\left(x_b/d\right) - \frac{x_b}{df\left(F^{-1}\left(x_b/d\right)\right)} + d\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} \right)\frac{\partial x_b}{\partial p_b} - x_b \quad \text{and} \tag{A.19}$$

$$\frac{\partial \bar{V}}{\partial y} = \left(p_a - p_b - F^{-1}\left(x_b/d\right) - \frac{x_b}{df\left(F^{-1}\left(x_b/d\right)\right)} + d\frac{\mathrm{d}t_a(x_b)}{\mathrm{d}x_b} \right)\frac{\partial x_b}{\partial y} + \frac{\mathrm{d}k(y)}{\mathrm{d}y}. \tag{A.20}$$

The difference parameters Δp and Δy are positive and also sufficiently small. Thus, by comparing formulae (A.10) and (A.19) and comparing (A.12) and (A.20), it can be seen that Steps 2.4 and 3.4 of the TESCO procedure are adopted to estimate $-\partial \bar{V}/\partial p_b$ and $-\partial \bar{V}/\partial y$ without resorting to the function g.

In Steps 2.1 to 2.7, the bus run frequency $y^{(n)}$ remains unchanged and the bus fare (or subsidy) p_b is updated in an approximate descent direction $G_p^{(m)}$ of $\bar{V}$ with respect to p_b until a minimum point $p_b^{(n+1)}$ is generated. After the iterative direction $G_p^{(m)}$ of the transit fare (or subsidy) is determined in Step 2.4, the iterative step size $\gamma_p^{(m)}$ is generated in Step 2.5 in the same way as the iterative step size $\alpha^{(n)}$ in the TESTO procedure in Section 2.3. The number m_{p0} in Step 2.5 records the iteration, in which the iterative direction changes for the first time. In each of the first $m_{p0}-1$ iterations, the iterative step size $\gamma_p^{(m)}$ takes the fixed value γ_{p0} to obtain the interval in which the optimal solution lies. In iteration m_{p0} and subsequent iterations, a bisection iterative method is used to find the optimal solution in the interval.

In Steps 3.1 to 3.7, given the transit fare (or subsidy) $p_b^{(n+1)}$, the bus run frequency y is updated in an approximate descent direction $G_y^{(m)}$ of $\bar{V}$ with respect to y until a minimum point $y^{(n+1)}$ is obtained. After the iterative direction $G_y^{(m)}$ of the bus run frequency is determined in Step 3.4, the iterative step size $\gamma_y^{(m)}$ is generated in Step 3.5 in the same way as the iterative step size $\beta_y^{(m)}$ in the TEPM procedure in Section 3.2. The number m_{y0} in Step 3.5 records the iteration, in which the iterative direction changes for the first time. In each of the first $m_{y0}-1$ iterations, the iterative step size $\gamma_y^{(m)}$ takes the fixed value γ_{y0} to obtain the interval of containing the optimal solution. In iteration m_{y0} and subsequent iterations, a bisection iterative method is used to find the optimal solution in the interval.

In this way, in each iteration of the TESCO procedure, the social/system cost $\bar{V}$ decreases by a certain value until the iterative trajectory gets to a globally or locally minimum point.

References

Ahn, K., 2009. Road pricing and bus service policies. Journal of Transport Economics and Policy 43 (1), 25–53.

Arnott, R., Yan, A., 2000. The two-mode problem: Second-best pricing and capacity. Review of Urban & Regional Development Studies 12 (3), 170–199.

Cantarella, G.E., Velonà, P., Watling, D.P., 2015. Day-to-day dynamics & equilibrium stability in a two-mode transport system with responsive bus operator strategies. Networks and Spatial Economics 15 (3), 485–506.

David, Q., Foucart, R., 2014. Modal choice and optimal congestion. Regional Science and Urban Economics 48, 12–20.

de Palma, A., Lindsey, R., Monchambert, G., 2017. The economics of crowding in rail transit. Journal of Urban Economics 101, 106–122.

Gonzales, E.J., 2015. Coordinated pricing for cars and transit in cities with hypercongestion. Economics of Transportation 4 (1–2), 64–81.

Gonzales, E.J., Daganzo, C.F., 2012. Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. Transportation Research Part B 46 (10), 1519–1534.

Guo, R.Y., Szeto, W.Y., 2018. Day-to-day modal choice with a Pareto improvement or zero-sum revenue scheme. Transportation Research Part B 110, 1–25.

Han, D., Yang, H., 2009. Congestion pricing in the absence of demand functions. Transportation Research Part E 45 (1), 159–171.

Haywood, L., Koning, M., 2015. The distribution of crowding costs in public transport: New evidence from Paris. Transportation Research Part A 77, 182–201.

Huang, H.J., 2000. Fares and tolls in a competitive system with transit and highway: The case with two groups of commuters. Transportation Research Part E 36 (4), 267–284.

Huang, H.J., 2002. Pricing and logit-based mode choice models of a transit and highway system with elastic demand. European Journal of Operational Research 140 (3), 562–570.

Huang, H.J., Tian, Q., Yang, H., Gao, Z.Y., 2007. Modal split and commuting pattern on a bottleneck-constrained highway. Transportation Research Part E 43 (5), 578–590.

Kraus, M., 2003. A new look at the two-mode problem. Journal of Urban Economics 54 (3), 511–530.

Li, M.Z.F., 2002. The role of speed-flow relationship in congestion pricing implementation with an application to Singapore. Transportation Research Part B 36 (8), 731–754.

Li, X., Yang, H., 2016. Dynamics of modal choice of heterogeneous travelers with responsive transit services. Transportation Research Part C 68, 333–349.

Li, Z., Hensher, D.A., 2011. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. Transport Policy 18 (6), 880–887.

Li, Z.C., Lam, W.H.K., Wong, S.C., 2012. Modeling intermodal equilibrium for bimodal transportation system design problems in a linear monocentric city. Transportation Research Part B 46 (1), 30–49.

Liu, W., Geroliminis, N., 2017. Doubly dynamics for multi-modal networks with park-and-ride and adaptive pricing. Transportation Research Part B 102, 162–179.

Liu, W., Li, X., Zhang, F., Yang, H., 2017. Interactive travel choices and traffic forecast in a doubly dynamical system with user inertia and information provision. Transportation Research Part C 85, 711–731.

Liu, W., Zhang, F., Yang, H., 2016. Managing morning commute with parking space constraints in the case of a bi-modal many-to-one network. Transportmetrica A: Transport Science 12 (2), 116–141.

Parry, I.W.H., Small, K.A., 2009. Should urban transit subsidies be reduced? American Economic Review 99 (3), 700-724.

Pel, A.J., Bel, N.H., Pieters, M., 2014. Including passengers' response to crowding in the Dutch national train passenger assignment model. Transportation Research Part A 66, 111–126.

Pels, E., Verhoef, E.T., 2007. Infrastructure pricing and competition between modes in urban transport. Environment and Planning A 39 (9), 2119–2138.

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

Tirachini, A., Hensher, D.A., 2012. Multimodal transport pricing: First best, second best and extensions to non-motorized transport. Transport Reviews 32 (2), 181–202.

Trench, W.F., 2003. Introduction to Real Analysis. Prentice Hall, New Jersey.

van den Berg, V.A.C., Verhoef, E.T., 2014. Congestion pricing in a road and rail network with heterogeneous values of time and schedule delay. Transportmetrica A: Transport Science 10 (5), 377–400.

van der Weijde, A.H., Verhoef, E.T., van den Berg, V.A.C., 2013. Competition in multi-modal transport networks: A dynamic approach. Transportation Research Part B 53, 31–44.

Wang, S., Zhang, W., Qu, X., 2018. Trial-and-error train fare design scheme for addressing boarding/alighting congestion at CBD stations. Transportation Research Part B 118, 318–335.

Wang, X., Yang, H., 2012. Bisection-based trial-and-error implementation of marginal cost

pricing and tradable credit schemes. Transportation Research Part B 46 (9), 1085–1096.
Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: Evidence and lessons from British experience. Transport Reviews 31 (3), 379–398.
Wu, W.X., Huang, H.J., 2014. Equilibrium and modal split in a competitive highway/transit system under different road-use pricing strategies. Journal of Transport Economics and Policy 48 (1), 153–169.
Yang, H., Meng, Q., Lee, D.H., 2004. Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions. Transportation Research Part B 38 (6), 477–493.
Yang, H., Xu, W., He, B.S., Meng, Q., 2010. Road pricing for congestion control with unknown demand and cost functions. Transportation Research Part C 18 (2), 157–175.
Zhang, F., Yang, H., Liu, W., 2014. The Downs–Thomson Paradox with responsive transit service. Transportation Research Part A 70, 244–263.
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. Transportation Research Part B 72, 77–92.