Title	Transit assignment: approach-based formulation, extragradient method, and paradox
Author(s)	Szeto, WY; Jiang, Y
Citation	Transportation Research Part B: Methodological, 2014, v. 62, p. 51-76
Issued Date	2014
URL	http://hdl.handle.net/10722/202644
Rights	NOTICE: this is the author's version of a work that was accepted for publication in Transportation Research Part B: Methodological. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Transportation Research Part B: Methodological, 2014, v. 62, p. 51-76. DOI: 10.1016/j.trb.2014.01.010

Transit Assignment: Approach-based Formulation, Extragradient Method, and Paradox

W.Y. Szeto, Y. Jiang

Department of Civil Engineering, the University of Hong Kong, Pokfulam Road, Hong Kong, PR China

Abstract

This paper uses the concept of approach proportion to propose a novel variational inequality (VI) formulation of the frequency-based transit assignment problem. The approach proportion is defined as the proportion of passengers leaving a node through its outgoing link. To solve the VI problem, an extragradient method with adaptive stepsizes is developed. Unlike the existing methods for solving the frequency-based transit assignment problem, the convergence of our method requires only the pseudomonotone and Lipschitz continuous properties of the mapping function in VI, and it is not necessary for the Lipschitz constant to be known in advance. A Braess-like paradox in transit assignment is also discussed, where providing new lines to a transit network or increasing the frequency of an existing line may not improve the system performance in terms of expected total system travel cost. Various numerical examples are given to illustrate some paradox phenomena and to test the performance of our proposed algorithm.

Keywords: frequency-based transit assignment, approach proportion, extragradient method, variational inequality, paradox

1. INTRODUCTION

The transit assignment problem has received considerable attention, as finding solutions for this issue is essential both for designing or managing transit networks and for evaluating transit system performance. Many transit assignment models have been developed, and some of the earliest can be traced back to Dial (1967) and Fearnside and Draper (1971). However, these early models did not consider the route choice problem of passengers at stops served by several competing lines—known as the common line problem.

The first model to handle the common line problem was developed by Chriqui and Robillard (1975). By assuming that passengers are willing to minimise their individual expected travel cost, a hyperbolic model was solved through finding an optimal set of attractive lines directly serving two locations. This idea was further generalised to the optimal strategy concept (Spiess, 1984; Spiess and Florian, 1989). According to this concept, a strategy associated with a node is defined by a set of rules that, when applied, allow a passenger to travel efficiently from that node to his/her destination. A strategy specifies a set of attractive lines at every node, and hence an ordered set of successor nodes. An optimal strategy is defined as a strategy that minimises the passenger's expected travel time. The behavioural assumption used by Spiess (1984) and Spiess and Florian (1989) was that passengers use their individual optimal strategies in travelling. Assuming that this is true, Spiess and Florian (1989) proposed a linear programming model to tackle the common line problem, and provided their proof that their model's dual solution satisfied the user equilibrium conditions.

Later, two modelling streams were derived from the abovementioned behavioural assumption by using different network representations: the hyperpath graph representation (Nguyen and Pallottino, 1988; Wu *et al*., 1994; Cominetti and Correa, 2001; Cortés *et al*., 2013), and the route-section representation (de Cea and Fernández, 1993; Lam *et al*., 1999; Li *et al*., 2009; Lam *et al*., 2002; Szeto *et al*., 2011, 2013). The hyperpath graph representation captures a strategy by using a hyperpath. The route-section representation aggregates common lines into sections, and a sequence of sections forms a route. Hence, each route in the route-section approach can be seen as a special case of travel strategy.

Although both of these modelling representations are based on the same behavioural assumption, they have different pros and cons. The merit of the hyperpath graph representation is that the optimal sets of attractive lines can be easily determined, but the cost involved is the requirement to create more boarding and alighting nodes. The route-section

representation always reduces the numbers of links required to form the network when the number of common lines is large. Moreover, this representation allows the transit assignment problem to be easily solved by using available algorithms. However, the optimal set of attractive lines on each section has to be determined before aggregating attractive common lines.

These two modelling streams adopt similar methods to handle the in-vehicle congestion issue. These methods can be classified into two approaches, namely the capacity constraint approach and the congestion cost function approach. The difference between these approaches is that passengers are forbidden to board a fully occupied transit vehicle in the capacity constraint approach, but they are permitted to do so in the congestion cost function approach. The capacity constraint approach is more realistic, but its resultant models are normally solved by the method of successive averages (MSA), which can guarantee convergence only under some conditions, and these conditions for convergence may not be satisfied by the models themselves. In addition, there may be no solution for the problems that result from insufficient capacity. One advantage of the congestion cost function approach is that a solution must exist to the resultant transit assignment problem under a very mild assumption (*e.g.*, Szeto *et al.*, 2013). However, the models that result from this approach may allow link flows to be greater than the link capacities, which is unrealistic.

Most of the abovementioned models are developed from the frequency-based approach, in which frequency is assumed to follow certain distributions to approximate the average waiting and travel times. However, during the last 20 years a schedule-based approach has been proposed for modelling detailed arrivals and departures (Tong and Wong, 1999; Poon *et al.*, 2004; Hamdouch *et al.*, 2011; Nuzzolo *et al.*, 2012). In general, the frequency-based approach is more suitable for strategic long-term planning, and the schedule-based approach can best be used for modelling daily operations.

Most of the abovementioned models have adopted link flows, path flows or hyperpath flows as decision variables. When congestion effects are considered, the resultant models are often solved by methods that require strong conditions to be satisfied for guaranteeing convergence. For example, the symmetric linear method (*e.g.*, Wu *et al.*, 1994) requires that the link cost function must be strictly monotone[1] for convergence. The methods of de Cea and Fernández (1993) (*i.e.*, the diagonalisation method) and of Szeto *et al.* (2013) (*i.e.*, self-

[1] A vector function $\boldsymbol{F}$ is strictly monotone on a non-empty set C if for all $\mathbf{x},\mathbf{y}\in C, \mathbf{x}\neq\mathbf{y}$, $(\mathbf{y}-\mathbf{x})^T\left(\mathbf{F}(\mathbf{y})-\mathbf{F}(\mathbf{x})\right)>0$.

adaptive projection and contraction algorithm with column generation) have assumed monotone[2] mapping to ensure convergence. Kurauchi *et al.* (2003), Cepeda *et al*. (2006), Sumalee *et al*. (2009, 2010), Schmöcker *et al*. (2011), and Cortés *et al*. (2013) all adopted the MSA, whose convergence requires the cost function to be strictly pseudo-contractive (Johnson, 1972). However, these conditions may not always be satisfied, especially when asymmetric link cost functions are used in transit assignment.

This paper proposes a link-based variational inequality (VI) formulation, which can be transformed into an approach-based formulation. This proposed formulation is based on the concept of approach proportion. To solve the approach-based formulation, we use an extragradient projection method, also known as the double projection method. The convergence of the algorithm only requires mild assumptions, *i.e*., pseudomonotone[3] and Lipschitz continuous properties. Moreover, it is not necessary to know the Lipschitz constant in advance. This algorithm, however, cannot be used directly to solve our proposed formulation, and some modifications of the cost and flow updating algorithms are required. Hence, we also propose a cost and a flow updating scheme.

The proposed formulation can also be used to evaluate the performance of a transit network design. In fact, one important application of any transit assignment model is to evaluate network design strategies, such as proposals to improve the system performance through new transit itineraries or adjustments to service frequency. To the best of our knowledge, little effort has been spent on investigating whether a paradox actually exists in transit assignment, even though there are many studies concerning paradoxes in traffic assignment, and there are similarities between traffic assignment and transit assignment in terms of formulation approaches. Only Cominetti and Correa (2001) have actually revealed a paradox, showing that the transit time will not be affected under a certain range of demand increments. However, unlike Cominetti and Correa's work, in which the changes occur at the demand side, our paper focuses on the changes in the supply side of transit networks, such as routes and frequency. A small network is created, and various scenarios are tested to investigate the existence of the paradox and the roles that different parameters or factors play in its occurrence. Our numerical results verify that providing a new transit line and increasing transit frequency may fail to improve, and may even deteriorate the network performance in

---

[2] A vector function $\boldsymbol{F}$ is monotone on a non-empty set C if for all $\mathbf{x},\mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}$, $(\mathbf{y}-\mathbf{x})^T\left(\mathbf{F}(\mathbf{y})-\mathbf{F}(\mathbf{x})\right) \geq 0$.

[3] A vector function $\boldsymbol{F}$ is pseudomonotone on a non-empty set C if for all $\mathbf{x},\mathbf{y} \in C$, $(\mathbf{y}-\mathbf{x})^T\mathbf{F}(\mathbf{x}) \geq 0$ implies that $(\mathbf{y}-\mathbf{x})^T\mathbf{F}(\mathbf{y}) \geq 0$.

terms of expected total system travel cost (including in-vehicle travel time cost and waiting time cost).

The contributions of this paper include the following:

- Identifying and illustrating the Braess-like paradox in transit assignment, by which providing a new transit line or increasing service frequency may not necessarily enhance the system performance.
- Identifying the factors that affect the occurrence of this paradox, and illustrating their effects via examples.
- Proposing link-based equilibrium conditions for transit assignment, and proposing a link-based VI formulation.
- Proposing a novel approach-based VI formulation to represent the frequency-based transit assignment problem, in which the congestion effect is tackled via the congestion cost function approach. This approach provides an alternative way to view and formulate transit assignment.
- Proposing use of an extragradient method with adaptive stepsizes to solve the VI problem. Compared with existing methods, the proposed algorithm requires only mild assumptions to converge.
- Proposing algorithms to update flows and costs for evaluations of the mapping function in the VI formulation, with discussions of the computational complexity involved.
- Demonstrating the effects of different parameters in the extragradient method via various experiments.

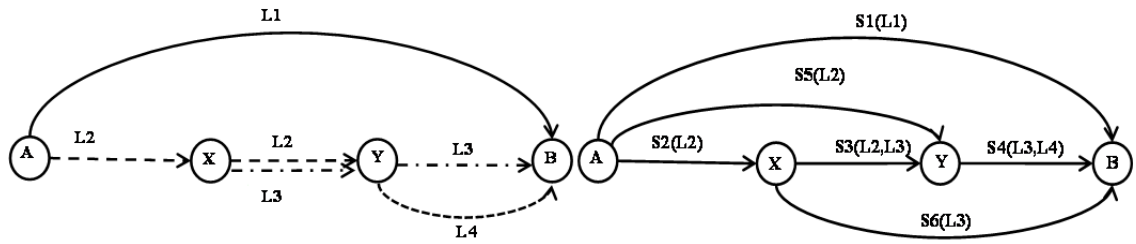
The remainder of this paper proceeds as follows: Section 2 introduces the route-section approach to representing transit networks, followed by the notations and assumptions used in this paper. The section then presents the path-based formulation and the link-based formulation, which is further modified into the link-based VI formulation. In conclusion, an approach-based VI formulation is derived. The solution algorithm is detailed in Section 3. Section 4 depicts a series of experiments to illustrate the transit paradox problem and tests the performance of the algorithm. Finally, Section 5 gives our conclusions.

2. FORMULATION

2.1. Network Presentation

We consider a general transit network consisting of a set of transit lines with a set of transit stops (nodes), where passengers can board, alight or transfer. This network is converted into a route-section network, as explained by de Cea and Fernández (1993). The idea of the route-section representation is to classify passengers waiting at transit stops (including origins and transfers) into different groups according to their next alighting nodes (which may be their next transfer locations or their destinations). For the passengers boarding at the same transit stop and alighting at the same stop, a link called a 'section' is created to connect the boarding and alighting stops. Different lines serving the same pair of nodes (*i.e.*, common lines) are then aggregated into one section.

To illustrate the route-section representation, a small transit network containing four nodes and four transit lines is given in Figure 1(a), and its route-section representation is given in Figure 1(b). In Figure 1(a), four bus lines are denoted as L1, L2, L3, and L4, and in Figure 1(b), six sections are denoted as S1, S2, S3, S4, S5, and S6. The notation S3(L2, L3) means that L2 and L3 are attractive lines in section S3. In Figure 1(a), passengers at node A who board L2 can only alight at either node X or node Y. Hence, two sections (S2 and S5) are created in Figure 1(b) to connect node A to node X, and node A to node Y. The line number in a pair of brackets next to the section name is the transit line in that section. Only L2 can take the passengers from node A to nodes X and Y directly. Therefore, L1 is inside the pairs of brackets next to S2 and S3. Another example is that passengers waiting at node X can only board either L2 or L3, and alight at node Y. Therefore, a section called S3, connecting X to Y, is created in Figure 1(b). This section contains the two common lines, and is based on the assumption that both L2 and L3 are attractive. The set of attractive lines in each section is determined using the method proposed by Chriqui and Robillard (1975). Only attractive lines are considered in this representation.



(a) A small transit network (b) Route section representation

Figure 1 A small transit network and its route-section representation

2.2. Notations

The following notations are used throughout this paper.

2.2.1. Sets

N	set of nodes or stops in the transit network;
L	set of lines in the transit network;
S	set of sections or approaches;
A_s	set of attractive lines in section s;
A_i^+ (A_i^-)	set of sections emanating from (going into) node i;
R, D, W	set of origins, destinations, and OD pairs, respectively;
Q^{rd}	set of paths between an OD pair $(r,d) \in W$;
Ψ^d	set of nodes following a topological ordering, with destination d as the last node and dimension $\lvert\Psi^d\rvert$.

2.2.2. Indices

r, d, l, p	indices of origin, destination, line, and path, respectively;
i, j, j'	indices of nodes;
s, s', m	indices of route sections;
λ_x^l	the xth stop on line l;
$\lambda'^l_{x(s)}$	the xth stop on line l, which is also the tail node of section s;
$\lambda''^l_{y(s)}$	the yth stop on line l, which is also the head node of section s;
$t(s)$, $h(s)$	the tail and head nodes of section s, respectively, where $s = (t(s), h(s))$;
ψ_u^d	the uth node in the topological set; $u = 1$ means that the node is the farthest away from destination d;
k, k', k''	iteration numbers.

2.2.3. Parameters

μ_T (μ_W)	value of travel time (waiting time);
t_s^l	in-vehicle travel time of line l on section s;
f_s^l	frequency of line l on section s;
w_s^l	relative frequency of line l on section s;

κ^l capacity of a single vehicle of line *l*;

α unit conversion parameter;

g_r^d demand of OD pair *rd*;

δ_s^m competing section indicator; $\delta_s^m = 1$, if section *m* is a competing section of section *s*; otherwise, $\delta_s^m = 0$;

ϑ_p^s element in the path-section incidence matrix. If route *p* contains section *s*, $\vartheta_p^s = 1$; otherwise $\vartheta_p^s = 0$;

a, b, n additional waiting time function parameters;

ϖ_s additional waiting time function parameter of section *s*;

n^l number of stops in line *l*;

$\nu, \mu, \lambda, \overline{\beta}$ parameters in the extragradient method;

β_k, τ_k stepsizes in iteration *k*.

2.2.4. *Decision variables*

v_s^d number of passengers travelling on section *s* toward destination *d*;

$\mathbf{v}$ vector of ($v_s^d, \forall s \in S, d \in D$) with dimension $|S| \times |D|$;

$\tilde{f}_p^{rd}$ path flow from origin *r* to destination *d* via path *p*;

$\mathbf{f}$ vector of ($\tilde{f}_p^{rd}, \forall p \in Q^{rd}, (r, d) \in W$) with dimension $|P| \times |W|$;

α_s^d proportion of passengers leaving node *t*(*s*) via approach *s* to destination *d*;

$\boldsymbol{\alpha}$ vector of ($\alpha_s^d, \forall s \in S, d \in D$) with dimension $|S| \times |D|$.

2.2.5. *Functions of decision variables or parameters*

c_s expected cost associated with section *s*;

t_s mean in-vehicle travel time on section *s*;

ω_s mean waiting time for passengers boarding the first arriving vehicle on section *s*;

ϕ_s additional waiting time for passengers boarding on section *s* due to insufficient capacity;

v_s number of passengers on section *s*;

v_{sl}^d number of passengers using line *l* on section *s* toward destination *d*;

$\overline{v}_s$ number of passengers on the competing sections of section *s*;

$\tilde{\pi}_p^{rd}$	expected travel cost from origin *r* to destination *d* via path *p*;
h	vector of ($\tilde{\pi}_p^{rd}, \forall p \in Q^{rd}, (r,d) \in W$) with dimension $\lvert P \rvert \times \lvert W \rvert$;
π^{id}	minimum expected travel cost between nodes *i* and *d*;
π_s^{id}	minimum expected travel cost from node *i* to node *d* via section or approach *s*;
$\boldsymbol{\pi}$	vector of ($\pi_s^{id}, \forall s \in S, i \in N, d \in D$) with dimension $\lvert S \rvert \times \lvert N \rvert \times \lvert D \rvert$;
q_i^d	number of passengers leaving node *i* toward destination *d*.

2.3. Assumptions

As in the literature (*e.g.*, Spiess and Florian, 1989; de Cea and Fernández, 1993), the following classical assumptions are made throughout this paper. A1) Passengers are assumed to arrive at transit stops randomly. A2) A passenger waiting at a transfer node considers a set of attractive lines before boarding, and he/she boards the first arriving bus if possible. A3) The waiting time for a transit line on a link is independent of the waiting times for other lines on the same section. A4) Vehicle headways are assumed to follow exponential distribution. A5) The passenger selects the transit route that minimises his/her expected travel cost. A6) The travel demand between each origin-destination (OD) pair in the system is assumed to be known and fixed. This assumption is reasonable for strategic planning when the day-to-day variation, especially during the peak hour period, is small or negligible. A7) For simplicity, the capacity of each transit vehicle is assumed to be the same. However, there is no conceptual difficulty in extending the formulation to a scenario in which vehicles of different capacities traverse different routes.

Based on the preceding route-section representation, notations, and assumptions, the cost components and the formulations are presented below.

2.4. Cost Components

The expected cost associated with section *s*, c_s is defined as

$$c_s = \mu_T t_s + \mu_W \omega_s + \mu_W \phi_s, \quad \forall s \in S. \tag{1}$$

The first term on the right-hand side represents the mean in-vehicle travel time cost. The second term is the mean waiting time cost, which represents the monetary value of the mean waiting time for the first vehicle to arrive. The third term denotes the perceived congestion

cost, which can be interpreted as the monetary value of the additional waiting time due to in-vehicle congestion. The following three sub-sections describe the individual cost components.

2.4.1. *Mean in-vehicle travel time cost*

The mean in-vehicle travel time cost of a section is equal to the product of the value of time and the mean in-vehicle travel time over that section. The mean in-vehicle travel time over a section is the weighted average of the in-vehicle travel times of all of the attractive lines in that section. Mathematically, the mean in-vehicle travel time t_s over section s can be expressed as

$$t_s = \sum_{l \in A_s} w_s^l t_s^l, \quad \forall s \in S, \tag{2}$$

where w_s^l is defined by

$$w_s^l = \frac{f_s^l}{\sum_{j \in A_s} f_s^j}, \quad \forall s \in S, l \in A_s. \tag{3}$$

2.4.2. *Mean waiting time cost*

The mean in-vehicle travel time cost of a section is the product of the waiting time required on a section and the value of the waiting time. The waiting time on a section is defined as the time that a passenger waits at a transit stop (*i.e.*, the tail node of a link) for the arrival of the first vehicle belonging to the set of attractive lines. Under assumptions A1) – A4), the mean waiting time can be expressed as

$$\omega_s = \frac{\alpha}{\sum_{l \in A_s} f_s^l}, \quad \forall s \in S. \tag{4}$$

In this case, the unit of frequency is vehicles/hour, and that of waiting time is in minutes, $\alpha = 60$ min/h.

2.4.3. *Perceived congestion cost*

The perceived congestion cost associated with section s models the additional waiting time cost due to in-vehicle congestion on the section. Such cost for the passengers arriving at node $t(s)$ and using section s is a function of the section flow, but this cost is also affected by the relative presence of two groups of passengers, as described in the literature.

One group is the group of passengers boarding before $t(s)$, using one or more attractive lines in section s, and alighting after $t(s)$. These are the passengers already on-board when the bus arrives at stop $t(s)$. They contribute to the flows on other sections, and directly affect the

waiting time of passengers who board at $t(s)$ and alight at $h(s)$. The greater the number of these passengers on these sections, the more congested the vehicles are, and the less remaining capacity there is for passengers boarding at $t(s)$. Hence, these sections are referred to as the competing sections of section s, and the flow on these competing sections is called the competing section flow of section s.

The other group of passengers consists of those arriving at stop $t(s)$ and using one or more attractive lines contained in section s, but who alight *after* $n(s)$. This group forms part of the flows on other sections, but also competes for the remaining capacity of one or more attractive transit lines with the passengers on section s (*i.e.*, they alight at $t(s)$ instead of after $t(s)$). Hence, these sections are also the competing sections of section s.

Recall that v_s is the flow, or the number of passengers per hour boarding the lines belonging to section s, and $\overline{v}_s$ represents the flows on the competing sections of section s. In that case, the flows on section s and its competing sections can mathematically be represented by

$$v_s = \sum_{d \in D} \sum_{l \in A_s} v_{sl}^d, \quad \forall s \in S, \text{ and} \tag{5}$$

$$\overline{v}_s = \sum_{d \in D} \sum_{m \in S, m \neq s} \delta_s^m \sum_{l \in A_s \cap A_m} v_{ml}^d, \quad \forall s \in S. \tag{6}$$

Equation (5) states that the flow on section s is obtained by aggregating all of the attractive line flows on section s. Equation (6) indicates that the competing section flow of section s is the sum of the attractive line flows on its competing sections.

The line flow v_{sl}^d is determined by

$$v_{sl}^d = v_s^d w_s^l, \quad \forall s \in S, l \in A_s, d \in D. \tag{7}$$

Equation (7) states that the flows on section s are distributed to the transit lines on that section based on the relative frequencies determined by equation (3).

To illustrate the concept of competing section flows and competing section indicators, we take section S3 in Figure 1(b) as an example. Node t(S3) refers to node X and h(S3) to node Y. The first group of passengers who compete with the passengers on S3 are the passengers on section S5, because section S5 starts before node X, ends after node X, and contains one of the attractive lines in S3 (*i.e.*, Line L2). Hence, S5 is a competing section of S3, and $\delta_3^5 = 1$. The second group of passengers who compete with the passengers on S3 is the group of those travelling on section S6 because S6 and S3 share the same boarding stop, and have a common attractive line (*i.e.*, Line L3). Hence, S6 is a competing section of S3, and $\delta_3^6 = 1$. After

identifying the competing sections, the competing section flow of S3 is obtained by combining these two groups of passengers. For example, $\overline{v}_3$ is calculated by $\overline{v}_3 = \sum_{d \in D} \left(v_{52}^d + v_{63}^d \right)$.

In addition, we should note that sections S3 and S6 are *mutually competing* (*i.e.*, $\delta_3^6 = \delta_6^3 = 1$). Both S3 and S6 emanate from the same nodes and contain L2. This implies that both the S3 and S5 passengers board L2 at the same stop. Hence, S3 is a competing section of S5, and vice versa, implying that the second condition of the competing section flow applies. However, sections S3 and S5 are not *mutually competing* (*i.e.*, $\delta_5^3 \neq \delta_3^5$) because $\delta_3^5 = 1$, but $\delta_5^3 = 0$. We conclude that $\delta_5^3 = 0$, as 1) section S3 do not start before node A, which is the tail node of S5 (*i.e.*, the passengers on S3 do not board transit lines before node A , and the S3 flow violates the first condition for competing section flow), and 2) S3 and S5 do not emanate from the same node, implying that the S3 flow violates the second condition for competing section flow.

Based on the preceding discussion, an additional waiting time function is developed and given in equation (8). This function is more general than that proposed by de Cea and Fernández (1993). Our equation introduces different weights to the flows on board and the flows waiting at the stop. Mathematically, the additional waiting time function for route section *s* is expressed as

$$\phi_s = \varpi_s \left(\frac{a v_s + b \overline{v}_s}{\sum_{l \in A_s} f_s^l \kappa^l} \right)^n, \quad \forall s \in S, \tag{8}$$

where the denominator is interpreted as the capacity of section *s*; calibration parameters *a*, *b*, ϖ_s and *n* are used to model different effects of various flows on additional waiting time and hence the congestion cost ($\mu_W \phi_s$), because the congestion cost due to waiting at stops may be higher than that due to in-vehicle congestion. These parameters are related to the passengers' perceptions concerning the level of congestion. A larger value means that the congestion level has a higher effect on the travel cost of passengers, leading to a higher congestion cost for a given ratio of flow to capacity. Based on equation (8), the congestion cost function can be expressed as $\mu_W \phi_s$.

2.5. Path-based User Equilibrium (UE) Formulation

2.5.1. Path cost

The UE can be defined using path flows and path costs. Path cost is a function of section flows, which in turn are a function of path flows. Given the path flow, $\tilde{f}_p^{rd}$, the section flows can then be determined by

$$v_s^d = \sum_{r\in R}\sum_{p\in Q^{rd}} \vartheta_p^s \tilde{f}_p^{rd}, \quad \forall s\in S, d\in D. \tag{9}$$

Based on the section flows, the expected section costs can be determined by (1) – (8). Given the expected section costs, the expected cost associated with path *p* between OD pair *rd* can be expressed as

$$\tilde{\pi}_p^{rd} = \sum_{s\in S} \vartheta_p^s c_s, \quad \forall p\in Q^{rd}, (r,d)\in W. \tag{10}$$

2.5.2. Path-based user equilibrium conditions

The path flow pattern is called the equilibrium route flow pattern if it satisfies the Wardropian conditions in transit networks, which are

$$\tilde{\pi}_p^{rd} \begin{cases} = \pi^{rd}, & \text{if } \tilde{f}_p^{rd} > 0; \\ \geq \pi^{rd}, & \text{if } \tilde{f}_p^{rd} = 0, \end{cases} \quad \forall p\in Q^{rd}, (r,d)\in W. \tag{11}$$

By definition, route flow must be non-negative and satisfy flow conservation at UE. These conditions are expressed as

$$\tilde{f}_p^{rd} \geq 0, \quad \forall p\in Q^{rd}, (r,d)\in W, \text{ and} \tag{12}$$

$$\sum_{p\in Q^{rd}} \tilde{f}_p^{rd} = g_r^d, \quad \forall (r,d)\in W. \tag{13}$$

By combining conditions (11) and (12), the path-based UE constraints can be represented as follows:

$$\begin{cases} (\tilde{\pi}_p^{rd} - \pi^{rd})\tilde{f}_p^{rd} = 0, \quad \forall p\in Q^{rd}, (r,d)\in W, \\ \tilde{\pi}_p^{rd} - \pi^{rd} \geq 0, \quad \forall p\in Q^{rd}, (r,d)\in W, \\ \tilde{f}_p^{rd} \geq 0, \quad \forall p\in Q^{rd}, (r,d)\in W. \end{cases} \tag{14}$$

2.5.3. Path-based VI formulation

The path-based UE transit assignment problem is to determine $\mathbf{f}^* = [\tilde{f}_p^{rd*}]$ to satisfy conditions (1) – (13) simultaneously. The superscript '*' refers to a solution of $\mathbf{f} = \left[\tilde{f}_p^{rd}\right]$ that

fulfils all of these conditions. This problem has been shown to be equivalent to the VI problem (see Florian and Spiess, 1983), that is, to determine an optimal vector $\mathbf{f}^*$ such that

$$\left(\mathbf{f}-\mathbf{f}^*\right)^{\mathrm{T}}\mathbf{h}\left(\mathbf{f}^*\right)\geq 0, \forall \mathbf{f}\in\Omega_f, \tag{15}$$

where $\Omega_f=\left\{\tilde{f}_p^{rd} \mid \tilde{f}_p^{rd}\geq 0,\ \forall(r,d)\in W, p\in Q^{rd},\ \sum_{p\in Q^{rd}}\tilde{f}_p^{rd}=g_r^d,\ \forall(r,d)\in W\right\}$ and $\mathbf{f}=\left[\tilde{f}_p^{rd}\right]$, $\mathbf{h}(\mathbf{f})=\left[\pi_p^{rd}\right]$.

2.6. Link-based User Equilibrium Formulation

The path-based formulation suffers from the computational burden of path enumeration when solving large network problems. To overcome this issue, one method is to use column generation techniques during the computation process. In that case, paths are only generated when necessary. The second method is to reformulate the problem into a link-based formulation, so that path enumeration is replaced by shortest path determination during the computation process. This section considers the second method.

2.6.1. Link-based equilibrium conditions

To reformulate the path-based transit assignment into a link-based assignment, we need the equivalent link-based UE conditions. In line with Ban *et al*. (2008), we propose the link-based user equilibrium conditions, which can be expressed as

$$\pi_s^{id}\begin{cases}=\pi^{id} & \text{if } v_s^d>0;\\ \geq\pi^{id} & \text{if } v_s^d=0,\end{cases}\qquad \forall i\in N, s\in A_i^+, d\in D, \tag{16}$$

where π_s^{id} and π^{id} are, respectively, defined by

$$\pi_s^{id}=\pi^{h(s)d}+c_s,\quad \forall i\in N, s\in A_i^+, d\in D, \text{ and} \tag{17}$$

$$\pi^{id}=\min_{s\in A_i^+}\pi_s^{id},\quad \forall i\in N, d\in D. \tag{18}$$

The function value π_s^{id} represents the minimum expected travel cost from node *i* to destination *d* via section *s*, and the value is obtained by adding the expected section cost to the minimum expected travel cost from the head node of section *s* to *d*. Condition (16) implies that if the flow entering section *s* from node *i* and going to destination *d* is positive (*i.e.*, $v_s^d>0$), then the expected travel cost from node *i* to *d* via section *s* equals the minimum expected path travel cost from *i* to *d*. In other words, section *s* is involved in the lowest

expected cost path from *i* to *d*. Otherwise, if no traveller enters section *s*, then the expected travel cost via *s* to *d* is not less than the minimum expected travel cost from *i* to *d*.

At UE, the following non-negativity constraint and flow conservation constraint must hold:

$$v_s^d \geq 0, \quad \forall s \in S, d \in D, \text{ and} \tag{19}$$

$$\sum_{s \in A_i^-} v_s^d + g_i^d = \sum_{s \in A_i^+} v_s^d, \quad \forall i \in N, d \in D. \tag{20}$$

By combining conditions (16) and (19), the link-based UE constraints can be represented in the following form:

$$\begin{cases} (\pi_s^{id} - \pi^{id}) v_s^d = 0, \quad \forall s \in A_i^+, i \in N, d \in D, \\ \pi_s^{id} - \pi^{id} \geq 0, \quad \forall s \in A_i^+, i \in N, d \in D, \\ v_s^d \geq 0, \quad \forall s \in A_i^+, i \in N, d \in D. \end{cases} \tag{21}$$

These link-based UE conditions are similar to those used for traffic assignment, except that only route sections containing the set of attractive lines are considered in the UE conditions for transit assignment. Therefore, for each OD pair, we can create a small network with only the route sections considered as attractive by the passengers being used to prove Proposition 1.

Proposition 1: The link-based UE conditions (21) imply the path-based UE conditions (14), and vice versa.

Proof. See Appendix I.

Proposition 1 basically means that if the link flow solution satisfies condition (21), then the corresponding path flow solution must also satisfy condition (14). The converse is also true.

2.6.2. Link-based VI formulation

The link-based UE transit assignment problem is to determine $\mathbf{v}^* = \left[v_s^{d*} \right]$ to satisfy constraints (1) – (8), (20), and (21). This link-based UE problem is shown to be equivalent to the VI problem of finding $\mathbf{v}^* = \left[v_s^{d*} \right]$ such that

$$\left(\mathbf{v} - \mathbf{v}^* \right)^{\mathbf{T}} \boldsymbol{\pi}\left(\mathbf{v}^* \right) \geq 0, \quad \forall \mathbf{v} \in \Omega_v, \tag{22}$$

where $\mathbf{v} = \left[v_s^d \right]$, $\boldsymbol{\pi}(\mathbf{v}) = \left[\pi_s^{id} \right]$, $\mathbf{v}^*$ stands for an UE solution of section flows, and $\Omega_v = \left\{ v_s^d \mid v_s^d \geq 0, \ \forall s, d, \ \sum_{s \in A_i^-} v_s^d + g_i^d = \sum_{s \in A_i^+} v_s^d, \ \forall i, d \right\}$. The proof of equivalence is given in Appendix II.

2.7. Approach-based Formulation

2.7.1. Approach-based equilibrium conditions

Alternatively, the above formulation can be reformulated using the concept of approach proportion, in which the *approach* of a node is defined by the route section coming out from that node, and an *approach proportion* is defined as the proportion of passengers leaving a node via the approach considered. Denote the approach proportion, α_s^d, where $s \in A_i^+$, as the proportion of passengers per hour leaving node *i* and going to destination *d* via route section *s*. Then, the proportion must satisfy the following conditions:

$$0 \le \alpha_s^d \le 1, \quad \forall s \in S, d \in D, \text{ and} \tag{23}$$

$$\sum_{s \in A_i^+} \alpha_s^d = 1, \quad \forall i \in N, d \in D. \tag{24}$$

Inequality (23) is the definitional constraint, and equation (24) states that the sum of all of the approach proportions coming out from node *i* equals 1.

Let q_i^d be the flow leaving node *i* and heading to destination *d*. Then, the section flow v_s^d can be replaced by $\alpha_s^d q_i^d$. Thereby, the link-based UE conditions (21) are equivalent to

$$\begin{cases} (\pi_s^{id} - \pi^{id})\alpha_s^d q_i^d = 0, \quad \forall s \in A_i^+, i \in N, d \in D, \\ \pi_s^{id} - \pi^{id} \ge 0, \quad \forall s \in A_i^+, i \in N, d \in D, \\ \alpha_s^d q_i^d \ge 0, \quad \forall s \in A_i^+, i \in N, d \in D. \end{cases} \tag{25}$$

Moreover, the flow conservation constraint (20) can be expressed as

$$\sum_{s \in A_i^-} \alpha_s^d q_{t(s)}^d + g_i^d = \sum_{s \in A_i^+} \alpha_s^d q_i^d, \quad \forall i \in N, d \in D. \tag{26}$$

By definition, the outflow q_i^d is non-negative; hence, condition (25) can be reduced to

$$\begin{cases} (\pi_s^{id} - \pi^{id})\alpha_s^d = 0, \quad \forall s \in A_i^+, i \in N, d \in D, \\ \pi_s^{id} - \pi^{id} \ge 0, \quad \forall s \in A_i^+, i \in N, d \in D, \\ \alpha_s^d \ge 0, \quad \forall s \in S, \forall d \in D. \end{cases} \tag{27}$$

Condition (27) implies that if the proportion of passengers per hour leaving node *i* and going to destination *d* via route section *s* is positive (*i.e.*, $\alpha_s^d > 0$), then the expected travel cost from node *i* to node *d* via section *s* is equal to the minimum expected travel cost between the two nodes. Moreover, if the approach proportion is zero (*i.e.*, $\alpha_s^d = 0$), then the expected travel cost via section *s* is greater than or equal to the minimum expected travel cost between nodes *i* and *d*. These conditions are consistent with the link-based UE conditions.

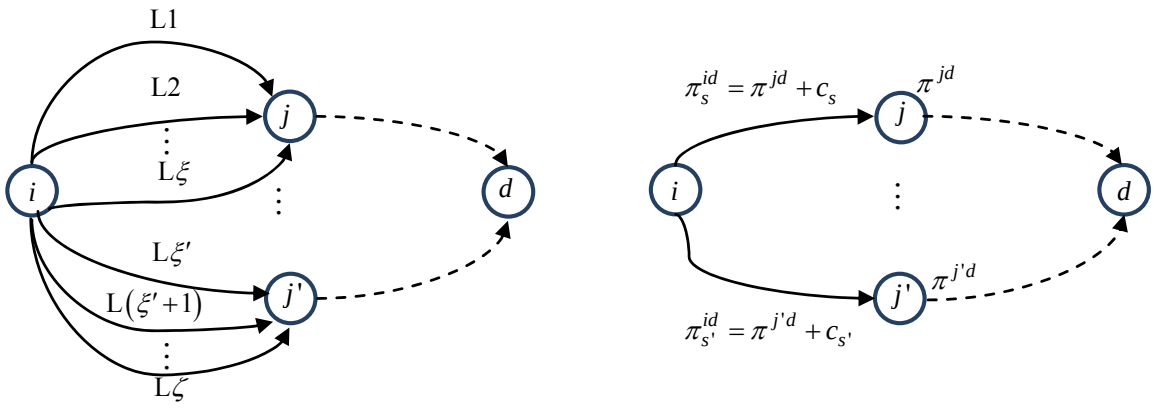
The approach-based UE condition (27) is further explained by Figure 2. Figure 2(a) is the original transit network, and Figure 2(b) is the corresponding route-section representation. The dotted line represents the minimum expected cost path connecting node *j* (*j'*) to destination *d*, and its cost, $\pi^{jd}\left(\pi^{j'd}\right)$, is marked next to node *j* (*j'*). Assuming that L1, L2, …, $\mathrm{L}\xi$ are the attractive lines connecting node *i* to node *j*, these lines are combined into one section, referred to as section *s* without loss of generality, with the expected section cost c_s as shown in Figure 2(b). Similarly, section s' is constructed to aggregate the set of attractive lines $\mathrm{L}\xi'$, $\mathrm{L}\left(\xi'+1\right)$, …, $\mathrm{L}\zeta$ connecting node *i* and node j'.

In Figure 2(b), there are two paths from node *i* to destination *d*. One path is via section or approach *s*, and the other is via approach s'. The proportions of these two approaches are denoted as α_s^d and $\alpha_{s'}^d$, respectively. The expected travel cost from node *i* to node *d* via approach *s* is obtained by adding the expected route-section cost c_s to the minimum expected path cost from the head node of section *s* (i.e., node *j*) to destination *d*. Similarly, the expected travel cost from node *i* to node *d* via s' is obtained by adding the expected route-section cost $c_{s'}$ to the minimum expected path cost from the head node of section s' (i.e., node j') to destination *d*.

Suppose that $\pi^{id} = \pi_{s'}^{id}$. That is, the lowest expected travel cost from node *i* to node *d* is equal to the expected travel cost associated with path $i - j' - d$. Then, this path is a lowest expected cost path, and approach s' is on the lowest expected cost path. Moreover, the approach proportion of approach s' must satisfy condition (27), because the term inside a pair of round brackets equals zero, and approach s' may carry flow, but the proportion must be between zero and one inclusively (*i.e.*, $1 \geq \alpha_{s'}^d \geq 0$) based on condition (23).

Under the preceding supposition, if approach *s* carries flow (*i.e.*, $1 \geq \alpha_s^d > 0$), then the expected travel cost from node *i* to node *d* via approach *s* must equal the minimum expected path cost from node *i* to destination *d* (*i.e.*, $\pi_s^{id} = \pi^{id}$), and approach *s* must be on another lowest expected cost path from node *i* to node *d*. (*i.e.*, $\pi_s^{id} = \pi_{s'}^{id} = \pi^{id}$). Otherwise, condition (27) is violated. If section *s* carries no flow (*i.e.*, $\alpha_s^d = 0$) under the preceding supposition, approach *s* may or may not be on one of the lowest expected cost paths. Approach *s* is on one of these paths only if the expected cost associated with path $i - j - d$ equals the expected cost associated with the lowest expected cost path $i - j' - d$. Nevertheless, in most cases, the

expected cost associated with path $i-j-d$ is larger than that associated with $i-j'-d$, and approach s is NOT on the lowest expected cost path $i-j'-d$.



(a) Transit network (b) Route section representation

Figure 2 Network representation of the approach-based formulation

2.7.2. Approach-based VI formulation

The approach-based transit assignment is to determine $\boldsymbol{\alpha}^* = \left[\alpha_s^{d*}\right]$ to satisfy constraints (1) – (8), (23), (24), (26) and (27). As with the link-based formulation, the approach-based formulation can be expressed as a VI: to determine $\boldsymbol{\alpha}^* = \left[\alpha_s^{d*}\right]$ such that

$$\left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\right)^{\mathrm{T}} \tilde{\boldsymbol{\pi}}\left(\boldsymbol{\alpha}^*\right) \geq 0, \quad \forall \boldsymbol{\alpha} \in \Omega_\alpha, \tag{28}$$

where $\tilde{\boldsymbol{\pi}}(\boldsymbol{\alpha}) = \left[\pi_s^{id}\right]$, $\boldsymbol{\alpha} = \left[\alpha_s^d\right]$, $\boldsymbol{\alpha}^*$ is the optimal solution vector, and $\Omega_\alpha = \left\{\alpha_s^d \mid 0 \leq \alpha_s^d \leq 1,\ \forall s,d,\ \sum_{s\in A_i^+} \alpha_s^d = 1,\ \forall d,\ \sum_{s\in A_i^-} \alpha_s^d q_{t(s)}^d + g_i^d = \sum_{s\in A_i^+} \alpha_s^d q_i^d, \forall i,d\right\}$. Note that $\tilde{\boldsymbol{\pi}}(\boldsymbol{\alpha}) = \left[\pi_s^{id}\right]$ can be replaced by $\tilde{\boldsymbol{\pi}}(\mathbf{v}) = \left[\pi_s^{id} - \pi^{id}\right]$, as $\sum_{s\in A_i^+} \alpha_s^d = \sum_{s\in A_i^+} \alpha_s^{d*} = 1$ and $\sum_i \sum_d \pi^{id} \sum_{s\in A_i^+} \left(\alpha_s^d - \alpha_s^{d*}\right) = 0$. The proof of equivalence between the approach-based problem formulation and VI (28) is similar to that between the link-based formulation and VI (22), and hence the proof for this last formulation is omitted.

For a solution to exist, the above VI problem requires (i) that the mapping function is continuous, and (ii) that Ω_α is a non-empty compact convex set (Theorem 1.4 in Nagurney, 1993). The continuity condition is satisfied, as π_s^{id} and π^{id} obtained by equations (17) and

(18) are continuous functions. The second condition is satisfied by the definition of the approach proportion (see (23) and (24)), and by the flow conservation condition (26). Therefore, for the proposed approach-based VI formulation, the existence of a solution is guaranteed. However, the uniqueness of the UE solution further requires that the mapping function is strict monotone. This requirement may not be fulfilled for the asymmetric, non-linear section cost function, in which case multiple solutions may be possible.

Although we use the term 'approach proportion,' the definition we use is different from that proposed by Bar-Gera (2002). We define 'approach proportion' based on outgoing links from nodes, whereas Bar-Gera's definition is based on incoming links. The reason for this redefinition is that by defining the proportion of flows using outgoing links as decision variables, the resulting formulation can easily capture the common line feature, and determine how passengers select transit lines from the set of attractive lines. As a result, in our model the decision variable, the sub-graph, the flow loading and cost updating schemes required in the solution process are all different, as we describe below.

3. SOLUTION ALGORITHM

3.1. Overview of the Algorithm

To solve the VI problem, we adopt a projection method that belongs to the class of extragradient methods introduced by Korpelevich (1976). The advantage of the extragradient method is that it does not require knowing the Lipschitz constant in advance. However, this projection method cannot be applied directly to solve the approach-based transit assignment problem. Therefore, a revised algorithm is proposed and its outline is presented as follows:

Step 0. Construct a route-section transit network (see Section 3.2).

Step 1. Set the iteration counter $k''=0$ and create an initial subnetwork $G^{k''}$ using the lowest expected cost path tree.

Step 2. Adopt the extragradient method to solve the transit assignment problem on the subnetwork $G^{k''}$ (see Section 3.3).

Step 3. Update the subnetwork. If $G^{k''}$ is a *UE bush*, then stop the algorithm; otherwise update $G^{k''}$ to $G^{k''+1}$, set $k''=k''+1$, and return to Step 2.

3.2. Network Construction Algorithm

A prerequisite of the solution algorithm is to construct a route-section network using real network information. To the best of our knowledge, this approach has not been previously

mentioned in the literature. Therefore, it is presented in this paper. The proposed algorithm contains two parts. The first part is to combine common lines into sections based on the bus line information. The second part is to identify the competing sections.

Step 0. Initialise parameters and sets. Set $\delta_s^m = 0, \forall s, m$; $A_s = \varnothing, \forall s$.

Step 1. Find and combine common lines into sections.

For each line $l \in L$,

For each pair of stops $\left\{ i, j \middle| i = \lambda_x^l, j = \lambda_y^l, \forall x, y = 1, 2, \cdots n^l, x \neq y \right\}$,

if $\neg\left(\exists s \in S : s = (i, j)\right)$, then

$S = S \cup \left(s = (i, j)\right)$,

else,

$A_s = A_s \cup l$, where $s = \left(i, j\right)$.

End if

Step 2. Identify competing sections.

If $A_s \cap A_m \neq \varnothing, \forall \left\{ s, m \middle| s, m = 1, 2, \cdots |S|, \text{ and } s \neq m \right\}$, then

if $t\left(s\right) = t\left(m\right)$, then

$\delta_s^m = 1, \delta_m^s = 1$.

else,

for each $l \in A_s \cap A_m$,

set $\left\{ x(s), x(m) \middle| \lambda_{x(s)}^{\prime l} = t\left(s\right), \lambda_{x(m)}^{\prime l} = t\left(m\right) \right\}$ and

$\left\{ y(s), y(m) \middle| \lambda_{y(s)}^{\prime\prime l} = h\left(s\right), \lambda_{y(m)}^{\prime\prime l} = h\left(m\right) \right\}$.

If $x(s) < x(m)$ and $x(s) \geq x(m)$, then $\delta_s^m = 1$;

If $x(m) < x(s)$ and $x(m) \geq x(s)$, then $\delta_m^s = 1$.

End if

End if

Step 1 scans all stops and creates the set of common lines for each section, which is used in Step 2 to identify competing sections by checking each pair of sections. The necessary condition of the competing sections is the existence of at least one common attractive line between two sections. Once such a condition is satisfied, two scenarios are distinguished:

a) if the two sections share the same boarding stop, *i.e.*, $t(s)=t(m)$, then they are mutually competing, *i.e.*, $\delta_s^m=1$ and $\delta_m^s=1$; and

b) if section *s* starts before and ends after section *m*, then *m* competes with *s*, and vice versa.

The above algorithm implicitly assumes that all lines are attractive on each section. However, the above algorithm can be easily modified to address a more general case in which not all lines are attractive. Hence, the assumption is not restrictive, as the set of attractive lines can be determined via, for example, the method used by Chriqui and Robillard (1975).

It is worthwhile mentioning that the network construction algorithm is not required in the hyperpath network representation, because there is no competing section in the hyperpath network. However, this representation requires more links and nodes than the route-section representation, especially if the number of common lines is large. It is unclear whether the hyperpath or the route-section representation can be solved quicker. Moreover, the congestion cost functions adopted in the two methods are often different. Hence, they require different algorithms with various convergence requirements to solve for solutions, and it may not be fair to directly compare the solution speeds of the two approaches with different convergence requirements.

3.3.Extragradient Method

In general, the extragradient method iteratively utilises two projection operators to make predictions and corrections until a convergence criterion is satisfied. Hence, this method is referred to as the double projection method. Accordingly, two stepsizes are required for the prediction and correction operators, respectively. Initially, Korpelevich (1976) set these two stepsizes to be identical and applied an Armijo-type scheme to update them, which resulted in a long computation time. To speed up the computation process, different rules for updating stepsizes have been proposed (He and Liao, 2002; Noor, 2003; Panicucci *et al*., 2007).

The extragradient method with adaptive stepsizes was first developed by He and Liao (2002), and has been shown to converge under the condition that the mapping function is pseudomonotone, which is a mild condition compared to the monotone requirement (Schaible, 1995). The major improvement of this method over the traditional double projection method is that the stepsizes are determined adaptively, leading to a faster convergence speed. Panicucci *et al*. (2007) adopted the improved extragradient method to solve the asymmetric path-based traffic assignment problem. However, as they solved the path-based problem, a

path set was initially generated using, for example, a k shortest path algorithm, and the path set was updated (if necessary) using a column generation heuristic. In contrast, the path set generation heuristic can be avoided when the improved extragradient method is used to solve the proposed approach-based formulation, although the convergence of the method is still based on the pseudomonotone requirement. The algorithmic steps are as follows:

Step 0. Initialise parameters and generate an initial solution $\boldsymbol{\alpha}^k$ by assigning all of the demand of each OD pair to the lowest expected cost path, where the superscript k refers to the solution obtained in iteration k. Set the iteration counter, $k=0$; set the initial stepsize of the prediction projection, $\beta_0 = 1.0$; select the parameters for updating the stepsizes of the prediction and correction, including $0<\bar{\beta}<1$, $\lambda=(0,2)$ and $0<\mu<\nu<1$; set the acceptable error $\varepsilon = 0.001$.

Step 1. Check the stopping criteria. Determine $\tilde{\boldsymbol{\pi}}(\boldsymbol{\alpha}^k)=[\pi_s^{id}]$ using the procedures described in Section 3.3.2. Then, check whether one of the following conditions are met:

a. $\max\left[\delta_s^{t(s)d}\left(\pi_s^{t(s)d}(\alpha_s^{d,k})-\pi^{t(s)d}(\alpha_s^{d,k})\right),\forall s,d\right]\le\varepsilon$, where $\delta_s^{t(s)d}$ equals 1 if section s carries flows (*i.e.*, flow proportions greater than zero), and equals zero otherwise;

b. The maximum allowable computation time is reached.

Step 2. Perform the prediction projection.

Calculate $\bar{\boldsymbol{\alpha}}^{k+1}=P_{\Omega_\alpha}\left[\boldsymbol{\alpha}^k-\beta_k\tilde{\boldsymbol{\pi}}\left(\boldsymbol{\alpha}^k\right)\right]$ using the method depicted in Section 3.3.1, where P_{Ω_α} represents the projection on the solution set Ω_α of the problem.

Step 3. Perform the correction projection.

Calculate the value of the corresponding mapping function $\tilde{\boldsymbol{\pi}}(\bar{\boldsymbol{\alpha}}^{k+1})$ (see Section 3.3.2).

If $r_k=\beta_k\dfrac{\left\|\tilde{\boldsymbol{\pi}}(\boldsymbol{\alpha}^k)-\tilde{\boldsymbol{\pi}}(\bar{\boldsymbol{\alpha}}^{k+1})\right\|}{\left\|\boldsymbol{\alpha}^k-\bar{\boldsymbol{\alpha}}^{k+1}\right\|}\le\nu$, then set

$$\mathbf{e}(\boldsymbol{\alpha}^k,\beta_k)=\boldsymbol{\alpha}^k-\bar{\boldsymbol{\alpha}}^{k+1},$$

$$\mathbf{d}\left(\boldsymbol{\alpha}^k,\beta_k\right)=\mathbf{e}(\boldsymbol{\alpha}^k,\beta_k)-\beta_k\left[\tilde{\boldsymbol{\pi}}(\boldsymbol{\alpha}^k)-\tilde{\boldsymbol{\pi}}(\bar{\boldsymbol{\alpha}}^{k+1})\right],$$

$$\tau_k=\lambda\beta_k\left[\frac{\mathbf{e}\left(\boldsymbol{\alpha}^k,\beta_k\right)^T\mathbf{d}\left(\boldsymbol{\alpha}^k,\beta_k\right)}{\left\|\mathbf{d}\left(\boldsymbol{\alpha}^k,\beta_k\right)\right\|^2}\right]\text{ (adaptive stepsize }\tau_k\text{),}$$

$$\beta_{k+1} = \begin{cases} \beta_k \dfrac{1}{\bar{\beta}}, & \text{if } r_k \le \mu, \\ \beta_k, & \text{otherwise,} \end{cases} \text{ (adaptive stepsize } \beta_k\text{),}$$

$\boldsymbol{\alpha}^{k+1} = P_{\Omega_\alpha}\left[\boldsymbol{\alpha}^k - \tau_k \tilde{\boldsymbol{\pi}}(\bar{\boldsymbol{\alpha}}^{k+1})\right]$ (see Section 3.3.1), and

$k = k+1$. Go to step 1.

Otherwise, reduce the stepsize β_k by $\beta_k = \beta_k \cdot \bar{\beta} \cdot \min(1, \dfrac{1}{r_k})$ and go to Step 2.

In Step 0, the initial solution is obtained by an all-or-nothing assignment. In Step 1, the maximum difference between the expected travel cost via the presently used approach and the minimum expected travel cost is taken as one of the stopping criteria for the extragradient algorithm. If the difference is small, then the solution is very close to an optimal solution. The second criterion is the maximum allowable computation time. If the maximum allowable computation time is reached, then the extragradient algorithm ceases without outputting an optimal solution. Steps 2 and 3 carry out the prediction and correction projections using adaptive stepsizes β_k and τ_k, respectively.

3.3.1. A linear projection method

The projection operation can be performed by many methods. In this section, a simple linear projection method (Panicucci *et al.*, 2007) is adopted, which makes use of the decomposable structure of the simplex constraint set (*i.e.*, $\sum_{s \in A_i^+} \alpha_s^d = 1$, $\alpha_s^d \ge 0, \forall i, d$). Given a vector $\mathbf{z} = \left[z_s^d\right] = \boldsymbol{\alpha}^k - \beta_k \tilde{\boldsymbol{\pi}}\left(\boldsymbol{\alpha}^k\right) \in R_+^{|A_i^+|}$ associated with the proportions of passengers leaving node $i = t(s)$ to destination d (i.e., the input to the prediction projection operator), or $\mathbf{z} = \boldsymbol{\alpha}^k - \tau_k \tilde{\boldsymbol{\pi}}(\bar{\boldsymbol{\alpha}}^{k+1})$ (i.e., the input to the correction projection operator), where $|A_i^+|$ is the dimension of the sections that emanate from node i, the following steps can be used to find the projection of $\mathbf{z}$ onto the simplex $\left\{\left[\alpha_s^d\right] \in R_+^{|A_i^+|} : \sum_{s \in A_i^+} \alpha_s^d = 1\right\}$:

Step 0. Initialise $k' = 0$, and set $\alpha_s^{d,k'} = z_s^d + \left(1 - \sum_{j \in A_i^+} z_j^d\right) / |A_i^+|, \quad \forall s \in A_i^+$.

Step 1. Check the stopping criterion.

If $\alpha_s^{d,k'} \ge 0 \ \ \forall s \in A_i^+$, then $\bar{\boldsymbol{\alpha}}^{k+1} = \left[\alpha_s^{d,k'}\right]$ or $\boldsymbol{\alpha}^{k+1} = \left[\alpha_s^{d,k'}\right]$, and stop.

Otherwise, create the index set $G = \left\{ g : \alpha_g^{d,k'} > 0 \right\}$.

Step 2. For all $s \in A_i^+$, compute

$$\alpha_s^{d,k'+1} = \begin{cases} 0, & \text{if } s \notin G, \\ \alpha_s^{d,k'} + \left(1 - \sum_{j \in A_i^+} \alpha_j^{d,k'} \right) / |G|, & \text{if } s \in G. \end{cases}$$

Set $k' = k'+1$, and return to Step 1.

The above algorithm performs very simple computations, and the projection of **z** onto the simplex $\left\{ \left[\alpha_s^d \right] \in R_+^{|A_i^+|} : \sum_{s \in A_i^+} \alpha_s^d = 1 \right\}$ can be found in most $|A_i^+|$ iterations.

3.3.2. Update the mapping function

After performing the linear projection method, the following two procedures are used to obtain the corresponding mapping function. The first procedure is the flow update and the other is the expected travel cost update. These two algorithms are quite similar in the sense that they both use the topological order to scan a network by a single pass. Given the topological order, a forward scan is used to update flows, followed by a backward scan to update the cost vector $\tilde{\boldsymbol{\pi}}(\boldsymbol{\alpha})$. To find the topological order, a depth-first search is adopted (Cormen *et al*., 2001).

The following are the algorithmic steps to update flow:

Flow update

Step 0. Initialisation. Set $v_s^d = 0, \forall s,\ d$.

Step 1. For each destination d,

for u = 1 to $|\Psi^d|$,

set $i = \psi_u^d$,

for $s = (i, j) \in A_i^+$,

$v_s^d = \alpha_s^d q_i^d$;

$q_j^d = q_j^d + v_s^d$.

Given the section flows, the section cost can be easily obtained following equations (1) – (8). Afterwards, the following subroutine is called to determine the elements of the cost vector $\tilde{\boldsymbol{\pi}}(\boldsymbol{\alpha})$ in condition (27):

Expected travel cost update

Step 0. Initialisation: set $\pi^{id} = \infty,\ \forall i,\ d;\ \pi^{dd} = 0, \forall d$.

Step 1. For destination d ,

for $u = \left|\Psi^d\right|$ to 1,

set $i = \psi_u^d$.

for $s = (j,i) \in A_i^-$,

$\pi_s^{jd} = \pi^{id} + c_s$;

$\pi^{jd} = \min(\pi^{jd}, \pi^{id} + c_s)$.

The above algorithm is similar to Dijkstra's shortest path algorithm, except that this algorithm starts from a destination node.

3.3.3. Computational complexity analysis

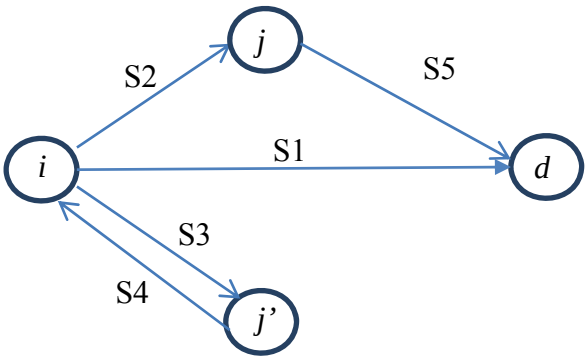
For a single destination, the running time of the flow updating algorithm is $O(V+E)$, where V stands for the total number of vertices (*i.e.*, nodes) and E for the number of edges (*i.e.*, route sections or links). To analyze the computational complexity, we consider the case of a single destination. The worst scenario of the first loop occurs when all of the nodes are contained in the topological order set. In such a condition, it costs $O(V)$ time to check all of the vertices. Meanwhile, because there is no cyclic flow, the worst scenario of the second loop is that all of the links are used, and the computational complexity is $O(E)$. By aggregating these two conditions, the computational complexity for the single destination case is $O(V+E)$. Similarly, it can be shown that the complexity of the single destination case to update expected travel cost is also $O(V+E)$. In practice, it is possible to reverse the network order to save computation time if the dimension of the set of destinations is larger than that of origins.

3.4. The Necessity of Subnetwork Updating

The double projection algorithm is performed on an acyclic subnetwork. The reason for using subnetworks is that the projection algorithm may produce cyclic paths if the network is

cyclic. To illustrate such an issue, a small example is arbitrarily created in Figure 3. In this network there are five sections, namely S1 – S5. At node *i*, there are three approaches, S1 – S3. Travellers at node *i* can select one of these approaches to travel to destination *d*. If S3 is selected, then the resultant *lowest expected cost path* to destination *d* (*i.e.*, either *i – j' – i – d* or *i – j' – i – j – d*) must be cyclic, with *i* visited twice. The demand and cost function settings are also shown in the figure. Assume that at iteration *k* all demands are assigned to the section with the lowest expected cost path, which is section S1. The resultant flow pattern is $\mathbf{v}^k = \begin{pmatrix} v_{\mathrm{S1}}^d = 100 \\ v_{\mathrm{S2}}^d = 0 \\ v_{\mathrm{S3}}^d = 0 \\ v_{\mathrm{S4}}^d = 0 \\ v_{\mathrm{S5}}^d = 0 \end{pmatrix}$. Accordingly, the approach proportion and the corresponding mapping funcion value are $\boldsymbol{\alpha}^k = \begin{pmatrix} \alpha_{\mathrm{S1}}^{id} = 1.0 \\ \alpha_{\mathrm{S2}}^{id} = 0.0 \\ \alpha_{\mathrm{S3}}^{id} = 0.0 \end{pmatrix}$ and $\tilde{\boldsymbol{\pi}}^k = \begin{pmatrix} \pi_{\mathrm{S1}}^{id} = 10010 \\ \pi_{\mathrm{S2}}^{id} = 20 \\ \pi_{\mathrm{S3}}^{id} = 30 \end{pmatrix}$. Note that the approach proportions at *j* and *j'* always equal one, because there is only one exit approach at these nodes, and hence we omit the two approach proportions in this analysis. Following the double projection algorithm, we compute $\boldsymbol{\alpha}^{k+1} = P_{\Omega_\alpha}\left[\boldsymbol{\alpha}^k - \beta_k \cdot \tilde{\boldsymbol{\pi}}\left(\boldsymbol{\alpha}^k\right)\right]$; thus, the input of the projection operator is $\boldsymbol{\alpha}^k - \beta_k \cdot \tilde{\boldsymbol{\pi}}\left(\boldsymbol{\alpha}^k\right) = \begin{pmatrix} 1.0 - 10010 \\ -20.0 \\ -30.0 \end{pmatrix}$. Using the linear projection algorithm, the vector of the updated approach proportions is $\boldsymbol{\alpha}^{k+1} = \begin{pmatrix} \alpha_{\mathrm{S1}}^{id} = 0.000 \\ \alpha_{\mathrm{S2}}^{id} = 0.505 \\ \alpha_{\mathrm{S3}}^{id} = 0.495 \end{pmatrix}$. This result indicates that a proportion of passengers select an approach that is on a cyclic path, because $\alpha_{\mathrm{S3}}^{id} > 0$.

However, it has been shown that in traffic assignment, the optimal solution is acyclic by origin (*i.e.*, no cyclic flow starting from any origin is found at optimality, and there must be no flows on cyclic paths at UE), even if the original network is cyclic. Similarly, it can be easily shown that the optimal solution for our proposed formulation is also acyclic by destination. This can be shown simply by reversing the directions of sections and changing destinations to origins (or vice versa), and then by using Lemmas 1-3 as proposed by Bar-Gera (2002).



Cost Functions

$c_{S1} = 10 + v_1^2$; $c_{S2} = 15 + v_2$

$c_{S3} = 5 + v_3$; $c_{S4} = 5 + v_4$

$c_{S5} = 5 + v_5$;

Demand: $g_i^d = 100$

Parameter: $\beta_k = 0.001$

Figure 3 An example of cyclic paths

Proposition 2: The optimal approach-based solution by destination is acyclic.

Proof. See Lemmas 1-3 in Bar-Gera (2002). □

Proposition 2 implies that discarding approaches that are on cyclic paths will not affect the optimal solution, because the optimal solution is acyclic. This proposition also implies that for each UE solution found there exists an acyclic network, called a UE bush. Accordingly, obtaining a UE solution on a whole network is equivalent to finding a UE bush and performing the assignment on the UE bush, which can be done by iteratively updating an initial acyclic subnetwork. The benefits of using subnetworks include 1) reducing the size of the network on which the projection algorithm is conducted, and 2) avoiding production of cyclic flows. Hence, the double projection algorithm can be performed more efficiently.

To start the algorithm, a shortest spanning tree is adopted as an initial subnetwork for each destination. Then we adopt the updating method as proposed by Nie (2010). First, unused sections with no passenger flows are deleted. Afterwards, selected sections are added to expand the subnetwork. The following two conditions are used to determine whether section s should be added to existing subnetwork $G^{k''}$.

1) $c_s + \pi^{jd} < \pi^{id}$, $\forall d \in D, s = (i, j)$ and $s \notin G^{k''}$

2) $x < y, \forall d \in D, s = (i, j), s \notin G^{k''}$ and $i = \psi_x^d$, $j = \psi_y^d$

Condition 1) requires that an added section promises to reduce the expected travel cost. Condition 2) guarantees that the updated network is still acyclic after adding the new section. (Proposition 3 in Nie (2010)). After obtaining a new subnetwork, the projection operation is performed again. The above procedure is repeated until no sections can be added, indicating that a UE bush is obtained.

4. NUMERICAL EXAMPLES

4.1. Paradox Examples

To illustrate the paradoxical phenomenon, a simple network is built. Figure 4 shows the route-section presentation. Three bus lines are designed, and denoted as L1, L2 and L3. Assuming that all of the lines are attractive, the corresponding route-sections are constructed following the network construction algorithm described in Section 3.2. The notation S3(L3) in Figure 4 means that L3 is an attractive line in section S3. Section S1(L1) is drawn by a dashed line, as it is assumed that L1 does not initially exist. Table 1 lists the required data. For illustration purposes, the path information is presented instead of the approach information, as the purpose of this example is to illustrate the existence of the paradox rather than the approach formulation. Three paths exist in the network. Take path 2 as example. This path connects OD pair AC, and passes sections S1 and S2, whereas section S1 contains L1 and section S2 contains L2 as the attractive lines. The headway of L1 varies in the following tests. As a result, path 2, which passes S1(L1), is not available until L1 is added (*i.e.*, the frequency of L1 is greater than 0). Unless specified, values of parameters are set as follows: the value of in-vehicle cost $\mu_T = 1.0$ dollar per hour; the value of waiting time $\mu_W = 2.0$ dollars per hour; the congestion function parameters are $\varpi_1 = \varpi_3 = 0.1$, $\varpi_2 = 0.3$, and $a = 1.0$; $n = 3$; bus capacity $\kappa^l = 120$ passengers per bus; the demand levels of the two OD pairs $g_A^C = g_B^C = 360$ passengers per hour.

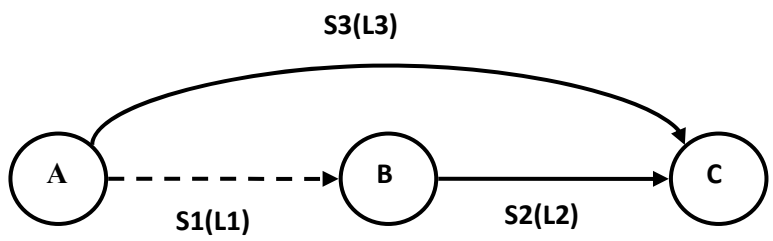


Figure 4 Small network

Table 1 Input data of the small network

(a) Path data

OD Pair	Path No.	Path
(1) A-C	1	S3(L3)
	2	S1(L1)-S2(L2)
(2) B-C	3	S2(L2)

(b) Line data

Line No.	Travel time (minutes)	Headway (minutes)	Frequency (buses/h)
1	9	-	-
2	3	10	6
3	14	24	2.5

4.1.1. Adding a new line to the transit network and increasing the frequency of an existing transit line

Three scenarios, as described below, are designed to demonstrate the paradox phenomenon.

- Scenario 1: without line L1;
- Scenario 2: L1 is provided, and its frequency is set to be 4.6 buses per hour;
- Scenario 3: L1 is provided, and its frequency varies from 3.4 to 5.0 buses per hour.

The results of expected total system travel cost (ETSTC) are summarised in Table 2.

Table 2 Summary of three scenarios

	Scenario 1	Scenario 2	Scenario 3
With S1(L1)	No	Yes	Yes
Frequency of L1 (buses/h)	-	4.6	3.4-5.0
ETSTC (dollars)	31508.5	31890.0	See Figure 5

Comparing scenario 1 with scenario 2, the ETSTC increases from 31508.5 dollars to 31890.0 dollars, demonstrating that providing a new line in a transit network may deteriorate the whole network's performance in terms of the ETSTC.

This situation is analogous to the Braess paradox in traffic assignment, in which adding a new link to a network can increase the ETSTC. In our case, the increase in the ETSTC happens because some passengers of OD pair 1 are diverted from L3 to L2, inducing more congestion cost associated with section 2. More importantly, the total increment of the congestion cost associated with section S2 outweighs the cost reduction associated with S3. This paradox occurs because the introduction of the new transit line does not take into account the selfish route choice behaviour of passengers and their response to the new service. This example illustrates the importance of considering these factors when introducing new transit services.

The results of scenario 3 are plotted in Figure 5, which illustrates the changes in the ETSTC with respect to the frequency of L1. In addition, to facilitate the explanation of the changes in the ETSTC, the equilibrium costs of the two OD pairs are plotted in Figure 6. According to these figures, we can identify four regions of frequency: 1) less than 3.8 buses/h

(and greater than 0 buses/h); 2) between 3.8 and 4.1 buses/h; 3) between 4.1 and 4.6 buses/h, and 4) more than 4.6 buses/h.

In the first region, in which the frequency is less than 3.8 buses/h, the ETSTC remains stable, as do the corresponding equilibrium costs of the two OD pairs in Figure 6. This result indicates that adding the new line does not affect the expected travel cost. Because L1 is not considered as an attractive line, none of the passengers at node A switch to section S2. Such a condition implies that the buses of L1 would be vacant, and the operator would lose money by providing the L1 services even with a low frequency.

In the second region, the ETSTC starts to decline. When the frequency is greater than 3.8 buses/h, the new line is considered an attractive line by the travellers departing at node A. By adding L1 to the set of attractive lines, the expected travel cost of travellers can be reduced. This result can also be seen in Figure 6, which shows that by increasing the frequency, the equilibrium travel cost for OD pair 1 is reduced. However, the equilibrium cost for OD pair 2 grows due to the additional congestion cost induced by the passengers selecting S1(L1). Nevertheless, the rate of cost increase for OD pair 2 is mild compared with the rate of cost decrease for OD pair 1. Therefore, on the condition that the demand levels are equal for the two OD pairs, the initial ETSTC still decreases, and no paradox is observed in this region. At the frequency of 4.09 buses/h, the rate of cost increase for OD pair 2 equals that of the cost decrease for OD pair 1. As a result, the ETSTC arrives at one bottom value.

The paradox occurs in the third region, where the frequency of L1 keeps increasing from 4.09 to 4.6 buses/h. The ETSTC soars to a peak of 31890.0 dollars. The equilibrium cost for OD pair 2 grows sharply, because more passengers are attracted to route 2 when the frequency is larger. This increase in passengers on route 2 exacerbates the congestion on section S3(L3) due to the BPR-type congestion cost function, by which costs grow rapidly as flow approaches capacity. Although the cost for OD pair 1 keeps decreasing, the curve tends to be flat. Hence, the increment in the ETSTC is mainly driven by the rise in the expected travel cost for OD pair 2.

In the last region, where the frequency of L1 is greater than 4.6 buses/h, the ETSTC starts to descend again. The congestion cost of S3(L3) stops growing, as all of the travellers between OD pair 1 have switched to section S1. After the frequency of L1 is greater than 4.8 buses/h, the ETSTC drops below 31508.5 dollars, which equals the ETSTC of the scenario without L1 (*i.e.*, scenario 1).

The above scenarios illustrate that the frequency paradox phenomenon can be observed within certain regions of the frequency setting. These scenarios also illustrate that there are

regions where the paradox can be avoided and the system performance can be improved. However, the regions associated with better system performance can be significantly different from each other in terms of flow distribution. When the increased frequency reaches the second region (*i.e.*, 3.8 to 4.1 buses/h), the passengers at node A are distributed into two lines, and in the last region (*i.e.*, greater than 4.6 buses/h), the passengers select only one line to board. From the operator's point of view, both the total fare revenues associated with these regions and the operating cost of providing the corresponding level of service vary greatly. These factors of revenue and operating cost will influence the strategy of bus companies in competitive markets such as Hong Kong.

In addition, it is worth mentioning that the occurrence of the paradox depends on both the initial condition and the increment in the frequency. For example, when the initial frequency is 3.6 buses/h, if the frequency increment is 0.3 bus/h, the system performance improves. If the frequency increment is 1.0 buses/h, the paradox occurs. However, if the initial frequency is 4.0 buses/h, a 0.3 bus/h increment in the frequency would induce the paradox, but a 1.0 bus/h increment in the frequency would not. The paradoxical region investigated and marked in the graphs is the region where a small increment in frequency can lead to the paradox.

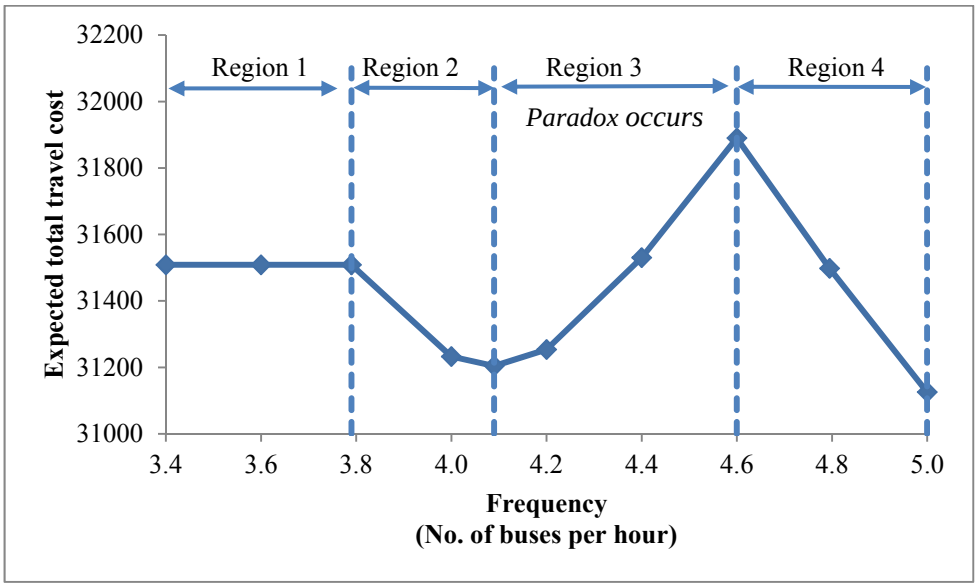


Figure 5 Expected total travel cost under various frequencies in scenario 3

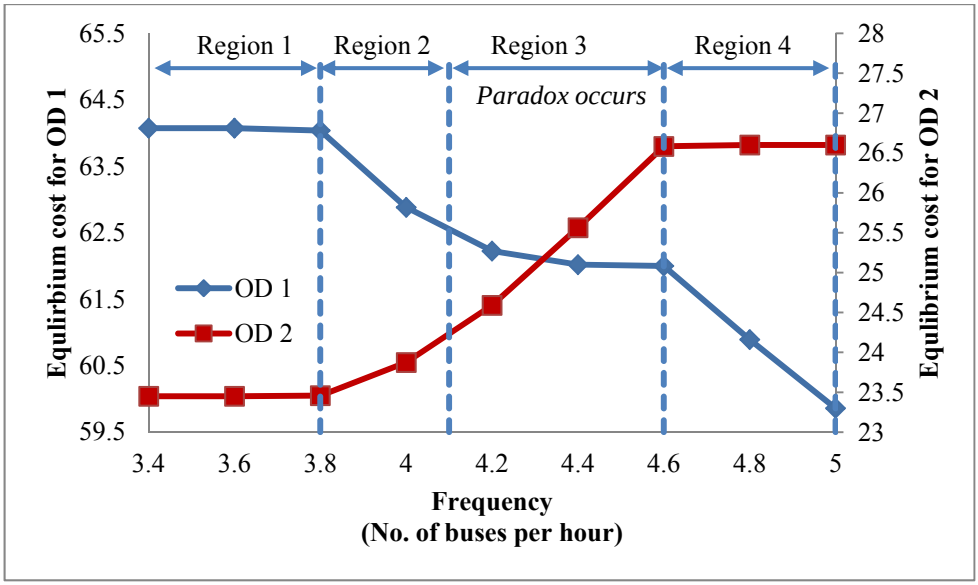


Figure 6 Equilibrium costs of OD pairs 1 and 2

4.1.2. Effects of the perceived congestion parameters

Figure 7 illustrates how different values in perceived congestion parameters influence the existence of the paradox. The perceived congestion parameter value, n varies from 1 to 3; a varies from 0.5 to 1.5; ϖ_2 increases from 0.2 to 0.4. In each case, the frequency of L1 changes from 3.4 to 5.0 buses/h, and the corresponding ETSTC is plotted. The observations are summarised as follows.

First, the values of the perceived congestion parameters affect the occurrence of the paradox. The paradox can be observed when $n = 1$ (Figure 7(a)), $a = 0.5$ (Figure 7(b)), or $\varpi_2 = 0.2$ (Figure 7(c)), and the paradox is inevitable when these values are large. The paradox appears because the route choice behaviour of passengers is not taken into an account when the frequency is adjusted. Increasing the frequency can reduce the waiting time for the line in question, but it also attracts passengers from other lines to use that line. Hence, the network-wide effect cannot be ignored when trying to improve the transit network by adjusting frequency.

Second, the ETSTC grows along with the increasing values of the perceived congestion parameters, up to the peak values for the ETSTC. The peak values are larger when the values of the parameters are larger. Figure 7 shows that the ETSTC increases as the values of n, a, and ϖ_2 increase, given a fixed frequency. This increase happens because, by definition, a larger value of the perceived congestion parameter implies a higher perceived congestion cost. For the same reason, the maximum ETSTC increases with the rising value of the perceived

congestion parameters. These patterns can be found whether the frequency is within the paradox region or not.

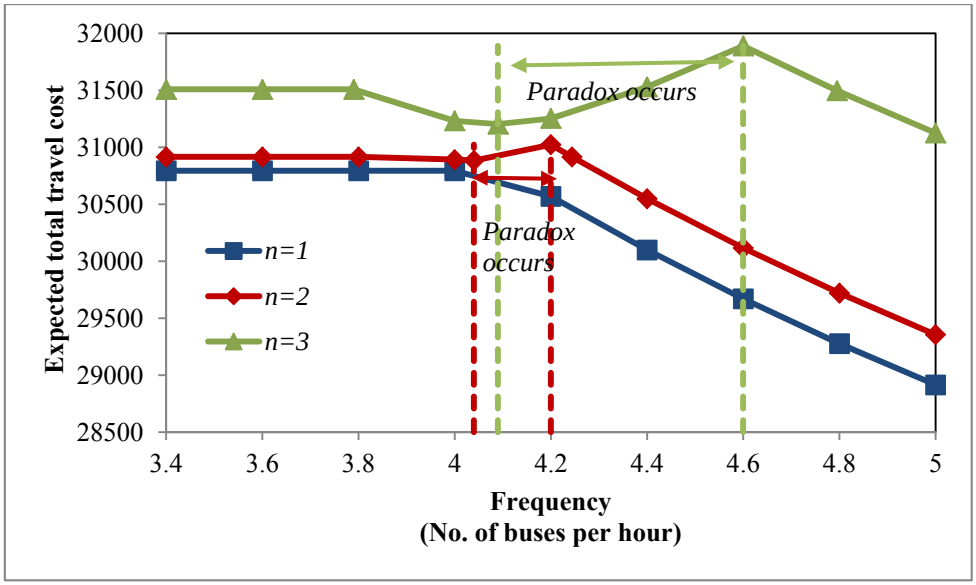
Third, the perceived congestion parameters affect the width of the paradox region. The width of the paradox region can be characterised by the difference between two frequency values. The first value is the frequency at which the ETSTC starts to increase, and the second is the finishing frequency at which the ETSTC starts to drop from its peak.

For the starting frequency, we are interested to observe that higher values of n and a can defer the occurrence of the paradox. For example, in Figure 7(b), when $a = 1.5$, the ETSTC starts to increase at 4.09 buses/h, but the starting frequency is 4.04 buses/h when $a = 1.0$. A similar phenomenon exists in Figure 7(a). However, in Figure 7(c), a higher value of ϖ_2 advances the occurrence of the paradox. The paradox emerges at 4.04 buses/h when $\varpi_2 = 0.3$, and at 4.02 buses/h and $\varpi_2 = 0.4$.

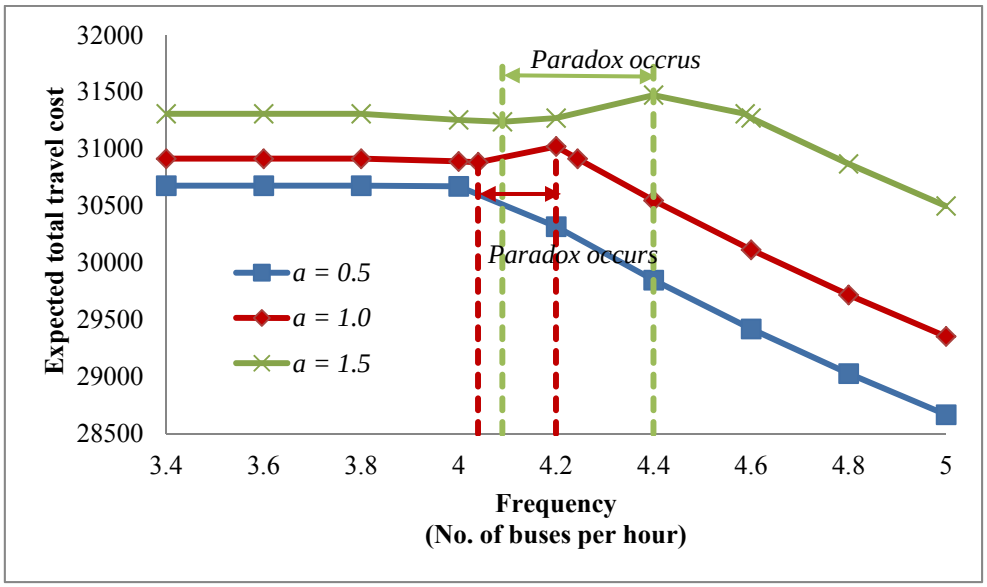
For the finishing frequency, the larger the perceived congestion parameter values are, the higher the finishing frequency. For example, in Figure 7(a), the paradox region ends at 4.6 buses/h when $n = 3$, but it ends at 4.2 buses/h when $n = 2$.

Despite these different effects of the perceived congestion parameters on the two frequencies, the inference mechanisms in determining these frequencies are the same. These mechanisms can be attributed to the changing equilibrium costs of the two OD pairs. If the equilibrium cost for OD pair 1 decreases at a rate that is less (greater) than the rate of increase for OD pair 2, the ETSTC increases (decreases) and the paradox occurs (vanishes). Thus, the starting (finishing) frequency is the point before which the increasing rate of equilibrium cost for OD pair 1 is less (greater) than the decreasing rate of the equilibrium cost for OD pair 2, and after which the increasing rate of the equilibrium cost for OD pair 1 is greater (less) than the decreasing rate of the equilibrium cost for OD pair 2. When the perceived congestion parameters vary, the rates of change in equilibrium costs are affected due to the various changes in perceived congestion cost. Consequently, the starting and finishing frequencies and hence the width of the paradox regions change.

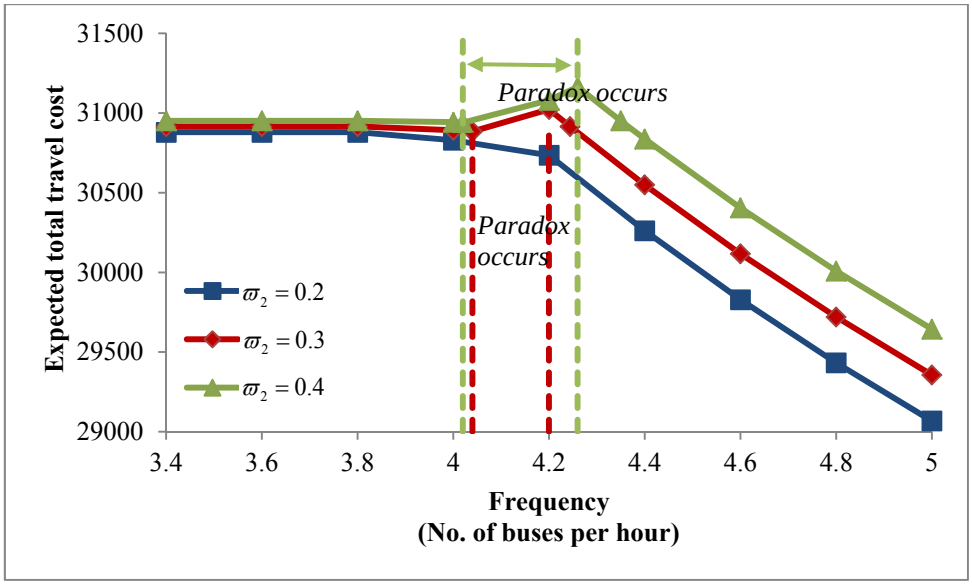
Our experiments indicate that the paradox region is wider when the values of perceived congestion parameters are larger. However, we expect that the starting and finishing frequencies of the paradox region and the width of each frequency region (including the paradox region) depend on the combination of these congestion parameters in general.



(a) Effect of *n*



(b) Effect of *a*



(c) Effect of ϖ_2

Figure 7 Effects of different perceived congestion parameter values

In general, the above experiments verify the importance of calibrating the perceived congestion parameters. These parameters are determined by various factors such as the layout of bus stops, the types of buses, the passengers' behaviour (in terms of whether they choose to join a queue or to push and squeeze to board a bus), and their tolerance for waiting and in-vehicle congestion. A field survey is needed to calibrate the perceived congestion parameters to further identify whether the paradox is avoidable during the transit network design process, or, if the paradox cannot be avoided, how frequently the service should be provided to overcome the paradox region effect.

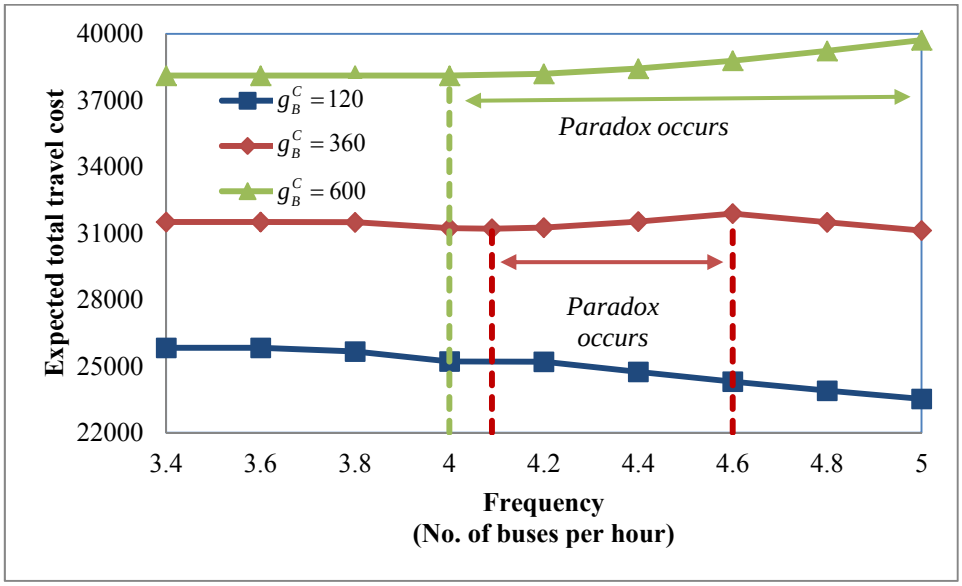


Figure 8 Effect of demand

4.1.3. Effect of demand

Figure 8 presents the changes in the ETSTC when the demand of OD pair 2 varies from 120 to 600 passengers/h. The paradox can be avoided when the demand is 120 passengers/h, but it occurs when the demand is between 360 and 600 passengers/h. This finding implies that the occurrence of paradox also depends on demand levels.

Surprisingly, the ETSTC keeps increasing when the frequency is larger than 4.0 buses/h when demand reaches 600 passengers/h. This result implies that whenever a line is added between nodes A and B, the system performance cannot be improved via providing adjustments in practical frequency (*i.e.*, less than 5.0 buses/h). Therefore, managers should consider alternative methods to enhance the system performance.

In general, we notice that the effect of varying demand is similar to that of variations in perceived congestion parameter values, as higher demand normally generates a higher congestion cost.

4.2. Performance of the Proposed Algorithm

4.2.1. The Sioux-Falls network

The proposed algorithm was tested with the Sioux-Falls network, shown in Figure 9. Ten itineraries were arbitrarily created. Table 3 describes the headway and stop sequence settings. We set 16 OD pairs, and the demand data are shown in Table 4. The algorithm was coded and compiled by Compaq Visual Fortran 6.6. The effects of different parameters were tested. In this study, we adopted the number of intermediate solutions evaluated as the criterion for measuring computation speed. An intermediate solution refers to one feasible solution $\boldsymbol{\alpha} = [\alpha_s^d]$ computed during the computation process, and the number of intermediate solutions evaluated means the total number of feasible solutions generated during the whole computation process. This criterion was used for two reasons. First, the accuracy of CPU time measurement provided in Fortran is only up to 0.0001. During this interval, many intermediate solutions could be evaluated. To address this issue, we counted the number of intermediate solutions evaluated. Second, the number of iterations cannot be adopted as the stopping criterion, as the iteration counter only accounts for the number of prediction projection operations. Unless specified otherwise, the parameters in the double projection method were set to the following: $\varepsilon = 0.001$, $\nu = 0.9$, $\mu = 0.7$, $\beta_0 = 1.0$, $\lambda = 1.9$. The maximum allowable computation time was set to 600 seconds.

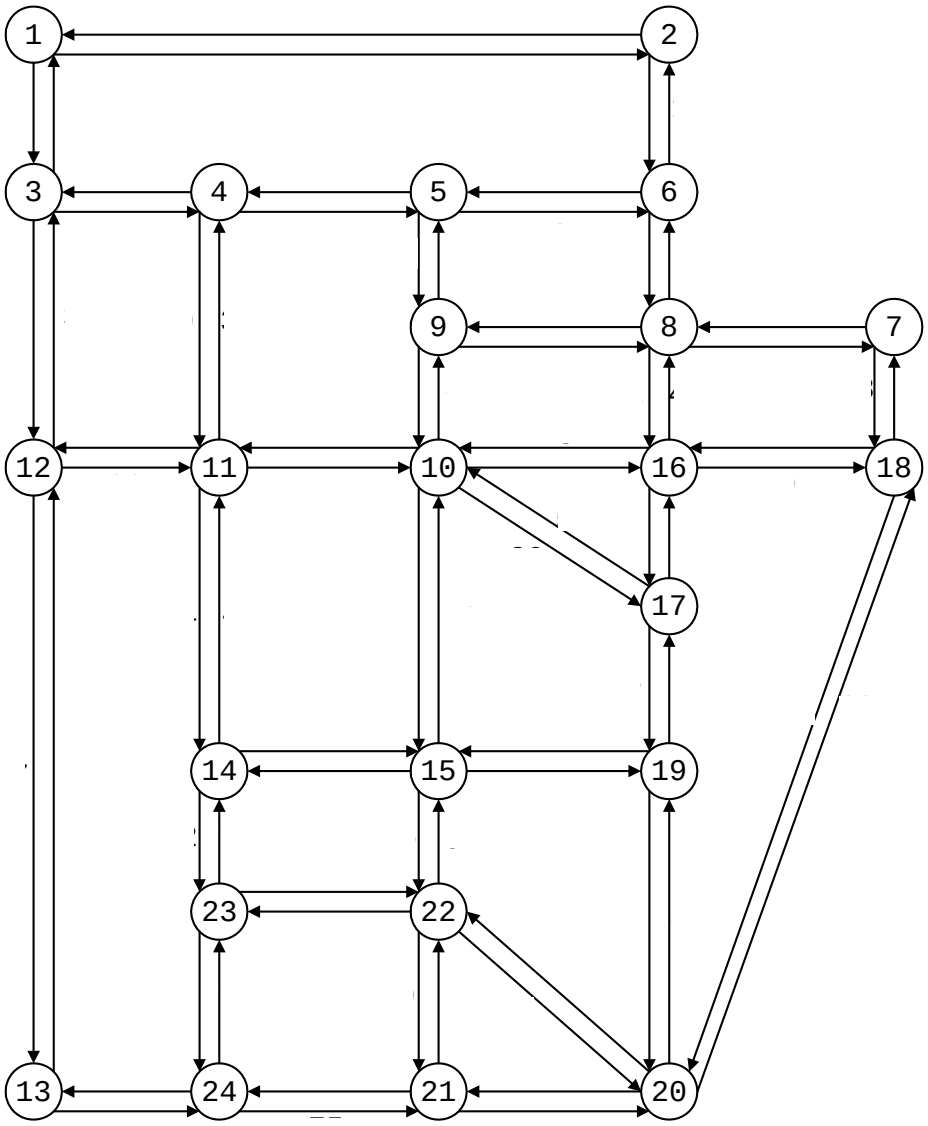


Figure 9 Sioux-Falls network

Table 3 Transit route setting of the Sioux-Falls network

Route No.	Headway (minutes)	Stop Sequence								
1	6	4	11	23	24					
2	6	1	3	12	13	24				
3	6	11	14	23	24	13				
4	5	8	20	21	22	23				
5	6	7	8	16	18	20				
6	6	14	15	19	20	22	23			
7	3	2	6	8	9	10	11	12		
8	3	4	5	9	10	17	19	20		
9	3	10	16	17	19	20	21	24		
10	3	1	3	4	5	9	10	15	19	20

Table 4 Demand data

OD pair	Demand (passengers/h)	OD pair	Demand (passengers/h)
(1-13)	500.0	(3-13)	500.0
(1-20)	500.0	(3-20)	500.0
(1-21)	500.0	(3-21)	500.0
(1-24)	500.0	(3-24)	500.0
(2-13)	400.0	(4-13)	400.0
(2-20)	400.0	(4-20)	400.0
(2-21)	400.0	(4-21)	400.0
(2-24)	400.0	(4-24)	400.0

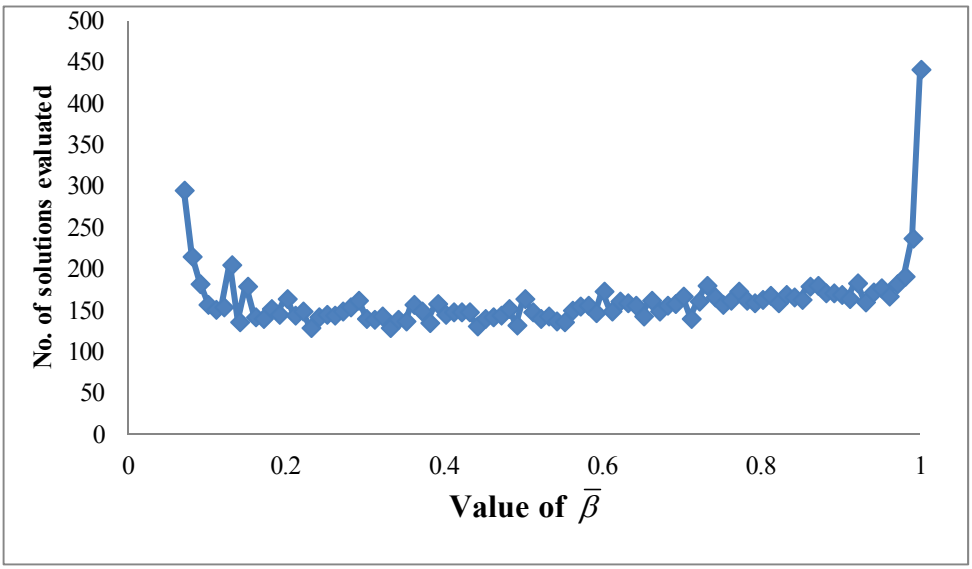


Figure 10 Effect of $\bar{\beta}$

Figure 10 shows the effect of $\bar{\beta}$, which is the parameter to adjust the stepsize of the prediction projection operator. The value of $\bar{\beta}$ increased from 0.01 to 1.00, and the corresponding numbers of solutions evaluated were plotted. When the value of $\bar{\beta}$ was between 0.9 and 1.0, the computation time was long, as the stepsize could only be reduced. However, when $0.01 \le \bar{\beta} < 0.1$, the stepsize became large and the computation time was also long. In particular, the algorithm failed to converge within the predefined stop time when

$0.01 \le \bar{\beta} < 0.06$. Between 0.1 and 0.9, the effect of $\bar{\beta}$ was minor. In general, the role of $\bar{\beta}$ is to control the range between two successive stepsizes. Ranges that are either too large or too small are not beneficial for the convergence speed. For our network, the minimum value occurred at $\bar{\beta} = 0.33$, which was adopted in the following tests.

Table 5 Effect of different combinations of μ and ν

ν	μ									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Average
0.1	-	-	-	-	-	-	-	-	-	-
0.2	238	-	-	-	-	-	-	-	-	238
0.3	191	187	-	-	-	-	-	-	-	189
0.4	189	165	141	-	-	-	-	-	-	165
0.5	180	167	151	130	-	-	-	-	-	157
0.6	201	207	168	134	133	-	-	-	-	169
0.7	212	228	180	132	129	125	-	-	-	168
0.8	188	255	170	145	158	143	143	-	-	172
0.9	203	187	174	172	192	148	130	138	-	168
Average	200	199	164	143	153	139	137	138	-	-

Table 5 presents the effects of different combinations of μ and ν. As convergence requires that $\mu < \nu$ (see He and Liao, 2002), the combinations that do not satisfy this condition are marked as '-'. For each value of ν, few intermediate solutions were generated when μ was large. To find a more general trend, the average numbers of solutions evaluated (*i.e.*, the last row and last column) are plotted in Figure 11. This figure clearly shows that the computation time reduces with the increments of ν and μ, and then remains at a stable level. We selected $\mu = 0.6$ and $\nu = 0.7$ for later tests.

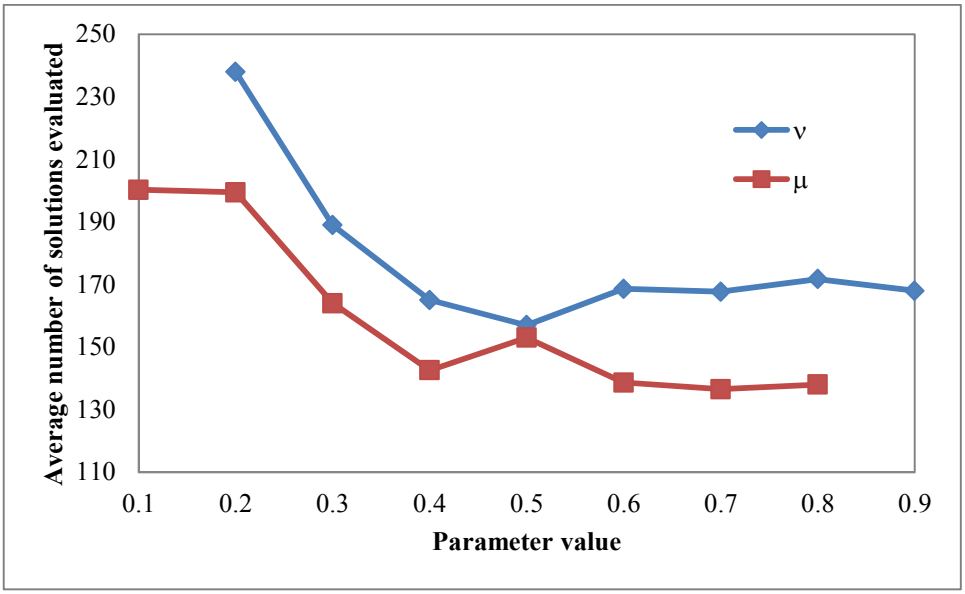


Figure 11 Average number of intermediate solutions evaluated

Figure 12 illustrates the effect of the parameter λ on the speed of convergence. In general, a large value of λ accelerates the projection method to converge. In our case, $\lambda = 1.8$ was chosen for the following tests, as it gave the minimum number of intermediate solutions evaluated.

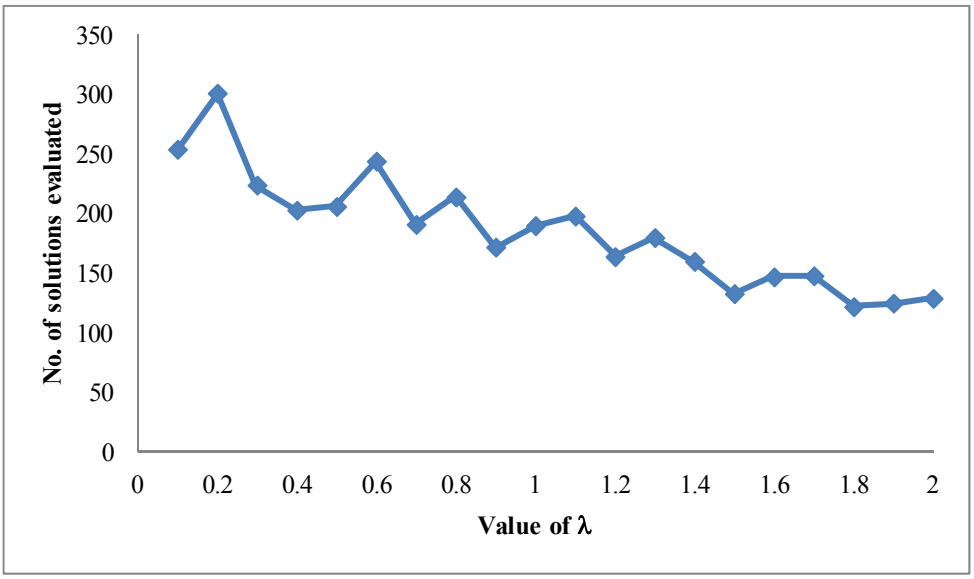


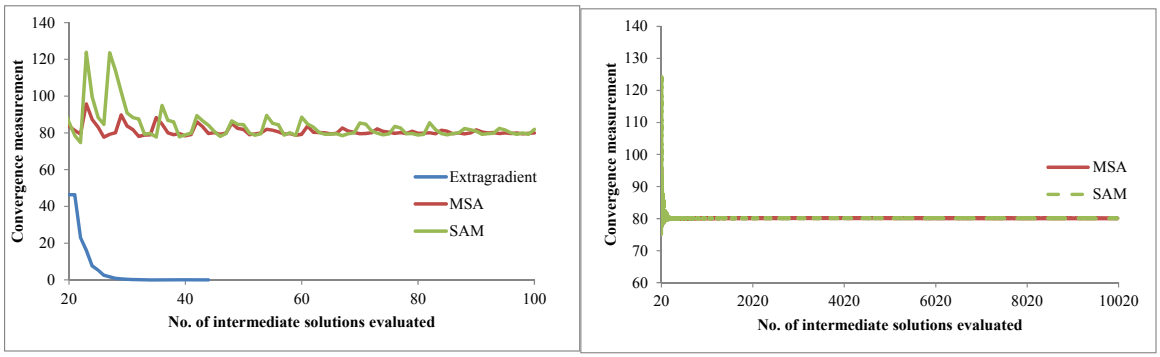
Figure 12 Effect of λ

4.2.2. Winnipeg Network

To show the efficiency of the proposed algorithm for solving large transit network assignment problems, the proposed method was tested using the Winnipeg network. The data

for this network were obtained from the base scenario in Emme3, which consists of 1067 nodes, 3647 OD pairs and 133 transit lines. As the original network is used for multimodal assignment, some OD pairs are not connected by transit lines. For simplicity, these OD pairs were eliminated. The proposed algorithm was compared to the self-regulated averaging method (SAM) developed by Liu *et al*. (2009). The SAM includes the method of successive averages (MSA) as a special case. In the SAM, two stepsize increment parameters are used for solution updates to increase the convergence speed. When both stepsizes are equal to 1, the SAM becomes equal to the MSA. Preliminary tests were carried out, which found that the best combination of the two stepsize increment parameters of the SAM was (1.0, 0.4).

The convergence curves are presented in Figure 13, from which we can see that the proposed projection method converges quickly after generating 44 intermediate solutions, but the SAM and MSA have longer tails. More importantly, both the SAM and MSA fail to converge, as is shown in Figure 13(b). In addition, we note that both curves are not smooth, probably because the mapping function is not strictly monotone (as the congestion cost function is asymmetric).



(a) Comparison of convergence (b) Convergence of the SAM and MSA

Figure 13 Comparison of convergence with the SAM and MSA

5. CONCLUSIONS

This paper develops a new formulation for transit assignment based on the concept of approach proportion. The problem is formulated as a VI. An extragradient method with adaptive stepsizes is adopted for solving the problem. Compared with the existing solution methods, the proposed method requires a mild assumption to converge. To implement the extragradient method, a simple linear projection operator is adopted which utilises the properties of the solution set. In addition, the paper proposes flow loading and cost updating algorithms, and discusses the computational complexity involved. The performance of the proposed algorithm is tested with the Sioux-Fall and Winnipeg networks, and the effects of

different parameters in the algorithm are demonstrated. The advantage of the solution algorithm is illustrated by comparing its convergence curve and speed with that of the SAM method.

This paper also illustrates a Braess-like paradox, in which adding a new line to a transit network or increasing the frequency of an existing transit line may worsen the network performance in terms of the ETSTC. To demonstrate this phenomenon, a small network example is created, and different scenarios are tested. The results illustrate that the occurrence of the paradox depends on the demand level, the congestion parameters, and the frequency setting. The results also show that the paradox can be avoided under certain frequency settings. These findings may call for a bilevel transit network design formulation to determine the optimal frequency to improve the system performance.

This paper opens up many research directions. For example, based on the bilevel frameworks of Szeto and Lo (2008), Lo and Szeto (2009), Miandoabchi *et al*. (2012a,b, 2013) and Zanjirani Farahani *et al*. (2013), the proposed model can be combined with the model of Szeto and Wu (2011) or Szeto and Jiang (2012) to develop various bilevel transit or multimodal network design models. The proposed model can also be extended to consider multiple user classes and other demand and supply uncertainties.

Acknowledgements

The work described in this paper was partially supported by a grant from the Central Policy Unit of the Government of the Hong Kong Special Administrative Region and the Research Grants Council of the Hong Kong Special Administrative Region, China (HKU7026-PPR-12), a grant (201111159056) from the University Research Committee, a grant from the National Natural Science Foundation of China (71271183), and a Research Postgraduate Studentship from the University of Hong Kong. The authors are grateful to the three referees for their constructive comments.

References

Ban, X.J., Liu, H.X., Ferris, M.C., Ran, B., 2008. A link-node complementarity model and solution algorithm for dynamic user equilibria with exact flow propagations. Transportation Research Part B 42(9), 823-842.

Bar-Gera, H., 2002. Origin-based algorithm for the traffic assignment problem. Transportation Science 36(4), 398-417.

Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: Characterization and computation of equilibria. Transportation Research Part B 40(6), 437-459.

Chriqui, C., Robillard, P., 1975. Common bus lines. Transportation Science 9(2), 115-121.

Cominetti, R., Correa, J., 2001. Common-lines and passenger assignment in congested transit networks. Transportation Science 35(3), 250-267.

Cormen, T.H., 2001. Introduction to Algorithms. The MIT Press, Cambridge, Mass.

Cortés, C.E., Jara-Moroni, P., Moreno, E., Pineda, C., 2013. Stochastic transit equilibrium. Transportation Research Part B 51, 29-44.

de Cea, J., Fernández, E., 1993. Transit assignment for congested public transport-systems: An equilibrium model. Transportation Science 27(2), 133-147.

Dial, R.B., 1967. Transit pathfinder algorithm. Highway Research Record 205, 67-85.

Dial, R.B., 2006. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. Transportation Research Part B 40(10), 917-936.

Fearnside, K., Draper, D., 1971. Public transport assignment—A new approach. Traffic Engineering and Control 12, 298-299.

Florian, M., Spiess, H., 1983. On binary mode choice/assignment models. Transportation Science 17(1), 32-47.

Hamdouch, Y., Ho, H.W., Sumalee, A., Wang, G., 2011. Schedule-based transit assignment model with vehicle capacity and seat availability. Transportation Research Part B 45(10), 1805-1830.

He, B., Liao, L.Z., 2002. Improvements of some projection methods for monotone nonlinear variational inequalities. Journal of Optimization Theory and Applications 112(1), 111-128.

Jackson, W.B., Jucker, J.V., 1982. An empirical study of travel time variability and travel choice behavior. Transportation Science 16(4), 460-475.

Johnson, G.G., 1972. Fixed points by mean value iterations. Proceedings of the American Mathematical Society 34(1), 193-194.

Korpelevich, G., 1976. The extragradient method for finding saddle points and other problems. Matecon 12, 747-756.

Kurauchi, F., Bell, M., Schmöcker, J.D., 2003. Capacity constrained transit assignment with common lines. Journal of Mathematical Modelling and Algorithms 2, 309-327.

Lam, W.H.K., Gao, Z., Chan, K., Yang, H., 1999. A stochastic user equilibrium assignment model for congested transit networks. Transportation Research Part B 33(5), 351-368.

Lam, W.H.K., Shao, H., Sumalee, A., 2008. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. Transportation Research Part B 42(10), 890-910.

Lam, W.H.K., Zhou, J., Sheng, Z.H., 2002. A capacity restraint transit assignment with elastic line frequency. Transportation Research Part B 36(10), 919-938.

Li, Z.C., Lam, W.H.K., Wong, S.C., 2009. The optimal transit fare structure under different market regimes with uncertainty in the network. Networks and Spatial Economics 9, 191-216.

Liu, H.X., He, X.Z., He, B.S., 2009. Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. Networks and Spatial Economics 9, 485-503.

Lo, H.K., Luo, X., Siu, B.W.Y., 2006. Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion. Transportation Research Part B 40(9), 792-806.

Lo, H.K., Szeto, W.Y., 2009. Time-dependent transport network design under cost-recovery. Transportation Research Part B 43(1), 142-158.

Miandoabchi, E., Daneshzand, F., Szeto, W.Y., Zanjirani Farahani, R., 2013. Multi-objective discrete urban road network design. Computers and Operations Research 40(10), 2429-2449.

Miandoabchi, E., Zanjirani Farahani, R., Dullaert, W. and Szeto, W.Y., 2012b. Hybrid evolutionary metaheuristics for concurrent multi-objective design of urban road and public transit networks. Networks and Spatial Economics 12(3), 441-480.

Miandoabchi, E., Zanjirani Farahani, R., Szeto, W.Y., 2012a. Bi-objective bimodal urban road network design using hybrid metaheuristics. Central European Journal of Operations Research 20(4), 583-621.

Nagurney, A., 1993. Network Economics: A Variational Inequality Approach. Kluwer Academic Publishers, Norwell, MA, USA.

Nguyen, S., Pallottino, S., 1988. Equilibrium traffic assignment for large scale transit networks. European Journal of Operational Research 37(2), 176-186.

Nie, Y.M., 2010. A class of bush-based algorithms for the traffic assignment problem. Transportation Research Part B 44(1), 73-89.

Noor, M.A., 2003. Extragradient methods for pseudomonotone variational inequalities. Journal of Optimization Theory and Applications 117(3), 475-488.

Nuzzolo, A., Crisalli, U., Rosati, L., 2012. A schedule-based assignment model with explicit capacity constraints for congested transit networks. Transportation Resarch Part C 20(1), 16-13.

Panicucci, B., Pappalardo, M., Passacantando, M., 2007. A path-based double projection method for solving the asymmetric traffic network equilibrium problem. Optimization Letters 1, 171-185.

Poon, M.H., Wong, S.C. and Tong, C.O., 2004. A dynamic schedule-based model for congested transit networks. Transportation Research Part B 38(4), 343-368.

Schaible, S., 1995. Generalized monotonicity-concepts and uses. In Variational Inequalities and Network Equilibrium Problems, F. Giannessi and A. Maugeri (eds.), pp. 289-299. Plenum, New York.

Schmöcker, J.D., Fonzone, A., Shimamoto, H., Kurauchi, F., Bell, M.G.H., 2011. Frequency-based transit assignment considering seat capacities. Transportation Research Part B 45(2), 392-408.

Siu, B.W.Y., Lo, H.K., 2008. Doubly uncertain transportation network: Degradable capacity and stochastic demand. European Journal of Operational Research 191(1), 166-181.

Spiess, H., 1984. *Contributions à la théorie et aux outils de planification de réseaux de transport urbain.* Ph.D. thesis, Département d'informatique et de recherche opérationnelle, Centre de recherche sur les transports, Université de Montréal.

Spiess, H., Florian, M., 1989. Optimal strategies: A new assignment model for transit networks. Transportation Research Part B 23(2), 83-102.

Sumalee, A., Tan, Z.J., Lam, W.H.K., 2009. Dynamic stochastic transit assignment with explicit seat allocation model. Transportation Research Part B 43(8-9), 895-912.

Sumalee, A., Uchida, K., Lam, W.H.K., 2010. Stochastic multi-modal transport network under demand uncertainties and adverse weather conditions. Transportation Research Part C 19(2), 338-350.

Szeto, W.Y., Jiang Y., 2012. Hybrid artificial bee colony algorithm for transit network design. Transportation Research Record 2284, 47-56.

Szeto, W.Y., Jiang, Y., Wong, K.I., Solayappan, M., 2013. Reliability-based stochastic transit assignment with capacity constraints: Formulation and solution method. Transportation Research Part C 35, 286-304.

Szeto, W.Y., Lo, H.K., 2008. Time-dependent transport network improvement and tolling strategies. Transportation Research Part A 42(2), 376-391.

Szeto, W.Y., Solayappan, M., Jiang, Y., 2011. Reliability-based transit assignment for congested stochastic transit networks. Computer-Aided Civil and Infrastructure Engineering 26(4), 311-326.

Szeto, W.Y., Wu, Y.Z., 2011. A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. European Journal of Operational Research 209(2), 141-155.

Tong, C.O., Wong, S.C., 1999. A stochastic transit assignment model using a dynamic schedule-based network. Transportation Research Part B 33(2), 107-121.

Wang, Y., Xiu, N., Wang, C., 2001. Unified framework of extragradient-type methods for pseudomonotone variational inequalities. Journal of Optimization Theory and Applications 111(3), 641-656.

Wu, J.H., Florian, M., Marcotte, P., 1994. Transit equilibrium assignment: A model and solution algorithms. Transportation Science 28(3), 193-203.

Zanjirani Farahani, R., Miandoabchi, E., Szeto, W.Y., Rashidi, H., 2013. A review of urban transportation network design problems. European Journal of Operational Research 229(2), 281-302.

Appendix I

This appendix proves that the link-based transit UE conditions (21) imply the path-based UE conditions (14), and vice versa.

Proof of Necessity

To prove necessity, we need to show that the link-based conditions (21) imply the path-based conditions (14). This proof is done by the following three steps.

Step 1. $(\pi_s^{id} - \pi^{id}) \cdot v_s^d = 0, \forall i, s, d$ implies that $(\tilde{\pi}_p^{rd} - \pi^{rd}) \cdot \tilde{f}_p^{rd} = 0, \forall r, d, p$.

Without loss of generality, we consider a path $p = \left(r = n_1 \overset{s_1}{-} n_2 \overset{s_2}{-} \cdots n_e \overset{s_e}{-} n_{e+1} \overset{s_{e+1}}{-} \cdots \overset{s_{n^p-1}}{-} n_{n^p} = d \right)$ in which all sections on this path carry flow (*i.e.*, $v_{s_e}^d > 0$), where $n_1, n_2 \cdots, n_{n^p}$ represent nodes on path *p*, and $s_1, s_2 \cdots,$ and s_{n^p-1} represent the sections emanating from nodes $n_1, n_2 \cdots,$ and n_{n^p-1}, respectively. As $v_{s_e}^d > 0$, the link-based condition $(\pi_{s_e}^{n_e d} - \pi^{n_e d}) \cdot v_{s_e}^d = 0$ requires that $\pi_{s_e}^{n_e d} - \pi^{n_e d} = 0$. By definition (17), we have $\pi^{n_{e+1} d} + c_{s_e} - \pi^{n_e d} = 0$. By rearranging this equation, we get $\pi^{n_e d} = \pi^{n_{e+1} d} + c_{s_e}$. This

recursive relationship implies that $\pi^{rd} = \sum_{e=1}^{n^p-1} c_{s_e} + \pi^{dd}$, where the left-hand side of the equation is the lowest expected travel cost between OD pair *rd*. For the right-hand side, the first term is the expected travel cost associated with path *p*, $\tilde{\pi}_p^{rd}$, and the second term equals zero by definition. We can therefore conclude that the expected travel cost of a used path *p* between an OD pair *rd* is equal to the minimum expected travel cost for that OD pair. This analysis can be repeated for other used paths between any OD pairs. We can therefore conclude that $(\pi_s^{id} - \pi^{id}) \cdot v_s^d = 0, \forall i, s, d$ implies $(\tilde{\pi}_p^{rd} - \pi^{rd}) \cdot \tilde{f}_p^{rd} = 0, \forall r, d, p$.

Step 2. $\pi_s^{id} - \pi^{id} \geq 0, \forall i, s, d$ implies that $\tilde{\pi}_p^{rd} - \pi^{rd} \geq 0, \forall r, d, p$.

Consider a path $p = \left(r = n_1 \xrightarrow{s_1} n_2 \xrightarrow{s_2} \cdots n_e \xrightarrow{s_e} n_{e+1} \xrightarrow{s_{e+1}} \cdots \xrightarrow{s_{n^p-1}} n_{n^p} = d \right)$. From the link-based condition $\pi_s^{id} - \pi^{id} \geq 0, \forall i, s, d$ and by definition (17), we have $\pi^{n_{e+1}d} + c_{s_e} - \pi^{n_e d} \geq 0$, $e = n_1, n_2 ..., n_{n^p-1}$. By adding all of these inequalities together, we have $\sum_{e=1}^{n^p-1} c_{s_e} - (\pi^{rd} - \pi^{dd}) \geq 0$. The first term is the expected travel cost associated with path *p*, $\tilde{\pi}_p^{rd}$, and the second term is the equilibrium cost between OD pair *rd*, given that $\pi^{dd} = 0$ by definition. By repeating such analysis on other paths, we can conclude that $\pi_s^{id} - \pi^{id} \geq 0, \forall i, s, d$ implies $\tilde{\pi}_p^{rd} - \pi^{rd} \geq 0, \forall r, d, p$.

Step 3. $v_s^d \geq 0, \forall s, d$ implies that $\tilde{f}_p^{rd} \geq 0, \forall r, d, p$.

Suppose that $v_s^d \geq 0$ holds, but $\tilde{f}_p^{rd} < 0$ for some path *p*. Consider a section *s* where there is only one path involved. As the flow on this path is negative, the flow on section *s* will be negative, contradicting the assumption. Hence, we conclude that $v_s^d \geq 0$ implies $\tilde{f}_p^{rd} \geq 0$.

Proof of Sufficiency

To prove sufficiency, we need to show that the path-based conditions (14) imply the link-based conditions (21). Again, the proof is done by the following three steps.

Step 1. $(\tilde{\pi}_p^{rd} - \pi^{rd}) \cdot \tilde{f}_p^{rd} = 0, \forall r, d, p$ implies that $(\pi_s^{id} - \pi^{id}) \cdot v_s^d = 0, \forall i, s, d$.

As $(\tilde{\pi}_p^{rd} - \pi^{rd}) \cdot \tilde{f}_p^{rd} = 0$ implies that if a path *p* between OD pair *rd* carries flow, the expected travel cost associated with path *p* is equal to the minimum expected travel cost for OD pair *rd*.

Consider a used path $p = \left(r = n_1 \overset{s_1}{-} n_2 \overset{s_2}{-} \cdots n_e \overset{s_e}{-} n_{e+1} \overset{s_{e+1}}{-} \cdots \overset{s_{n^p-1}}{-} n_{n^p} = d \right)$. The path-based conditions imply that for every node $n_e = n_2, n_3, \cdots, n_{n^p-1}$, subpath $\tilde{p}_e = \left(n_e \overset{s_e}{-} n_{e+1} \overset{s_{e+1}}{-} \cdots n_{n^p} = d \right)$ is the minimum expected travel cost path from node n_e to destination *d*. The subpath optimality condition can be proven by contradiction. Suppose that *p* is the minimum expected travel cost path from origin *r* to destination *d*, but $\tilde{p}_e$ is not the minimum expected travel cost path from node n_e to destination *d*. This condition implies that there is a lower expected cost path, $\overline{p}_e$, from n_e to destination *d*. As a result, a path is formed by $\overline{p}_e$ and subpath $r = n_1 \overset{s_1}{-} n_2 \overset{s_2}{-} \cdots n_e = d$ that connects *r* and *d*, and is shorter than path *p*, which contradicts the assumption. Therefore, we conclude that subpath $\tilde{p}_e$ is the minimum expected travel cost path from n_e to destination *d*.

The expected travel cost associated with subpath $\tilde{p}_e$ is

$$\pi^{n_e d} = \sum_{q=e}^{n^p-1} c_{s_q} .$$

Similarly, the expected travel cost from n_{e+1} to *d* is

$$\pi^{n_{e+1} d} = \sum_{q=e+1}^{n^p-1} c_{s_q} .$$

Based on the above two equations, we can obtain

$$\pi^{n_e d} - \pi^{n_{e+1} d} = \sum_{q=e}^{n^p-1} c_{s_q} - \sum_{q=e+1}^{n^p-1} c_{s_q} = c_{s_e} .$$

Rearranging the above equation, we have

$$\pi^{n_e d} - \left(\pi^{n_{e+1} d} + c_{s_e} \right) = 0 .$$

By definition, $\pi^{n_{e+1} d} + c_{s_e} = \pi_{s_e}^{n_e d}$. Hence, the above equation can be reduced to $\pi_{s_e}^{n_e d} - \pi^{n_e d} = 0$. As the flow on path *p* is greater than zero ($\tilde{f}_p^{rd} > 0$), the flow on each section of this path must be greater than zero (*i.e.*, $v_{s_e}^d > 0$), in which case $v_{s_e}^d > 0$ must occur simultaneously with condition $\pi_{s_e}^{n_e d} - \pi^{n_e d} = 0$. As this proof can be applied to any arbitrarily

minimum expected travel cost path of any OD pair, we conclude that $(\tilde{\pi}_p^{rd} - \pi^{rd}) \cdot \tilde{f}_p^{rd} = 0, \forall r, d, p$ implies $(\pi_s^{id} - \pi^{id}) \cdot v_s^d = 0, \forall i, s, d$.

Step 2. $\tilde{\pi}_p^{rd} - \pi^{rd} \geq 0, \forall r, d, p$ implies that $\pi_s^{id} - \pi^{id} \geq 0, \forall i, s, d$.

Consider two paths, namely $p' = \left(r = n_1 \xrightarrow{s_1} n_2 \xrightarrow{s_2} \cdots n_e, \ldots, n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \xrightarrow{s_{e'+1}} \cdots \xrightarrow{s_{n^p-1}} n_{n^p} = d \right)$ and $p = \left(r = n_1 \xrightarrow{s_1} n_2 \xrightarrow{s_2} \cdots n_e \xrightarrow{s_{e''}} n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \xrightarrow{s_{e'+1}} \cdots \xrightarrow{s_{n^p-1}} n_{n^p} = d \right)$. Path p' is the minimum expected travel cost path between OD pair *rd*, whereas *p* consists of section $s_{e''} = (n_e, n_{e'})$ and two subpaths of *p*: $\left(r = n_1 \xrightarrow{s_1} n_2 \xrightarrow{s_2} \cdots n_e \right)$ and $\left(n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \xrightarrow{s_{e'+1}} \cdots \xrightarrow{s_{n^p-1}} n_{n^p} = d \right)$. These two subpaths are connected by section $s_{e''} = (n_e, n_{e'})$. As path p' is the minimum expected travel cost path between OD pair *rd*, the expected travel cost associated with p' is $\tilde{\pi}_{p'}^{rd} = \tilde{\pi}^{rd}$. Hence, the difference in the expected travel cost between *p* and p' can be represented as $\tilde{\pi}_p^{rd} - \tilde{\pi}_{p'}^{rd} = \tilde{\pi}_p^{rd} - \tilde{\pi}^{rd}$. Because of the path-based condition $\tilde{\pi}_p^{rd} - \pi^{rd} \geq 0, \forall r, d, p$, the difference $\tilde{\pi}_p^{rd} - \tilde{\pi}^{rd}$ is non-negative. The difference is actually the expected travel cost difference between the two non-overlapping sub-paths because

$\tilde{\pi}_{p'}^{rd} - \tilde{\pi}^{rd} = \tilde{\pi}_{p'}^{rd} - \tilde{\pi}_p^{rd} = \left(\sum_{q=1}^{e-1} c_{s_q} + c_{s_{e''}} + \sum_{q=e'}^{n^p-1} c_{s_q} \right) - \left(\sum_{q=1}^{e-1} c_{s_q} + \sum_{q=e}^{e'-1} c_{s_q} + \sum_{q=e'}^{n^p-1} c_{s_q} \right) = c_{s_{e''}} - \sum_{q=e}^{e'-1} c_{s_q} \geq 0$. The term $\sum_{q=e}^{e'-1} c_{s_q}$ represents the minimum expected travel cost from n_e to $n_{e'}$ on the lowest expected travel cost path p', and this term can be replaced by $\left(\sum_{q=e}^{n^p-1} c_{s_q} \right) - \left(\sum_{q=e'}^{n^p-1} c_{s_q} \right)$, where the first bracket term equals $\pi^{n_e d}$ because of subpath optimality for $\left(n_e, \ldots, n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \xrightarrow{s_{e'+1}} \cdots \xrightarrow{s_{n^p-1}} n_{n^p} = d \right)$, and the second bracket term equals $\pi^{n_{e'} d}$ because of subpath optimality for $\left(n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \xrightarrow{s_{e'+1}} \cdots \xrightarrow{s_{n^p-1}} n_{n^p} = d \right)$. Therefore, we can obtain $c_{s_{e''}} - (\pi^{n_e d} - \pi^{n_{e'} d}) \geq 0$. Rearranging the above equation, we obtain $\pi^{n_e d} - (c_{s_{e''}} + \pi^{n_{e'} d}) \geq 0$. As the sum inside the round brackets is the expected travel cost associated with subpath

$\left(n_e \xrightarrow{s_{e''}} n_{e'} \xrightarrow{s_{e'}} n_{e'+1} \xrightarrow{s_{e'+1}} \cdots \xrightarrow{s_{n^p-1}} n_{n^p} = d \right)$ (*i.e.*, $c_{s_{e''}} + \pi^{n_{e'}d} = \pi^{n_e d}_{s_{e''}}$), we have $\pi^{n_e d}_{s_{e''}} - \pi^{n_e d} \geq 0$. The above proof can be applied to any section $s \in A_i^+, \forall i \in N$ in a network. Hence, we can conclude that $\tilde{\pi}_p^{rd} - \pi^{rd} \geq 0, \forall r, d, p$ implies $\pi_s^{id} - \pi^{id} \geq 0, \forall i, s, d$.

Step 3. $\tilde{f}_p^{rd} \geq 0, \forall r, d, p$ implies that $v_s^d \geq 0, \forall s, d$.

The non-negativity condition of route flows (*i.e.*, $\tilde{f}_p^{rd} \geq 0, \forall r, d, p$) implies the non-negativity condition of section flows, as section flow is the sum of route flows on that section, and the sum of non-negative real numbers must be non-negative.

This completes the proof. □

Appendix II

In this appendix, we prove that the link-based UE problem (1) – (8), (20) and (21) is equivalent to the VI formulation (22). In other words, we prove that the link-based flow pattern satisfies (1) – (8), (20) and (21) if and only if the flow pattern satisfies the VI formulation (22). As the solutions to both problems must satisfy (1) – (8), and (20) by definition, we only need to prove the equivalence between the UE conditions (21) and VI (22).

Proof of Necessity

We need to prove that the UE constraints (21) imply VI (22). For any section $s \in A_i^+, i \in N$, a feasible section flow heading to $d \in D$ must be non-negative:

$v_s^d \geq 0$.

Multiplying the above inequality and the second condition of (21), we have

$\left(\pi_s^{id} - \pi^{id}\right) v_s^d \geq 0.$

We add the first equation in constraints (21) to the above inequality and obtain

$$\left(\pi_s^{id} - \pi^{id}\right) v_s^d - \left(\pi_s^{id} - \pi^{id}\right) v_s^{d*} \geq 0, \quad \forall i \in N, d \in D, s \in A_i^+. \tag{29}$$

where v_s^{d*} represents the flow on section *s* toward destination *d* at user equilibrium.

In summing up (29) over all of the links, nodes and destinations, it follows that

$$\sum_i \sum_d \sum_{s \in A_i^+} \left(\pi_s^{id} - \pi^{id}\right)\left(v_s^d - v_s^{d*}\right) \geq 0,$$

which is VI (22).

Proof of Sufficiency

We need to prove that any solution to the VI problem satisfies the link-based UE conditions. According to the constraint set Ω_v, the optimal solution must be $v_s^{d*} \geq 0$. Hence, the third condition of UE constraints (21) must be satisfied by the optimal solution of VI (22). By definition, $\pi_s^{id} - \pi^{id} \geq 0$ must be satisfied. Therefore, the second condition of UE constraints (21) must be satisfied, and we only need to prove that

$$v_s^{d*}\left(\pi_s^{id} - \pi^{id}\right) = 0, \forall i \in N, d \in D, s \in A_i^+ .$$

If we assume that $v_{s'}^{d'*}\left(\pi_{s'}^{i'd'} - \pi^{i'd'}\right) = 0$ is not satisfied for a specific section $s' \in A_{i'}^+, i' \in N$, and for a specific flow heading to $d' \in D$, but the VI solution satisfies $v_s^{d*}\left(\pi_s^{id} - \pi^{id}\right) = 0$ for all other cases, then we have $v_{s'}^{d'*}\left(\pi_{s'}^{i'd'} - \pi^{i'd'}\right) > 0$ and

$$\sum_i \sum_d \sum_{s \in A_i^+} v_s^{d*}\left(\pi_s^{id} - \pi^{id}\right) > 0 . \tag{30}$$

We can also find a feasible solution $\mathbf{v} = [v_s^d \mid v_s^d \in \Omega_v]$ by the following method. First, for each node *i*, we can always find one section s'' that is on the minimum expected travel cost path to destination *d*. It follows that

$$\pi_{s''}^{id} - \pi^{id} = 0, \forall i \in N, d \in D .$$

Then, for each node *i*, we can assign all flows heading to *d* and passing through this node to that section and obtain $v_{s''}^d > 0$. As $\pi_{s''}^{id} - \pi^{id} = 0$, $v_{s''}^d\left(\pi_{s''}^{id} - \pi^{id}\right) = 0, \forall i \in N, d \in D$ for s'' on the minimum expected travel cost path to destination *d*. The other sections, s''', are not on the minimum expected travel cost path to *d*. Hence, $v_{s'''}^d = 0$, and $v_{s'''}^d\left(\pi_{s'''}^{id} - \pi^{id}\right) = 0, \forall i \in N, d \in D$. For both cases, we have

$$v_s^d\left(\pi_s^{id} - \pi^{id}\right) = 0, \forall i \in N, d \in D, s \in A_i^+ . \tag{31}$$

Summing (31) over all links, nodes and destinations, it follows that

$$\sum_i \sum_d \sum_{s \in A_i^+} v_s^d \left(\pi_s^{id} - \pi^{id}\right) = 0 . \tag{32}$$

If we subtract (32) from (30), then we obtain

$$\sum_i \sum_d \sum_{s \in A_i^+} \left(v_s^d - v_s^{d*}\right)\left(\pi_s^{id} - \pi^{id}\right) < 0 .$$

The above condition contradicts the VI formulation (22). Therefore, an optimal solution of the VI problem (22) must satisfy the conditions (21). This completes the proof.□