3. Network Equilibria

Annals of Operations Research 44 (1993) 245–260

A simplicial decomposition method for the transit equilibrium assignment problem

Jia Hao Wu and Michael Florian

Centre de Recherche sur les Transports, Université de Montréal, C.P. 6128, Succursale A, Montréal, Canada H3C 3J7

A transit equilibrium assignment problem assigns the passenger flows on to a congested transit (public transportation) network with asymmetric cost functions and a fixed origin–destination matrix. This problem which may be formulated in the space of hyperpath flows, is transformed into an equivalent problem in the space of total arc flows and an auxiliary variable. A simplicial decomposition algorithm is developed and its convergence is proved under the usual assumptions on the cost functions. The algorithm requires relatively little memory and its efficiency is demonstrated with computational results.

Keywords: Transit equilibrium assignment, variational inequalities, simplicial decomposition.

1. Introduction

A transit equilibrium assignment problem (TEAP) assigns the passenger flows on to a congested transit (or public transportation) network with asymmetric cost functions and a fixed origin–destination matrix. In a congested transit network, the passengers' travel costs depend on the passenger flows. The passengers' travel behaviour is assumed to satisfy the user equilibrium principle [17]. The transit network consists of zone centroids, where passenger trips are generated and attracted, the underlying road network and the transit lines that use it. The zone centroids, the road network and the transit lines are used to generate a general network in which there are four kinds of arcs: walk arcs, wait arcs, in-vehicle arcs and transfer / alight arcs.

The transit assignment problem has attracted the interest of many researchers and practitioners. In many cities the transit services are essential to ensuring the mobility of the inhabitants, either due to increasing road congestion or the unavailability of the private car mode, such as in developing countries. In order to provide a suitable service, one has to understand the route choices that passengers make in the transit network.

The transit assignment (equilibrium) problem has been studied for more than a decade. Many models and algorithms were proposed for its solution. Two modelling aspects are relevant when one studies the TEAP. The first one concerns the assumption about the passenger route choice in the uncongested case, while the second aspect is related to how the level of congestion in the transit network affects the passengers' travel behaviour. Different ways of approaching these modelling aspects lead to various models. A good survey of the models used for uncongested transit networks is given by Spiess and Florian [16]. In this case, several assumptions have been proposed. The "all-or-nothing" assignment is a traditional assignment model which assumes that the passengers select shortest paths based on travel costs that are independent of passenger flows. The multipath assignment assumes that several paths are found first and the passengers are assigned to paths in some proportion of their costs. Not all of these methods consider the passengers' behaviour at a bus stop carefully. In fact, to model the choices made by the passengers waiting at a bus stop, that is shared by several transit lines, all the models assume a certain behaviour. These choices are called strategies. Spiess [15] and Spiess and Florian [16] formulated a transit assignment model for an uncongested (i.e., fixed travel time) transit network, based on the concept of an optimal strategy, as a linear programming problem, and solved it by a polynomial time algorithm. In this model, the passenger flows that correspond to the optimal strategy may use more than one elementary path. Similarly, De Cea and Fernandez [3] used a minimal route concept to formulate the transit assignment problem as a linear program.

Nguyen and Pallottino [11] characterized and studied in detail the notion of the strategy, by using a graph theoretical approach, and developed the concept of a hyperpath. More precisely, a hyperpath from p to q is an acyclic subgraph which is the union of a subset of elementary paths from p to q. To each arc of the subgraph is associated a conditional probability that it is traversed by a passenger who arrives at the beginning node of the arc. A strategy corresponds to a hyperpath, i.e., a subnetwork containing all the arcs of the corresponding strategy. Spiess's model assumes that, at a bus stop, passengers always board the first vehicle to arrive from the set of attractive lines of the hyperpath at that stop. The demand at a node is then assigned to the hyperpath in proportion to the fixed frequencies of the lines retained in the attractive set of the hyperpath at the node.

Although this class of models can be useful and efficient to study transit systems where the impact of congestion effects on the route choice is low, they are unsuitable in the simulation of transit flows in some cities where high congestion levels are so severe that they affect the passenger route choices. This is the second aspect to consider.

Gendreau [7] and De Cea and Fernandez [4] extended the TEAP to include the consideration of congestion effects not only in the in-vehicle cost but also on the waiting time. The first one models the waiting times by using a queueing theoretical approach on an expanded network, while the second models the waiting time to be a direct function of the flow-dependent effective frequency. However, only an algorithm for a simplified version of the model is presented. Spiess and Florian [16] gave a general version of a nonlinear model in which in-vehicle travel times (or generalized travel costs) are increasing functions of passenger flows (called "discomfort functions"). The main limitations of these approaches are that waiting times at stops are not affected by congestion and there are no cross effects of flows on the cost functions. A similar formulation was also proposed by Nguyen and Pallottino [11]. They showed that, by explicitly defining the flows and structures of the hyperpath, the standard equilibrium traffic assignment model could be adapted to transit networks by substituting the hyperpath flow space for the path flow space. As in Spiess and Florian's model, the waiting times are not flow dependent. It is noted that this adaption is carried in the hyperpath flow space (not in the total arc flow space). In this sense, the transit equilibrium assignment model is different from the traditional traffic equilibrium model. We note that Spiess and Florian [16] adapted the linear approximation method for a symmetric transit equilibrium assignment problem with flow dependent arc times.

Recently, Wu et al. [18] formulated a transit equilibrium assignment model (WFM) as a variational inequality problem and solved it by using a linearized approximation method in the hyperpath flow space. Since the WFM model is defined in the hyperpath space, the space required to store the hyperpath information is large, even though a restriction strategy is used to store only the necessary hyperpaths. This space increases as the number of zone centroids and the size of the transit network increase.

This paper is devoted to the development of a method for the transit equilibrium assignment problem. First, the original variational inequality is converted to an equivalent variational inequality problem in a new variable space; then a simplicial decomposition method [10,14] is used to solve it. The advantages of this approach for the large scale transit equilibrium assignment are the following:

(1) One need only store the limited number of extreme points in the space of total arc flows plus one auxiliary variable (total waiting time due to the frequency delay). This requires much less memory than the WFM algorithms, which operate in the space of hyperpath flows.
(2) The solution method for the resulting quadratic subproblems is a one-dimensional search method and an extreme point is generated by a solution of a shortest hyperpath problem.

However, one potential difficulty is that the cost function, in the new variable space, may not be strongly monotone, a condition which is required for the convergence of the simplicial decomposition method if an iterative method is used for the restricted variational inequality problem. In the following, we will show that only the strong monotonicity of the cost functions in the space of the total arc flows, and not in the new variable space, is required for the convergence of the simplicial decomposition algorithm.

The rest of the paper is organized as follows. Section 2 presents the basic definitions and the formulations of the transit equilibrium assignment problem. In section 3, we give the basic simplicial decomposition algorithm for the equivalent transit equilibrium assignment problem and a discussion of the steps of the algorithm. Section 4 presents the method used to solve the restricted variational inequality problem. In section 5, we report some computational results. Finally, section 6 concludes the paper.

2. Basic definitions and formulations

In the following we assume that all the vectors are column vectors.

As shown in the WFM model, the transit equilibrium assignment problem can be formulated as the following variational inequality problem: find $h^* \in \Omega_h$ such that

$$\mathbf{VIP}(\Omega_h, S): \qquad S(h^*)^{\mathrm{T}}(h - h^*) \geqslant 0, \quad h \in \Omega_h . \tag{1}$$

Here $S(h) \in \mathbb{R}^m$ is a vector of hyperpath costs; $h \in \mathbb{R}^m$ is a vector of hyperpath flows; $h^* \in \mathbb{R}^m$ is a vector of equilibrium hyperpath flows and Ω_h is the feasible region satisfying

$$\sum_{k \in K_i} h_k = g_i, \qquad i \in I ,$$

$$h_k \geqslant 0, \qquad k \in K_i, i \in I ,$$

where h_k is the flow along hyperpath k; i denotes the ith origin–destination pair (p, q); I is the set of all such indexes i; g_i is the travel demand from origin p to destination q, $(p, q) = i$; and K_i is the set of hyperpaths connecting the ith origin–destination pair.

It is known [11,18] that

$$S(h^*)^{\mathrm{T}}(h - h^*) = s(v^*)^{\mathrm{T}}(v - v^*) + W^{\mathrm{T}}(h - h^*) , \tag{2}$$

where $s(v) = (s_a(v))$ is a vector of arc costs, $v = \delta h$ (δ is the arc–hyperpath incidence matrix) and $W = (w_k)$ is a vector of hyperpath waiting times that depends on the frequencies of the transit lines, i.e.,

$$v_a = \sum_{i \in I} \sum_{k \in K_i} \delta_{ak} h_k$$

and

$$w_k = \sum_{a \in A} \delta_{ak} \frac{1}{\sum_{a' \in FS_k(t(a))} f_{a'}} ,$$

where v_a is the total flows on arc a; w_k is the waiting time on the hyperpath k due to the frequency delay in the general network; A is the set of arcs in the general network; f_a is the bus frequency associated with arc a; $t(a)$ is the tail of arc a and $FS_k(t)$ is the set of forward arcs at the node t.

Equation (2) appears to use explicitly the hyperpath related information. However, this is unnecessary. Consider the following new variable space $\hat{\Omega}$ defined by

$$\hat{\Omega} = \left\{ \begin{pmatrix} v \\ u \end{pmatrix} : \begin{pmatrix} v \\ u \end{pmatrix} = \begin{pmatrix} \delta \\ W^{\mathrm{T}} \end{pmatrix} h, \forall h \in \Omega_h \right\}. \tag{3}$$

We denote $\hat{\delta} = \left(\delta, W^{\mathrm{T}}\right)^{\mathrm{T}}$. Then we have the following result.

PROPOSITION 2.1

Assume that the set $\Omega_v = \{v : v = \delta h, \forall h \in \Omega_h\}$ is a bounded polyhedron. The new variable space $\hat{\Omega}$ is also a bounded polyhedron.

Proof

First, we show that $\hat{\Omega}$ is a polyhedron. According to (3), we know that $\hat{\Omega}$ is the region obtained from a polyhedron via a linear transformation $\hat{\delta}$. Therefore $\hat{\Omega}$ is a polyhedron. Since we assume that $\Omega_v = \{v : v = \delta h, \forall h \in \Omega_h\}$ is bounded, W is a nonnegative constant vector and Ω_h is bounded, then $\hat{\Omega}$ is also bounded. □

By definition, $\hat{\Omega}$ is also a convex set and it can be expressed as a linear combination of its extreme points, i.e.,

$$Co(e_1, e_2, \ldots) = \hat{\Omega},$$

where Co denotes the convex hull of the extreme points $(e_1, e_2, \ldots)$ of $\hat{\Omega}$.

According to (3), we can see that every distinct extreme point in $\hat{\Omega}$ corresponds to a distinct extreme point in Ω_h. The reverse is not necessarily true since $\hat{\delta}^{\mathrm{T}} \hat{\delta}$ may be singular. Thus a set of extreme points in Ω_h may correspond to a single point in $\hat{\Omega}$.

Proposition 2.1 indicates that the problem can be formulated in the space of total arc flows v and the auxiliary variable u, by using successive hyperpath computations to generate extreme points. The space $\hat{\Omega}$ is gradually constructed by taking the convex hull of the generated points of $\hat{\Omega}$. Now we consider the following proposition.

PROPOSITION 2.2

The variational inequality problem (1) is equivalent to finding $(v^*, u^*)^{\mathrm{T}} \in \hat{\Omega}$ such that

$$s(v^*)^{\mathrm{T}}(v - v^*) + (u - u^*) \geqslant 0, \quad \forall \begin{pmatrix} v \\ u \end{pmatrix} \in \hat{\Omega}, \tag{4}$$

where $\hat{\Omega}$ is the new variable space.

Proof

Consider $(v^*, u^*)^{\mathrm{T}} \in \hat{\Omega}$ which satisfies

$$s(v^*)^{\mathrm{T}}(v - v^*) + (u - u^*) \geqslant 0, \quad \forall \begin{pmatrix} v \\ u \end{pmatrix} \in \hat{\Omega},$$

It follows that there exists at least an $h^* \in \Omega_h$ such that

$$s(\delta h^*)^{\mathrm{T}}(\delta h - \delta h^*) + W^{\mathrm{T}}(h - h^*) \geqslant 0, \quad \forall h \in \Omega_h ,$$

which again implies that

$$S(h^*)^{\mathrm{T}}(h - h^*) + W^{\mathrm{T}}(h - h^*) \geqslant 0 .$$

Conversely, suppose that we have $h^* \in \Omega_h$ such that

$$S(h^*)^{\mathrm{T}}(h - h^*) + W^{\mathrm{T}}(h - h^*) \geqslant 0, \quad \forall h \in \Omega_h .$$

This means that

$$(\delta^{\mathrm{T}} s(v^*))^{\mathrm{T}}(h - h^*) + u - u^* \geqslant 0 .$$

It follows then that

$$s(v^*)^{\mathrm{T}}(v - v^*) + (u - u^*) \geqslant 0, \quad \forall \begin{pmatrix} v \\ u \end{pmatrix} \in \hat{\Omega} . \qquad \square$$

The following proposition is a direct consequence of a well-known theorem in variational inequality theory [10], i.e.,

PROPOSITION 2.3

The variational inequality problem (4) is equivalent to the following problem:

$$\min_{(v,u)^{\mathrm{T}} \in \hat{\Omega}} G\begin{pmatrix} v \\ u \end{pmatrix} = -s(v)^{\mathrm{T}}(\hat{v} - v) - (\hat{u} - u) , \tag{5}$$

where $(\hat{v}, \hat{u})^{\mathrm{T}}$ is a solution of the following linear program:

$$\min_{(x,y)^{\mathrm{T}} \in \hat{\Omega}} s(v)^{\mathrm{T}}(x - v) + (y - u)$$

and $G(v^*, u^*)^{\mathrm{T}} = 0$ if and only if $(v^*, u^*)^{\mathrm{T}}$ is the solution of the variational inequality problem (4).

G is called the gap function and is used as a stopping criterion in an algorithm that will be discussed below.

3. The basic simplicial decomposition algorithm

In the following, we give a simplicial decomposition (SDA) and discuss in detail the computations required at each of its steps.

ALGORITHM SDA

Step 1 *(Initialization)*: find an extreme point $(e_v^1, e_u^1)^{\mathrm{T}}$ of $\hat{\Omega}$ and define

$$E^1 = \begin{pmatrix} E_v^1 \\ E_u^1 \end{pmatrix} = (e^1) = \begin{pmatrix} e_v^1 \\ e_u^1 \end{pmatrix}.$$

Let $\bar{G}^1 = \infty$, β be a positive parameter and $l = 1$. Let $\hat{\Omega}^l$ be the convex hull of the columns of E^l and ϵ^l be an error such that $\epsilon^l \to 0$ as $l \to \infty$. $(v^1, u^1)^{\mathrm{T}} = (e_v^1, e_u^1)^{\mathrm{T}}$.

Step 2 *(Solve the restricted variational inequality problem)*: find $(v^l, u^l)^{\mathrm{T}} \in \hat{\Omega}^l$ such that

$$s(v^l)^{\mathrm{T}}(v - v^l) + (u - u^{l+1}) \geqslant -\epsilon^l, \quad \forall \begin{pmatrix} u \\ v \end{pmatrix} \in \hat{\Omega}^l. \tag{6}$$

Step 3 *(Solve the shortest hyperpath)*:

$$\begin{pmatrix} e_v \\ e_u \end{pmatrix} = \arg \min_{(v,u)^{\mathrm{T}} \in \hat{\Omega}} s(v^l)^{\mathrm{T}} v + u. \tag{7}$$

Step 4 *(Stopping criterion)*: **if** $G(v^l, u^l)^{\mathrm{T}} = 0$, **then** stop. $(v^l, u^l)^{\mathrm{T}}$ is the solution.

Step 5 *(Update E)*: **if** $G(v^l, u^l)^{\mathrm{T}} \geqslant \bar{G}^l - \beta$, **then** $E^{l+1} = E^l \cup \{e_v, e_u\}^{\mathrm{T}}$; **if** $G(v^l, u^l)^{\mathrm{T}} < \bar{G}^l - \beta$, **then** $E^{l+1} = E^l \backslash D^l \cup \{e_v, e_u\}^{\mathrm{T}}$;
where D^l is the columns of E^l with zero weight in the expression of $(v^l, u^l)^{\mathrm{T}}$ as a convex combination of columns of E^l. Let

$$\bar{G}^{l+1} = \min\left\{ \bar{G}^l, G^l \begin{pmatrix} v^l \\ u^l \end{pmatrix} \right\}.$$

$l = l + 1$. Goto step 2.

In step 1, the $(e_v^1, e_u^1)^{\mathrm{T}}$ can be found by the solution of a shortest hyperpath problem where the arc costs are chosen to be $s(0)$. Step 2 will be discussed in section 4. In step 3, as in step 1, the shortest hyperpath problem can be solved in polynomial time (see [11, 16]).

Let us consider the convergence of the simplicial decomposition algorithm (SDA) for the equivalent variational inequality problem. We define the steps which compute the shortest hyperpath problems as major cycles (also called the subproblems) and the steps of solutions of the restricted variational inequality problem as minor cycles.

PROPOSITION 3.1

If each restricted variational problem in the minor cycles is well defined, then the SDA converges to a unique solution $(v^*, u^*)^{\mathrm{T}}$ of the equivalent variational inequality problem.

Proof

For the convergence, see [10]. For the uniqueness, see [18]. □

The restricted variational inequality problem is in fact well defined which can be seen in section 4.

4. Solving the restricted variational inequality problem

In this section, we give a solution method for the restricted variational inequality problem (6). Its solution requires a sequence of affine variational inequality problems in $\hat{\Omega}^l$. A possible algorithm, which we denote the linearized method, is the following. Let $B(v)$ be a symmetric and positive definite matrix. For an initial solution $(v^0, u^0)^{\mathrm{T}} \in \hat{\Omega}^l$, solve $(v^{k+1}, u^{k+1})^{\mathrm{T}} = H(v^k, u^k)^{\mathrm{T}}$ for $k = 0, 1, 2, \ldots,$ where $H(v^k, u^k)^{\mathrm{T}}$ is the solution of the following quadratic programming problem:

$$\min_{(v,u)^{\mathrm{T}} \in \hat{\Omega}^l} s(v^k)^{\mathrm{T}}(v - v^k) + \frac{1}{2\alpha}(v - v^k)^{\mathrm{T}} B(v^k)(v - v^k) + (u - u^k), \tag{8}$$

where α is a positive number. Let λ^k satisfy $(v^k, u^k)^{\mathrm{T}} = E^l \lambda^k$ and $\Omega^l_\lambda = \{\lambda | \sum_{i=1}^l \lambda_i = 1, \lambda_i \geqslant 0, i = 1, \ldots, l\}$. The quadratic program (8) is equivalent to

$$\min_{\lambda \in \Omega^l_\lambda} \hat{s}(\lambda^k)^{\mathrm{T}}(\lambda - \lambda^k) + \frac{1}{2\alpha}(\lambda - \lambda^k)^{\mathrm{T}} \hat{B}(\lambda^k)(\lambda - \lambda^k), \tag{9}$$

where $\hat{s}(\lambda^k)^{\mathrm{T}} = s(E^l_v \lambda^k)^{\mathrm{T}} E^l_v + E^l_u = s(v^k)^{\mathrm{T}} E^l_v + E^l_u$ and $\hat{B}(\lambda^k)$ is assumed to be symmetric positive definite and is an approximation of $(E^l_v)^{\mathrm{T}} B(v^k) E^l_v$. It is noted that (i) $(E^l_v)^{\mathrm{T}} E^l_v$ is a nonsingular matrix, since all the columns in E^l_v are linearly independent; (ii) $\hat{B}(\lambda^k)$ is positive definite if and only if $B(v^k)$ is positive definite.

Several procedures have been proposed to solve the problem (9). Here we use the method developed by Dussault et al. [5], where the problem is solved by a sequence of separable quadratic programs each solvable by one-dimensional search algorithms [8,9]. In our implementation, we define the linearized method as a projection method, if $\hat{B}(\lambda^k)$ is a constant diagonal and positive definite matrix, and a linearized Jacobi method if $\hat{B}(\lambda^k)$ is the diagonal of $E_v^{\mathrm{T}} \nabla s(v^k) E_v$. Thus, the problem (9) is solved by a sequence of the one-dimensional search computations.

Now we consider the global convergence of the linearized method for the restricted variational inequality problem and show that the restricted variational inequality problem is well defined.

THEOREM 4.1 (CONVERGENCE THEOREM OF THE LINEARIZED METHOD)

Assume that $\hat{\Omega}_v$ is nonempty closed, compact and convex. Let $s(v)$ be a continuous mapping from $\mathbb{R}^m$ into itself. Let $B(v)$ be a symmetric positive definite matrix from $\mathbb{R}^m$ to $\mathbb{R}^m \times \mathbb{R}^m$. Then there exists a symmetric positive definite matrix $\bar{B}$ with

$B(v) - \bar{B}$ positive semi-definite for all $v \in \Omega_v$. If there is a parameter $0 < \theta < 1$ such that, for all $(v', u')^{\mathrm{T}}, (v'', u'')^{\mathrm{T}} \in \hat{\Omega}^l$,

$$\|\bar{B}^{-1}\{\alpha[s(v') - s(v'')] - B(v')(v' - v'')\}\|_{\bar{B}} \leqslant \theta \|v' - v''\|_{\bar{B}}, \tag{10}$$

then the sequence of iterates $\{v^k, u^k\}^{\mathrm{T}}$ generated by the linearized method converges to the unique solution of the restricted variational inequality problem for any initial feasible vector $(v^0, u^0)^{\mathrm{T}} \in \hat{\Omega}^l$.

Proof

The existence of $\bar{B}$ is obvious since $\bar{B}$ can be selected such that

$$\bar{B} = \lambda_{min} I,$$

where λ_{min} is the minimum eigenvalue of $B(v)$ and I is a corresponding identity matrix.

Since $(v^k, u^k)^{\mathrm{T}}$ and $(v^{k+1}, u^{k+1})^{\mathrm{T}}$ are the solutions of the quadratic program (8), we have the following two inequalities

$$\left[s(v^{k-1}) + \frac{1}{\alpha} B(v^{k-1})(v^k - v^{k-1})\right]^{\mathrm{T}} (v - v^k) + (u - u^k) \geqslant 0, \quad \forall \begin{pmatrix} v \\ u \end{pmatrix} \in \hat{\Omega}^l \tag{11}$$

and

$$\left[s(v^k) + \frac{1}{\alpha} B(v^k)(v^{k+1} - v^k)\right]^{\mathrm{T}} (v - v^{k+1}) + (u - u^{k+1}) \geqslant 0, \quad \forall \begin{pmatrix} v \\ u \end{pmatrix} \in \hat{\Omega}^l. \tag{12}$$

By replacing v and u in (11) and (12) by v^{k+1} and u^{k+1} and v^k and u^k respectively, and by adding them up and rearranging the terms, we obtain

$$\begin{aligned}
&(v^{k+1} - v^k)^{\mathrm{T}} \frac{1}{\alpha} B(v^k)(v^{k+1} - v^k) \\
&\leqslant [s(v^{k-1}) + \frac{1}{\alpha} B(v^{k-1})(v^k - v^{k-1}) - s(v^k)]^{\mathrm{T}} (v^{k+1} - v^k) \\
&\leqslant \|\bar{B}^{-1}\{s(v^{k-1}) - s(v^k) - \frac{1}{\alpha} B(v^{k-1})(v^{k-1} - v^k)\}\|_{\bar{B}} \|v^{k+1} - v^k\|_{\bar{B}},
\end{aligned}$$

where the last inequality follows from the fact that $a^{\mathrm{T}} b = a^{\mathrm{T}} \bar{B}^{-1} \bar{B}^{1/2} \bar{B}^{1/2} b \leqslant \|\bar{B}^{1/2} \bar{B}^{-1} a\|_2 \|\bar{B}^{1/2} b\|_2 \leqslant \|\bar{B}^{-1} a\|_{\bar{B}} \|b\|_{\bar{B}}$. Thus we have

$$\begin{aligned}
&(v^{k+1} - v^k)^{\mathrm{T}} B(v^k)(v^{k+1} - v^k) \\
&\leqslant \|\bar{B}^{-1}\{\alpha[s(v^{k-1}) - s(v^k)] - B(v^{k-1})(v^{k-1} - v^k)\}\|_{\bar{B}} \|v^{k+1} - v^k\|_{\bar{B}} \\
&\leqslant \theta \|v^{k-1} - v^k\|_{\bar{B}} \|v^{k+1} - v^k\|_{\bar{B}} \quad \text{(using (10))}.
\end{aligned} \tag{14}$$

However, the assumption that $B(v) - \bar{B}$ is positive semi-definite implies that

$$(v^{k+1} - v^k)^{\mathrm{T}} \bar{B}(v^{k+1} - v^k) \leqslant (v^{k+1} - v^k)^{\mathrm{T}} B(v^k)(v^{k+1} - v^k). \tag{16}$$

Combining (15) and (16) and using the definition of $\| \ \|_{\bar{B}}$, we obtain

$$\|v^{k+1} - v^k\|_{\bar{B}} \leqslant \theta \|v^k - v^{k-1}\|_{\bar{B}}.$$

Since $\hat{\Omega}^l$ is compact and convex, the algorithm is contracting in v and hence it follows that

$$\lim_{k \to \infty} v^{k+1} = \hat{v}.$$

Given $\hat{v}$, according to (3), we may find a corresponding $\hat{h}$ (not necessarily unique) and eventually $\hat{u}$.

From (12), we obtain

$$s(\hat{v})^{\mathrm{T}}(v - \hat{v}) + (u - \hat{u}) \geqslant 0, \quad \forall \begin{pmatrix} v \\ u \end{pmatrix} \in \hat{\Omega}^l.$$

By definition, $(\hat{v}, \hat{u})^{\mathrm{T}}$ is a solution of the restricted variational inequality problem. According to Nguyen and Pallottino [11] and Wu et al. [18], it is unique in $\hat{\Omega}^l$. □

It is noted that, if $B(v) = D$ (D is a fixed symmetric and positive definite matrix), $s(v)$ is strongly monotone and Lipschitz continuous, the linearized method is convergent for a sufficiently small α [13].

5. Computational results

In this section, we give computational results for the solution method applied to three networks. Let us first recall several important notions in the modelling of the transit equilibrium assignment problem.

As mentioned before, the road network (for modelling purposes) consists of (i) zone centroids where the transit travel demand is generated and attracted, (ii) nodes representing bus stops and road intersections and (iii) roads or streets connecting those nodes and (pedestrian) connectors linking those nodes and centroids. A transit line is defined by its itinerary as a set of nodes that it passes through, and attributes such as its speed and frequency. The transit travel demand g_i is the number of trips from origin p to destination q (where $i = (p, q)$). The road network and the set of transit lines can be used to generate a general network where there are four types of arcs:

- walk arcs – the arcs connecting the zone centroids and the nodes of the general network;
- wait arcs – the arcs modelling the waiting of passengers at bus stops for the corresponding transit lines;

- in-vehicle arcs – the arcs corresponding to the line segments (every two consecutive nodes on a transit line define a transit segment); and
- transfer / alight arcs – the arcs from the transit lines to bus stops.

The procedure of generating such a general network is not discussed here. However, the size of the general network is of the order of the number of pedestrian connectors plus three times that of transit line segments. For the transit equilibrium assignment, we know that two attributes are associated with every arc a: cost function $s_a(v)$ and a frequency f_a. The cost functions for these arcs may be defined based on theoretical and empirical considerations. In the following, for the purpose of numerical tests, we present a possible cost structure for the applications, as discussed in Wu et al. [18].

5.1. WALK ARCS

We assume that the costs on the walk arcs are independent of flows and given by $s_a = \alpha_1 t_a$, where α_1 is a positive parameter and t_a is a constant walk time; f_a is ∞.

5.2. WAIT ARCS

The costs on the wait arcs are dependent on both the flows that correspond to passengers waiting for the bus (v_a) and the direct flows (v_d) corresponding to through passengers that share the bus capacity with the waiting flows. See fig. 1 where the direct flow is defined as

$$v_d = v_c - v_e \, .$$

The function form chosen is as follows:

$$s_a(v_a, v_d) = \alpha_2[(v_a + \beta_2 v_d)/K_b]^\rho \, , \tag{17}$$

where α_2, β_2 and ρ are positive parameters and K_b is the bus capacity corresponding to the in-vehicle arc b. According to the definition, we have $v_b = v_a + v_d$, where v_b is the in-vehicle flows and (17) may be rewritten as

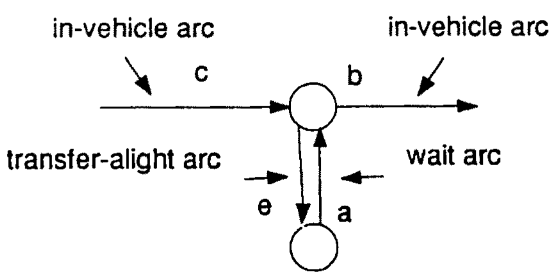

Fig. 1. An illustration of the arcs at nodes.

$$s_a(v_a, v_b) = \alpha_2[((1-\beta_2)v_a + \beta_2 v_b)/K_b]^\rho\,, \tag{18}$$

with $0 \leqslant \beta_2 \leqslant 1$. In this case, f_a represents the bus frequency on arc b.

5.3. IN-VEHICLE ARCS

Here we use the same notations as in fig. 1. The costs of the in-vehicle arcs s_b are composed of two parts: in-vehicle travel costs t_b, which are independent of flows, and discomfort costs, which depend on the direct flows (v_d) and the passengers (v_a) waiting for the bus. Thus we propose the following relation:

$$s_b(v_d, v_a) = \alpha_3 t_b + \beta_3[(v_d + \gamma_3 v_a)/K_b]^\rho\,, \tag{19}$$

where $\alpha_3, \beta_3, \gamma_3$ and ρ are the positive parameters associated with the in-vehicle arc. K_b is the capacity of the bus on arc b. Since $v_d = v_b - v_a$, we have

$$s_b(v_a, v_b) = \alpha_3 t_b + \beta_3[(v_b + (\gamma_3 - 1)v_a)/K_b]^\rho\,. \tag{20}$$

f_b is ∞ for in-vehicle arcs.

5.4. TRANSFER/ALIGHT ARCS

We assume that the arc costs are independent of the arc flows and given by a constant walk time t_a, i.e., $s_a = \alpha_4 t_a$ (here α_4 is positive) and f_a is ∞.

We note that K_b is a multiple of the time period considered, the corresponding bus frequency and the bus capacity. We implemented the SDA in Fortran 77 on a SUN-SPARC 2 workstation and ran it on three following networks. The first network (Small) is a four zone small test network cited from Wu et al. [18]; the second network (Hefei) is the network of the city of Hefei, China, and the third network (Winnipeg) is the network of the city of Winnipeg, Canada. Table 1 contains the dimensions that characterize the three networks.

We define EG to be number of extreme points generated, EL to be the remaining number of extreme points, CPU as the total cpu time mainly for the computation of the restricted variational inequality problems and the shortest hyperpaths. The computed values, GAP and RGAP, are the absolute and the relative gaps in the sense of variational inequality theory respectively. ATC is the average travel costs. V/C is the congestion measure, which is calculated only for the in-vehicle arcs by

Table 1
Size of the tested neworks.

Networks	Zones	Nodes	Directed arcs	Transit	Nodes in GN	Direct arc in GN	Total demand
Small	4	9	16	6	27	44	599
Hefei	80	180	398	16	444	904	324,685
Winnipeg	150	1067	2975	67	5,253	13,149	18,007

Table 2
Parameters used for the tested networks.

Network	Cost function parameters									SDM parameters	
	α_1	α_2	β_2	α_3	β_3	γ_3	α_4	π	ρ	α	β
Small	1	1	0.20	1	1	1.20	1	0.7	2	6	0
Hefei	1	4	0.30	6.0	5.0	1.20	1	0.7	3	8.5	0
Winnipeg	99	1.3	0.10	1.0	0.2	1.20	1	0.7	2	5	0

dividing the flows by the capacity of the buses over the period studied (K_b). We computed the average (AVE), the minimum (MIN) and the maximum (MAX) of this congestion measure.

The algorithm used for the restricted variational inequality problem is the linearized Jacobi method. Since α may be different for each restricted variational

Table 3
Comparison of the WFM and SDA algorithms on two tested networks.

EG	Small				Hefei			
	SDM		WFM		SDM		WFM	
	GAP	RGAP	GAP	RGAP	GAP	RGAP	GAP	RGA
2	211.28	2.39E − 2	27185.2	6.70E − 1	785160.7	1.09E − 2	158436256	0.605
3	75.12	8.61E − 3	915.48	9.96E − 2	392010.3	5.49E − 3	2449305	0.0336
4	12.18	1.41E − 3	588.64	6.49E − 2	336111.3	4.72E − 3	473501	6.59E − 3
5	3.78	3.38E − 4	143.97	1.66E − 2	234252.4	3.79E − 3	401170	5.60E − 3
6	1.70	1.39E − 4	82.72	9.47E − 3	118644.6	1.67E − 3	299879	4.19E − 3
7	0.179	1.51E − 5	44.13	5.11E − 3	74371.5	1.05E − 3	271223	3.79E − 3
8	0.0183	2.52E − 6	23.16	2.68E − 3	16615.3	2.34E − 4	228018	3.19E − 3
9	0.0153	1.77E − 6	15.81	1.83E − 3	11633.7	1.64E − 4	195095	2.73E − 3
10	–	–	9.60	1.11E − 3	11035.6	1.56E − 4	176963	2.48E − 3
11	–	–	–	–	5066.7	7.90E − 5	160881	2.25E − 3
12	–	–	–	–	5801.0	8.18E − 5	149984	2.10E − 3
13	–	–	–	–	7141.8	1.01E − 4	137882	1.93E − 3
14	–	–	–	–	5353.7	7.55E − 5	130925	1.84E − 3
15	–	–	–	–			122034	1.71E − 3
CPU (sec)	0.73		1.18		58.62		115.73	
Average V/C	0.87		0.87		0.33		0.33	
Min V/C	0.10		0.10		0.02		0.02	
Max V/C	1.79		1.79		1.64		1.64	
Average travel cost	8618.16		8622.82		218.44		219.35	

Table 4
Computational results at different congestion levels for small network.

V/C			Initial		Final		EG	EL	ATC	CPU (sec)
Ave	Min	Max	GAP	RGAP	GAP	RGAP				
1.68	0.24	3.58	423.99	2.65E − 2	4.58	2.95E − 5	9	7	25.96	0.99
0.87	0.10	1.79	211.28	2.38E − 2	1.53E − 2	1.77E − 6	9	7	14.39	0.72
0.46	0.04	0.92	7.73	1.13E − 3	2.33E − 3	3.42E − 7	9	8	11.36	0.68
0.40	0.04	0.84	2.17	3.26E − 4	9.26E − 2	1.39E − 5	9	8	11.07	0.96

inequality problem, we adopt the following heuristic rule for the implementation. Starting with any given value of α (large), if the sequence is contracting, then we continue. If it is not, we reduce the value of α, i.e., $\alpha = \pi\alpha$, where $0<\pi<1$. The shortest hyperpath algorithm is implemented with a heap data structure.

Table 2 contains the parameters of the cost functions and for the SDA. We note that all the cost functions used are asymmetric. In table 3, we give the comparison of the results obtained by applying the WFM algorithm, in the space of hyperpath flows, and this SDA for the Small and Hefei networks. We could not execute the WFM algorithm for the Winnipeg network due to the excessive space requirements for the hyperpaths. For these tested networks, the SDA outperforms the WFM algorithm both in terms of the cpu time and the convergence measures (GAP and RGAP). For the WFM algorithm, the gap and relative gap at EG are calculated for a solution computed based on the (EG-1) extreme points, while, for the SDA, they are calculated for a solution based on the extreme points retained at the previous iteration. In tables 4, 5 and 6, we give the results obtained for the Small, Hefei and Winnipeg networks respectively with SDA at different congestion levels. These congestion levels were created by adjusting the bus capacity. These results include the initial gap (GAP) and the relative gap (RGAP), the final gap (GAP) and the relative gap (RGAP), the number of extreme points generated (EG) and kept (EL) in the final solution, the average travel cost (ATC) and the total cpu time in seconds. For the same number of iterations, the final gap and the relative gap are higher for the more congested problems, which is to be expected. It is interesting to note that the number of extreme points left in the final solution is less for the congested problems than for the uncongested problems.

Table 5
Computational results at different congestion levels for Hefei network.

V/C			Initial		Final		EG	EL	ATC	CPU (sec)
Ave	Min	Max	GAP	RGAP	GAP	RGAP				
0.69	0.04	3.28	15,325,507	0.128	274,117	2.53E − 3	19	10	334.33	75.06
0.33	0.02	1.64	785,161	1.095E − 2	2,217.9	3.12E − 5	19	11	218.45	81.04
0.16	0.01	0.82	70,492	1.07E − 3	161.76	2.47E − 6	19	17	201.32	79.55
0.08	0.00	0.41	9,988.9	1.55E − 4	13.61	2.11E − 7	19	18	198	80.9

Table 6
Computational results at different congestion levels for Winnipeg network.

V/C			Initial		Final		EG	EL	ATC	CPU (sec)
Ave	Min	Max	GAP	RGAP	GAP	RGAP				
1.59	0.00	7.29	505,796.8	0.17	28,707.5	1.03E − 3	15	10	154.16	1450.1
0.79	0.00	3.90	38,061.0	1.77E − 2	2,155.5	9.96E − 4	15	7	120.12	1439.9
0.52	0.00	2.75	10,669.2	5.23E − 3	106.1	5.25E − 5	15	9	112.27	1430.9
0.39	0.00	8.27	3,797.0	1.93E − 3	43.12	2.19E − 5	15	10	109.06	1439.4
0.31	0.00	1.92	1,926.6	1.94E − 4	43.00	2.22E − 5	15	12	107.45	1448.5

6. Conclusions

In this paper, we developed a simplicial decomposition algorithm (SDA) for the transit equilibrium assignment problem. Although the problem is formulated in the hyperpath space, this problem can be reformulated in the space of total arc flows and the auxiliary variable. The new problem is then solved by using the simplicial decomposition approach. The resulting algorithm is a computationally tractable way of solving large scale transit equilibrium assignment problems. A proof of the convergence of the SDA is also given, which follows from the work of Lawphongpanich and Hearn [10] and Pang and Yu [14].

The size of the general network for the transit equilibrium assignment is much larger than that for the road traffic equilibrium assignment. Thus the problem is a typical large scale problem. The computations of the shortest hyperpath take more cpu time when the general network is getting larger, which is to be expected. Since the computation of the shortest hyperpath must be carried out for all the destinations, the use of parallel computation implementation seems attractive.

Acknowledgement

The authors would like to thank the editor and two referees for their valuable comments, which greatly improved the exposition of the paper.

References

[1] D.P. Bertsekas and E.M. Gafni, Projection methods for variational inequalities with application to the traffic assignment problem, Math. Progr. Study 17 (1982) 139–159.
[2] S. Dafermos, Traffic equilibrium and variational inequalities, Transp. Sci. 14 (1980) 42–54.
[3] J. De Cea and E. Fernández, Transit assignment to minimal routes: an efficient new algorithm, Traffic Engineering Control 30 (1989) 491–494.
[4] J. De Cea and E. Fernández, Equilibrium assignment on congested transit network, Publication 584, CRT, Université de Montréal (1988).

[5] J.-P. Dussault, J.A. Ferland and B. Lemaire, Convex quadratic programming with one constraint and bounded variable, Math. Progr. 36 (1986) 90–104.
[6] M. Florian and H. Spiess, The convergence of diagonalization algorithms for asymmetric network equilibrium problems, Transp. Res. B 16B (1982) 447–483.
[7] M. Gendreau, Etude approfondée d'un model d'équilibre pour l'affectation des passagers dans les réseaux de transport en commun, Publication 384, CRT, Univerité de Montréal (1984).
[8] M. Held, P. Wolfe and H.P. Crowder, Validation of subgradient optimization, Math. Progr. 6 (1974) 62–88.
[9] J.L. Kennington, *Algorithms for Network Programming* (Wiley, 1980).
[10] S. Lawphongpanich and D.W. Hearn, Simplicial decomposition of the asymmetric traffic assignment problem, Transp. Res. B 18 B (1984) 123–133.
[11] S. Nguyen and S. Pallottino, Equilibrium traffic assignment for large transit networks, Publication 494, CRT, Université de Montréal (1985).
[12] S. Nguyen and S. Pallottino, Hyperpaths and shortest hyperpath, Quadero no. 19 (1986).
[13] J.S. Pang and D. Chan, Iterative methods for variational and complementary problem, Math. Progr. 24 (1982) 284–313.
[14] J.S. Pang and C.S. Yu, Linearized simplicial decomposition methods for computing traffic equilibria on networks, Networks 14 (1984) 427–438.
[15] H. Spiess, Contributions à la théorie et aux outils de planification des réseaux de transport urbain, Publication 382, CRT, Université de Montréal (1984).
[16] H. Spies and M. Florian, Optimal strategies: A new assignment model for transit network, Transp. Res. B 23 B (1989).
[17] J.G. Wardrop, Some theoretical aspects of road traffic research, Proc. Inst. Civil Eng., Part II (1952) 325–378.
[18] J.H. Wu, M. Florian and P. Marcotte, Transit equilibrium assignment: a model and solution algorithm, Publication 721, CRT, Université de Montréal (1991).