A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria

M. Cepeda [a,1], R. Cominetti [b,*,2], M. Florian [c,3]

[a] *Departamento de Ingeniería Industrial, Universidad de Concepción, Chile*
[b] *Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Universidad de Chile, Chile*
[c] *Département d'IRO and Centre de Récherche sur les Transports, Université de Montréal, Canada*

Abstract

This paper concerns a frequency-based route choice model for congested transit networks, which takes into account the consequences of congestion on the predicted flows as well as on the expected waiting and travel times. The paper builds on the results presented in Correa [Correa, J., 1999. Asignación de flujos de pasajeros en redes de transporte público congestionadas. Engineering thesis, U. de Chile, Santiago] and Cominetti and Correa [Cominetti, R., Correa, J., 2001. Common-lines and passenger assignment in congested transit networks. Transportation Science 35(3), 250–267], extending these to obtain a new characterization of the equilibria which allows us to formulate an equivalent optimization problem in terms of a computable gap function that vanishes at equilibrium. This new model formulation can deal with flow dependent travel times and is a generalization of the previously known strategy (hyperpath) based transit network equilibrium models. The approach leads to an algorithm which has been applied successfully on

* Corresponding author. Tel.: +56 2 6784462; fax: +56 2 6883821.
E-mail address: rcominet@dim.uchile.cl (R. Cominetti).

[1] Partially supported by Dirección de Investigación, Universidad de Concepción. Benefited from NSERC—Canada and MIDEPLAN—Chile support for doctoral studies.
[2] Supported by Fondap Matemáticas Aplicadas, CONICYT—Chile.
[3] Partially supported by NSERC—Canada.

large scale networks. Computational results for transit networks originating from practice demonstrate the applicability of the proposed approach.

Keywords: Transit assignment; Congested networks; Equilibrium; Algorithms

1. Introduction

The planning of urban transit services relies on the use of transit assignment models for predicting the way in which transit travellers choose routes from their origins to their destinations. While much progress has been achieved in the past two decades, the issue of how to model the consequences of congestion on the predicted flows in transit networks has received relatively little attention. The purpose of this paper is to build on the results of Correa (1999) and Cominetti and Correa (2001), and extend these to the formulation of a solvable model for large scale transit networks. This topic was the subject of the doctoral thesis of Cepeda (2002).

It is worthwhile to review the contributions made to the study of transit route choice. The earliest methods for finding paths in transit networks—such as Dial (1967), Fearnside and Draper (1971) and Le Clercq (1972)—recognized that waiting time at stops served by several lines was an important aspect of this problem, and proposed various heuristic ways to combine waiting and travel times in computing shortest paths. The seminal paper of Chriqui and Robillard (1975) introduced the notion that, on a simple network of one origin and one destination, passengers can select a subset of attractive lines and board the first one of these that arrives at a stop in order to minimize the expected sum of waiting plus travel times.

The ideas of Chriqui and Robillard were extended to general transit networks in two ways. Spiess (1984) and Spiess and Florian (1989) introduced the notion of strategy, which is a choice of an attractive set of lines at each decision point; that is, at each node where boarding occurs. The resulting model and algorithm achieve the minimization of the expected value of the total travel time which includes access, wait and in-vehicle time. Nguyen and Pallotino (1988) provided a graph theoretic interpretation of a strategy as an acyclic directed graph, and denoted it as a hyperpath. These models considered congestion aboard the vehicles by associating discomfort functions with each segment of a transit line, so that the resulting equilibrium models can be solved by standard algorithms for convex minimization. However, the waiting times are underestimated since they do not consider the fact that in a period of heavy congestion passengers may not be able to board the first vehicle to arrive at a stop.

The results of Chriqui and Robillard were used in a different way by De Cea and Fernández (1989) in a transit assignment model based on a restricted notion of strategy which allows choices among multiple lines at a given stop only if they all share the next stop to be served (for a comparison with the strategy approach see De Cea et al., 1988). This model was extended by De Cea and Fernández (1993) to heuristically incorporate the effects of congestion at bus stops and aboard the vehicles, leading to an asymmetric equilibrium model which is solved by the Jacobi method. The model works on a large augmented graph so that the computational effort is significant, and the resulting flows often exceed the capacity of the vehicles. A model in which passengers that can not board are routed through spill-links was proposed in Kurauchi et al. (2003) and was tested on a small example. This model considers the risk-aversion of passengers to overcrowded stations and combines the computation of common-line strategies with a Markovian approach in which the boarding probability is determined by the residual capacity of the transit vehicles. On the other hand, the issue of the capacity of transit services is equally important for schedule-based transit route choice models. A relevant contribution is that of Tong and Wong (1999) who develop a stochastic dynamic model which is applied to the Hong Kong underground services. A survey of contributions to schedule-based approaches to transit route choice can be found in the proceedings of a conference held in Ischia and edited by Nuzzolo and Wilson (2004).

For the modeling of frequency-based transit services, a formal study of congestion at bus stops based on queueing theory was initiated by Gendreau (1984) who was the first to formulate a general transit assignment model with congestion. For more recent results on the waiting processes at bus stops see Bouzaïene-Ayari et al. (2001) and Cominetti and Correa (2001), as well as the recent thesis by Cepeda (2002). As a consequence of these studies it became clear that the congestion at bus stops does not only increase the waiting times but it also affects the flow share of each attractive line. In the case of bus lines with independent exponential interarrivals, the flow split is proportional to the so-called *effective frequency*, that is to say, the inverse of the waiting time of each line. The stop models are called *semi-congested* if they consider only the increase of waiting times, and *full-congested* if they also include the effects on the flow split. Wu et al. (1994) considered a semi-congested transit network model in which the time required to board a vehicle increases with flow, but the distribution of flows among attractive lines is done in proportion to the nominal frequencies. Bouzaïene-Ayari (1996) and Bouzaïene-Ayari et al. (1995) extended the latter to a full-congested model which combines a fixed point problem in the space of arc flows with a variational inequality in the space of hyperpath flows. An algorithm reminiscent of the method of successive averages is also proposed in these works but the combinatorial character of hyperpaths seem to limit its applicability to small networks. Moreover, the congestion functions are assumed finite and the arc travel times are strongly monotone, excluding relevant cases such as constant travel times or waiting times based on queuing theory which tend to infinity as the flows approach the line capacities.

More recently, Cominetti and Correa (2001) analyzed a full-congested version of the common-lines problem of Chriqui and Robillard, and used it to develop a frequency-based transit equilibrium model which can deal with general arc travel times as well as more realistic waiting time functions with asymptotes at bus capacity. The latter are introduced by considering effective frequency functions that vanish when the flows exceed the capacity of the line. Although they establish the existence of a network equilibrium, it fails to propose an algorithm to compute it. Nevertheless, since the model is stated as a fixed point in the space of arc flows only, it opens the way to deal with large scale networks.

The purpose of the present paper is precisely to describe an alternative formulation for the full-congested frequency-based transit equilibrium model proposed in Cominetti and Correa (2001), and to derive from it a solution algorithm for large scale networks. The paper is organized as follows. The common-lines problem is revisited in Section 2 and the alternative formulation for the general network model is described in Section 3. Some special cases of the model are discussed in Section 3.3 and in the appendix at the end of the paper. The solution algorithm, as well as some comments on the difficulty of the problem and numerical tests on real size networks are given in Section 4.

2. The congested common-line problem revisited

Let us recall the common-line problem with congestion described in Cominetti and Correa (2001) (Fig. 1). Consider a simple network consisting of an origin O connected to a destination D by a finite set of bus lines $A = \{a_1, \ldots, a_n\}$. Each line $a \in A$ is characterized by a constant in-vehicle travel time $t_a \in \mathbb{R}$ and a smooth *effective frequency* function $f_a : [0, \bar{v}_a) \to (0, \infty)$ with $f'_a(v_a) < 0$ and $f_a(v_a) \to 0$ when $v_a \to \bar{v}_a$. The effective frequency is supposed to be decreasing in order to reflect the increment in waiting time induced by an augmentation of flow, and the constant $\bar{v}_a > 0$ (eventually $\bar{v}_a = \infty$ for some links including the walking links) is called the *saturation flow* of the line. The case of flow-dependent travel times as well as more general effective frequency functions will be considered in Section 3.

For the purpose of travelling from O to D, each passenger selects a non-empty subset of lines $s \subseteq A$, called the *attractive lines* or *strategy*, boarding the first incoming bus from this set with available capacity. Thus, the total flow $x \geqslant 0$ splits among all possible strategies $s \in \mathscr{S}$ so that $x = \sum_{s \in \mathscr{S}} h_s$ where $h_s \geqslant 0$ denotes the flow on strategy $s \in \mathscr{S}$. Assuming that a passenger using strategy s boards line $a \in s$ with probability $\pi_a^s = f_a(v_a) / \sum_{b \in s} f_b(v_b)$, it turns out that each strategy-flow vector $h = (h_s)_{s \in \mathscr{S}}$ induces a unique vector of line-flows $v = v(h)$ through the system of equations

$$\text{(E)} \quad v_a = \sum_{s \in \mathscr{S}:\ a \in s} h_s \frac{f_a(v_a)}{\sum_{b \in s} f_b(v_b)} \quad \forall a \in A.$$

This line-flow vector v determines in turn the expected transit time of each strategy

$$\text{(T)} \quad T_s(v) \triangleq \frac{1 + \sum_{a \in s} t_a f_a(v_a)}{\sum_{a \in s} f_a(v_a)}.$$

Invoking Wardrop's principle, a strategy-flow vector $h \geqslant 0$ with $\sum_{s \in \mathscr{S}} h_s = x$ is said to be an *equilibrium* iff all strategies carrying flow are of minimal time, that is to say,

$$\text{(W)} \quad h_s > 0 \Rightarrow T_s(v(h)) = \tau(v(h)),$$

where $\tau(v) \triangleq \min_{s \in \mathscr{S}} T_s(v)$. The set of equilibria is denoted H_x, while V_x stands for the set of induced line-flows $v(h)$ corresponding to all $h \in H_x$.

Notice that (E) and (T) presuppose that the arrivals of the different lines are independent and exponentially distributed. This assumption is somewhat restrictive, but to date this seems the only

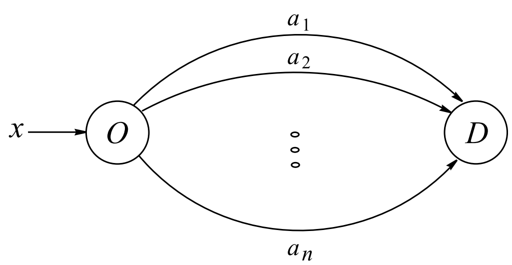

Fig. 1. The common-line problem.

arrival distribution which is analytically tractable. For details on the modeling assumptions behind these equations, as well as for possible analytic expressions for $f_a(v_a)$, the reader is referred to Bouzaïene-Ayari et al. (2001), Cominetti and Correa (2001) and Gendreau (1984).

The set V_x was characterized in Cominetti and Correa (2001) as the optimal solution set of an equivalent optimization problem which implies the existence of a constant $\alpha_x \geqslant 0$ such that $v \in V_x$ if and only if $v \geqslant 0$ with $\sum_{a\in A} v_a = x$ and

$$(\mathrm{C}_0) \qquad \frac{v_a}{f_a(v_a)} \begin{cases} = \alpha_x & \text{if } t_a < \tau(\hat{v}(\alpha_x)), \\ \leqslant \alpha_x & \text{if } t_a = \tau(\hat{v}(\alpha_x)), \\ = 0 & \text{if } t_a > \tau(\hat{v}(\alpha_x)), \end{cases}$$

where $\hat{v}_a(\alpha)$ is the inverse function of $v_a \mapsto v_a/f_a(v_a)$. Our first result provides a simpler description of V_x which will be used in Section 3 to derive an optimization problem that characterizes the equilibrium in the case of general transit networks. Formally the condition looks similar to (C_0) but it avoids the functions $\hat{v}_a(\alpha)$. Note also that the constant α_x which is common to all $v \in V_x$ is now allowed to change for each $v \in V_x$ (though in fact it will not!).

Theorem 2.1. *$v \in V_x$ if and only if $v \geqslant 0$ with $\sum_{a\in A} v_a = x$ and there exists $\alpha \geqslant 0$ such that*

$$(\mathrm{C}_1) \qquad \frac{v_a}{f_a(v_a)} \begin{cases} = \alpha & \text{if } t_a < \tau(v), \\ \leqslant \alpha & \text{if } t_a = \tau(v), \\ = 0 & \text{if } t_a > \tau(v). \end{cases}$$

This result states that—for a simple network of one OD pair and non-overlapping common lines—all the arcs with travel time strictly less than the equilibrium time must carry an amount of flow which induces the same waiting time α, so that they are in some sense *evenly congested*. Arcs with travel time equal to the equilibrium time must have waiting time at most α, while arcs with larger travel times are not used. This characterization leads to an equivalent formulation of Wardop's user equlibrium for the transit route choice problem in a general network, as will be shown in the next section. The proof of the theorem uses the following characterization of optimal strategies (see Lemma 1.2 and Corollary 1.1 in Cominetti and Correa, 2001).

Lemma 2.1. *Let $\check{s}(v) \triangleq \{a:\ t_a < \tau(v)\}$ and $\hat{s}(v) \triangleq \{a:\ t_a \leqslant \tau(v)\}$. Then for each $s \in \mathscr{S}$ the following are equivalent*:

(a) $T_s(v) = \tau(v)$,
(b) $t_a \leqslant T_s(v) \leqslant t_b \quad \forall a \in s,\ b \notin s$,
(c) $\check{s}(v) \subseteq s \subseteq \hat{s}(v)$.

Proof of Theorem 2.1

($\Rightarrow$) Let $v \in V_x$—so that (C_0) holds—and take $\alpha = \alpha_x$. According to Theorem 1.2 in Cominetti and Correa (2001), one has $\tau(v) = \tau(\hat{v}(\alpha_x))$, so that condition (C_1) follows at once from (C_0).

($\Leftarrow$) To prove the converse, let us assume that the vector v satisfies (C_1) for some $\alpha \geqslant 0$. It will be shown that $\tau(v) = \tau(\hat{v}(\alpha))$. Indeed, let $s = \check{s}(v)$ so that Lemma 2.1 implies that $\tau(v) = T_s(v)$

and $t_a \leqslant T_s(v) \leqslant t_b$ for all $a \in s$, $b \notin s$. Now, from (C_1) one gets $v_a = \hat{v}_a(\alpha)$ for all $a \in s$ and since $T_s(\cdot)$ depends only on these flows it follows that $T_s(v) = T_s(\hat{v}(\alpha))$. Hence $t_a \leqslant T_s(\hat{v}(\alpha)) \leqslant t_b$ for all $a \in s$, $b \notin s$, and using Lemma 2.1 once again one gets $T_s(\hat{v}(\alpha)) = \tau(\hat{v}(\alpha))$. Putting together these equalities it follows that $\tau(v) = \tau(\hat{v}(\alpha))$ as claimed. Now, it was already noticed that (C_1) implies $v_a = \hat{v}_a(\alpha)$ if $t_a < \tau(v) = \tau(\hat{v}(\alpha))$, and similarly one obtains $v_a \leqslant \hat{v}_a(\alpha)$ if $t_a = \tau(\hat{v}(\alpha))$ and $v_a = 0$ if $t_a > \tau(\hat{v}(\alpha))$. These estimates, combined with $\sum_{a \in A} v_a = x$, imply

$$\sum_{a \in \check{s}(\hat{v}(\alpha))} v_a \leqslant x \leqslant \sum_{a \in \hat{s}(\hat{v}(\alpha))} v_a$$

so that Theorem 1.2 in Cominetti and Correa (2001) yields $\alpha = \alpha_x$ and $v \in V_x$ as was to be proved. $\square$

When modeling general transit networks many arcs have no waiting time and are assigned infinite frequencies. In order to extend the model and cover this possibility, it suffices to adopt the convention that all computations are performed by first replacing the infinite frequencies by a common finite value f which is then considered to tend to ∞. In particular, if a strategy s contains one or more arcs with infinite frequency the transfer time $T_s(v)$ turns out to be the simple average of the travel times t_a of the infinite frequency arcs in s; while the boarding probabilities are uniform among these arcs and zero on the rest (if any). Theorem 2.1 still holds in this more general setting, provided that the third case of condition (C_1) is interpreted as "$v_a = 0$ if $t_a > \tau(v)$". For details see Section 3.6 in Cepeda (2002).

3. The network equilibrium model

The transit network is represented by a directed graph $G = (N, A)$ which is built as follows (Fig. 2) (see Spiess and Florian, 1989). Let $N_s \subseteq N$ be a set of nodes representing the bus stops in the network. Each bus line l is represented by a set of line-nodes $N_l \subseteq N$ which correspond to the sequence of bus stops served by the line. Each line-node in N_l connects to the corresponding stop-node in N_s through boarding and alight arcs, as well as to the next line-node in the sequence through an on-board arc (also called line segment). Eventually, one may consider walk arcs connecting directly a pair of nodes in N_s.

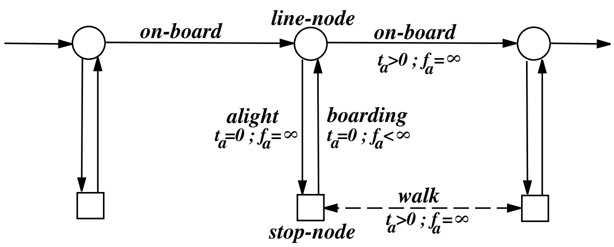

Fig. 2. Representation of a transit network.

The set of destinations is denoted $D \subseteq N$, and for each $d \in D$ and every node $i \neq d$ a fixed demand $g_i^d \geqslant 0$ is given. Typically the demands g_i^d are strictly positive only at nodes i corresponding to stop-nodes, that is to say the bus stops where users wait for service, but we do not impose any restriction. The set $\mathscr{V} = [0, \infty)^{A \times D}$ denotes the space of arc-destination flow vectors v with non-negative entries $v_a^d \geqslant 0$, while $\mathscr{V}_0$ is the set of feasible flows $v \in \mathscr{V}$ such that $v_a^d = 0$ for all $a \in A_d^+$ (i.e. no flow with destination d exits from d) and satisfying the flow conservation constraints

$$g_i^d + \sum_{a \in A_i^-} v_a^d = \sum_{a \in A_i^+} v_a^d \quad \forall i \neq d. \tag{1}$$

Here $A_i^+ = \{a : \ i_a = i\}$ and $A_i^- = \{a : \ j_a = i\}$ are the forward and backward stars of node $i \in N$, with i_a and j_a denoting respectively the tail and head nodes of arc $a \in A$.

Every arc $a \in A$ has an associated travel-time function $t_a : \mathscr{V} \to [0, \infty)$ and an effective frequency $f_a : \mathscr{V} \to [0, \infty]$, both of which are continuous with $t_a(\cdot)$ bounded and $f_a(\cdot)$ either identically ∞ or everywhere finite. In the latter case, for each $d \in D$ we assume that $f_a(v) \to 0$ when $v_a^d \to \infty$ with $f_a(v)$ strictly decreasing with respect to v_a^d as long as $f_a(v) > 0$. Notice that $f_a(v)$ may take the value 0, which allows to model waiting times that explode to infinity beyond the line capacity.

Typically, one assigns infinite effective frequencies $f_a(\cdot) \equiv \infty$ to all the arcs except the boarding arcs (where the waiting processes occur), while for travel times one takes $t_a(\cdot) \equiv 0$ on all arcs except for the on-board and walk arcs which are assigned a positive travel time function $t_a(\cdot) > 0$. However, one may also assign a positive travel time to the alight and boarding arcs, in order to represent the times spent on embark/disembark operations or the effect of the bus fare on the user choices. We stress that the functions $t_a(v)$ and $f_a(v)$ are allowed to depend on *all* the flows in the network. This is particularly relevant for the effective frequencies since the waiting times do not only depend on the boarding flows and operational characteristics of the lines (capacity, nominal frequency, interarrival distribution) but also on the on-board flows which consume part of the line capacity. In fact, as proved in Cepeda (2002), the situation is much more complex as congestion may be affected in cascade by the waiting processes occurring at the upstream nodes, which are in turn affected by the flows on other lines. The analysis of such complex relations is beyond the scope of this paper, and therefore we isolate the difficulty by proving our results for very general functions $f_a(v)$ without reference to a specific model and allowing much freedom for modeling the congestion.

3.1. The notion of network equilibrium

The intuitive idea behind the notion of a network equilibrium is the following (see Fig. 3). Consider a passenger heading towards destination d and reaching an intermediate node i in his trip. To exit from i he can use the arcs $a \in A_i^+$ to reach the next node j_a. By taking the arc travel times $t_a(v)$ and the transit times $\tau_{j_a}^d$ from j_a to d as fixed, the decision faced at node i is a common-line problem with travel times $t_a(v) + \tau_{j_a}^d$ and effective frequencies corresponding to the services operating on the arcs $a \in A_i^+$. The solution of this common-line problem determines the transit time τ_i^d from i to d, which can then be used recursively to solve the upstream nodes. It is important to stress that

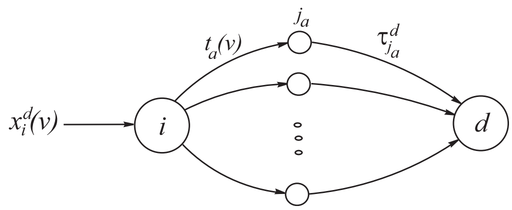

Fig. 3. The *i*-to-*d* common-line problem.

every node i must be considered as a potential origin even if the initial demand g_i^d is zero, since it may receive transfer flows coming from other nodes and heading towards d.

Since all the variables τ_i^d and v_a^d must be determined at the same time, the transit network model is stated as a set of simultaneous common-line problems (one for each pair i, d) coupled by flow conservation constraints. More precisely, for each $v \in \mathscr{V}$ the flow entering node i with destination d is defined as $x_i^d(v) \triangleq g_i^d + \sum_{a \in A_i^-} v_a^d$, while the *time-to-destination* functions $\tau_i^d(v)$ are the unique solution of the generalized Bellman equations (see Nguyen and Pallotino, 1988; Spiess and Florian, 1989; Cominetti and Correa, 2001):

$$\begin{cases} \tau_d^d = 0, \\ \tau_i^d = \min\limits_{s \in \mathscr{S}_i} \frac{1+\sum_{a \in s}[t_a(v)+\tau_{j_a}^d]f_a(v)}{\sum_{a \in s} f_a(v)} \quad \forall i \neq d, \end{cases}$$

where $\mathscr{S}_i$ stands for the set of non-empty subsets $s \subseteq A_i^+$. The quantity $t_a^d(v) \triangleq t_a(v) + \tau_{j_a}^d(v)$ represents the minimal time to destination d when using arc a. Let $V_i^d(v)$ denote the set of *local equilibrium flows* corresponding to a common-line problem defined by the arcs A_i^+, with total flow $x_i^d(v)$, constant travel times $t_a^d(v)$, and diagonal effective frequencies $f_a^d(\cdot)$ obtained from $f_a(\cdot)$ by considering it as a function of the flow v_a^d alone, and keeping the other variables fixed. A flow v is a *global* equilibrium if it is a local equilibrium with respect to itself, that is to say, a fixed point for the above set-valued maps. More precisely,

Definition 3.1. A feasible flow $v \in \mathscr{V}_0$ is called a *network equilibrium* iff for all $d \in D$ and $i \neq d$ the flows $(v_a^d)_{a \in A_i^+}$ belong to $V_i^d(v)$. The set of network equilibrium flows will be denoted $\mathscr{V}_0^*$.

Taking into account the common-line model described in the previous section and introducing strategy-flow variables h_s^d, one may directly restate the above definition in the following terms: the vector $v \in \mathscr{V}$ is an equilibrium iff for all $d \in D$ and $i \neq d$ there exist $\tau_i^d \in \mathbb{R}$ and strategy-flows $h_s^d \geqslant 0$ for $s \in \mathscr{S}_i$ satisfying the flow equations

$$\sum_{s \in \mathscr{S}_i} h_s^d = x_i^d(v),$$

$$v_a^d = \sum_{s \ni a} h_s^d \frac{f_a(v)}{\sum_{b \in s} f_b(v)} \quad \forall a \in A_i^+,$$

and the Wardrop's equilibrium conditions

$$T_s^d(v) \begin{cases} = \tau_i^d & \text{if } h_s^d > 0, \\ \geqslant \tau_i^d & \text{if } h_s^d = 0, \end{cases}$$

where the cost of strategy $s \in \mathscr{S}_i$ is given by

$$T_s^d(v) = \frac{1 + \sum_{a \in s}[t_a(v) + \tau_{j_a}^d]f_a(v)}{\sum_{a \in s} f_a(v)}.$$

The existence of an equilibrium may be derived from Kakutani's fixed point theorem under appropriate compactness conditions, which in this case are complicated by the fact that the waiting times may present asymptotes. Theorem 2.1 in Cominetti and Correa (2001) shows that the set of equilibria $\mathscr{V}_0^*$ is non-empty provided that every node $i \in V$ can be connected to each destination $d \in D$ by a path formed by arcs with infinite frequency (e.g. a pedestrian path), that there exist constants $\bar{t}_a < \infty$ such that $t_a(v) \leqslant \bar{t}_a$ for all $v \in \mathscr{V}$, and for each arc a and any fixed value of v_a^d the quantity $f_a(v)$ is maximal when all the remaining flows are 0. The result does not mean that at equilibrium such pedestrian arcs will be used, which is in fact unlikely unless the bus network has a saturated bottleneck with insufficient capacity that forces some flow along walk arcs. In any case, the present paper is not concerned with conditions for the existence of an equilibrium which is taken for granted.

3.2. Characterization of the equilibrium

A direct application of Theorem 2.1 gives the following characterization of network equilibria.

Theorem 3.1. *$v \in \mathscr{V}_0^*$ iff $v \in \mathscr{V}_0$ and there exist numbers $\alpha_i^d \geqslant 0$ such that for all $d \in D$ and $i \neq d$*

$$\frac{v_a^d}{f_a(v)} \begin{cases} = \alpha_i^d & if\ t_a^d(v) < \tau_i^d(v), \\ \leqslant \alpha_i^d & if\ t_a^d(v) = \tau_i^d(v), \\ = 0 & if\ t_a^d(v) > \tau_i^d(v), \end{cases}$$

where $t_a^d(v) \triangleq t_a(v) + \tau_{j_a}^d(v)$.

This characterization allows to derive a gap function with the property that each of its minimizers corresponds precisely to a network equilibrium.

Theorem 3.2. *For all $v \in \mathscr{V}_0$, $d \in D$ and $i \neq d$ the following inequality holds*:

$$\sum_{a \in A_i^+} [t_a(v) + \tau_{j_a}^d(v)]v_a^d + \max_{a \in A_i^+} \frac{v_a^d}{f_a(v)} \geqslant \tau_i^d(v) \sum_{a \in A_i^+} v_a^d \tag{2}$$

and moreover $v \in \mathscr{V}_0^*$ iff $v \in \mathscr{V}_0$ and all these inequalities are satisfied as equalities.

Proof. Although the proof given below holds for finite as well as infinite frequencies, the reader may assume for simplicity that all the functions $f_a(v)$ are finite. Moreover, the functional dependencies of t_a, f_a, τ_i^d and t_a^d on the vector v are omitted, and Δ is written for the expression on the left of (2). Fix $v \in \mathscr{V}_0$, $d \in D$, $i \neq d$, and consider the vector $(h_s^d)_{s \in \mathscr{S}_i}$ computed by the algorithm

$$\text{(A)} \quad \begin{cases} h_s^d \leftarrow 0 \text{ for all } s \in \mathscr{S}_i; \\ \textbf{while} \text{“} s \triangleq \{a \in A_i^+ : v_a^d > 0\} \neq \phi \text{”} \\ \quad h_s^d \leftarrow \min\left\{v_a^d[\sum_{b \in s} f_b]/f_a : a \in s\right\} \\ \quad v_a^d \leftarrow v_a^d - h_s^d f_a/[\sum_{b \in s} f_b] \quad \text{for all } a \in s \\ \textbf{end}. \end{cases}$$

This iteration decomposes v_a^d for $a \in A_i^+$ as

$$v_a^d = \sum_{s \ni a} h_s^d \frac{f_a}{\sum_{b \in s} f_b} \tag{3}$$

and since $\sum_{a \in A_i^+} \sum_{s \ni a} \equiv \sum_{s \in \mathscr{S}_i} \sum_{a \in s}$ one gets

$$\sum_{a \in A_i^+} v_a^d = \sum_{s \in \mathscr{S}_i} h_s^d.$$

Moreover, it is easy to check that v_a^d / f_a is maximal iff a belongs to every s generated by (A) (i.e. all s such that $h_s^d > 0$), in which case the sum on the right hand side of (3) may be extended to all $s \in \mathscr{S}_i$ and therefore

$$\max_{a \in A_i^+} \frac{v_a^d}{f_a} = \sum_{s \in \mathscr{S}_i} \frac{h_s^d}{\sum_{b \in s} f_b}. \tag{4}$$

Replacing (3) and (4) into the expression for Δ and exchanging the order of summation it results that

$$\Delta = \sum_{s \in \mathscr{S}_i} h_s^d \frac{1 + \sum_{a \in s} [t_a + \tau_{j_a}^d] f_a}{\sum_{a \in s} f_a} \geqslant \sum_{s \in \mathscr{S}_i} h_s^d \tau_i^d = \tau_i^d \sum_{a \in A_i^+} v_a^d$$

hence proving (2). These inequalities become equalities iff for all s such that $h_s^d > 0$ one has

$$\frac{1 + \sum_{a \in s} [t_a + \tau_{j_a}^d] f_a}{\sum_{a \in s} f_a} = \tau_i^d,$$

that is to say, if all the strategies s generated by (A) are optimal. By Lemma 2.1 this is equivalent to the fact that s must contain every arc such that $t_a^d < \tau_i^d$ and no arc with $t_a^d > \tau_i^d$. By a previous observation, this corresponds exactly to the condition that v_a^d / f_a is maximal if $t_a^d < \tau_i^d$ and 0 if $t_a^d > \tau_i^d$, and then according to Theorem 3.1 it follows that (2) hold as equalities iff v is a network equilibrium. $\square$

This result shows that the equilibrium flows $v \in \mathscr{V}_0^*$ are the global minima of the function

$$\sum_{\substack{d \in D \\ i \neq d}} \left[\sum_{a \in A_i^+} [t_a(v) + \tau_{j_a}^d(v)] v_a^d + \max_{a \in A_i^+} \frac{v_a^d}{f_a(v)} - \tau_i^d(v) \sum_{a \in A_i^+} v_a^d \right]$$

with optimal value equal to 0. Re-arranging terms, the problem of finding equilibrium flows may be restated in the form

$$(\mathrm{P}) \quad \underset{v \in \mathscr{V}_0}{\text{Minimize}} \; G(v),$$

where

$$G(v) \triangleq \sum_{d \in D} \left[\sum_{a \in A} t_a(v) v_a^d + \sum_{i \neq d} \max_{a \in A_i^+} \frac{v_a^d}{f_a(v)} - \sum_{i \neq d} g_i^d \tau_i^d(v) \right].$$

A natural approach for computing a full-congested network equilibrium is therefore to minimize the gap function $G(v)$. Now, since descent algorithms may be trapped by local minima it would be useful to prove that no such local minima exist. Although it is unlikely that such a result holds in general, the appendix at the end of the paper shows that this is the case for the common-lines problem.

Despite the possibility of having local minima, the fact that the minimum value of $G(v)$ is known to be 0 allows to use this gap function for monitoring the progress of a minimization algorithm and to derive a stopping rule as described in Section 4. Before proceeding, let us use (P) to compare the full-congested network equilibrium with two previous strategy-based transit assignment models: the linear cost and the convex cost models. It is worth mentioning that the linear cost model will be used in Section 4 as a subroutine in the solution of (P).

3.3. Two special cases

Two special cases of (P) are particularly interesting, as they correspond exactly to the models studied by Spiess (1984) as well as Nguyen and Pallotino (1988). The first one, called the uncongested case, is characterized by constant travel times $t_a(v) \equiv t_a$ and constant effective frequencies $f_a(v) \equiv f_a$. The second case, called semi-congested, only takes the functions $f_a(v) \equiv f_a$ to be constant, whereas the functions $t_a(v)$ are used to model the congestion effects. In both cases problem (P) becomes

$$\operatorname*{Min}_{v \in \mathscr{V}_0} \sum_{d \in D} \left[\sum_{a \in A} t_a(v) v_a^d + \sum_{i \neq d} \max_{a \in A_i^+} \frac{v_a^d}{f_a} - \sum_{i \neq d} g_i^d \tau_i^d(v) \right]$$

which can be rewritten as

$$(\mathrm{P}_1) \quad \operatorname*{Min}_{v \in \mathscr{V}_0} \quad \sum_{d \in D} \left[\sum_{a \in A} t_a(v) v_a^d + \sum_{i \neq d} w_i^d - \sum_{i \neq d} g_i^d \tau_i^d(v) \right]$$
$$\text{s.t.} \quad v_a^d \leqslant w_i^d f_a \quad \forall d \in D, \ i \neq d, \ a \in A_i^+.$$

It is clear that for any optimal solution and any pair (i, d), the constraints $v_a^d \leqslant w_i^d f_a$ must be satisfied as equality for at least one $a \in A_i^+$; since otherwise, it would be possible to reduce the value of the objective function by reducing the corresponding variable w_i^d.

In the uncongested case the functions $\tau_i^d(v)$ become constant, and so does the third term in the objective function. Hence problem (P_1) reduces exactly to the linear program found by Spiess (1984)

$$(\mathrm{PL}) \quad \operatorname*{Min}_{v \in \mathscr{V}_0} \quad \sum_{d \in D} \left[\sum_{a \in A} t_a v_a^d + \sum_{i \neq d} w_i^d \right]$$
$$\text{s.t.} \quad v_a^d \leqslant w_i^d f_a, \quad d \in D, \ i \neq d, \ a \in A_i^+,$$

where w_i^d represents the waiting time at node i of all passengers going to destination d. In this setting, the time-to-destination variables τ_i^d are the dual variables of the linear program, and the third term of the objective function in (P_1) is the objective value of the dual so that the condition

$$\sum_{d\in D}\left[\sum_{a\in A} t_a v_a^d + \sum_{i\neq d} w_i^d - \sum_{i\neq d} g_i^d \tau_i^d\right] = 0$$

is just the strong duality theorem of linear programming.

Consider now the case where congestion is introduced only through segment crowding functions in such a way that the cost of arc a is an increasing function of the total flow $v_a = \sum_{d\in D} v_a^d$ on that arc. As shown in Spiess (1984), the transit equilibrium may be found as the solution of the following equivalent convex cost minimization problem

$$\text{(PC)} \quad \underset{v\in\mathscr{V}_0}{\text{Min}} \quad \sum_{a\in A}\int_0^{v_a} t_a(x)\,\mathrm{d}x + \sum_{\substack{d\in D\\ i\neq d}} w_i^d$$

$$\text{s.t.} \quad v_a^d \leqslant w_i^d f_a, \quad d\in D,\ i\neq d,\ a\in A_i^+,$$

$$v_a = \sum_{d\in D} v_a^d, a\in A.$$

Setting $w_i^d(v) = \max_{a\in A_i^+} v_a^d/f_a$, the optimality conditions for this problem are

$$\sum_{a\in A} t_a(v_a)(v_a' - v_a) + \sum_{\substack{d\in D\\ i\neq d}} [w_i^d(v') - w_i^d(v)] \geqslant 0 \quad \text{for all } v'\in\mathscr{V}_0$$

so that an optimal solution satisfies

$$\sum_{a\in A} t_a(v_a)v_a + \sum_{\substack{d\in D\\ i\neq d}} w_i^d(v) - \sum_{\substack{d\in D\\ i\neq d}} g_i^d \tau_i^d(v) = 0$$

and therefore (PC) can be reformulated as

$$\underset{v\in\mathscr{V}_0}{\text{Min}} \sum_{d\in D}\left[\sum_{a\in A} t_a(v_a)v_a^d + \sum_{i\neq d}\max_{a\in A_i^+}\frac{v_a^d}{f_a} - \sum_{i\neq d} g_i^d\tau_i^d(v)\right]$$

which is again a special case of (P). Notice that problem (P) is not directly equivalent to (PC) but rather to the corresponding optimality conditions. In this sense, despite the fact that (PC) applies in a more restrictive setting it has an advantage over (P) since the latter need not be a convex program.

4. A solution algorithm

Our attempts to solve problem (P) directly as an optimization problem by using descent methods were not successful. Among the algorithms tried there was a projected generalized gradient method similar to the one used by Constantin (1993). This approach was not pursued since the tests carried out on a small network which permitted an analytical expression of the projected gradient exhibited an extremely slow convergence; and moreover it seems difficult to establish the convergence of the method. From a different perspective, since the relation between strategy-flows and arc-flows is non-linear, the attempts to formulate the problem as a variational inequality and to use numerical methods for VI's were not successful either. The model could be written as a quasi-variational inequality (see pp. 64–66 in Cepeda, 2002), but this approach did not lead to an algorithm.

Now, since the optimal value of (P) is known, a simple alternative is to use a heuristic minimization method and to evaluate the deviation from optimality of the computed solutions by using the value of the gap function $G(v)$. One of the simplest and most robust schemes for solving equilibrium problems is the well known method of successive averages which has been extensively used in transportation applications. At each iteration, the method computes a transit network equilibrium for the linear network obtained by fixing the travel times and the frequencies at the values determined by the current flows, and then updates these flows by averaging the previous iterate and the newly computed solution. As mentioned earlier, an equilibrium for the linear cost network can be found by solving the linear program (PL), which amounts to compute a shortest hyperpath for each destination $d \in D$ and then determining the corresponding arc-destination flow vector. Optimal hyperpaths may be computed with the method proposed by Spiess (1984) and Spiess and Florian (1989) or the hyperpath-Dijkstra method described in Cominetti and Correa (2001). Hence, choosing $\alpha_k \in (0,1)$ with $\alpha_k \to 0$ and $\sum_{k=0}^{\infty} \alpha_k = \infty$, the method can be described as

Algorithm MSA
Initialize: find $v^0 \in \mathscr{V}_0$ and set $k \leftarrow 0$
while $G(v^k) > \epsilon$
- compute $t_a = t_a(v^k)$ and $f_a = f_a(v^k)$
- compute shortest hyperpaths for each $d \in D$
- determine the induced flows $\hat{v}_a^d$
- update $v^{k+1} = (1 - \alpha_k)v^k + \alpha_k \hat{v}$
- set $k \leftarrow k + 1$

end
Stop: v^k is a solution with gap $G(v^k) \leqslant \epsilon$.

The computation of the gap $G(v^k)$ requires the storage of the flow variables v_a^d for every destination $d \in D$ and all arcs $a \in A$. As the network and the number of destinations grows, the storage requirement grows accordingly but even for large networks this remains within the capabilities of standard computers.

A starting point $v^0 \in \mathscr{V}_0$ may be obtained by performing an ‘‘all-or-nothing’’ assignment using shortest hyperpaths computed with constant times $t_a = t_a(0)$ and frequencies $f_a = f_a(0)$. It may happen that this flow violates the capacity constraints reflected in the fact that $f_a(v^0) = 0$ for some arcs, and then (PL) could become infeasible in the following iteration. Alternatively, one could think of initially computing a multi-commodity flow with capacity constraints but this may be expensive and moreover there is no guarantee that feasibility will be maintained along subsequent iterations. Infeasibility will occur for instance if there is insufficient capacity to carry all the demand for travel on the transit network. To avoid this problem and to allow the use of an ‘‘all-or-nothing’’ initialization, we assume the existence of a subnetwork which is not subject to saturation: for each node i and every destination d there is a path joining i to d with arcs such that $f_a(v) > 0$ for all $v \in \mathscr{V}$. This holds for instance if the network contains a pedestrian subgraph with infinite capacity connecting every node to each destination. Otherwise this condition may be enforced by truncating the effective frequency functions as $\tilde{f}_a(v) = \max\{f_a(v), \epsilon\}$. Under such a non-saturation condition the iterates v^k may still violate the capacity constraints at some stage, but (PL) will remain feasible and its solution will only allow small or null flows along the over-saturated arcs so that the averaging process in MSA will sequentially reduce the corresponding flows until they eventually become capacity feasible. Of course this will not occur if the network is so saturated and the waiting times are so large that some flow along pedestrian arcs is required at equilibrium. In the latter case MSA will find an equilibrium solution with large increases in pedestrian flows and/or with over-saturated transit line segments, revealing the corridors of the transit network that require additional capacity.

4.1. Numerical experiments

The MSA algorithm with an "all-or-nothing" initialization was implemented as a macro procedure within the EMME/2 software package (see INRO Consultants Inc. (1998)). The latter provides a module that computes an optimal hyperpath with linear costs and fixed frequencies on a general transit network, and which is used to solve the linearized subproblems that compute $\hat{v}$ in the algorithm stated above.

In the numerical tests we considered constant travel times $t_a(v) = t_a$, and effective frequencies equal to ∞ except on the boarding arcs where

$$f_a(v) = \begin{cases} \mu\left[1 - \left(\frac{v_a}{\mu c - v_{a'} + v_a}\right)^{\beta}\right] & \text{if } v_{a'} < \mu c, \\ 0 & \text{otherwise} \end{cases}$$

with v_a representing the flow boarding at the stop (summed over all destinations, i.e. $v_a = \sum_{d\in D} v_a^d$), and similarly $v_{a'}$ is the on-board flow right after the stop (notice that $v'_a \geqslant v_a$). The parameter μ denotes the nominal frequency of the corresponding line and c is the physical capacity of the buses, so that $\mu c - v_{a'}$ is the expected residual capacity after the stop, and this functional form explicitly incorporates the capacity constraint $v'_a < \mu c$. The effective frequencies are internally truncated as $\tilde{f}_a(v) = \max\{f_a(v), \frac{1}{999}\}$ so that the largest headway is 999 min or 16.7 h. This number is high enough if one considers that in an urban environment the typical travel times by transit are of the order of 20 min–1 h. The computed flows may then exceed the capacity during the first iterations but if a capacity feasible solution exists it will be found in subsequent steps of the computations since newly computed strategies will not contain the transit line segments that have very high waiting times. The algorithm will find an approximate equilibrium or it will determine that a capacity feasible solution does not exist.

4.1.1. A small example

Consider the small network with three centroids denoted A, B, C in Fig. 4, and demands of 10 trips from A to B, 10 trips from B to C, and 100 trips from A to C.

There are two transit lines: the "express" line connects A–C by one segment 20 km long, and the "local" line connects A–B–C by two consecutive segments of 10 km each. The capacity of

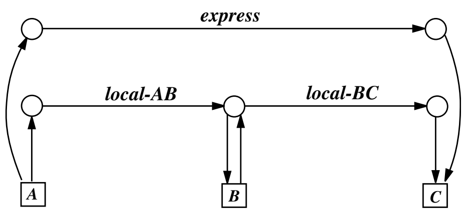

Fig. 4. Small network.

buses is 20 pax/bus, dwell time at stops is 0.01 min, and the effective frequency functions are taken as above with exponent $\beta = 0.2$. The speed, frequency and total capacity of the services are summarized in the following table:

Transit line	Speed (km/h)	Frequency (bus/h)	Capacity (pax/h)
Express	50	16	320
Local	30	6	120

The local demands (A to B and B to C) have only one alternative each which is to take the local line, while the demand from A to C has three possible strategies: express (E), local (L), express + local (EL). In fact, since strategy L is always dominated by EL, the only rational strategies are E and EL (i.e. passengers will always board the express line if there is available space). The initial "all-or-nothing" assignment yields the intuitive result that all the trips use the most direct transit line as shown below:

Segment	Travel time	Link volume	Load factor
Express	24.01	100	0.31
Local-AB	20.01	10	0.08
Local-BC	20.01	10	0.08

However, the load factor of the express line is so high that the total time (wait + travel) of strategy E is 42.08 min. This value is larger than the 40.02 min travel time of the local line from A to C, which then becomes competitive. Hence, the 100 demand from A to C should also consider the combined strategy EL. Splitting the demand among these two strategies will reduce the load of the express line, and the equilibrium is reached when the total time of E equals the travel time of the local line, which occurs with 47.2 flow units on E and 52.8 on EL. The latter splits proportionally to the effective frequencies between the express line (37.1 units) and the local line (15.7 units) yielding the following line loads:

Segment	Travel time	Link volume	Load factor
Express	24.01	84.3	0.26
Local-AB	20.01	25.7	0.21
Local-BC	20.01	25.7	0.21

with an equilibrium travel time of 40.02 min for both the E and EL strategies. This equilibrium time coincides with the travel time of the local line from A to C, and will remain exactly the same if the demands are increased by a small amount (though the flow shares will of course change). In this example both the initial all-or-nothing assignment and the equilibrated solution satisfy the capacity constraints. If the demand from A to C is increased to 350 the initial "all-or-nothing" assignment gives a flow of 350 on the express line which is not capacity feasible, while the equilibrium yields

Segment	Travel time	Link volume	Load factor
Express	24.01	260.5	0.81
Local-AB	20.01	99.5	0.83
Local-BC	20.01	99.5	0.83

In this case the congestion at node A is so high that strategy E is no longer used and all the AC demand takes the combined strategy EL with an equilibrium time of 97.36 min. Notice the apparently counter-intuitive fact that the local line ends up slightly more saturated than the express line. This is not contradictory though: the only way to reduce the load on the local line would be to shift some flow from the strategy EL to the strategy E, but since the total time of E is 117.04 min it is unreasonable to expect that any user would refuse to board a local bus arriving at the stop which offers a 40.02 min ride to destination C. Of course the local strategy L is still worse with a total time of 312.13 min from A to C.

4.1.2. Larger networks

MSA was also tested on some medium and large instances. The transit networks used for the numerical tests presented below originate from the cities of Stockholm, Winnipeg and Santiago, Chile. The origin–destination matrices correspond to real estimated demands.

4.1.2.1. Stockholm and Winnipeg. The transit network of the City of Stockholm consists of six modes, six transit vehicle types, 185 centroids, 395 regular nodes, 2079 directional links, 76 transit lines and 1096 transit line segments. In the case of the City of Winnipeg, the transit network consists of 154 centroids, 903 regular nodes, 2975 directional links, 133 transit lines and 4338 transit line segments. A walking speed of 3 km/h was taken for the pedestrian arcs. Fig. 5 shows the evolution of the first 70 MSA iterations, by plotting the *relative gap*, that is to say, the gap $G(v^k)$ expressed as a percentage of the total travel time $\sum_{\substack{d\in D\\ i\neq d}} g_i^d \tau_i^d(v^k)$.

In both cases, the convergence towards 0 is relatively fast and smooth, reaching a value close to 0.25% after 70 iterations. A closer look reveals that in the case of Stockholm there is not enough capacity to accommodate the demand and the problem is infeasible. Indeed, the maximum value of the flow-to-capacity ratio over all the line segments, which must be less than 1 if capacities are respected, drops from 4.77 to 1.5 stabilizing at this value after 10 iterations. Simultaneously, the number of over-saturated line segments having a ratio larger than one, reduces from 45 arcs in the first iteration to four arcs at iteration 10, oscillating thereafter around this value without ever reaching a feasible flow. This indicates the existence of a bottleneck due to an overestimated demand or because some line capacities are too small. This situation does not occur in the case of

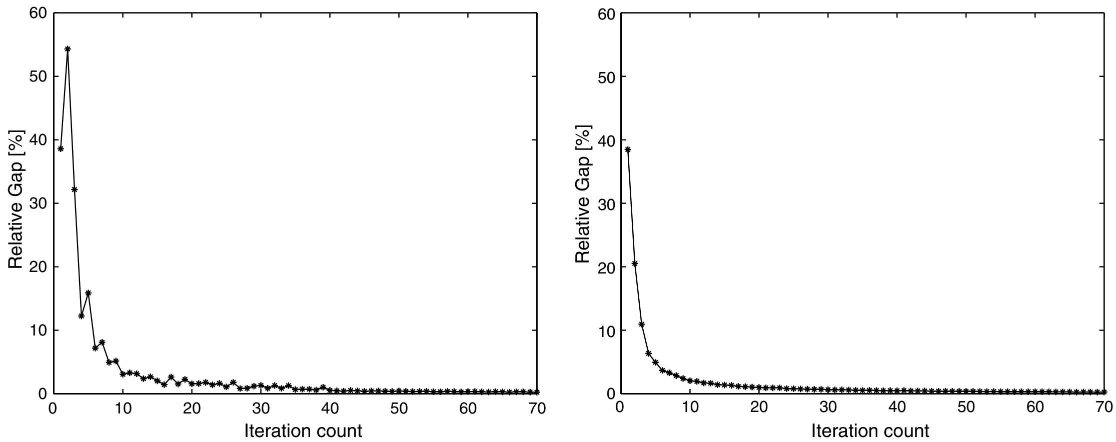

Fig. 5. MSA convergence for Stockholm (left) and Winnipeg (right).

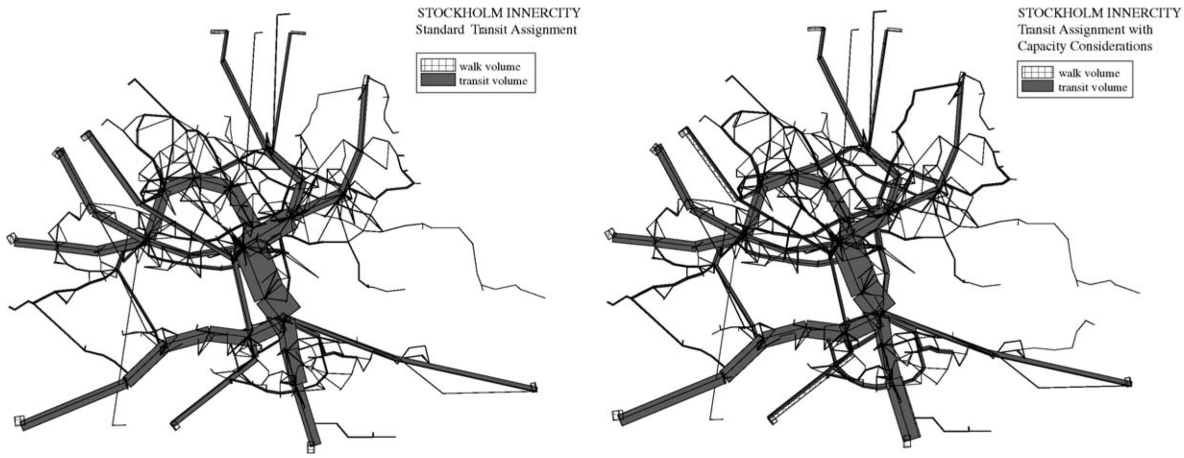

Fig. 6. Standard transit assignment (left) vs capacitated equilibrium assignment (right).

Winnipeg where the maximum flow-to-capacity ratio drops from 2.12 to 0.97, while the over-saturated transit line segments are reduced from 221 to 0, stabilizing at these feasible values after nine iterations. Fig. 6 compares the standard transit assignment obtained in the first iteration for Stockholm with the flow pattern that results after equilibration, while Fig. 7 shows the difference between both. The over-saturated line segments may be easily identified by looking at the corridors which present high pedestrian flows, allowing to determine the source of the conflict.

4.1.2.2. The Santiago equilibrium model. The network of the City of Santiago has 409 centroids, 1808 nodes, 11 331 directional links, 1116 transit lines and 52 468 transit line segments. There are 11 modes including five pure transit modes and four combined modes (such as auto-metro, etc.).

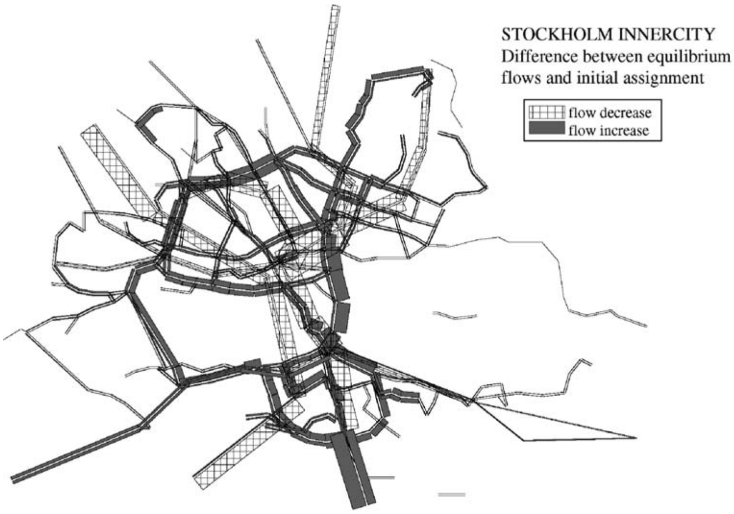

Fig. 7. Difference between equilibrium flows and initial assignment.

The transportation planning model includes all the demand computation steps and the network assignment for 13 socioeconomic classes and three trip purposes. The model is quite complex and may be referred to in De Cea and Fernández (2001) or Florian et al. (2002). The critical part of the transit network was the underground (metro) which was consistently over simulated. The MSA algorithm was embedded in the complex equilibration procedure that solves the model, after which the resulting metro flows satisfied the capacity constraints. Fig. 8 illustrates the flows on

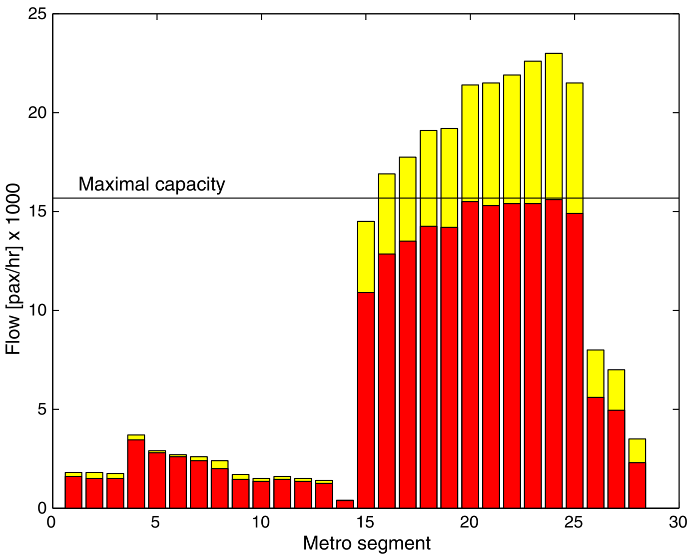

Fig. 8. Flows on metro line 5.

the different segments of metro line 5 at the beginning of MSA (light) and after 20 iterations (dark).

5. Conclusion

The developments presented in this paper open the way for the use of a new transit assignment model in congested networks where it is important to respect the capacity of the services offered. The theoretical analysis presented provides a solid foundation for a simple and robust algorithm since the nearness to an exact equilibrium solution is easily quantified. The empirical convergence of the algorithm raises the interesting issue of devising a rigorous convergence proof which is still an open research question.

Appendix. Global optimality

When dealing with equilibrium problems it is useful to have a gap function whose global minimizers characterize the solutions. However, this fact does not guarantee that minimization of the gap function will lead necessarily to an equilibrium because descent methods may be trapped at local minima. To prove the usefulness of a gap function it is important to establish that every local minimizer is also a global one and therefore an equilibrium. Proving such a result is usually very difficult and the gap function $G(v)$ in Section 3 is no exception. However we can prove this result for the special case of common-lines, namely a two-node network linked by finitely many arcs $a \in A$ with constant travel times $t_a(v) = t_a$ and decreasing diagonal effective frequencies $f_a(v) = f_a(v_a)$. As a matter of fact, in this case one can show that every stationary point is automatically an equilibrium and hence a global optimum. In order to prove the latter, let us consider a common-line problem with total demand $x > 0$ so that (P) becomes

$$(\tilde{\mathrm{P}}) \quad \operatorname*{Min}_{v} \quad \sum_{a \in A} t_a v_a + \max_{a \in A} \frac{v_a}{f_a(v_a)} - x \min_{s \in \mathscr{S}} T_s(v)$$
$$\text{s.t.} \quad v_a \geqslant 0; \quad \sum_{a \in A} v_a = x$$

with

$$T_s(v) = \frac{1 + \sum_{a \in s} t_a f_a(v_a)}{\sum_{a \in s} f_a(v_a)}.$$

Let us rewrite this problem as

$$(\tilde{\mathrm{P}}) \quad \operatorname*{Min}_{(v,\alpha,\tau)} \quad \sum_{a \in A} t_a v_a + \alpha - x\tau$$
$$\text{s.t.} \quad 0 \leqslant v_a \leqslant \alpha f_a(v_a) \quad \forall a \in A,$$
$$\sum_{a \in A} v_a = x,$$
$$\tau \leqslant T_s(v) \quad \forall s \in \mathscr{S}$$

and introduce multipliers $\mu \in \mathbb{R}$, $\beta_a \geqslant 0$, $\gamma_a \geqslant 0$ and $h_s \geqslant 0$ in order to build the Lagrangian

$$\mathscr{L} \triangleq \sum_{a\in A} t_a v_a + \alpha - x\tau + \mu\left[x - \sum_{a\in A} v_a\right] - \sum_{a\in A}\beta_a v_a + \sum_{a\in A}\gamma_a[v_a - \alpha f_a(v_a)] + \sum_{s\in\mathscr{S}} h_s[\tau - T_s(v)].$$

A feasible point (v,α,τ) for (P) is a KKT point if there exist multipliers (μ,β,γ,h) satisfying the complementarity conditions

$$h_s > 0 \Rightarrow T_s(v) = \tau, \tag{5}$$

$$\beta_a > 0 \Rightarrow v_a = 0, \tag{6}$$

$$\gamma_a > 0 \Rightarrow v_a = \alpha f_a(v_a), \tag{7}$$

and the stationarity conditions $\partial\mathscr{L}/\partial\tau = 0$, $\partial\mathscr{L}/\partial\alpha = 0$ and $\partial\mathscr{L}/\partial v_a = 0$, which correspond to

$$\sum_{s\in\mathscr{S}} h_s = x, \tag{8}$$

$$\sum_{a\in A}\gamma_a f_a(v_a) = 1, \tag{9}$$

$$t_a - \mu - \beta_a + \gamma_a[1 - \alpha f_a'(v_a)] = \sum_{s\ni a}\frac{h_s}{\sum_{b\in s} f_b(v_b)}[t_a - T_s(v)]f_a'(v_a). \tag{10}$$

To simplify the notation, the dependence of f_a and f_a' on v_a is omitted. Also set $\eta = \mu - \tau$ and

$$\alpha_a = \sum_{s\ni a}\frac{h_s}{\sum_{b\in s} f_b(v_b)}.$$

Then, taking into account (5), Eq. (10) may be rewritten as

$$(t_a - \tau)[1 - \alpha_a f_a'] = \beta_a - \gamma_a[1 - \alpha f_a'] + \eta. \tag{11}$$

Proposition 5.1. *If (v,α,τ) is a KKT point for $(\tilde{\mathrm{P}})$ then v is a common-line equilibrium.*

Proof. The constraints of (P) imply $\tau \leqslant \min_s T_s(v)$. Now, from (8) it follows that there exists $s \in \mathscr{S}$ with $h_s > 0$ so that (5) implies $T_s(v) = \tau$ and consequently $\tau = \min_{s\in\mathscr{S}} T_s(v)$. The result will follow from Theorem 2.1 if it is shown that $\eta = 0$, since in this case (11) combined with the complementarity conditions imply

$$\begin{cases} t_a < \tau \Rightarrow \gamma_a > 0 \Rightarrow v_a/f_a = \alpha, \\ t_a > \tau \Rightarrow \beta_a > 0 \Rightarrow v_a/f_a = 0. \end{cases}$$

In order to show that $\eta = 0$ note that β_a and γ_a may not be positive simultaneously (otherwise one would have $\alpha = 0$, so that $v_a = 0$ for all a and then $x = 0$). Also, every $s \in \mathscr{S}$ with $h_s > 0$ is such that $T_s(v) = \tau$, hence $T_s(v)$ is minimal and according to Lemma 2.1 s contains all the arcs such that $t_a < \tau$ and no arc with $t_a > \tau$. It follows that

$$\begin{cases} t_a < \tau \Rightarrow \alpha_a = \bar{\alpha}, \\ t_a > \tau \Rightarrow \alpha_a = 0 \end{cases}$$

with $\bar{\alpha} = \sum_{s\in\mathscr{S}} \frac{h_s}{\sum_{b\in s} f_b(v_b)}$. Finally, remark that

$$\sum_{a\in A} \alpha_a f_a = \sum_{a\in A} \sum_{s\ni a} \frac{h_s}{\sum_{b\in s} f_b} f_a = \sum_{s\in\mathscr{S}} \sum_{a\in s} \frac{h_s f_a}{\sum_{b\in s} f_b} = \sum_{s\in\mathscr{S}} h_s = x = \sum_{a\in A} v_a$$

which gives

$$0 = \sum_{a\in A} \left[\alpha_a - \frac{v_a}{f_a}\right] f_a. \tag{12}$$

Let us show that $\eta = 0$ by contradiction.

Suppose first $\eta < 0$ and let $s_0 = \{a\colon v_a > 0\}$. For $a \in s_0$ the value of β_a equals 0 and then (11) yields $t_a < \tau$ and $\gamma_a < (\tau - t_a)[1 - \bar{\alpha} f'_a]/[1 - \alpha f'_a]$. Also, by noting that $\gamma_a = 0$ for $a \notin s_0$ (since $\gamma_a > 0$ implies $v_a = \alpha f_a > 0$), and by using (9) one gets

$$1 = \sum_{a\in s_0} \gamma_a f_a < \sum_{a\in s_0} (\tau - t_a) f_a \frac{[1 - \bar{\alpha} f'_a]}{[1 - \alpha f'_a]}. \tag{13}$$

Now, $\tau \leqslant T_{s_0}(v)$ implies $\sum_{a\in s_0} (\tau - t_a) f_a \leqslant 1$, which combined with (13) leads to

$$0 < \sum_{a\in s_0} (\tau - t_a) f_a \left(\frac{[1 - \bar{\alpha} f'_a]}{[1 - \alpha f'_a]} - 1 \right),$$

or equivalently

$$0 < (\alpha - \bar{\alpha}) \sum_{a\in s_0} \frac{(\tau - t_a) f_a f'_a}{[1 - \alpha f'_a]}.$$

Since each term in this last sum is negative it follows that $(\alpha - \bar{\alpha}) < 0$. By substituting this back into (12) and by using the fact that $\alpha_a = \bar{\alpha}$ for $a \in s_0$ and $v_a = 0$ for $a \notin s_0$, the following contradiction is obtained

$$0 = \sum_{a\in s_0} \left[\bar{\alpha} - \frac{v_a}{f_a}\right] f_a + \sum_{a\notin s_0} \alpha_a f_a \geqslant (\bar{\alpha} - \alpha) \sum_{a\in s_0} f_a > 0.$$

Suppose now that $\eta > 0$ and let $s_1 = \{a\colon t_a \leqslant \tau\}$. For $a \in s_1$ Eq. (11) implies that $\gamma_a > 0$, and therefore one has $v_a/f_a = \alpha$, $\beta_a = 0$ as well as $\gamma_a > (\tau - t_a)[1 - \bar{\alpha} f'_a]/[1 - \alpha f'_a]$. These conditions, combined with (9) give

$$1 \geqslant \sum_{a\in s_1} \gamma_a f_a > \sum_{a\in s_1} (\tau - t_a) f_a \frac{[1 - \bar{\alpha} f'_a]}{[1 - \alpha f'_a]}.$$

By Lemma 2.1 we have $\tau = T_{s_1}(v)$ which implies $\sum_{a\in s_1} (\tau - t_a) f_a = 1$. This leads to

$$0 > \sum_{a\in s_1} (\tau - t_a) f_a \left(\frac{[1 - \bar{\alpha} f'_a]}{[1 - \alpha f'_a]} - 1 \right)$$

and further

$$0 > (\alpha - \bar{\alpha}) \sum_{a \in s_1} \frac{(\tau - t_a) f_a f_a'}{[1 - \alpha f_a']}$$

which yields $(\alpha - \bar{\alpha}) > 0$. Going back to (12) once again and by using now the fact that $v_a/f_a = \alpha$ for $a \in s_1$ and $\alpha_a = 0$ for $a \notin s_1$, one reaches the contradiction

$$0 = \sum_{a \in s_1} [\alpha_a - \alpha] f_a - \sum_{a \notin s_1} v_a \leqslant (\bar{\alpha} - \alpha) \sum_{a \in s_1} f_a < 0.$$

These two contradictions prove that $\eta = 0$, thus completing the proof. □

Corollary 5.1. *If* (v, α, τ) *is a local minimum for* $(\tilde{\mathrm{P}})$ *then it is a global minimum and* v *is a common-line equilibrium.*

Proof. It is easy to check that (v, α, τ) satisfies the Mangasarian–Fromovitz constraint qualification so that it is a KKT point. Proposition 5.1 then implies that v is an equilibrium, and then Theorem 2.1 implies that v is a global optimum with cost equal to zero. □

References

Bouzaïene-Ayari, B., 1996. Affectation statique des passagers dans les réseaux de transport en commun: modélisation et algorithme de résolution. Ph.D. Thesis, Département d'Informatique et Récherche Opérationnelle, Publication 96-12, CRT, U. de Montréal.

Bouzaïene-Ayari, B., Gendreau, M., Nguyen, S., 1995. An equilibrium-fixed point model for passenger assignment in congested transit networks. Technical Report CRT-95-57, U. de Montréal.

Bouzaïene-Ayari, B., Gendreau, M., Nguyen, S., 2001. Modeling bus stops in transit networks: a survey and new formulations. Transportation Science 35 (3), 304–321.

Cepeda, M., 2002. Modèle d'équilibre dans les réseaux de transport en commun: le cas des capacités explicites des services. Ph.D. Thesis, Département d'Informatique et Récherche Opérationnelle, Publication 2002-43, CRT, U. de Montréal.

Chriqui, C., Robillard, P., 1975. Common bus lines. Transportation Science 9, 115–121.

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

Constantin, I., 1993. L'optimisation de fréquences d'un réseaux de transport en commun. Ph.D. Thesis, Département d'Informatique et Récherche Opérationnelle, Publication 881, CRT, U. de Montréal.

Correa, J., 1999. Asignación de flujos de pasajeros en redes de transporte público congestionadas. Engineering thesis, U. de Chile, Santiago.

De Cea, J., Fernández, E., 1989. Transit assignment to minimal routes: an efficient new algorithm. Traffic Engineering and Control 30, 491–494.

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

De Cea, J., Fernández, E., 2001. ESTRAUS: a simultaneous equilibrium model to analyse and evaluate multimodal urban transportation system with multiple user classes. In: 9th WCTR Proceedings.

De Cea, J., Bunster, J.P., Zubieta, L., Florian, M., 1988. Optimal strategies and optimal routes in public transit assignment models. Traffic Engineering and Control 29, 520–526.

Dial, R.B., 1967. Transit pathfinder algorithms. Highway Research Record 205, 67–85.

Fearnside, K., Draper, D.P., 1971. Public transport assignment—a new approach. Traffic Engineering and Control, 298–299.
Florian, M., Wu, J.H., He, S., 2002. A multi-class multi-mode variable demand network equilibrium model with hierarchical logit structures. In: Gendreau, M., Marcotte, P. (Eds.), Transportation and Network Analysis: Current Trends, pp. 119–133.
Gendreau, M., 1984. Etude approfondie d'un modèle d'équilibre pour l'affectation de passagers dans les réseaux de transports en commun. Ph.D. Thesis, Département d'Informatique et Récherche Opérationnelle, Publication 384, CRT, U. de Montréal.
INRO Consultants Inc., 1998. EMME/2 User's Manual Release 9.
Kurauchi, F., Bell, M.G.H., Schmöcker, J.D., 2003. Capacity constrained transit assignment with common lines. Journal of Mathematical Modeling and Algorithms 2 (4), 309–327.
Le Clercq, F., 1972. A public transport assignment model. Traffic Engineering and Control, 91–96.
Nguyen, S., Pallotino, S., 1988. Equilibrium traffic assignment in large scale transit networks. European Journal of Operational Research 37 (2), 176–186.
Nuzzolo, A., Wilson, N.H.M. (Eds.), 2004. Schedule-Based Dynamic Transit Modeling. Theory and Applications. Operations Research/Computer Science Interfaces Series, Band 28. Springer-Verlag, Berlin.
Spiess, H., 1984. Contribution à la théorie et aux outils de planification des réseaux de transport urbains. Ph.D. Thesis, Département d'Informatique et Récherche Opérationnelle, Publication 382, CRT, U. de Montréal.
Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. Transportation Research Part B 23 (2), 83–102.
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.
Wu, J., Florian, M., Marcotte, P., 1994. Transit equilibrium assignment: a model and solution algorithms. Transportation Science 28 (3), 193–203.