Pergamon

0969-6016(94)00023-9

Int. Trans. Opl Res. Vol. 2, No. 2, pp. 149–164, 1995
Elsevier Science Ltd.
0969-6016/95 $9.50 + 0.00

Optimizing Frequencies in a Transit Network: a Nonlinear Bi-level Programming Approach

ISABELLE CONSTANTIN and MICHAEL FLORIAN
Université de Montréal, Canada

We consider the problem of optimizing the frequencies of transit lines in an urban transportation network. The problem is formulated first as a nonlinear nonconvex mixed integer programming problem and then it is converted into a bi-level Min–Min nonconvex optimization problem. This problem is solved by a projected (sub)gradient algorithm, where a (sub)gradient is obtained at each iteration by solving the lower level problem. Computational results obtained with this algorithm are presented for the transit networks of the cities of Stockholm, Sweden, Winnipeg, Man., Canada and Portland, OR, U.S.A.

Key words: optimization of transit network, bi-level programming, projected gradient, urban transportation

INTRODUCTION

The optimization of frequencies is an important aspect of the design of a transit network. Yet, it is not well integrated in traditional models used in the transit planning process. Until now it has received relatively scant attention in the literature and has sometimes been studied in conjunction with the design of transit line itineraries. Furthermore, none of the models proposed so far (see e.g. Hasselström (1981) and Schéele (1977)) consider the travellers' behavior regarding route choice in a satisfactory manner.

We define, for the purpose of the model developed in this paper, the optimal transit line frequencies to be those which minimize the total expected travel and waiting time on the network, while satisfying fleet size constraints as well as lower bound constraints. We formulate the problem as a mixed integer nonlinear program. This problem has an ill-behaved objective function, which is neither convex nor concave, and binary variables in the constraints. We show that it is advantageous to reformulate the problem as a Min–Min nonlinear bi-level program. Then we adapt the theory of Danskin (1967) for Max–Min problems, for developing feasible descent directions obtained by solving the second-level problem. The latter turns out to be a transit assignment (route-choice) problem with fixed frequencies. This problem was studied by Spiess (1984) and also reported in Spiess and Florian (1989). This model is based on the concept of 'optimal strategies', and can be solved very efficiently by a label-setting algorithm which has polynomial time complexity.

The solution method that we develop for solving the Min–Min nonlinear bi-level problem is of the gradient projection type. Since the objective function is not continuously differentiable, the gradient does not exist everywhere. Thus, at some steps of the algorithm, a subgradient may be projected in order to obtain a direction of descent. The particular structure of the problem permits the determination of the projecting matrix analytically, without requiring any matrix inversion. An optimal step size is computed at each iteration by using an Armijo–Goldstein rule (see Luenberger (1984)). It may be shown that the algorithm converges towards a local minimum.

The paper is organized as follows: in the next section, we present a brief literature review of previous work done on optimizing transit line frequencies and related models. Section 3 is dedicated to the model formulation, while in Section 4 we give the theoretical results that characterize the structure of the model (no proofs are given in this paper, but they can be found in Constantin (1993)). The solution algorithm is presented in Section 5. Section 6 presents a generalization of the model and Section 7 contains numerical results obtained with three real transit networks.

Correspondence: Prof. Michael Florian, Centre de recherche sur les transports et Département d'informatique et de recherche opérationnelle, Université de Montréal, C.P. 6128, Succursale A, Montréal, Canada H3C 3J7

PREVIOUS CONTRIBUTIONS

As remarked in the Introduction, relatively few articles may be found in the literature on the topic of optimizing the frequency of transit lines. These are Schéele (1977), Grega and Jörnsten (1979), Furth and Wilson (1981) and LeBlanc (1988). On the other hand, there are several studies on the more general question of the improvement or reorganization of a transit company. In these studies, the optimization of the transit line frequencies appears as a subproblem of a more general one. It is worthwhile to cite Lampkin and Saalmans (1967), Silman *et al.* (1974), Rosello (1976), Dubois *et al.* (1979) and Mandl (1980).

Other authors addressed the problem of determining, simultaneously, transit line frequencies and other decision variables. Holroyd (1967) considered the design of a transit network on a *grid* network by determining simultaneously the location and frequency of the lines. Byrne proposed two models for the same problem on a radial network (Byrne, 1975) and for parallel lines (Byrne, 1976). Hasselström (1981) sought to determine simultaneously the *itinerary* and the frequency of transit lines on a general network. He also proposed a model which optimizes simultaneously the *type of vehicle* and the frequency assigned to transit lines. Stephanedes and Kwon (1988) consider the *fare* as a decision variable, as well as the line frequencies. Talley (1989) proposes a procedure to determine the *speed* and the line frequencies.

Table 1 provides a chronological review of the main models mentioned above. While we do not examine the contributions of these authors in detail, we intend to do so in a forthcoming paper.

Table 1. Chronological review of some models

Year and authors	Variables (other than frequencies)	Model	Demand
1967 Lampkin and Saalmans	Timetable, itinerary	Min. travel time s.t. fleet constraints, segment capacity	fixed
1969 Jewell		Min. waiting s.t. fleet constraints	fixed
1974 Silman *et al.*	itinerary	Min. travel time + discomfort s.t. fleet constraints	fixed
1976 Rosello	itinerary	Min. travel time s.t. fleet constraints	modal choice
1977 Schéele		Min. travel time s.t. fleet constraints and capacity (normative model)	variable trip distribution
1979 Dubois *et al.*	itinerary	Min. travel time s.t. fleet constraints	modal choice
1979 Grega and Jörnsten		Min. waiting time s.t. fleet constraints and link capacity	direct demand functions
1980 Mandl	itinerary	Min. waiting time s.t. fleet constraints	fixed
1981 Furth and Wilson		Max. consumer surplus + no. of passengers s.t. fleet constraints, budget, segment capacities and lower bounds	direct demand functions
1981 Hasselström	itinerary, vehicle type	Max. consumer surplus s.t. fleet constraints (or budget) and segment capacities	direct demand functions
1986 Mizokami and Kawakami		Max. consumer surplus s.t. network equilibrium and link capacities	modal choice
1988 LeBlanc		Min. no. of auto trips + weighted sum of frequencies s.t. lower bounds	modal choice
1988 Stephanedes and Kwon	fare	Max. no. of passengers s.t. deficit, or Min. deficit s.t. ridership	logit model
1989 Talley	speed	Min. operating cost + passengers' cost s.t. deficit	direct demand functions

NOTATION AND MODEL FORMULATION

We consider a transit network $G = (N,A)$ that consists of origin and destination nodes, and intermediate nodes which define transit line segments. The links describe all possible travel on the network. The network representation is that presented in Spiess (1984), which we refer to as the *generalized* network. In this network each intermediate node of the physical transit network is duplicated to represent the *stops* of each line which is incident to the node. Links a, $a \in A$, correspond to boarding, in-vehicle, alighting and *walk* links. Each link has two attributes: a travel time (or travel cost) t_a and a frequency f_a. For walk links, which do not incur any waiting, the frequency is infinite. The *boarding* links connect a node to a stop; their travel time is zero and their frequency corresponds to that of the line to which they access. The *in-vehicle* links connect two consecutive stops of a line; their travel time is that of the time taken by the transit vehicle to travel between these two stops and their frequency is infinite. The *alighting* links connect a stop to a node; their travel time is zero and their frequency is infinite. Finally, the *walk* links connect a node to another node; their travel time is a function of their length and of the average walk speed, and their frequency is infinite. An example of a simple transit network and the corresponding generalized network are given in Figs 1 and 2, respectively.

While we do not present the transit route choice model of Spiess (1984) in detail here, we refer the reader to the published paper of Spiess and Florian (1989). This model is based on the notion of the *optimal strategy*. A strategy may be interpreted to be an acyclic directed subgraph of the generalized network, corresponding to the choices made by travellers on the transit network. In order to formulate the model we use the following notation:

$i, i \in I$ a node of the generalized network;
$a, a \in A$ a link of the generalized network;

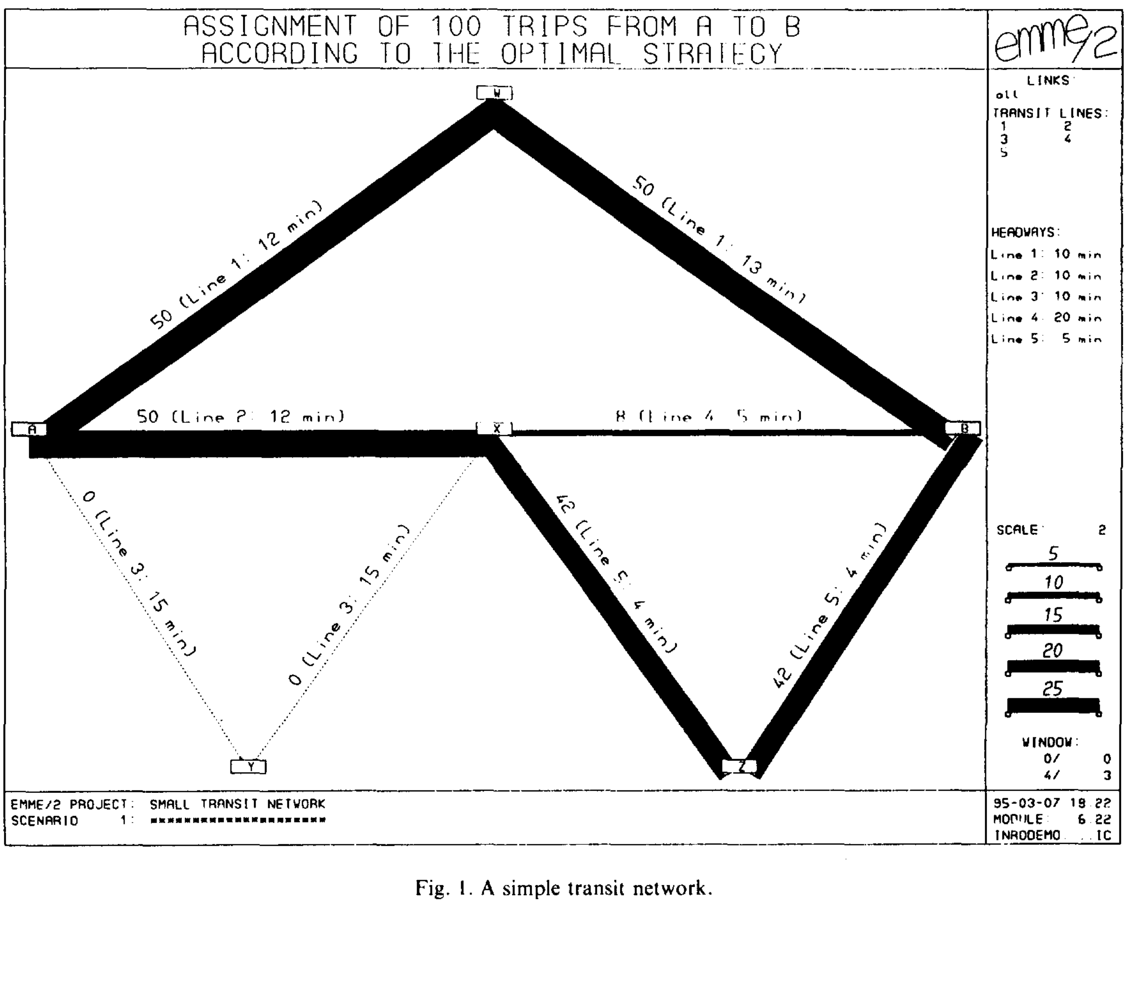


Fig. 1. A simple transit network.

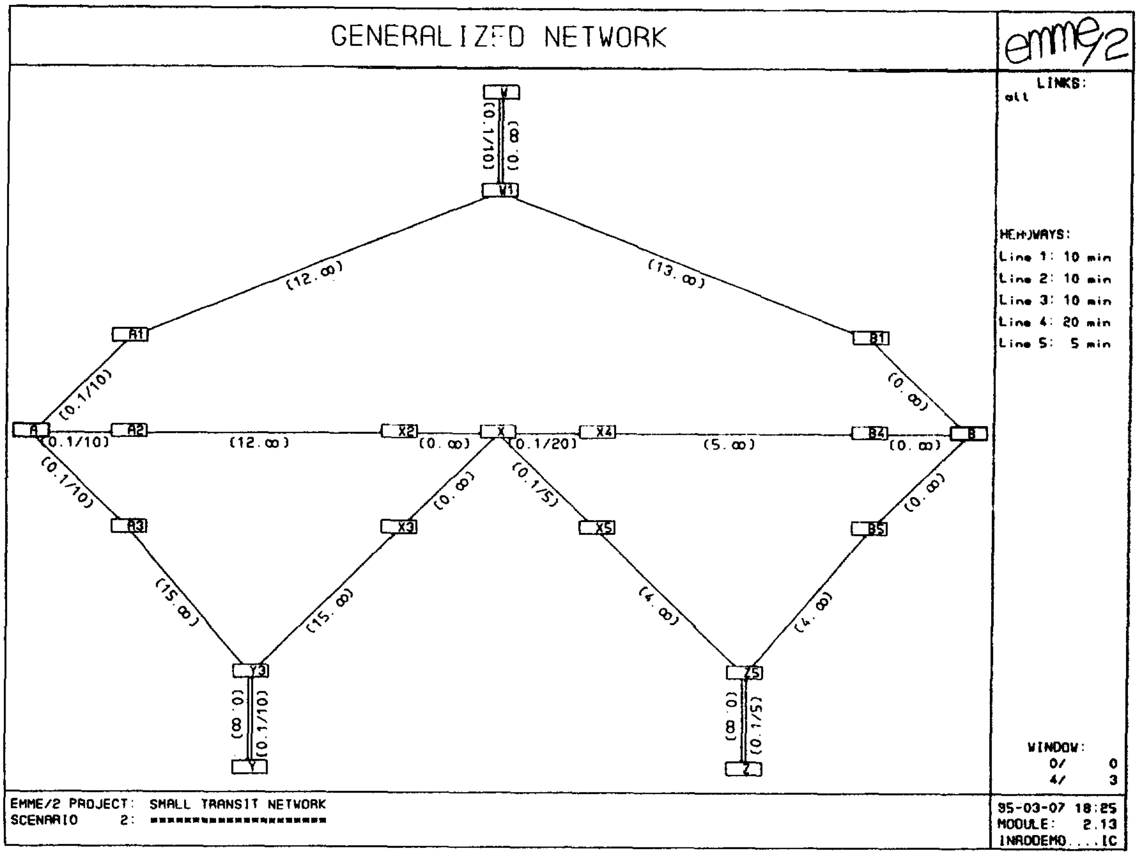


Fig. 2. A corresponding generalized network.

$a,a \in A_i^+$	the forward star of node i;
$\bar{A} \subseteq A$	a strategy, which is an acyclic subgraph;
$a \in \bar{A}$	the links belonging to a strategy;
$g_i, i \in I$	the travel demand from node i to a destination;
$v_a, a \in A$	the volume of transit passengers on link a.

According to the assumptions made in the model, the waiting time at a node i for a strategy $\bar{A}$ is that it is given by

$$W(\bar{A}_i^+) = \frac{\alpha_i}{\sum_{a \in \bar{A}_i^+} f_a}, \ \alpha_i > 0, \ i \in I, \tag{1}$$

and the proportion of passengers that board link a to leave node i is

$$P_a(\bar{A}_i^+) = \frac{f_a}{\sum_{a' \in \bar{A}_i^+} f_{a'}}, \ a \in \bar{A}_i^+, \ i \in I. \tag{2}$$

The parameter α is a *wait time* factor. A value of $\alpha = 1$ corresponds to an exponential distribution of the interarrival times of the vehicles with a mean $1/f_a$. A value of $\alpha = 0.5$ corresponds to an approximation of constant interarrival time $1/f_a$. In the following, for the simplicity of the exposition, we shall assume that $\alpha = 1$.

Since the optimal strategy $\bar{A}$ is not known a priori the formulation of the transit route choice model requires the introduction of binary variables x_a defined as

$$x_a = \begin{cases} 0 & \text{if } a \notin \bar{A}, \\ 1 & \text{if } a \in \bar{A}. \end{cases} \tag{3}$$

The total waiting time at a node i, w_i, is then

$$w_i = \frac{\sum_{a \in A_i^+} v_a}{\sum_{a \in A_i^+} x_a f_a}, \; i \in I. \tag{4}$$

Thus, assuming that transit passengers seek to minimize the expectation of their travel and wait times, the transit route choice model is formulated as

$$\text{Min} \sum_{a \in A} t_a v_a + \sum_{i \in I} w_i \tag{5}$$

subject to

$$\sum_{a \in A_i^+} v_a - \sum_{a \in A_i^-} v_a = g_i, \; i \in I, \tag{6}$$

$$v_a = x_a f_a w_i, \; a \in A_i^+, \; i \in I, \tag{7}$$

$$v_a \geq 0, \; a \in A, \tag{8}$$

$$x_a \in \{0,1\}, \; a \in A. \tag{9}$$

The constraints (7) may be relaxed to

$$v_a \leq f_a w_i, \; a \in A_i^+, \; i \in I. \tag{10}$$

The resulting linear programming model may be solved by an algorithm of polynomial complexity (see Spiess (1984), Spiess and Florian (1989)).

This algorithm was implemented in the EMME/2 software package (INRO (1992)) which is widely used in North America, Europe, Asia, South America and Australia.

The objective that we have chosen for the formulation of the line frequency optimization is that of offering the best service to the transit passengers, while respecting a constraint on total expenditures (such as the total number of transit vehicles operated).

The decision variables are the line frequencies, f, which are assumed to be continuous variables. If the number of vehicles assigned to a line, or the headway, must be integer values, the solution obtained would have to be adjusted manually. The constraint on the total number of vehicles operated may be formulated as

$$t^{\mathrm{T}} f \leq N,$$

where t, $t > 0$, is the vector times of the transit lines and N is the number of available vehicles.

In order to simplify the presentation of the model, we consider a generalized transit network, such as described previously. Also we shall assume, for the simplicity of the exposition, that each line has only one segment and that there is only one destination. As we shall show later, the model may be easily extended for multiple segment transit lines and multiple destinations.

In principle, each transit line considered is represented by a boarding link, an in-vehicle link and an alighting link as illustrated in Fig. 3.

This representation may be simplified as follows. The node A0 and its two incident links may be replaced by a single link N0–A1 with attributes $(0 + t_l, f_l)$. The node A1, the link A1–N1 and the new link N0–A1 may be replaced by a single link N0–N1 which has as attributes $(t_l + 0, f_l)$. Figure 4 illustrates this simplified representation.

The initial formulation of the frequency optimization model is:

$$\text{Min} \sum_{a \in A} t_a v_v + \sum_{i \in I} w_i \tag{11}$$

subject to

$$\sum_{a \in A_i^+} v_a - \sum_{a \in A_i^+} v_a = g_i, \; i \in I, \tag{12}$$

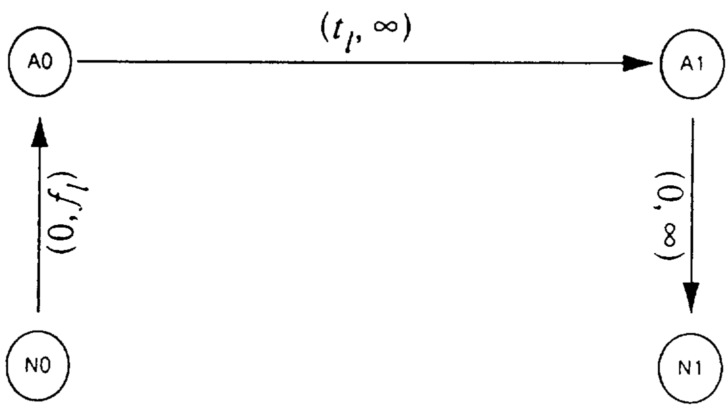


Fig. 3. The representation of a one segment transit line.

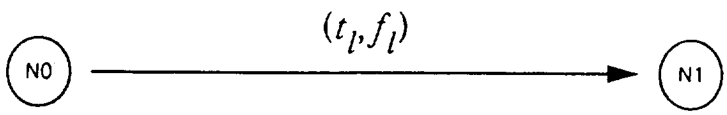


Fig. 4. Simplified representation of a one segment transit line.

$$v_a = x_a f_a w_i,\ a \in A_i^+,\ i \in I, \tag{13}$$

$$v_a \geq 0,\ a \in A, \tag{14}$$

$$x_a \in \{0,1\},\ a \in A, \tag{15}$$

$$\sum_{a \in A} t_a f_a \leq N, \tag{16}$$

$$f_a \geq \underline{f}_a,\ a \in A. \tag{17}$$

This is a nonlinear programming problem with binary variables. The variables of the problem (11)–(17) may be subdivided into those which are controlled by the supplier of the transit services, the frequencies, f, and those which are determined by the travellers' route choice, x, v and w. We will say that $f \in F$ if the constraints (16) and (17) are satisfied, and that $x,v,w \in V$ if the constraints (12), (14) and (15) are satisfied. In this form, the problem appears to be intractable since the objective function is neither convex nor concave and there are binary variables in the constraints. The natural subdivision of the decision variables suggest reformulation of the model as

$$\min_{\substack{f \in F \\ x,v,w \in V}} \sum_{a \in A} t_a v_a + \sum_{i \in I} w_i \tag{18}$$

subject to

$$v_a = x_a f_a w_i,\ a \in A_i^+,\ i \in I. \tag{19}$$

Geoffrion (1969) suggests the transformation of problems of this type by *projection on the space of the external variables* (the frequencies in this model). By carrying out this projection, the problem becomes

$$\min_{f \in F} \left[\min_{x,v,w \in V} \sum_{a \in A} t_a v_a + \sum_{i \in I} w_i \text{ subject to } v_a = x_a f_a w_i,\ a \in A_i^+,\ i \in I \right]. \tag{20}$$

More explicitly, it takes the form of a *bi-level programming problem*

$$\min_{f} \left\{ \min_{x,v,w} \sum_{a \in A} t_a v_a + \sum_{i \in I} w_i \right\} \tag{21}$$

subject to (12)–(15): subject to (16) and (17).

This reformulation appears to be natural in our context. The external problem corresponds to the decision of the supplier of transit services: to determine the optimal frequencies subject to the fleet size constraint and lower bounds, in order to minimize the total travel and waiting time. The internal, or lower level problem, is a transit route choice problem with *fixed frequencies*. Hence, it is now possible to eliminate the binary variables by relaxing the constraint (13). The resulting problem is

$$\operatorname*{Min}_{f}\left\{\operatorname*{Min}_{v,w} \sum_{a\in A} t_a v_a + \sum_{i\in I} w_i\right\} \tag{22}$$

subject to

$$\sum_{a\in A_i^+} v_a - \sum_{a\in A_i^-} v_a = g_i,\ i\in I, \tag{23}$$

$$v_a \le f_a w_i,\ a\in A_i^+,\ i\in I, \tag{24}$$

$$v_a \ge 0,\ a\in A, \tag{25}$$

subject to

$$\sum_{a\in A} t_a f_a \le N \tag{26}$$

$$f_a \ge \underline{f}_a,\ a\in A. \tag{27}$$

SOME MATHEMATICAL PROPERTIES OF THE MIN–MIN FORMULATION

The bi-level programming formulation (22)–(27) has two special properties. The outer and inner level problems share the same objective function. Furthermore, it is a Min–Min formulation which is easier to solve than the initial single level program (11)–(17). However, the results available from the study of Min–Max problems may be easily adapted for devising an algorithm for its solution. Descent directions may be obtained for the outer problem by solving the inner level problem.

Let $\varphi(x)$, $x\in\mathbb{R}^n$, be the function to be minimized. The direction γ is a descent direction if the directional derivative

$$D_\gamma\varphi(x) = \gamma\cdot\nabla\phi(x) = \sum_{i=1}^{n}\gamma_i\frac{\partial\phi(x)}{\partial x_i} < 0. \tag{28}$$

Since the objective function (22) is not continuously differentiable, we adapt Danskin's (1967) results derived for Max–Min problems which state that the directional derivative of φ in the direction $\gamma\left(\sum_{i=1}^{n}\gamma_i^2 = 1\right)$ at x, exists and is given by

$$D_\gamma\varphi(x) = \operatorname*{Min}_{y\in\hat{Y}(x)}\sum_{i=1}^{n}\gamma_i\frac{\partial\mathscr{F}(x,y)}{\partial x_i}, \tag{29}$$

where $\varphi(x) = \mathrm{Min}_{y\in Y}\mathscr{F}(x,y)$, Y is a compact topological space, $\mathscr{F}(x,y)$ is a continuous function in (x,y) with continuous partial derivatives $\dfrac{\partial\mathscr{F}(x,y)}{\partial x_i}$. $\hat{Y}(x)\subseteq Y$ is the set of optimal solutions of the inner problem for a given x.

For the problem that we consider, the function φ is defined as

$$\varphi(f) = \operatorname*{Min}_{v,w\in W}\mathscr{F}(f,v,w) = \operatorname*{Min}_{v,w\in W}\sum_a t_a v_a + \sum_{i\in I} w_i, \tag{30}$$

where W is defined by the constraints of the inner problem. $\mathscr{F}$ is continuous in v and w, however it is not an *explicit* function of the frequencies f. In order to obtain the partial derivatives $\partial\mathscr{F}/\partial f_a$ we make use of the sensitivity analysis of the inner problem. The dual of this problem is

$$\text{Max} \sum_{i \in I} g_i u_i \tag{31}$$

subject to

$$u_j + t_a + \mu_a \geq u_i,\ a = (i, j) \in A, \tag{32}$$

$$\sum_{a \in A_i^+} f_a \mu_a = 1,\ i \in I, \tag{33}$$

$$\mu_a \geq 0,\ a \in A. \tag{34}$$

The dual variables $-u_i$ are associated with the conservation of flow constraints (6) and μ_a to the flow subdivision constraints (10). The optimality conditions for the primal-dual pair of problems are the primal feasibility (6), (8), (10), the dual feasibility (31)–(33) and the complementary slackness conditions

$$\mu_a(v_a - f_a w_i) = 0,\ a \in A_i^+,\ i \in I, \tag{35}$$

$$v_a(u_i - u_j - t_a - \mu_a) = 0,\ a = (i, j) \in A. \tag{36}$$

Given an optimal primal solution $\hat{v}$, $\hat{w}$, the dual optimal variables $\hat{u}$, $\hat{\mu}$ may be constructed as follows (Spiess (1984)):

$$\hat{v}_a = \begin{cases} f_a \hat{w}_i \text{ if } a \in \bar{A}_i^+,\ i \in I, \\ 0 \quad \textit{if } a \notin \bar{A}; \end{cases} \tag{37}$$

$$\hat{u}_i = \begin{cases} \dfrac{1 + \sum_{a \in \bar{A}_i^+} f_a(\hat{u}_j + t_a)}{\sum_{a \in \bar{A}_i^+} f_a} & \text{if } i \in I \backslash \{r\}, \\ 0 & \text{if } i = r; \end{cases} \tag{38}$$

$$\hat{\mu}_a = \begin{cases} \hat{u}_i - \hat{u}_j - t_a \text{ if } a = (i,j) \in \bar{A}, \\ 0 \qquad\qquad \text{if } a \notin \bar{A}. \end{cases} \tag{39}$$

The partial derivatives may be computed by using the following proposition.

Proposition 1

Let $(\hat{v},\hat{w})$ be an optimal primal solution of the inner problem and $(\hat{u},\hat{\mu})$ the corresponding dual optimal solution. The partial derivatives of the objective function of the inner problem, $\mathscr{F}(f,\hat{v},\hat{w})$ are given by

$$\frac{\partial \mathscr{F}(f,v,w)}{\partial f_a} = -\hat{\mu}_a \hat{w}_i,\ a \in A_i^+,\ i \in I. \tag{40}$$

Proof

See Constantin (1993).

These partial derivatives are continuous and uniformly bounded since $\hat{\mu}_a$ are the dual variables of a feasible linear program. The transit route choice problem (5)–(9) admits a feasible solution if the network $G = (N,A)$ is strongly connected and the frequencies are nonzero, that is $f_a > 0$, $a \in A$. The partial derivatives may be computed while the inner problem is being solved with very little computational overhead. Hence, it follows that the direction of descent for the outer problem is

$$d = -\underline{\nabla}\varphi = [-\partial\mathscr{F}(f,\hat{v},\hat{w})/\partial f_a,\ a \in A] = [\hat{\mu}_a \hat{w}_i,\ a \in A_i^+,\ i \in I],$$

where $(\hat{v},\hat{w}) \in W(f)$ is an optimal solution of the inner problem. This direction has an intuitive interpretation. It is nonnegative, by the nonnegativity of μ_a (39) and v_a (37) as well as the definition of w_i (4). Thus, in order to decrease the total transport time, one should increase all the frequencies of all the lines which carry flow.

The direction $-\underline{\nabla}\varphi(f)$ must be projected on the *active constraints* at f, in order to determine if the current solution may be improved upon. We adapt the *gradient projection* method (Rosen (1960)) and present it in a simplified form based on Luenberger's (1984) text.

The direction opposite to the gradient is projected on the subspace defined by the active constraints, which are all the equality constraints and the inequality constraints which are satisfied as equality in the current solution. If the projected direction $d \neq 0$, a step carried out in this direction decreases the objective function. If the projected direction $d = 0$, a best solution was found for the current set of active constraints. If all the *multipliers* associated to the active constraints are nonnegative, $\lambda \geq 0$, the current solution satisfies necessary conditions for optimality. Otherwise, the objective function may be further decreased by removing, from the set of active constraints, a constraint which has a corresponding negative multiplier.

For the problem which we consider, the active constraints includes a subset of the lower bound constraints and the fleet size constraint, $\sum_{a \in A} t_a f_a = N$. When this constraint is active, the direction $-\underline{\nabla}\varphi$ is not feasible and it must be projected. The following proposition provides the analytical expression of the projected direction as well as that for the multipliers associated with the active constraints.

Proposition 2

Consider the set of active constraints

$$\sum_{a \in A} t_a f_a = N, \tag{41}$$

$$-f_a = -\underline{f}_a, \; a \in \underline{A}. \tag{42}$$

The direction obtained by the projection of the gradient on the subspace of these constraints is

$$d = [d_a, a \in A] \text{ where } d_a = \begin{cases} -\underline{\nabla}\varphi_a - \rho t_a & \text{if } a \notin \underline{A}, \\ 0 & \text{if } a \in \underline{A}, \end{cases} \tag{43}$$

and the multipliers associated with the active constraints are

$$\lambda = \begin{bmatrix} \rho \\ \underline{\nabla}\varphi_a + \rho t_a, \; a \in \underline{A} \end{bmatrix}, \tag{44}$$

where

$$\rho = -\frac{\sum_{a' \notin \underline{A}} t_{a'} \underline{\nabla}\varphi_{a'}}{\sum_{a' \notin \underline{A}} t_{a'}^2}. \tag{45}$$

Proof

See Constantin (1993).

We can now proceed to state the proposed solution algorithm for solving (22)–(27).

THE ALGORITHM FOR SOLVING THE MIN-MIN PROBLEM

We state the algorithm in a general form, before giving its interpretation and discussing its convergence.

Line Frequency Optimization Algorithm

Step 0. Initialization

- Determine a set of feasible line frequencies: $f^0 \leftarrow f \in F$.
- Initialize the iteration count: $k \leftarrow 0$.

Step 1. Search for a descent direction

- Solve the transit route choice problem: $\hat{v}(f^k),\hat{w}(f^k)$.
- Compute a (sub)gradient based on the optimal strategy: $\underline{\nabla}(f^k,\hat{v}(f^k),\hat{w}(f^k))$.
- Compute the projected direction: $d^k \leftarrow -P^k\underline{\nabla}(f^k,\hat{v}(f^k),\hat{w}(f^k))$.
- If $|d^k| > \varepsilon_0$: go to Step 2; otherwise go to Step 3.

Step 2. Computation of the new line frequencies

- Compute the *maximal* step size: $\bar{\alpha}^k$.
- Compute the *optimal* step size $\hat{\alpha}^k$: $0 \leq \hat{\alpha}^k \leq \bar{\alpha}^k$.
- Compute new line frequencies: $f^{k+1} \leftarrow f^k + \hat{\alpha}^k d^k$.
- Update the iteration count: $k \leftarrow k + 1$.
- Return to Step 1.

Step 3. Stopping test

- Compute the multipliers associated to the active constraints: λ^k.
- If $\lambda^k \geq 0$, then the line frequencies f^k correspond to a stationary point: STOP.
- If $\lambda^k \not\geq 0$, then remove from the active set the constraint with the most negative multiplier: recompute the projected direction d^k, and return to Step 2.

Given a projected gradient direction d, the maximal step size $\bar{\alpha}^k$ is determined by the expression

$$\bar{\alpha}^k = \min_{\substack{a \notin A \\ d_a^k < 0}} \left\{ \frac{\underline{f}_a - f_a^k}{d_a^k} \right\}, \tag{46}$$

which ensures that the lower bounds on the frequencies are not violated. Note that the algorithm generates only frequencies such that all the resources available are used. This is justified by the fact that the fleet size constraint is active for any local minimum of the function $\varphi(f)$ (as is proved in Constantin (1993)).

The computation of the optimal step size in Step 2 is carried out with Armijo's (1966) rule, which requires α not to be *too large*, i.e.

$$\phi(\alpha) \leq \phi(0) + \varepsilon\phi'(0)\alpha, \tag{47}$$

and not *too small*, i.e.

$$\phi(\eta\alpha) > \phi(0) + \varepsilon\phi'(0)\eta\alpha, \tag{48}$$

for $0 < \varepsilon < 1$ and $\eta > 1$.

In order to evaluate $\phi(\alpha)$ exactly, we would have to solve the inner problem for each value of α. In order to avoid this computational burden, $\phi(\alpha)$ is evaluated by assuming that the optimal strategy computed for the current inner problem does not change. It may be shown (see Constantin (1993)) that this procedure always ensures that the $\hat{\alpha}^k$ found results in a decrease of the objective function.

The algorithm may be interpreted as the sequence of actions taken by the supplier of transit services and the travellers: the supplier provides a set of line frequencies, the travellers react to these frequencies and the supplier uses the information on the travellers' choice to improve the services offered.

As proven in Constantin (1993), this algorithm converges to a *stationary point* of the line frequency optimization problem. This is due to the fact that the objective function of the outer problem is neither convex nor concave. However, as Rosen (1960) pointed out, such stationary points are likely to be *local minima* since they are found by a descent method.

A GENERAL FORMULATION OF THE LINE FREQUENCY OPTIMIZATION MODEL

The basic model presented in the previous sections of the paper may be easily generalized to transit lines that consist of multiple segments, to networks with multiple destinations and to vehicle fleets that consist of several types of vehicles.

In order to formulate the more general model we introduce

$$\delta_{al} = \begin{cases} 1 \text{ if link } a \text{ belongs to line } l; \\ 0 \text{ otherwise.} \end{cases}$$

Hence, the frequency of a link a is

$$f_a = \sum_{l \in L} \delta_{al} f_l, \; a \in A', \tag{49}$$

where A' is the set of boarding links. The travel time of a line l, t_l, is the sum of its in-vehicle links

$$t_l = \sum_{a \in A} \delta_{al} t_a, \; l \in L.$$

The corresponding waiting times w_i are given by

$$w_i = \frac{\sum_{a \in A_i'^+} v_a}{\sum_{a \in A_i'^+} x_a f_a}, \; i \in I.$$

Let $\mathscr{V}$ be the set of vehicle types available and L_v the subset of lines which use vehicle v. We assume that each transit line is assigned a single vehicle type. By letting g_i^r be the demand from node i to destination r, v_a^r be the flow of link a to destination r and w_i^r, the waiting time at node i to destination r, the more general formulation is stated below.

$$\underset{f}{\text{Min}} \; \underset{v,w}{\text{Min}} \sum_{r \in R} \left(\sum_{a \in A} t_a v_a^r + \sum_{i \in I} w_i^r \right) \tag{50}$$

subject to

$$\left.\begin{aligned} &\sum_{a \in A_i^+} v_a^r - \sum_{a \in A_i^-} v_a^r = g_i^r, \; i \in I \\ &v_a^r \le \left(\sum_{l \in L} \delta_{al} f_l \right) w_i^r, \; a \in A_i'^+, \; i \in I \\ &v_a^r \ge 0, \; a \in A \end{aligned}\right\} r \in R \tag{51}$$

subject to

$$\left.\begin{aligned} &\sum_{l \in L_v} t_l f_l \le N_v \\ &f_l \ge \underline{f}_l, \; l \in L_v \end{aligned}\right\} v \in \mathscr{V}. \tag{52}$$

Fortunately, the structure of the problem remains unchanged and only the computation of the gradient, its projection and the line search require modifications. All the theoretical results hold, although the proofs are more evolved.

COMPUTATIONAL RESULTS

The algorithm presented in Section 5 was implemented as a nonstandard module of the EMME/2 software package and numerical experimentation was carried out with the transit networks of the cities of Stockholm (Sweden), Winnipeg (Canada) and Portland (U.S.A.).

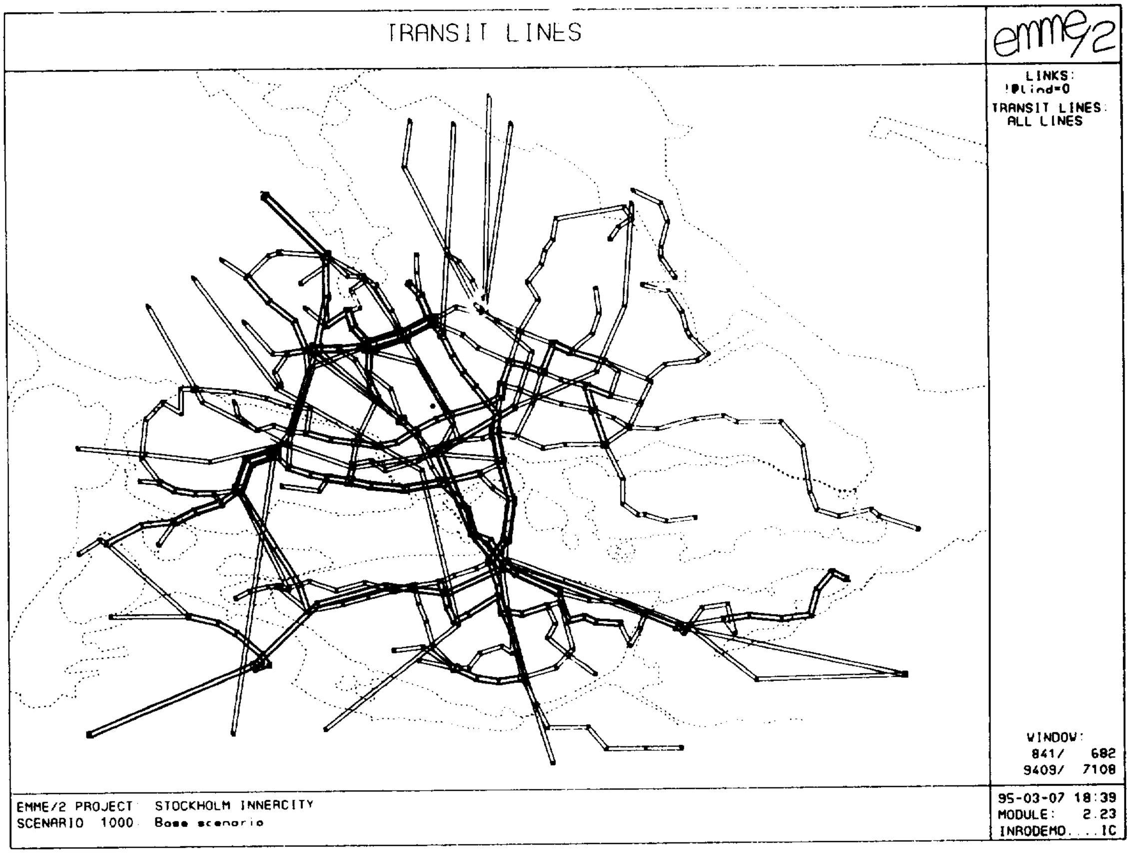


Fig. 5. The transit network of Stockholm (Sweden).

The Stockholm network has 31 bus lines, 4 metro lines, 1 suburban rail line and 2 other lines which are relatively old suburban rail lines. The fleet is composed of 6 vehicle types, including three different types of buses. The Winnipeg network consists of 67 bus lines that operate with two different types of buses. Most are regular buses and some are smaller buses which are used to operate feeder lines. The Portland network consists of 112 bus lines and 3 light rapid transit (LRT) lines. The bus fleet is composed of regular buses (70%) and articulated buses (30%). The transit networks of these cities are shown in Figs 5, 6 and 7.

The computational results depend on the initial solution provided to start the frequency optimization algorithm. We tried the following three initial solutions: the *actual* line frequencies; all lines that use the same vehicle type are assigned the *same frequency*; all lines that use the same vehicle are assigned the *same number of vehicles*.

Figure 8 illustrates the reduction in the mean travel time vs the number of iterations for these three networks. The empirical convergence rate seems to be slightly influenced by the initial solution. A 'better' initial solution requires less iterations to reach the local optimum found. However, the 'best' initial solution is not the same one for the three networks.

Table 2 indicates the influence of the initial solution on the mean travel time before and after the optimization (T^0, T^K, where K is the number of iterations carried out). While the difference between the mean travel times differ, for the three initial solutions chosen, the value of the mean travel times after the optimization indicates that the algorithm converges either to the same local minimum or to local minima which are almost equivalent in terms of this criterion.

CONCLUSION

In this paper, we have presented a model for the transit line frequency optimization problem. This model seems particularly sound because it takes into account, in an explicit and realistic way, the

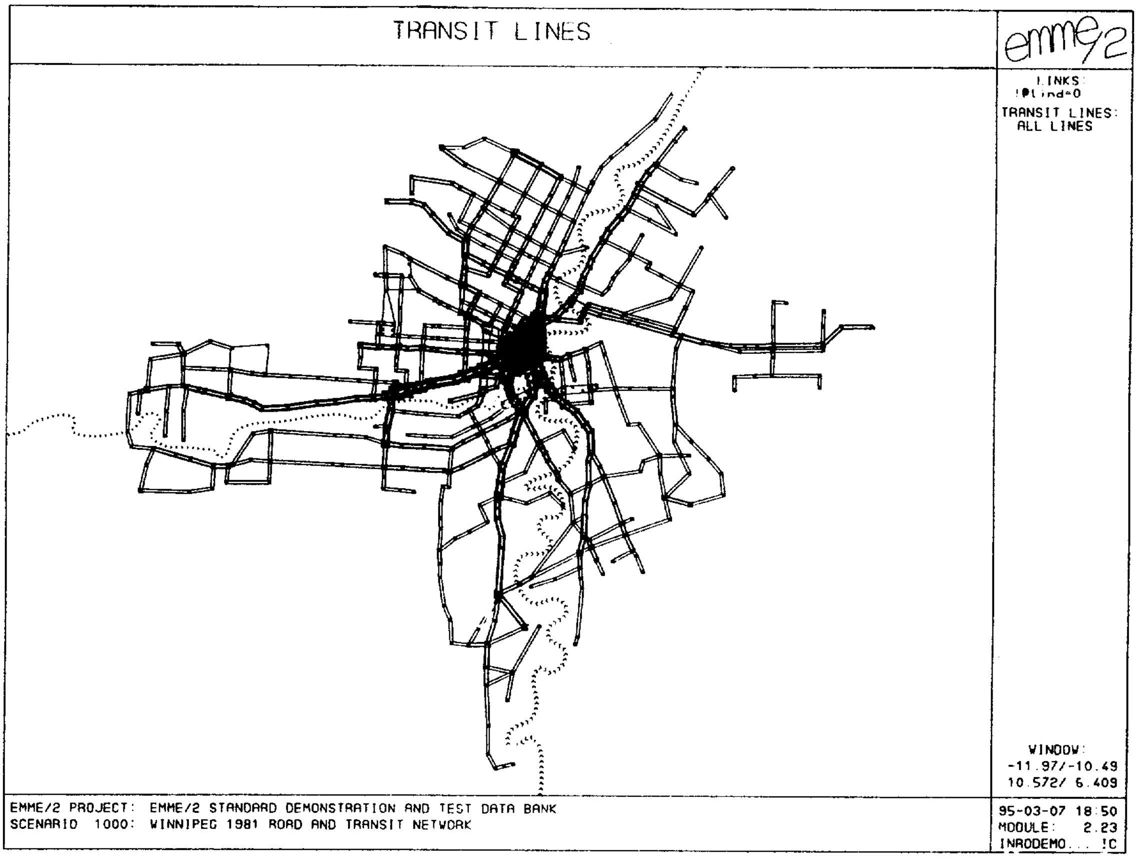


Fig. 6. The transit network of Winnipeg (Canada).

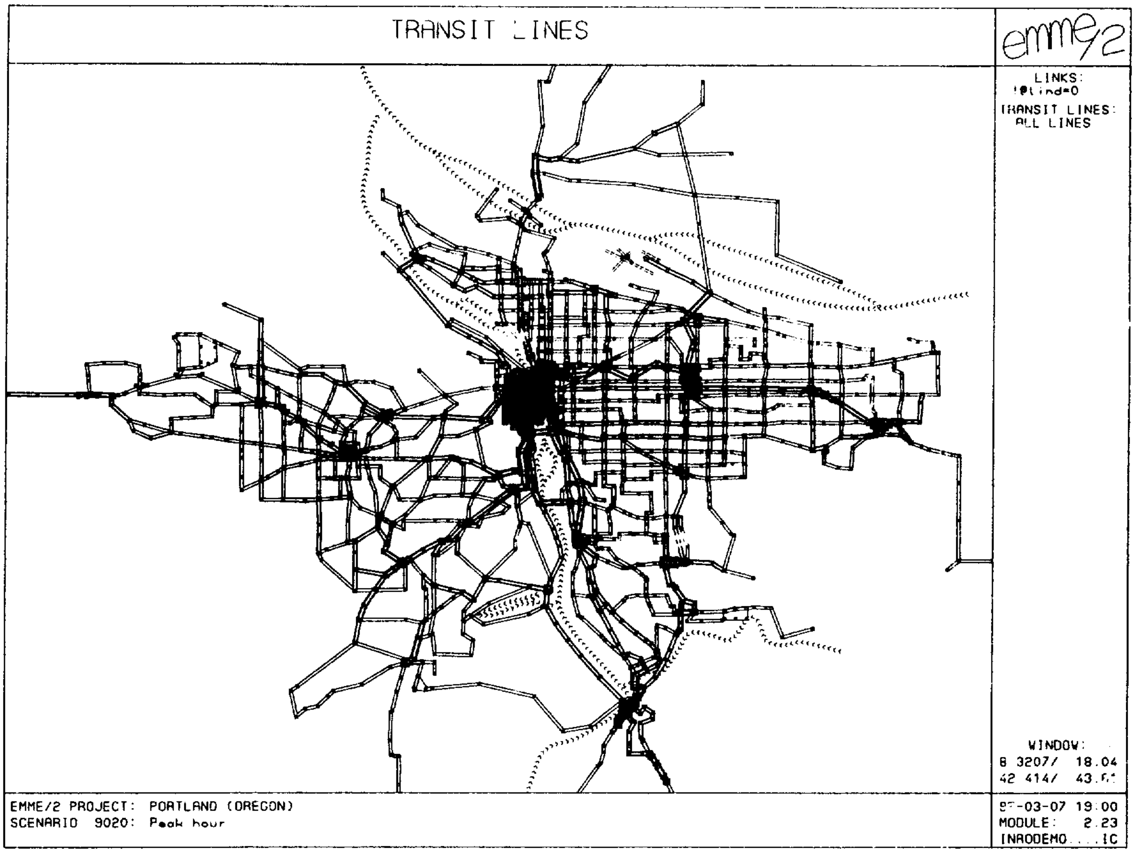
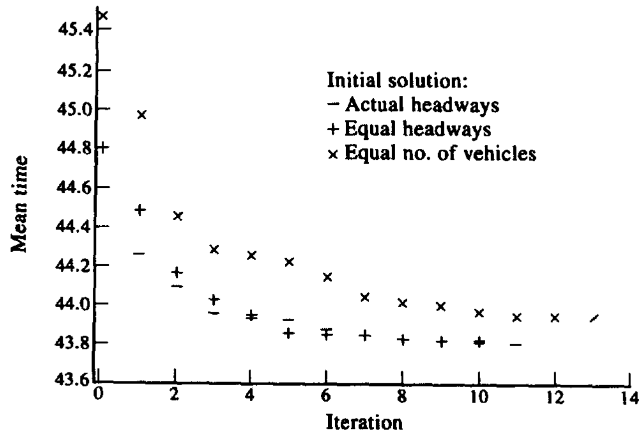
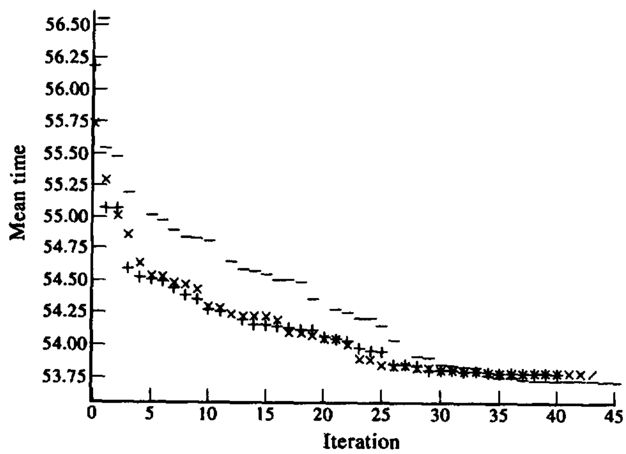


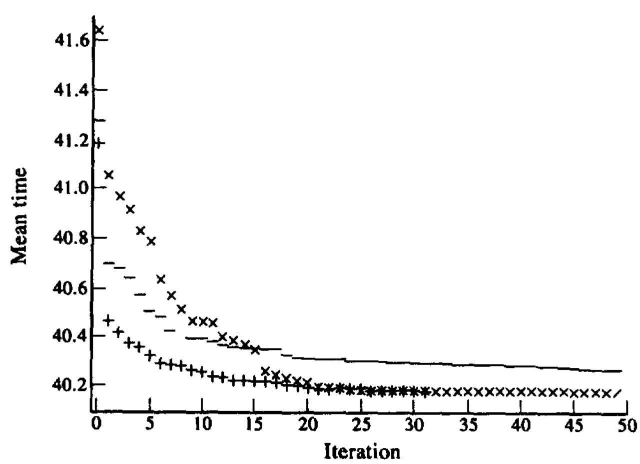
Fig. 7. The transit network of Portland (U.S.A.).



(a)



(b)



(c)

Fig. 8. Reduction in the mean travel time vs the number of iterations for (a) the Stockholm, (b) the Winnipeg and (c) the Portland networks.

Table 2. Influence of the initial solution value on the mean travel times

	Stockholm			Winnipeg			Portland		
Initial solution	T^0	T^K	K	T^0	T^K	K	T^0	T^K	K
Actual headway	44.36	43.82	11	56.54	53.72	45	41.28	40.27	50
Equal headway	44.80	43.83	10	56.18	53.77	40	41.19	40.18	31
Equal number of veh.	45.56	43.95	13	55.73	53.77	43	41.64	40.18	50

route choice that travellers face on a transit network. We have also presented an efficient algorithm for solving the proposed model that can be applied to real, large-scale transit networks. The results obtained on three such networks are very promising.

The model described in this paper may be used for short-term planning, when the fleet size by vehicle type is fixed and known. However, the model may also be used for medium term planning, by considering various envisaged values for the fleet size and studying the resulting optimal frequencies for each of these values, such as in parametric analysis.

The model has also been extended to take into account the capacity of vehicles. Essentially, the capacity of the vehicles is considered to be a constraint for the supplier of the transit services. That is, the supplier must ensure that there is sufficient capacity to carry all the demand without exceeding this capacity. The resulting model is more complex than the one described in this paper. It may be solved (see Constantin (1993)), by an adaptation of the same (sub)gradient projection method. This extension of the model is the topic of a forthcoming paper.

REFERENCES

Armijo, L. (1966). Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, Vol. 16, pp. 1–3.

Byrne, B. F. (1975). Public transportation line positions and headways for minimum user and system cost in a radial case. *Transportation Research*, Vol. 9, pp. 97–102.

Byrne, B. F. (1976). Cost minimizing positions, lengths and headways for parallel public transit lines having different speeds. *Transportation Research*, Vol. 10, pp. 209–214.

Constantin, I. (1993). L'optimisation des fréquences d'un réseau de transport en commun. Ph.D. thesis, Département d'informatique et de recherche opérationnelle, Centre de recherche sur les transports, Université de Montréal, Montréal (Canada), Publication 881.

Danskin, J. M. (1967). *The Theory of Max–Min and its Application to Weapons Allocation Problems.* (Econometrics and Operations Research VI). New York: Springer.

Dubois, D., Bel, G. & Llibre, M. (1979). A set of methods in transportation network synthesis and analysis. *Journal of Operational Research Society*, Vol. 30, pp. 797–808.

Furth, P. G. & Wilson, N. H. M. (1981). Setting frequencies on bus routes: theory and practice. *Transportation Research Record*, Vol. 818, pp. 1–7.

Geoffrion, A. M. (1969). *Elements of Large-scale Mathematical Programming.* Rapport R-481-PR. Santa Monica, CA: Rand, p. 14–17.

Grega, W. & Jörnsten, K. O. (1979). *A Solution Method for the Bus Allocation Problem.* Rapport technique LiTH-MAT-R-79-43. Linköping (Sweden): Linköping Institute of Technology.

Hasselström, D. (1981). *Public transportation planning: a mathematical programming approach.* Ph.D. thesis, University of Göteborg, Göteborg (Sweden).

Holroyd, E. M. (1967). The optimum bus service: a theoretical model for a large uniform urban area. In *Vehicular Traffic Science*, pp. 308–328.

INRO Consultants Inc. (1992). *EMME/2: User's Manual.* Montréal (Québec): INRO.

Jewell, W. S. (1969). Allocation of route service in a transportation network. *Information Processing*, Vol. 68, pp. 1468–1470. Amsterdam: North-Holland.

Lampkin, W. & Saalmans, P. D. (1967). The design of routes, service frequencies and schedules for a municipal bus undertaking: a case study. *Operations Research Quarterly*, Vol. 18, pp. 375–396.

LeBlanc, L. J. (1988). Transit system network design. *Transportation Research B*, Vol. 22, pp. 383–390.

Luenberger, D. G. (1984). *Linear and Nonlinear Programming.* Reading, MA: Addison-Wesley.

Mandl, C. E. (1980). Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research*, Vol. 5, pp. 396–404.

Mizokami, S. & Kawakami, S. (1986). Method for determining bus schedules under demand-performance equilibrium. Research for Tomorrow's Transport Requirements. In *Proceedings of World Conference on Transportation Research*, Vol. 2, pp. 1713–1728. Vancouver (Canada).

Rosello, X. (1976). An heuristic algorithm to generate an urban buses network. Advances in Operations Research. In M. Roubens (ed.) *Proceedings of EURO II Conference.* Stockholm (Sweden): North-Holland, pp. 409–419.

Rosen, J. B. (1960). The gradient projection method for nonlinear programming. Part I. linear constrains. *Journal Society Industrial Applied Mathematics*, Vol. 8, pp. 181–217.

Schéele, S. (1977). A mathematical programming algorithm for optimal bus frequencies. Ph.D. thesis, Linköping Institute of Technology, Linköping (Sweden).

Silman, L. A., Barzily, Z. & Passy, U. (1974). Planning the route system for urban buses. *Computer & Operations Research*, Vol. 1, pp. 201–211.

Spiess, H. (1984). Contributions à la théorie et aux outils de planification des réseaux de transport urbain. Ph.D. thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal (Canada).

Spiess, H. & Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. *Transportation Research B*, Vol. 23, pp. 83–102.

Stephanedes, Y. J. & Kwon, E. (1988). Optimization strategies for transit systems in urban corridors. *Transportation Research Record*, Vol. 1165, pp. 75–85.

Talley, W. K. (1989). Optimization of bus frequency and speed of service: a system approach. *Logistics and Transportation Review*, Vol. 25, pp. 139–156.