Abstract: In this paper some mathematical programming models are exposed in order to set the number of services on a specified system of bus lines, which are intended to assist high demand levels which may arise because of the disruption of Rapid Transit services or during the celebration of massive events. By means of this model two types of basic magnitudes can be determined, basically: a) the number of bus units assigned to each line and b) the number of services that should be assigned to those units. In these models, passenger flow assignment to lines can be considered of the system optimum type, in the sense that the assignment of units and of services is carried out minimizing a linear combination of operation costs and total travel time of users. The models consider delays experienced by buses as a consequence of the get in/out of the passengers, queuing at stations and the delays that passengers experience waiting at the stations. For the case of a congested strategy based user optimal passenger assignment model with strict capacities on the bus lines, the use of the method of successive averages is shown.

Key–Words: public transportation modeling, congestion, mathematical programming

1 Introduction

Setting properly the required services to attend transport demand taking into account available resources in urban public transportation networks is a key aspect in order to keep their good performance as well as to ensure users confidence in public transportation as a valid alternative.

Models for the overall design of transit networks or simply for some management aspects of public transport lines which take into account demand in the design process, have an intrinsic relationship with passenger transit assignment models. Such assignment models can be classified in a first approach by two criterions: a) static or dynamic and b) frequency based or time table based. Within the classical passenger transit assignment models under the concept of strategy, the classical work in [12] must be cited. This initial model is unable to take into account congestion in public transportation systems. It has not been until very recently that these strategy-based models have been able to reflect how effective frequencies may be altered by congestion ([5], [1], [8]). Frequency setting models have been formulated using transit assignment schemas based on strategies and time table based. Using a strategy based assignment model under a static approach, the works in [6] and in [10] must be taken into account. For the case of lines under strict time table, assignment models that must be cited are those implemented in the commercial package EMME and others very recently developed such as those in [7] and in [11]. In this paper two service setting models in [3] and in [4] are described which are able to reflect the effects of congestion under a static approach for public transportation lines intended for emergency situations or for supporting special events. In these two models the underlying passenger assignment schema is a non-strategy based user optimum and a congested shortest route choice, respectively. Also the formulation in variational inequalities in [2] for the congested transit assignment model in [5] is briefly described and some numerical results are presented for a variant of the model with sharp capacity constraints.

2 Notation and network model

In this section a unified notation is presented for all the models under discussion. The transit network is represented by means of a directed graph \( G = (N,A) \), where \( N \) is the set of nodes and \( A \) is the set of links. The number of trips from origin \( p \) to destination \( q \) will be denoted by \( g_{p,q} \). By \( C \subseteq N \) it will be denoted the subset of nodes representing centroids or trip attraction/generation points. Nodes in \( C \) may have a role of origin zones where a trip starts and/or simultaneously the role of destinations where a trip ends. By \( W = \{ (p,q) \in C \times C \mid g_{p,q} > 0 \} \) it is denoted the set of active origin-destination pairs \( \omega = (p,q) \).
on the network. The set of destinations in the network for a given origin $p \in C$ shall be denoted by $Q(p) = \{ q \in C \mid \exists (p, q) \in W \}$ and the set of origin nodes for a fixed destination $q \in C$ shall be denoted by $P(q) = \{ p \in C \mid (p, q) \in W \}$. For a node $i \in N$, the set of emerging links will be denoted by $E(i)$ and the set of incoming links by $I(i)$. The representation of transit lines will be done in the form of an expanded network, as in [12] (see figure 1 below).

Figure 1: Internal representation of the transit network assumed in the model.

By $v^d_a$ it will be denoted the flow at link $a \in A$ corresponding to the origin-destination pair $\omega \in W$. Then the following notation will be used for the various types of vector flows and origin-destination volumes:

- $v^d_i = (\ldots, v^d_1, \ldots; a \in E(i)) \in \mathbb{R}^{I(E(i))}$, $i \in N$, $d \in D$ is the vector of flows with destination $d$ at emerging links of node $i$.
- $v^d_i = \sum_{a \in E(i)} v^d_a$ is the total inflow through node $i \in N$ with destination $d \in D$.
- $v^d = (\ldots, v^d_1, \ldots; i \in N) \in \mathbb{R}^{I[A]}$, $d \in D$. $v = (\ldots, v^d_1, \ldots; d \in D) \in \mathbb{R}^{I[D]}$.
- $v = \sum_{d \in D} v^d \in \Pi^{I[A]}$. Vector of total flows on links and $v_a = \sum_{d \in D} v^d_a$, $a \in A$.
- $g^d = (\ldots, g^d_1, \ldots; i \in O(d)) \in \mathbb{R}^{I[D]}$, $d \in D$. $g = (\ldots, g^d_1, \ldots; d \in D) \in \Pi^{I[W]}$.

The feasibility set for the congested transit equilibrium problem can be formulated as $V = \bigotimes_{d \in D} V^d$, being each set $V^d$ defined as:

$$V^d = \left\{ v^d \in \mathbb{R}^{I[A]} \mid \sum_{a \in E(i)} v^d_a - \sum_{a \in I(i)} v^d_a = g^d_a, \quad \sum_{a \in I(d)} v^d_a = \sum_{i \in O(d)} g^d_i, v^d_a = 0, i \in N_d, \forall a \in E(d) \right\}$$

The polyhedron of total link flows $v$ is $V = \{ v \in \mathbb{R}^{I[A]} \mid v = \sum_{d \in D} v^d, \forall d \in V^d \}$. Because of the finite capacity of vehicles, boarding of passengers may not happen at the first arriving vehicle seen by the passenger. Mean waiting times for a boarding, or inverse of effective frequencies, shall be denoted by $\sigma_a(\cdot) = 1/f_a(\cdot)$. Travel times on links are given by functions $t_a(v)$, $a \in A$ which are finite on $V$. The subset of nodes for which emerging links exist with a finite efft on $V$ will be denoted by $N = \{ i \in N \mid \exists a \in E(i), f_a(\cdot) < \infty \}$. The sets $\bar{N}_d = N \setminus \{ d \}$, $d \in D$ and $\bar{A} = \{ a \in A \mid \exists i \in N, a \in E(i) \}$ will be also used. For nodes $i \in N$, the subset of emerging links with finite effective frequency will be denoted by $E(i)$. Line segments as well as pedestrian, transfer and non transit facilities shall be represented by links $a \in A$ with either constant or flow dependent travel time functions $t_a(\cdot)$ and infinite frequencies, $f_a = +\infty$. This apply also for links $a \in I(i)$, $i \in N$, representing alighting at stops.

3 Frequency setting models

3.1 A user equilibrium based service setting model

The first model by Codina and Marín [3], model [SUE] below, is oriented to set the number of services when passengers have a behavior characterized by two facts: a) no recommendation or regulation is made on the assignment from passengers to lines b) at each stop they choose a transit line accordingly to a route from their origin to their destination that they consider as optimal. The design model can be stated as a bilevel programming in which the lower level is an asymmetric traffic assignment problem. Asymmetries in costs come from the fact that passenger delays at stations waiting for a bus line to arrive depend not only on passenger’s flow arriving at the station to board on that line but also on the unit’s occupancy of that line arriving at the station. The upper level objective function is composed by two terms. The first one evaluates the operational costs of assigning units to a line plus the operational costs of bus services. The second cost is proportional to the total time spent by all passengers. The coefficient $\theta$ can be considered as the social cost of time.

In the formulation of model [SUE] below, $S^*(z)$ is the solution set of an asymmetric traffic model that can be stated as a variational inequality (V.I.): Find $v^* \in V$ so that $T(v^*, z)^T (v - v^*) \geq 0$, $\forall v \in V$. This V.I. which makes up the lower level problem, is parametrized by the number of services $\zeta^\ell$ assigned at each bus line $\ell \in L$. The number of services plays
the role of a parameter for the links in the expanded network modeling passenger flows on line $\ell$.

**[SUE]**

Min $n, z, v$ 

$$ \sum_{\ell \in L} (\xi^n \ell + \gamma^z \ell) + \theta v^T T(v, z) $$

s.t. 

$$ v \in S^*(z) $$

A0-1 

$$ \sum_{\ell \in L} n^\ell \leq p, n^\ell \geq 0, n^\ell \in \mathbb{Z}, \ell \in L $$

A0-2 

$$ H^\ell n^\ell \geq \zeta^\ell C_t(v, z), \ell \in L $$

A0-3 

$$ 0 \leq z^\ell \leq \lambda^\ell \hat{f} H, \lambda^\ell \in \{0, 1\}, \ell \in L $$

A0-4 

$$ z_\ell \geq \lambda^\ell \frac{H}{h_{\max}}, z^\ell \in \mathbb{Z}, \ell \in L $$

Model [SUE] was solved by means of the simulated annealing algorithm on the expanded transit network of figure 3 and with a passenger’s demand given in table 1. Figure 2 shows the evolution of the objective function for 2000 iterations of S.A. algorithm with low temperature. Execution time on a HP laptop with 2Gb took $\tilde{1}h15\text{min}$ for 2000 iterations problems. In the computational experiences, the V.I., once the number of services were set, was solved using a diagonalization algorithm using a maximum of 500 iterations for each run of the diagonalization algorithm. A technique for reducing the number of iterations of this algorithm was used resulting in 25% savings in CPU time. As it can be seen from the figure, good objective function values for model SUE above were reached at a much earlier iteration than the 2000-th one. Runs with high temperature provided much worse computational results requiring almost all the 2000 iterations in order to reach very similar objective function values.

**3.2 A non-linear congested shortest path based service setting model**

For the case of special services set in order to alleviate disruptions, it is difficult to impose to the passengers of a given o-d pair a splitting amongst several routes as a policy oriented to follow a system-optimum behavior. Instead it is easier to recommend a single route to be followed by all passengers of a given o-d pair. The recommended route should be optimal and should take into account congestion effects. Because of congestion, non-linearities appear and the model is similar to a non-linear shortest path choice problem and the objective function of the design model might minimize total costs. For this case, model [SS] below was developed by Codina et al. in [4].

**[SS]**

Min $n, z, v, \tau, \lambda$ 

$$ \sum_{\ell \in L} (\xi^n \ell + \gamma^z \ell) + \theta \sum_{a \in A} e_a T_a(v, z) + \theta \sum_{\ell \in L, b \in \Pi_\ell} a_{(\ell, b)}(v, z) $$

s.t. 

$$ v \in \mathcal{V} $$

R0-1 

$$ \tau^\omega_a \leq 1, \tau^\omega_a \in \{0, 1\}, \left\{ \begin{array}{l} a \in \hat{E}(i) \end{array} \right\} i \in N $$

R0-2 

$$ v^\omega_a \leq M \tau^\omega_a, a \in A \setminus A_G, \omega \in W $$

Qb0 

$$ \sum_{\ell \in L_b} z^\ell \leq \hat{Z}_b(v, z), b \in \hat{N}_G $$

Qp0-1 

$$ e_a + v_{\tau(a)} \leq c_{z^\ell} $$

Qp0-2 

$$ \sum_{\ell \in L_a} \zeta_a(v, z) \leq \frac{H}{\eta_b} N_b \text{pax} $$

(3)

It consists of the minimization of total costs, as in previous model [SUE], but being these expressed conveniently in order to handle bulk service type queueing models for passengers at stations. The first term includes operational costs for setting and operation of services and the second plus the third one are in total the total travel time. The third term is made up by functions $\zeta$ for modeling queueing time of passengers at stations, whereas the second term includes times at links of the expanded network excluding queueing of passengers at stations. Routing considerations appear reflected in constraints R0-1, R0-2 of the formulation, where binary decision variables $\tau^\omega_a$ indicate which of the boarding links in the expanded network,

![Figure 2: Evolution of objective function in model SUE using the simulated annealing algorithm.](image-url)
outgoing from a station, must be chosen by passengers with origin-destination pair \( \omega \). The model also includes constraints (Qb0) in order to reflect the capacity of a station in terms of maximum number of incoming buses per hour that the facility is able to admit taking into account the spillback of buses queuing for boarding/alighting operations and also, the maximum number of passengers that can be standing at a station, queuing for boarding (constraint Qp0–2). Constraints Qp0–1 impose a limitation in the boarding flow \( v_a \) at a boarding link \( a \) in a station \( b \) according to the number of services \( z^a \) of line \( \ell \) to which the link belongs, the bus capacity \( c \) and the average number of passengers \( v_{x(a)} \) on buses of line \( \ell \) arriving at station \( b \).

Model [SS] is of the nonlinear mixed integer type and several optimization techniques are currently on essay in order to solve it. Function \( \zeta \) has been determined using simulations with bulk-service queues and a convex piecewise approximation has been developed, resulting into an approximate model. A heuristic technique for obtaining suboptimal solutions has been developed showing a very good computational performance. It consists of freezing values of nonlinear functions appearing in model [SS] based on flows \( v \) and number of services \( z \) at previous iteration. In this way a mixed integer linear programming problem appears at each iteration which can be solved efficiently using CPLEX for medium size networks. This linear integer problem preserves the network structure avoiding non-linearities and shall be referred to as model [SSlin].

**Heuristic algorithm for Model SS**

0. (a) Calculate initial values for the number of services and an initial value for \( \text{Q}^0_a \) at a station. Evaluate approximate line cycle lengths \( \bar{c}^0_a \) for each line and also initial bus service times, \( \kappa^0_b \) initial values for bus waiting times at stations \( w_{qb}^0 \), \( b \in \mathcal{N}_G \), so that an initial value for the maximum number of services allowable at a station, \( \bar{Z}^0_b \) can be evaluated using function \( \bar{Z}_b(\cdot, \cdot) \), i.e. \( \bar{Z}^0_b = \bar{Z}_b(\kappa^0_b, w_{qb}^0) \). Also, determine initial link travel times \( \bar{T}^0_a \) accordingly.

(b) Solve model [SSlin] for parameters \( \text{T}^0_a, \bar{Z}^0_a, \text{Q}^0_a, \zeta^0 \) so that flows and number of services \( (v^{1}, z^{1}) \) are obtained. Set \( \nu = 0 \)

At iteration \( \nu + 1 \):

When considered convenient superscript + is used to denote \( \nu + 1 \) and superscript - is used to denote \( \nu \)

1. Calculate new packet service times \( \kappa^b_{\nu+1} \), waiting times of buses at stations, \( w_{qb}^0(\nu + 1) \), and maximum number of services entering at each station \( \bar{Z}^b_{\nu+1} \) using an MSA step \( \alpha_{\nu} = 1/(\nu + 2) \):

\[
\begin{align*}
\kappa^b_{\nu+1} &= \kappa^b_{\nu} + \alpha_{\nu} \left( \kappa^b_{\nu} + z^+ - \kappa^b_{\nu} \right), \\
w_{qb}^0(\nu + 1) &= w_{qb}^0 - \alpha_{\nu} \left( w_{qb}^0(\nu + z^+) - w_{qb}^0 \right), \\
\bar{Z}^b_{\nu+1} &= \bar{Z}^b - \frac{\alpha_{\nu}}{\bar{Z}^b_{\nu} + w_{qb}^0} - \bar{Z}^b_{\nu}(\nu)
\end{align*}
\]

Evaluate new line cycles \( \zeta^b_{\nu+1} = C^b_{\nu}(v^{\nu}(1), z^{\nu}(1)), \ell \in L \), waiting time per passenger and per service at stations \( \text{P}^b_{\nu+1} = \text{P}_b(z^{\nu}(1)) \) and link travel times \( T^{\nu+1}_a \) as follows:

\[
\begin{align*}
\bar{T}^a_{\nu+1} &= t^0_a + b^0 + \kappa_{\nu,a}(v^{\nu}, z^+) + w_{qb}^0, \\
\text{if } a &= (j_{\ell}(b), j_{\ell}'(b')), \ell \in L, b, b' \in \Pi_{\ell} \\
\bar{T}^a_{\nu+1} &= \bar{T}^a_{\nu}(v^{\nu}, z^+) \text{ otherwise}
\end{align*}
\]

2. Solve approximate mixed linear integer model SSlin for parameters \( T^{\nu+1}, \bar{Z}^{\nu+1}_b, \text{P}^{\nu+1}_b, \zeta^{\nu+1} \) and obtain flows \( v^{\nu+2} \) and number of services \( z^{\nu+2} \). Let \( \nu = 1 + 1 \) and return to 1.

The algorithm stops when, at a predetermined number \( r \) of consecutive iterations, the number of services assigned to bus lines do not change \( (z^{\nu+r} = \ldots = z^{\nu+r+1}) \) and also, during these \( r \) iterations, flows \( v \) and total delays \( \zeta \) have little fluctuation \( (\| v^{\nu+r+1} - v^{\nu+s} \|_2 \leq \epsilon_v \) and \( \| \zeta^{\nu+r+1} - \zeta^{\nu+s} \|_2 \leq \epsilon_\zeta, s = 1, 2, \ldots, r - 1) \).

The proposed algorithm, although heuristic in nature, is supported by the fact that is based on the method of successive averages (MSA), which is applied to a number of variational models in continuous variables and it is known to converge.

**4 Congested transit assignment models**

Strategy based transit assignment models used in modeling passenger flows in regular lines of urban public transportation do not reflect congestion effects until very recently. Because of that frequency setting or service setting models which take into account congestion when passengers follow strategies have not yet
been developed. A classical uncongested model is that ofSpiess [12], which can be formulated as a linear program. Based on the results of Cominetti and Correa in [5], Cepeda et al. in [1] prove that their strategy based congested network equilibrium transit notion is equivalent to the minimization of the following non-convex, non-differentiable gap function $G_{\text{CCF}}(v)$

$$G_{\text{CCF}}(v) = \sum_{d \in D} \sum_{a \in A} v_a^d f_a(v) +$$

$$+ \sum_{i \in N_d} \max_{a \in E(i)} \left\{ \frac{v_a^d}{f_a(v)} \right\} - \sum_{i \in N_d} g_i^d \hat{x}_i^d(v) \right]$$

(6)

over the feasible set of destination flow vectors $V$, i.e. solutions of the congested transit equilibrium model are also global minima of the problem $\min_{v \in \mathcal{V}} G_{\text{CCF}}(v)$. Let now consider the polytope $S = \bigotimes_{d \in D} \bigotimes_{i \in \mathcal{N}_d} S^d_i$ and $S_i = \left\{ \alpha \in \mathbb{R}^{E(i)}_+ \mid \sum_{a \in E(i)} \alpha_a = 1 \right\}$ associated to node $i \in \mathcal{N}$. In [2] it is proved that solving this problem is equivalent to the following variational inequality (VI):

(VI) Find $(v, \zeta) \in \mathcal{V} \times S$ so that:

$$0 \in T^d(v, \zeta^d) + N_{\mathcal{V}_d}(v^d), \quad d \in D$$

$$0 \in -x_i^d(v) + N_{S^d_i}(\zeta_i^d), \quad d \in D, \quad i \in \mathcal{N}_d$$

(7)

where in (7), $N_{\mathcal{V}_d}(-)$ and $N_{S^d_i}(-)$ denote the normal cones on sets $\mathcal{V}^d$ and $S^d_i$ respectively at a point $(.)$. $\zeta_i^d = (\ldots, \zeta_i^d, \ldots; \alpha \in \hat{E}(i)), i \in \mathcal{N}_d$. $d \in D$ and $T^d(v, \zeta^d)$ are defined as $T^d(v, \zeta^d) = (\ldots, \Psi_d^d(v, \zeta_i^d), \ldots; \alpha \in A)$ and functions $\Psi_d^d$ are defined as $\Psi_d^d(v, \zeta_i^d) = t_a(v) + \sigma_a(v)\zeta_i^d$, if $a \in \hat{E}(i)$ and $\Psi_d^d(v, \zeta_i^d) = t_a(v)$ if $a \in \hat{E}(i) \setminus \hat{E}(i)$

In [2] previous results are also extended to the case of sharp capacity constraints on bus lines either explicitly or implicitly imposed by effective frequency functions $\sigma_a(v)$ and the MSA (method of successive averages) specialized for the congested strategy based transit assignment problem in [1] can be easily adapted for this case. Next subsection shows some computational results for this case.

4.1 Some computational results for the capacituated transit assignment problem

The transit network for this example is made up of eight transit lines and its expanded transit network is shown in figure 3. Effective frequency functions for boarding links are of the type $f_a(v) = 0.2(1 - \rho^2(v))$ and $\rho_a(v) = v_a/c - v_{m(a)}$. Capacity $c$ at boarding links is 9600 passengers for a period of 3 hours. Link travel times are given in [2]. Boarding links $(i,j)$ are those whose $i$-node is either 1, 2, 3 or 4. Demands in passengers for a 3 hours period are shown in table 1 below. This matrix has been uniformly augmented by a factor $\tau$ in order to conduct computational experiments

Table 1: O-D Trip table for a period of 180 minutes. Last row and column are average arrival and departure rates of passengers at bus stops.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total pax/min</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2011</td>
<td>22097</td>
<td>368</td>
<td>24476</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>170</td>
<td>0</td>
<td>0</td>
<td>3466</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>170</td>
<td>0</td>
<td>0</td>
<td>4106</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>170</td>
<td>0</td>
<td>0</td>
<td>4706</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>2011</td>
<td>22097</td>
<td>368</td>
<td>24476</td>
</tr>
</tbody>
</table>

Figure 3: Expanded network model for the example.

For this test network, the MSA algorithm in [1] with implicit capacity constraints behaves well for $\tau = 1.0, 1.2$. For $\tau = 1.3$, iterates violating capacity constraints appear during the run of the algorithm, although it finally converges to the solution. For $\tau = 1.6, 2.0$ and larger values, it shows unstable behaviour because capacity infeasible iterates appear too often. Results comparing algorithm in [1]
(implicit capacity constraints) and the algorithm using explicit capacity constraints are shown in table 2. Also the self-regulated MSA step in [9] has been used for this example with \( \tau = 2 \) for the MSA algorithm in [1] and for the algorithm using explicit capacities.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>#iter</th>
<th>( \text{lowest} \hat{g}_A )</th>
<th>( \hat{g}_B )</th>
<th>( \text{lowest} \hat{g}_B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6</td>
<td>1000</td>
<td>- 8.85E-04</td>
<td>-</td>
<td>8.36E-05</td>
</tr>
<tr>
<td>1.6</td>
<td>4000</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1.6</td>
<td>50000</td>
<td>- 2.43E-04</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2.0</td>
<td>4000</td>
<td>0.008</td>
<td>0.0233</td>
<td>1.95E-05</td>
</tr>
<tr>
<td>2.0</td>
<td>50000</td>
<td>1.95E-03</td>
<td>8.30E-04</td>
<td>7.68E-04</td>
</tr>
<tr>
<td>2.0</td>
<td>50000(*)</td>
<td>1.25E-03</td>
<td>2.15E-05</td>
<td>1.91E-05</td>
</tr>
</tbody>
</table>

Table 2: Comparison between algorithm with (B) and without (A) explicit capacity constraints. \( \hat{g} \) = relative gap. (*) use of self-regulated MSA step in [9].

5 Conclusions

Service setting models for public transportation lines in congested situations have been presented under two different passenger transit assignment approaches. The first model assumes that passengers make a choice accordingly to a user equilibrium principle following no recommendation but without assuming possible strategies. The second model assumes that passengers follow a recommendation based on a shortest congested route. Both models have been formulated as nonlinear mixed integer programming problems. The first one by means of simulated annealing and the second one by means of an ad hoc developed heuristic. Also, the congested transit assignment model based on strategies developed in [5] and its formulation in V.I. in [2] has been briefly introduced. Computational results using an MSA algorithm have been presented on a small test network with strict capacities.

Acknowledgements: Research supported under Spanish Research Projects TRA2008-06782-C02-02 and TRA2011-27791-C03-01/02

References:


