A model of urban public transport system design

Luigi dell'Olio
Universidad de Burgos

Luis De Grange
Fernández and De Cea Ingenieros Ltda.

Jose Luis Moura
Universidad de Cantabria

Hernan Gonzalo
Universidad de Burgos

Angel Ibeas
Universidad de Cantabria

Synopsis

This paper exposes a model of bus-stop location and frequency optimization in land public transport.
The aim of the problem-solving project is to minimize the total social cost involved in the operation of the transport system, including the costs of the provision of services (incurred by the operators), the travel costs (incurred by the users of the system) and the external costs incurred by the users of other transport means (as for example, car users) as well as the cost derived from the construction of stops.
The Optimization Problem is approached as a bi-level-type problem of full mathematical programming. We are defining a total cost function of the system which must be minimized, subject to technological restrictions, at the upper level, while at the lower level, a behaviour model for the system users is defined. Route choice on road network is made by private transport users according to Wardrop's first principle, in an attempt to minimize individual costs. As far as public transport users are concerned, it is also assumed that they choose the route which minimizes their total individual travel costs on the service network.
Given the difficulties that exist in finding a unique solution to this problem, a heuristic algorithm is proposed in order to provide adequate solutions, without necessarily guaranteeing the uniqueness of those solutions.

A model of urban public transport system design

It is known that the problem of design of urban public transport system is the most difficult to be solved in the transport sector. The international experience in various cities shows that public transport is gradually losing share to private modes. It is for this reason that researchers have made a big effort in recent years to focus on the problem of design. There are specific studies focused on improving frequency (Furth and Wilson, 1981; Kocur and Hendrickson, 1986; Ceder, 1994; Constantin and Florian, 1995), while others are particularly focused on improving of the configuration of routes (Newell, 1979; Ceder and Wilson, 1986).

These studies, although adequate, propose optimization models of only one level and do not often cover adequately the interaction existing between supply and demand (Ziyou Gao, Huijun Sun, Lian Long Shan, 2004)

Demand is generally connected with the private sector while supply is related to the public one. Demand adjusts its activities to the existing supply with the aim of minimizing travel time, while supply seeks to maximize the system's performances. In line with this, it is necessary to model both sectors at the same time.

The problem of urban public transport system design can be approached as a two-level non-cooperative game (Stackelberg Game). In this game there is a planner, who determines the characteristics of the transport system, and the system users, who tend to minimize the generalized travel cost, providing a flow pattern for this system. The generalized travel cost includes all the elements considered by an individual when traveling, as for example the time it takes them to reach the system walking, waiting time, travel time of the vehicle, comfort of the journey, fares, safety, etc.

Initially, at the first stage, the planner defines the topological structure of services, establishing the layouts and public transport technologies to be used at each service (Physical Design Problem). At a second stage, the planner determines the operational characteristics of the system, as for example the frequency and the optimal capacity of each service (Operational Design Problem). At the second level we find the system users, who, in connection with this service structure, generate a flow profile of the transport services proposed. This second level is normally modeled through a behavior model which is useful for predicting the assignment of users to the analyzed public transport structure. (Noranbuena I.J., 2002)

There is at present a varied bibliography on transport network design using bi-level programming techniques (Yang and Michael, 1998; Yang, 1997; Wong and Yang, 1997; Yang and Bell, 1997)

The study outlined here applies a bi-level optimization model with the aim of estimating the optimal service frequencies provided thus obtaining the optimal location of bus stops.

THE PROBLEM OF BUS-STOP LOCATION AND FREQUENCY OPTIMIZATION IN PUBLIC TRANSPORT

The problem of bus-stop location and Frequency Optimization in Public Transport Networks can be posed as a bi-level mathematical programming problem.

A solution must be found to the minimization of the total social cost involved in the operation of a transport system, including the costs of providing the services, incurred by operators, the travel costs, incurred by the system users, and the external costs, incurred by the users of other transport means (for example, car users) and those incurred by the general population of the city (for example, pollution).

The main constraints of the problem are that flows are generated by the decisions made by the users in order to satisfy their travel needs, given a specific public transport system, in other words, a service supply with its particular routes and operation frequencies.

A discrete approach is used in the location of bus stops and a choice of potential locations for these stops is studied. We have chosen this kind of approach because the results obtained are as near as possible to real life, as it is known that not all points of a public transport route are appropriate to locate bus stops, either due to lack of space or regulatory restrictions which would be very difficult to be modeled in the event of considering a continuous approach.

The frequency optimization variable is regarded as continuous. The use of a continuous programming approach does not constitute in practice a significant obstacle to the representation of the real phenomenon which is to be modeled, but, on the contrary, the hypothesis that the variables which represent the frequencies of public transport services are continuous, leads to important algorithmic and computational advantages (Fernández y Elton, 1996). The computational requirements of the full programming approach make this highly inefficient and of little application in full-size cases.

In general terms, if a continuous formulation is used in the decision variables of the problem, it will be possible to obtain an optimal discrete solution straight from the continuous solution, by using easy approximation rules. According to the results obtained in Fernández and Elton (1996), once the related approximations were made, the solutions obtained in most of the cases analyzed in the problem of design of infrastructure networks, under both approaches, were identical.

PROPOSED MODEL

As mentioned above, the problem of bus-stop location and Frequency Optimization in Public Transport Networks can be posed as a bi-level mathematical programming problem. At the upper level, we have minimized the total social cost involved in the operation of the transport system. Special attention is given to the total travel cost (CTV) and the operation cost - dependant on the location and (n) number of bus stops and on the service frequency of each route (f_l) - as well as the cost of construction of stops - dependant on the number of stops that will be needed. All of these being subject to capacity restrictions in buses (2) and stations (3).

Upper level:

$$Min \quad Z = CTV(n, f_l) + CO(n, f_l) + CC(n) \tag{1}$$

$$\text{s. to} \quad f_l \geq \frac{NTP_l}{K_b} \quad \forall l \tag{2}$$

$$f_s \leq \min\left(\frac{3600}{TO_{k'}}\right) \quad \forall s \tag{3}$$

Where:

CTV = Total travel costs
CO = Operation costs
CC = Costs of construction of stops
f_l = Frequency of line l.
NTP_l = Total number of passengers in the busiest section of line l in the time range studied.
$TO_{k'}$ = Occupation time of station k' expressed in seconds.

At the lower level, we consider a model of assignment to public transport. The model of equilibrium assignment in public transport networks used in the formulation, requires the definition of a more complex network represented by graph $G' = (\overline{N}, S)$, where *S* is the set of network links which comprises *route sections* and access links. A route section is a portion of a route between two consecutive transfer nodes, which has a set of lines equally *attractive* for the associated users (see De Cea and Fernández, 1993).

Lower Level:

$$Min \quad \sum_{s \in S} \int_0^{V_s} c_s(x)dx \tag{4}$$

s. to

$$
\begin{aligned}
&\sum_{r\in R_w} h_r = T_w && \forall w \in W \\
&\sum_{r\in R} \delta_{sr} h_r = V_s && \forall s \in S \\
&v_l^s = \frac{f_l \cdot V_s}{f_s} && \forall l \in B_s, \forall s \in S \\
&h_r \geq 0 && \forall r \in R
\end{aligned}
\tag{5}
$$

Where:
W: Set of origin-destination pairs O-D.
w: Element of set W, in which $w = (i, j)$, with i, j centroids.
T_w: Total number of journeys between the O-D w pair for public transport users.
l: Rate of a public transport line.
R: Set of routes available for public transport users.
r: Rate of a public transport route.
R_w: Set of public transport routes associated with the O-D w pair.
$\overline{h}_r$: Passenger flow in public transport on route r.
s: Rate of a route section in public transport.
S: Set of route sections available for public transport users.
c_s : Travel cost for public transport users on route section s.
δ_{sr} : Element of the route - route section incidence matrix: it takes value 1 if route r passes by s and 0 in the remaining cases.
V_s: Passenger flow in route section s.
v^l_s: Flow of passengers using line l in route section s.
f_l: Service frequency in line l.
f_s: Total frequency in route section s.

In order to represent the fact that travelers select a subset of lines L_1 (subset of *attractive* lines) to travel from A to B, taking into account at the same time the capacity restriction for vehicles, the above situation is modeled as follows:

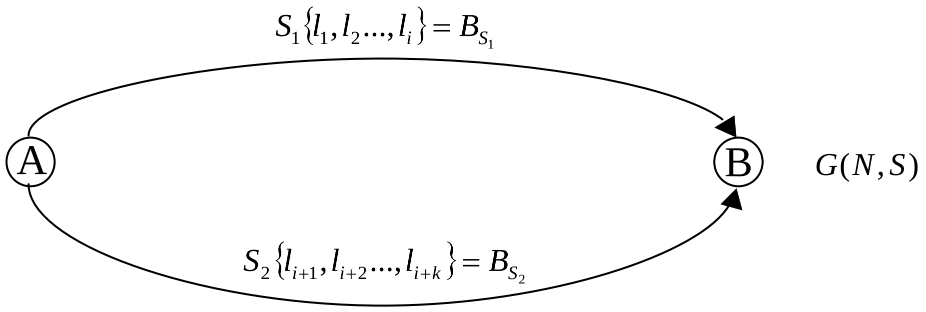

Figure 1: Links of Public Transport Network $G' = (\overline{N}, S)$

The first *public transport network,* S_1, represents the set of *fast lines* B_{S_1} including the set of lines which minimize the generalized travel time (cost) without considering the capacity restriction for vehicles, in other words, it is the set of common lines determined by the Chriqui algorithm (see Chriqui, 1974, and Chriqui and Robillard, 1975). The second public transport arch, S_2, represents the set of *slow lines*, B_{S_2} .

Consequently, there may be more than one link in the public transport network (route section) by joining a pair of given nodes. The first link, as mentioned above, contains the "fastest" lines of all the routes which join this pair of nodes (set of common lines). If certain lines have not been included within this set, the Chriqui algorithm is again applied and a second link, parallel to the previous one, which comprises the new set of common lines, is formed. The process continues until all lines joining the pair of nodes are associated to some route section. As the route sections which include the "fastest" lines get congested, the route sections including the "slowest" lines begin to attract users. This assumption, while separating the assignment problem from the problem of selection of common lines, thus simplifying the equilibrium problem at user level, simplifies the problem formulation for frequency optimization. Really it is only when line frequencies are known, that it is possible to determine which services exactly belong to the set of "fast lines" and to the "slowest lines". In any case, the solution algorithm proposed does not take the network as fully defined and whenever a vector of frequencies is obtained at the upper level, the lines which make up each one of the route sections of the public transport network are determined. Only after this is completed can we proceed to carry out the corresponding assignment.

BASIC ASSUMPTIONS

The selection of route is made by public transport users according to Wardrop's first principle with the aim of maximizing individual utilities (by minimizing total travel cost).
It is assumed that public transport users may choose the route that minimizes their total travel cost (walking time + fare + waiting time + in bus travel time). It is also assumed that the public transport system has a limited capacity and therefore, the travel cost increases as passenger flow increases. Thus, as some routes become congested, passengers will use alternative routes that are attractive to them.

The cost functions for the route sections that make up the public transport network are as follows:

$$c_s = \overline{t}_s + \left(\frac{\alpha}{d_s}\right) + \beta \cdot \varphi_s\left(\frac{V_s + \tilde{V}_s}{K_s}\right) \tag{6}$$

where $\overline{t}_s$ is travel time on the vehicle plus the fare, d_s is the total frequency in the route section s, α and β are calibration parameters, K_s is the capacity in the route section s, V_s is the total number of passengers in the route section s and $\tilde{V}_s$ is the flow that competes for the same capacity. φ_s function must be such that c_s is strictly monotonous in V_s. BPR-type functions are normally used:

$$\varphi_s = \left(\frac{V_s + \tilde{V}_s}{K_s}\right)^n \tag{7}$$

It can be observed that the cost functions in the route sections of the public transport network are not separable. The cost function c_s of a route section of public transport depends on the passenger flow in other route sections and not just its own flow, as well as on the flow of private transport vehicles in the travel links concerned.

USERS' BEHAVIOR MODELS: AN OVERVIEW

The various algorithms that solve the Problem of Design of Public Transport Networks must interact with the system **users' behavior models** with the aim of properly assessing each one of the solutions analyzed by the design algorithm.

Behavior models enable us to simulate the travel decisions made by the transport system users and to determine the equilibrium flows (passengers and vehicles) on the networks concerned. Depending on the type of analysis to be carried out and the hypotheses made about the impact that the modification in service frequency would have on modal distribution and share of journeys, a variety of behavior models will be necessary.

In fact, when making a medium-long-term analysis, it is necessary to assume that the design of the public transport network (in general, layouts and optimal frequencies) may affect users' decisions at different levels i.e.: destination, mode and route choices. In line with this, the equilibrium flows must be determined by using a simultaneous distribution-modal share-assignment equilibrium model.

If this type of analysis requires considering the transport mode and route choices, then, the behavior model to be used is a simultaneous modal share-assignment equilibrium model so that the effects of redesigning the public transport system are internalized within those choices. In this case, we make the distribution among the public transport modes as a result of a process of modal choice (it is assumed that frequency changes do not affect either travel distribution or the modal share between public and private transport).

If it is assumed in the analysis that no changes occur either in travel distribution or in modal share as a consequence of the redesign of the public transport system and, subsequently, the matrices of the journeys of the diverse analyzed models remain fixed, equilibrium flows are determined by using an equilibrium allocation model that will be called *bimodal,* when considering private transport-public transport interactions, or multimodal, when only one public transport network with multiple modes is considered.

As regards public transport, a matrix of total journeys (pure and combined modes) allocated on a multimodal network of services, is used. Therefore, travel distribution among the different modes is determined from the allocation of the matrix on that network.

The equilibrium allocation model in public transport networks used when formulating the frequency optimization problem, corresponds to a model proposed by De Cea and Fernández (1993). This model enables us to solve the problem of allocation to congested public transport networks. Congestion is associated with waiting that users have to undergo at stations when the number of journeys approximates the capacity of the system

The model assumes that users choose - among all the route choices connecting a specific pair of nodes on the public transport network - that route which minimizes their total travel time (cost) (fare + travel time on the vehicle + waiting time + walking time).

On the other hand, it is assumed that the system has a limited capacity so that, the greater the number of system users the greater travel times become. It is assumed that the congestion phenomenon occurs at the stations, as travelers must wait depending on the total capacity of all the lines and on the number of passengers who wish to use those lines to travel.

Once a passenger boards a vehicle at the station, travel time on that vehicle will only be determined by the congestion level that may exist on the road network (flow of private transport and public transport vehicles).

As far as the set of lines available to travel is concerned, the model assumes that, between each pair of nodes of the public transport network, there is a set of “common lines” or equally attractive lines for the users. Thus, at each station, passengers will take into account the set of common lines supplied to make their journeys and board the first vehicle with available capacity included in that set.

SOLUTION ALGORITHM

In order to solve the bi-level mathematical programming problem, the following heuristic algorithm is exposed:

Step 1 – A vector of feasible frequencies f_l is generated and a distance d_z between stops which may satisfy the constraints of the problem at the upper level.

Step 2 – The optimization problem is solved at the lower level; in other words, the origin-destination matrix of public Transport is allocated and the V_l^* equilibrium flows obtained.

Step 3 – The feasible frequencies f_l and the equilibrium flows V_l^* are entered in the objective function at the upper level, and the objective function *Z* is evaluated.

Step 4 – The Hooke-Jeeves algorithm is used to evaluate the new f_l frequencies and the new d_z distances between stops, then returning to step 2.

The Hooke-Jeeves algorithm is repeated until no optimal frequencies and distances are found.

HOOKE-JEEVES ALGORITHM

For the Hooke-Jeeves algorithm, no special attribute is required of the objective function. The algorithm does not require this function to be convex nor an explicit analytical expression of its derivatives with regards the decision variables in the problem. It does, however, require the function to be continuous and evaluable for any feasible value of the variables. Any continuous function may be worked with, no matter how strange the form, because of the way in which the algorithm works.

Although these characteristics offer an advantage from the point of view of applicability of the algorithm, they can create a disadvantage as regards efficiency because they do not use the regularity characteristics that the objective function may produce, such as, convexity of the function in each of the decision variables.

On the other hand, this algorithm does not guarantee a global optimum unless the problem to be resolved is strictly convex. With public transport network design, because of the characteristics of the objective function, the Hooke-Jeeves algorithm only guarantees a local optimum. Therefore, to obtain a global optimum, it needs to be applied on the basis of different initial solutions with different values for the parameters involved.

Description of the algorithm:

The Hooke-Jeeves algorithm basically involves the repetition of two stages, which are:
Exploratory search through each of the coordinates of the solution space in order to find a good local descent direction (reduction in the value of the objective function).
Pattern movement, which consists of an advance in the direction determined in the first stage.

a) Exploratory search in coordinates stage
At this stage, the algorithm searches for a point indicating a good local movement direction. This is performed by increasing the value of one of the variables by a pre-set amount ("delta") and evaluating the objective function at this new point.
If the value of the objective function is less than that obtained at the starting point, the new point is accepted and the next variable is analyzed. However, if the value of the objective function at the new point is greater than at the starting point, the original value of the variable is reduced by "delta" and the objective function is evaluated again. If the value of the latter is less than at the starting point, the new point is accepted and the next coordinate is analyzed. If the value of the objective function is greater than at the new point, the variable is returned to its original value, plus "delta", and the next coordinate is examined.

This process is repeated with each of the decision variables in the problem.

Once the process has been completed, we need to check that the value of the objective function at the point obtained in the exploratory search is lower than at the best base point found up to that moment. If it is lower, the new point is set as the base point and we can move on to the next stage of the method. If the exploratory search was unsuccessful, the value of the "delta" parameter is reduced and a new search stage is performed, this time starting at the best base point obtained up to that moment.

b) *Stage of advance* in the direction indicated by the exploratory search (Pattern movement)
If a good result was obtained in the first stage of the method, i.e. a point was found at which the value of the objective function was lower than the best base point, an advance is made in the direction indicated by the latter and the new point that was found. The length of this advance is determined by the difference between these two points multiplied by a pre-set value ("alpha").

Once an advance has been made in the direction determined in the first phase, we can carry out another exploratory search from the new point. This is a new iteration of the Hooke-Jeeves algorithm. The point found in the previous exploratory phase is maintained as the base point, which is used as the comparison criterion for the result of the new search.

Algorithm stopping criteria
The value of the "delta" parameter is often used as the algorithm stopping criterion. Each time the value of this parameter is reduced, it is compared with its tolerance (minimum possible value) and, if it is lower, the algorithm is stopped.

The value of the "alpha" parameter is usually greater than the "delta" parameter because the exploratory phase only searches for a good local direction for movement whereas the second stage of the algorithm requires us to advance in this direction.

The "delta" and "alpha" parameters are empirical so their values must be determined by tests, taking into account the speed of algorithm convergence and the solutions obtained for different values of these parameters.

A second stopping criterion that can be used involves specifying the maximum number of iterations to be made (a complete iteration includes the exploratory search phase and the advance in the direction determined in it).

Algorithm convergence
If the objective function of the problem is strictly convex, then the Hooke-Jeeves method is clearly capable of obtaining the global optimum. However, if the function has local optimums, the algorithm can fall into one of these and the values of the parameters may not be sufficient to exit it.

To increase the possibilities of obtaining the global optimum of the problem, we can re-apply it by choosing random starting points and varying the values of the parameters.
Although the Hooke-Jeeves algorithm can obtain “good” solutions, it has a low convergence speed and hence a high consumption of computational resources, which are due to the characteristics of the method’s operation: the algorithm makes exploratory advances, increasing and decreasing the value of each of the decision variables in the problem. Each of these advances requires an evaluation of the objective function and hence solution of an equilibrium problem in transport networks, a process that takes up more computational time.

To preclude this problem, Abdulaal and Le Blanc (1979) suggested a modification of the Hooke-Jeeves method, which consists of supposing that small variations in the values of decision variables do not produce significant variations in equilibrium flows. This allows us to conduct the exploratory stage with constant flows and hence limit the process of allocation to the base points determined after a complete iteration of the algorithm. This simplification enables us to obtain very similar results to those obtained using the original algorithm – albeit with considerable computational savings – particularly for solving problems with real transport networks.

CONCLUSIONS

This article illustrates a problem in public transport system design, based on the problem of locating bus stops and optimizing the frequency of lines along a given route.
In this analysis, we have seen that the two problems cannot be isolated because the location of bus stops is based on the frequency of the buses serving the different routes, and frequency is also based on the number and location of bus stops.
Therefore, the problem with the design of the public transport system for urban areas was approached as a two-level non-cooperative game (Stackelberg game), where the upper level minimizes social costs (user costs + operating costs + construction costs) and the lower level (public transport or multi-modal allocation model) describes a model of user behavior, which is assumed to follow Wardrop’s first principle.
The heuristic solution algorithm proposed here obtains the optimal solution through successive approximations that consist of updating the frequencies of the different lines (using the Hooke-Jeeves algorithm) and the location of bus stops (minimizing the objective function of the upper level) between one iteration and the next, so that the objective function of the upper level is “displaced” towards its minimum values.

To consider the interaction between demand and supply between one iteration and the next, it is assigned to public transport (or bimodal depending on the case). Thus, each iteration can have equilibrium flows that are consistent with the calculated frequencies and location of bus stops.

Lastly, we can confirm that a different approach is used here to that commonly used in the literature, although it is actually more correct than traditional methods. The only disadvantage of this method is that we cannot guarantee the uniqueness of the solution as in many cases, due to the non-linearity of the problem, it will in all likelihood depend on the initial feasible solutions with which the algorithm was begun.
However, as in all these types of study, we can confirm that obtaining a local optimum is more than sufficient for solving this type of problem.

ACKNOWLEDGEMENTS

The authors would like to thank Professors E. Fernandez and J. de Cea of the Pontificia Universidad Católica of Santiago de Chile and the engineer M. Barquín de Fernandez and Cea Ingenieros Ltda for their excellent advice.

BIBLIOGRAPHY

Abdulaal, M. and L.J. LeBlanc (1979) “Methods for combining modal split and equilibrium assignment models”, *Transportation Science*, 13] 292-314.
BaRquin, M., Diseño Operacional de Redes de Transporte Público: Formulación Matemática y Algoritmos de Solución. *Master of science thesis.* Pontificia Universidad Católica de Santiago de Chile.
Ceder, A., Wilson, N., 1986. Bus network design. *Transportation Research B* 20, 331–334.
Ceder, A., 1994. Bus frequency determination using passenger count data. *Transportation Research A* 18, 439–453.

Chiriqui, C. (1974) “Reseaux de transport en commun: Les prolemes decheminement et d’acces”, *Center of Transport Research* ,University of Montreal, Publication No. 11.
Chriqui, C. and P. Robilland (1975) “Common bus lines”, *Transportation Science*, 9: 115-121.
Constantin, I., Florian, M., 1995. Optimizing frequencies in a transit network: a nonlinear bi-level programming approach. *International Transactions in Operational Research* 2, 149–164.
De Cea, J. and J. E. Fernandez (1993) “Transit assignment for congested public transport systems: An equilibrium model”, *Transportation Science*, 27: 133-147.
Furth, P., Wilson, N.H.M., 1981. Setting frequency on bus routes: Theory and practice. *Transportation Research Record* 818, 1–7.
Kocur, G., Hendrickson, C., 1986. Design of local bus service with demand equilibrium. *Transportation Science* 16, 149–170.
Newell, G., 1979. Some issues relating to the optimal design of bus routes. *Transportation Science* 13, 20–35.
Norambuena, I. J., 2002. Diseño optimo de sistemas de transporte urbano. *Master of science thesis*. Pontificia Universidad Católica de Santiago de Chile.
Wong, S.C., Yang, H., 1997. Reserve capacity of a signal-controlled road network. *Transportation Research B* 30, 397–402.
Yang, H., 1997. Sensitivity analysis for the elastic-demand network equilibrium problem with application. *Transportation Research B* 31, 55–70.
Yang, H., Bell, M.G.H., 1997. Traffic restraint, road pricing and network equilibrium. *Transportation Research B* 31, 303–314.
Yang, H., Michael, G.H.B., 1998. Models and algorithm for the road network design: a review and some new development. *Transport Review* 18, 257–278.
Ziyou Gao, Huijun Sun, Lian Long Shan, 2004, A continuous equilibrium network design model and algorithm for transit systems. *Transportation Research Part B* 38, 235–250