From Strategic Modelling of Urban Transit Systems to Golden Rules for their Design and Management

Fabien Leurent, François P. A. Combes, Rob van Nes

▶ **To cite this version:**

Fabien Leurent, François P. A. Combes, Rob van Nes. From Strategic Modelling of Urban Transit Systems to Golden Rules for their Design and Management. 2020. hal-02462463

HAL Id: hal-02462463
https://hal.science/hal-02462463

Preprint submitted on 31 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From Strategic Modelling of Urban Transit Systems to Golden Rules for their Design and Management

Fabien Leurent ([1]), François Combes, Rob Van Nes
Université Paris Est, Laboratoire Ville Mobilité Transport, Ecole des Ponts ParisTech

Abstract

The paper provides a synthetic, "strategic" model of transit systems in urban areas that features out a set of modes, quality of service and terminal access, demand and network usage by users' trips, with some hints of spatial heterogeneity. The model encompasses technical relationships relating fleet size and design parameters such as infrastructure length and station spacing, to frequency, commercial speed and access distance, hence to wait time, running time and access/egress time. Economic features are modelled, too: generalized costs to individual users, demand elasticity, supply costs and system welfare. The model can be used for synthetic statistical description of real-world systems as well as for economic analysis and the assessment of given system states against theoretical references.

After introducing the model elements and relationships, we put forward a causal diagram that synthesizes the system under study and constitutes the model architecture. We then turn to mathematical analysis to formalize (i) the determination of a system state on the basis of a supply plan, technical relationships and demand behaviour, (ii) the optimisation of system welfare with respect to the action levers on the supply side. Next, for an uncongested system we establish theoretical conditions for both an optimum system state under fixed demand and a second best optimum under variable demand and tariffs. Three "golden rules" for transit network design and management are established, namely (i) balancing the rolling stock costs and the users' costs of waiting time, (ii) balancing the station costs plus the value to users of the dwelling part of their in-network times, against the users' costs of "longitudinal" access times, (iii) balancing the full supply costs and the users' costs of "transversal" access times. Furthermore, the existence and uniqueness of a System Optimum state are proven and a solution scheme is provided.

1. Introduction

1.1 Background

A transportation system is purported to serve the mobility needs of its customers. It is made up of one or several technical "modes" that are spread through the space in a given territory, mostly in network form in the urban setting. According to the general economic theory of welfare, such a technical system ought to be planned so as to serve travel demand in an optimal way i.e. by maximizing the net surplus of its users (Dupuit, 1844), which amounts to the difference between their gross surplus and the travel costs which they incur. In fact, the user bears both money costs that include prices paid for service access and use plus specific expenses (e.g. vehicle costs in private transport) and also the money equivalent of the time spent in travelling (Beckmann *et al*. 1956). In turn, the service prices should reflect the

[1] Corresponding author: fabien.leurent@enpc.fr. Ecole des Ponts ParisTech, Laboratoire Ville Mobilité Transport, 6-8 avenue Blaise Pascal, Champs sur Marne 77455 Marne la Vallée Cedex

production costs that are borne by the suppliers. Externalities such as scale or scope economies on the supply side, or impacts on the environment (such as pollutant emissions and contribution to greenhouse gases), can be taken into account by modifying the welfare function of collective surplus (Quinet and Vickerman, 2004).

To apply these general principles to the management of a given system, it is required to build a specific economic model that represents both the "local" demand and the "local" supply – here "local" refers to the territory. The specific setting must be modelled in terms of modal techniques, spatial features and behavioural features of the economic agents. Furthermore, the action levers available for system management need be identified and characterized. The multi-layered structure of a transportation system, from infrastructure to services passing by vehicles, lends itself to a variety of action levers. Infrastructural schemes either roadway- or railway-based or modal / intermodal stations, as well as transit services and demand-oriented mobility policies are traditionally addressed by using a Travel Demand Model. Such a TDM depicts the demand side in terms of passenger and/or freight flows and their respective behaviour concerning the choice of mode, path, departure time, destination place etc., and also the supply side in terms of network routes and services (Ortuzar and Willumsen, 2004). Such a model enables mobility analysts to incept and test projects of infrastructure and/or services in an incremental way, as well as pricing policies, in order to assess alternative planning scenarios and to aid decision-making in system planning. The objective function used in scenario assessment involves both demand surplus and supply surplus, since in most cases the relationship between usage prices and production costs is not straightforward (De Rus, 200x).

There are also network design models in which planning schemes are generated automatically, typically by adding links and nodes of infrastructure and/or services to a reference network (REF). In this stream, each scheme is evaluated on the basis of an objective function that involves some demand model (which amounts to a simplified TDM) and above all a technical and economic model of the production, its means and their costs.

Every supply scenario is evaluated on the basis of reference unit costs, e.g. the cost of a new road link obtains by multiplying its length by a cost per unit distance that depends on the infrastructural type and local set-ups. Such unit costs are evaluated in an average way by reference to past experience for analogous cases: their nature is basically statistical.

A second kind of system management models addresses traffic regulation at the level either of a junction node, or a line or corridor, or a sub-network or a full network; there, the action levers consist mainly in capacity set-ups – e.g. to share the time of green at a signalled roadway junction between the traffic streams via its branches. Other regulation levers consist in speed regulation, right-of-way assignment of some traffic lanes, access regulation, dynamic pricing and dynamic traffic information (e.g. Papageorgiou).

To sum up, the above-mentioned models deal with system management at two levels: system planning with manual or automated system design versus traffic regulation.

A third kind of management models may be identified concerning the composition of vehicle fleets: notably so to address the types of motors, energy vectors and environmental performance. The related vehicle-fleet models are generally built on a standalone basis prior to connecting them to a TDM at one or two stages (about network assignment and maybe also users' decisions of motorization).

Thus it appears that the complex structure of transportation systems is mirrored in management models that pertain each to one layer out of the three, Infrastructure, Vehicles, Services & operations. However, multi-layer policies are in order to manage the system in an

integrated way, by involving long-term, planning-based levers as well as short-term, regulation-based ones and by considering each technical mode in connection to the other modes rather than on a standalone basis. In a given territory, both the local authorities and the network operators are interested in such an integrated approach.

Among the issues of strategic importance, let us mention (i) the multimodal quality of service supplied to the users, (ii) the multimodal environmental impacts, (iii) the financial costs to the community, especially so in contexts of diminishing financial resources. Among the diverse modes, a key issue is the "relevance area" of each mode: in the urban setting, is the mode better suited for more or less dense areas, under which particular technical set-ups – e.g. station spacing along a transit line?

1.2 Objective: a synthetic model for strategic planning and theoretical properties

From the previous discussion, let us conclude that there is room for a strategic model of transportation systems in order to provide synthetic overview, economic evaluation and some guidelines for system optimization. Our very aim in this paper is to bring about such a model in the field of urban passenger transit.

The model captures (i) some spatial features, from trip endpoints and regions in space to distances between origins and destinations, passing by the lengths of way sections, (ii) modal techniques of urban passenger transit, each with specific technical factors including vehicle capacity, service frequency, running speed, station spacing, and a technical production function linking these factors to the resulting quality of service in terms of run time and wait time, (iii) service users and their trips, with individual generalized costs based on tariffs and travel times, plus some demand elasticity to cost, (iv) on the supply side, cost functions and the consideration of policy packages acting on design levers – infrastructure length and station spacing – as well as fleet size and tariffs.

Thus it is a technical as well as economic model, and also a strategic model as it enables one to analyse planning trade-offs between modes and regions, in a simplified yet still meaningful way. We submit the model to an analytical study of system optimality: by mathematical derivation of optimality conditions between the action levers and the dependent state variables, we obtain generic properties that characterize the optimal design and management of an urban transit system.

1.3 Method

Previous works in the field of strategic models for transportation planning fall into two streams. First, some TDM have been designed as "strategic models" by reducing the set of Traffic Analysis Zones (TAZ) to at most some dozens of "macro-zones", and similarly simplifying the roadway and / or transit networks to aggregate characteristics such as the line length of sub-network by mode in each TAZ. Then, by zone and mode the sub-network length and the local trip flow may be related to local speed by an aggregate function. The system of trip flows has been modelled as flow exchanges between neighbouring zones (REF) or as origin-destination (O-D) flows to be assigned to local and modal sub-networks by an average, aggregated network assignment (REF). Such models are mainly used to simplify the presentation to decision-makers of the detailed results obtained using a finer TDM (LAET).

In the second stream, the set of demand zones is even more simplified into a couple of regions, of which the sub-networks by mode are also simplified and aggregated further than in the first stream. Van Nes (2002)…

Each technical factor…

Our own work pertains to the latter stream of research. We build upon Van Nes's model of urban transit, keeping the three technical relationships of (i) terminal access time, (ii) service frequency and (iii) service commercial speed, by adding the diversity of regions, a set of transit modes and traffic phenomena, and by dealing with intermodal trip chains between the regions.

To curb the additional complexity, we have invested in systems analysis, "statistical" representation and mathematical formulation. We put the issue of system optimization as a mathematical program of constrained optimization, of which we derive the first order optimality conditions. At this stage, the linkage between strategic modelling and economic theory becomes very clear: our strategic model under its abstract formulation constitutes a theoretical economic model of multimodal transit in an urban area.

1.4 Structure

The rest of the paper is organized in six sections with roadmap as follows. Section 2 provides the macroscopic, statistical representation of space, supply, demand and usage that are the basic system features. In Section 3, the system structure is laid out and the interactions between the components are modelled as causal relationships of technical and / or economic nature. Section 4 completes the modelling framework by identifying the impacts and stating the values and costs of primary interest in the system, namely demand surplus and supply costs. Then, Section 5 provides a mathematical formulation of the model and uses it to analyse not only the determination of a system state under exogenous supply plan but also the optimization of system welfare by setting up an adequate plan. For the uncongested case in which traffic variations exert no effects on supply performance or demand behaviour, Section 6 provides an economic theory of system optimization, with some characteristic properties of an optimal state that can be called "golden rules" for system design and management. Lastly, Section 7 offers a discussion and some conclusions.

In a companion paper, the model is applied to the case of Greater Paris as an instance.

Table 1: Notation

z region (i.e. macro-zone in urban area), with set Z

A_z area of region z

m transit mode, with set M

$r = (m, z)$ supply component, with set R

L_r infrastructure length or line length of component r, with unit cost c_{Lr} by period H

σ_r number of stations in component r, with unit cost $c_{\sigma r}$ by period H

S_r station spacing in component r

v_r commercial speed stations in component r

φ_r service frequency

β_r headway regularity factor

N_r fleet size, with unit cost c_{Nr} by period H

d_{Ar} average walking distance between trip endpoint and nearest station

χ_r shape parameter of component r

i O-D pair belonging to set I, with origin region z_i^+ and destination region z_i^-

w_i walking speed.

t_{ri}^X (resp. d_{ri}^X) the time (resp. distance) in stage X and component r by trip in segment i, with user's specific value of time γ_i^X.

θ_i money equivalent of travel time by trip

τ_i network tariff

g_i generalized cost for individual user of segment i.

D_i demand function

Q_i trip flow during period H

I_{ri} (resp. O_{ri}) input (resp. output) flow of segment i from region to mode (during period)

$M_{rs}^i = \overline{M}_{rs}^i + \hat{M}_{rs}^i$ number of i-bound trips transferring from r to s, $\hat{M}_{rs}^i$ with line / mode change and $\overline{M}_{rs}^i$ without.

A_r^i (resp. B_r^i) the number of alightings (resp. boardings) in r due to segment i.

Λ_r^i (resp. λ_r^i) the number of legs (resp. average length of them) in r due to segment i.

x_{ri}^X assignment factor of segment i to component r and stage X in {A/E, R, W, T}

ω_r dwelling time per vehicle run and station in r, with reference value ω_r^0

a_r (resp. b_r) the number of alightings (resp. boardings) by vehicle run and station in r.

k_r^{AB} exchange capacity per time unit by vehicle in r

$\underline{k}_r$ (resp. $\bar{k}_r$) seated (resp. standing) capacity by vehicle in r

y_r (resp. ρ_r) passenger flow (resp. density) by vehicle (underlined / barred for seated / standing)

η_r average track reservation in r during period

$\hat{V}_r$ function linking reference running speed to track reservation

E_r^X Effect factor of characteristic time in stage X and component r

SR Minimum supply revenue to achieve under Variable demand

$\overline{\mu} = 1 + \mu_B$ dual parameter, with μ_B multiplier associated to budget constraint

SC System cost function

£ Lagrangian function of System Optimization program

μ_{Xr} dual variable associated to the technical constraint that determines primal variable X_r

2. The statistical representation of supply and demand

Our objective here is to model the essential technical and economic features of supply and demand in the system, in a both meaningful and sufficiently aggregate manner. As our treatment involves the aggregate description of collections of entities such as modal stations in a given region of space, we refer to it as "statistical". We shall first provide the representation of space (§ 2.1), then that of transport supply (§ 2.2), followed by that of travel demand (§ 2.3). Next, we focus on two basic interactions between supply and demand, respectively the usage of transit supply by demand (§ 2.4) and the issue of access distance between trip end points and transit stations (§ 2.5).

2.1 The representation of space

A whole urban area is taken as study area, delimited by a given perimeter. It is divided in regions, indexed by $z \in \mathrm{Z}$ the set of zones, each one with its own ground area, A_z. Within the area, every transit user performs their activities in given locations, of which the particular points constitute endpoints for both the trips prior and next to the activity. Instead of locating particular points accurately, we shall only identify their regions and deal with the issue of network access in a specific, statistical way (Cf. § 2.5).

Similarly, modal stations are characterized by their regions and modes, whereas their individual locations are dealt with statistically. Concerning points along transit routes, we do not need to identify them – the only requirement being to model transfers between regions and/or lines or modes (Cf. § 2.4).

2.2 Transport supply

Apart from walking that is addressed in a statistical way of its own (Cf. subsection on network access), let us describe the supply of transit services that belong to a set M of transit modes indexed by m. These may represent the Bus, the Tram, the Metro, or Suburban Rail.

A key notion is that of a "supply component", or "network component" that is a mode and region pair denoted by $r = (m, z)$ with set R. We model transit supply basically by component, each of which is characterized by the following attributes:

+ infrastructure length or line length L_r adds up the lengths of rail tracks or road lanes that are used in region z by the transit routes of mode m. Each line track / lane section is counted only once: by assumption it serves the two directions of traffic. Line length is the component's key feature concerning the coverage of space.

+ as stations provide local access, the number σ_r of them is relevant. As it is counted at the line level then aggregated at the sub-network level, the total number represents a number of "stops" i.e. of incidences between lines and stations.

+ between consecutive stations along a line served by transit routes, the spacing influences the running times of the vehicles. Let us consider an aggregate spacing, denoted S_r, averaged over the line elements in the sub-network.

+ commercial speed v_r depicts service efficiency and is averaged over the vehicle runs serving the component. More precisely, it is the harmonic average resulting from vehicles' travel times over the runs

+ service frequency, denoted φ_r, is counted by station and traffic direction. It influences passenger wait time on service access.

+ parameter β_r is a regularity factor for headway intervals, with value of 1 for perfectly regular headways or 2 in the absence of memory.

+ fleet size N_r denotes the number of vehicles involved in service operations in the component. It is a technical parameter of much economic significance since vehicle fleet constitutes a major part in the system investment costs.

The statement of production costs is provided in Section 4. As for now, let us emphasize that the notion of a component is modeled very much like a single transit route in the theory of traffic assignment to a transit network (Andreasson *et al.*, 2016). In practice, it is often the case that a given line links together several regions: this is modeled here by splitting it into line sections according to location; the service duty fulfilled by line vehicle in a regional sector is attributed to the region's vehicle times and vehicle fleet, again in an abstracted way.

Concerning time periods, we refer here to a peak period in which service frequency is maximized, which determines fleet sizing.

2.3 Travel demand

Again as in a transit assignment model, let us represent travel demand as a set of trips made by travellers within the urban area. Origin-destination (O-D) pairs of regions are indexed by i and constitute a set denoted I. The origin region (resp. destination region) of i is denoted z_i^+ (resp. z_i^-).

Each individual user making one trip has a given walking speed, w, and specific sensitivity to access distance (code A), transfer distance or time (code T), wait time (code W) and run time (code R). Denote by X a generic code for trip stage, with t_i^{X} (resp. d_i^{X}) the associated time (resp. distance) in the trip: the user's specific value of time is denoted γ_i^{X}. Taking into account the O-D network tariff, τ_i, as well as the money equivalent of spent times, the resulting generalized cost is denoted g_i.

At the aggregate level and during the period under study, the trips are characterized by O-D pair i using the following attributes:

+ trip flow denoted Q_i (so the flow rate is Q_i / H during the period).

+ by component r, the input flow from region to mode (resp. output flow) is denoted I_{ri} (resp. O_{ri}): it takes value 0 if the component's region differs from the region of trip origin (resp. destination), whereas for $r=(m,z)$ it holds that $\sum_{m\in\mathrm{M}} I_{(m,z)i} = Q_i$ if $z = \mathrm{z}_i^+$ (resp. $\sum_{m\in\mathrm{M}} O_{(m,z)i} = Q_i$ if $z = \mathrm{z}_i^-$).

+ by pair (r,s) of components, the number of trips with O-D pair i passing from r to s is denoted $M_{rs}^i = \overline{M}_{rs}^i + \hat{M}_{rs}^i$, in which $\overline{M}_{rs}^i$ accounts for transfers without line change whereas $\hat{M}_{rs}^i$ accounts for transfers with line or mode change.

+ $A_r^i \equiv O_r^i + \hat{M}_{rR}^i$ where $\hat{M}_{rR}^i \equiv \sum_{s\in\mathrm{R}} \hat{M}_{rs}^i$ is the number of alightings in r due to segment i.

+ $B_r^i \equiv I_r^i + \hat{M}_{Rr}^i$ where $\hat{M}_{Rr}^i \equiv \sum_{s\in\mathrm{R}} \hat{M}_{sr}^i$ is the number of boardings in r due to segment i.

These definitions enable us to deal with travel demand in an aggregate, macroscopic way, while paying some attention to space and especially O-D pairs. The associated indicators can be derived by e.g. statistical analysis of a household travel survey for the urban area.

We may consider trip flows Q_i that respond to quality of service and fare on the basis of the generalized cost, via a demand function denoted D_i.

2.4 The usage of supply by demand

Each individual trip contributes to the aggregate indicators of its O-D pair. To depict its path along the transit network, let us get again inspired by traffic assignment and decompose the path into legs, i.e. by walking sub-path or by line section between two stations of vehicle entry and exit. To account for transferring between regions along a "concrete" leg, we shall extend the definition of a leg by considering fictive stations at the points where the line crosses the frontier between two neighbouring regions.

Let us denote:

+ Λ^i_r the number of legs travelled in component r by O-D pair i.

+ λ^i_r the average travelled distance along such a leg.

+ $T^i_r = \Lambda^i_r.\lambda^i_r$ the traffic of demand segment i in component r.

Flow conservation is assumed to hold in the following way:

+ by origin (resp. destination) region, $I^i_R = Q_i$ (resp. $O^i_R = Q_i$).

+ between neighbouring regions r and s, the leg flow associated to segment i is proportional to the segment flow and constitutes both the leg flow going from r to s, and that coming in s from r.

We also take demand behaviour concerning mode and path choice as exogenous, by assuming constant ratios I^i_r / Q_i, O^i_r / Q_i, M^i_{sr} / Q_i (also with hats), Λ^i_r / Q_i and B^i_r / Q_i. Such ratios will be hereafter denoted as x^X_{ri} or x^X_{rsi} by stage X in {A, E, T, R, W} along the trip sequence. Of course, this postulate imposes the topological structure of the utilized paths and makes it strongly dependent on the structure of the services supplied in the reference scenario of model application.

2.5 Network access

The access distance between trip endpoints and network stations plays an important, though external, role in transit quality of service. Three principles are postulated here:

(1) that trip endpoints are evenly distributed across space,

(2) that each user chooses the "nearest to endpoint" station of the selected mode in his region,

(3) that the walk pathways make up a Manhattan grid.

These principles enable us to derive the statistical distribution of walking distance in the following way. Let x (resp. y) denote the "longitudinal" (resp. "transversal") distance between endpoint and station along the transit line (resp. perpendicular to it). From principle (2), x is uniformly distributed in $\left[0, \frac{1}{2}S\right]$ (recall that S denotes station spacing).

Let also ℓ denote the line stretch spanning the region area A in the dimension along the line axis, n the number of such stretches in the region. Similarly, in the other dimension of space,

let $h = \mathrm{A}/\ell$ be the spanning length (since $\mathrm{A} = h.\ell$ under sufficiently rectangular region shape) and m the number of stretches (Cf. figure 1): then y is uniformly distributed in $\left[0, \frac{1}{2}Y\right]$ wherein $Y = h/n$ is the spacing between neighboring parallel lines in that dimension.

For a mode that spans the region in one dimension only, it holds that $n.\ell = L$ so that

$$Y = \frac{\ell}{L}.\frac{\mathrm{A}}{\ell} = \frac{\mathrm{A}}{L}. \tag{2.1}$$

In this case we set the shape parameter χ_r to a value of 1 for the component.

For a mode that spans the region in two dimensions with rectilinear enough lines, then $n.\ell + m.h = L$: if $n = h/Y$ and $m = \ell/Y$ then $Y = 2\,\mathrm{A}/L$. Let then $\chi_r = 2$.

Whatever the case, the average access distance along the Manhattan grid satisfies the following functional relationship with respect to the component station spacing S_r, line stretch L_r, shape parameter χ_r:

$$d_{Ar} \equiv \mathrm{E}[\mathrm{x}+\mathrm{y}] = \frac{1}{4}(S_r + \frac{\chi_r.\mathrm{A}_r}{L_r}). \tag{2.2}$$

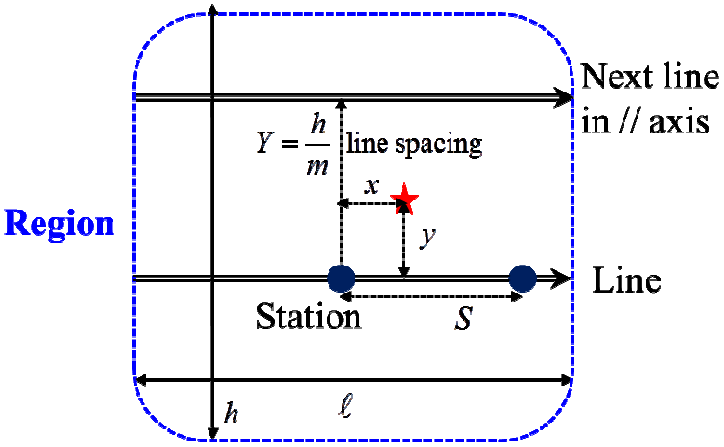

Fig. 1. Access distance between trip endpoints and component stations.

3. Causal relationships and model structure

So far we have focused on basic definitions to describe the subsystems of Supply, Quality of Service and Demand on the basis of quantitative variables, which are "state variables" in the system model. We shall now link together the state variables by causal relationships that depict the influences between the variables as factors or outcomes. Each relationship involves one or several influences of variables as factors, onto one dependent variable.

The set of variables, taken as nodes, together with the set of influences, taken as oriented links from factor to outcome, constitute a causal network that is the logical structure of our model. We shall illustrate the causal sub-network by subsystem, concerning respectively the supply side (§ 3.1), Quality of service (§ 3.2), Usage and demand (§ 3.3), Traffic interactions (§ 3.4). Lastly, putting together the sub-networks obtains the overall model structure (§ 3.5).

3.1 Supply side

There are four causal relationships between the state variables that characterize a given mode in a given region: they determine respectively (1) station spacing, (2) access distance, (3) commercial vehicle speed, (4) service frequency.

First, station spacing S_r depends on both the line length, L_r, and the number of stations, σ_r, since it satisfies that $S_r.\sigma_r = L_r$. Thus,

(Supply-1) $$S_r = \frac{L_r}{\sigma_r}. \tag{3.1}$$

Second, the accessibility of places to a given station at the region level is characterized by the average access distance to nearest station in the mode-region component: from (2.2), we have

(Supply-2) $$d_{Ar} = \frac{1}{4}(S_r + \frac{\chi_r.\mathrm{A}_r}{L_r}). \tag{3.2}$$

Third, the effective commercial speed that applies on average to all legs in component r, v_r, depends on the free-flow running speed, denoted V_r, and the dwell time by station, denoted ω_r. By unit of distance, there are on average $1/S_r$ stations hence dwellings per service runs, so that

(Supply-3) $$\frac{1}{v_r} = \frac{1}{\mathrm{V}_r} + \frac{\omega_r}{S_r}. \tag{3.3}$$

Fourth, service frequency φ_r stems from line length L_r, commercial speed v_r and the size N_r of the employed fleet. In a period of given duration H, the cycle time of a fictive, average vehicle for a return trip along the full component length amounts to

(Supply-4a) $$C_r = 2\frac{L_r}{v_r}. \tag{3.4}$$

There are $\varphi_r.\mathrm{H}$ such cycles runs accomplished by the N_r vehicles, so that

$$N_r.\mathrm{H} = \varphi_r.\mathrm{H}.C_r, \tag{3.5}$$

Which leads to

(Supply-4b) $$\varphi_r = \frac{N_r}{C_r} = \frac{N_r v_r}{2L_r}. \tag{3.6}$$

Taking χ_r, V_r and ω_r as exogenous parameters, the three variables L_r, σ_r and N_r are basic in the sense that they determine the other variables while being mutually independent. Figure 2 depicts the influences on the supply side in a network structure. It also shows the transfer distance d_{rs}^{T} that applies to transfer steps in individual trips, which is here taken on average and as exogenous parameter.

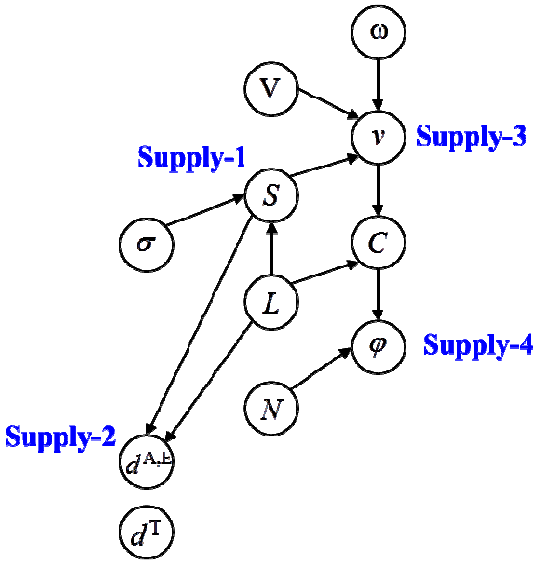

Fig. 2. Causal relationships on the supply side.

3.2 Quality of service

A trip made by an individual user travelling on O-D pair i involves access from origin point to network, egress from network to destination point, as well as running along line legs and waiting for each transit leg, plus transfer time for each (effective) transfer. The trip quality of service can be characterized by four kinds of user's time spent on each kind of step, namely (1) Access or egress, (2) Waiting, (3) Running, (4) Transfer.

First, access time $t^{\mathrm{A}}_{r,i}$ (resp. egress time $t^{\mathrm{E}}_{r,i}$) depends on access distance d^{A}_r and also walking speed w_i:

(Quality-1)
$$t^{\mathrm{A}}_{r,i} = \frac{d^{\mathrm{A}}_r}{w_i} \text{ and } t^{\mathrm{E}}_{r,i} = \frac{d^{\mathrm{A}}_r}{w_i}. \tag{3.7}$$

Second, the average wait time prior to boarding at a given station, denoted t^{W}_r, depends on service frequency φ_r in a straightforward way:

(Quality-2)
$$t^{\mathrm{W}}_r = \frac{\beta_r}{2\varphi_r}. \tag{3.8}$$

This formula holds for $\beta_r = 1$ depicting perfectly regular service headways between successive vehicles. Reciprocally, for "memoryless" service with exponentially distributed headways, it holds that $t^{\mathrm{W}}_r = 1/\varphi_r$, which justifies value $\beta_r = 2$ to depict that case.

Third, leg run time $t^{\mathrm{R}}_{r,i}$ involves leg length, λ^i_r, and commercial speed, v_r, in a simple way: on average

(Quality-3)
$$t^{\mathrm{R}}_{r,i} = \frac{\lambda^i_r}{v_r}. \tag{3.9}$$

Fourth, let us denote $t^{\mathrm{T}}_{rs,i}$ an average transfer time. For an i-bound user transferring from component r to component s, it depends typically on the walk distance d^{T}_{rs} and the walking speed of the users, w_i:

(Quality-4) $$t^{\mathrm{T}}_{rs,i} = \frac{d^{\mathrm{T}}_{rs}}{w_i}\,. \tag{3.10}$$

Relationships (Quality-1-4) depict quality of service in terms of user times. They are illustrated on Figure 3.

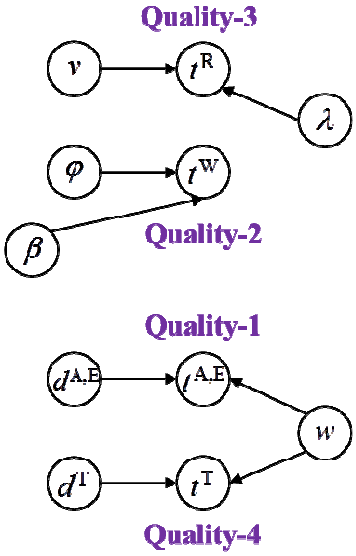

Fig. 3. Causal relationships for Quality of service.

3.3 Demand side

In microeconomic theory, the "demand function" relates the volume of demand for a desired good to its price and, eventually, quality factors. As concerns transportation, quality of service plays a major role so that the dominant model involves a demand function in two steps: first, the evaluation of a "generalized cost" that synthesizes the price and the quality of service factors into a single value meaningful to the individual user, second, linking the generalized cost to the demand volume by a specific relationship.

Here we shall keep to this two-step demand model. Denoting by γ^{X}_i the value of time associated to each time unit spent in a given state X in {A/E, W, T, R} (standing for Access/Egress, Wait, Transfer, Running, respectively) by a user on segment i, on average to such users the generalized cost of a trip amounts to

(Demand-1) $$g_i = \tau_i + \gamma^{\mathrm{A}}_i t^{\mathrm{A}}_{\mathrm{R}i} + \gamma^{\mathrm{E}}_i t^{\mathrm{E}}_{\mathrm{R}i} + \gamma^{\mathrm{T}}_i t^{\mathrm{T}}_{\mathrm{RR},i} + \gamma^{\mathrm{W}}_i t^{\mathrm{W}}_{\mathrm{R}i} + \gamma^{\mathrm{R}}_i t^{\mathrm{R}}_i\,. \tag{3.11}$$

Wherein τ_i is the tariff and the following times are average quantities:

$$t^{\mathrm{A}}_{\mathrm{R}i} = \left(\sum_{r\in\mathrm{R}} I^i_r\, t^{\mathrm{A}}_{ri}\right)/\,Q_i\,. \tag{3.12}$$

$$t^{\mathrm{E}}_{\mathrm{R}i} = \left(\sum_{r\in\mathrm{R}} O^i_r\, t^{\mathrm{E}}_{ri}\right)/\,Q_i\,. \tag{3.13}$$

$$t^{\mathrm{W}}_{\mathrm{R}i} = \left(\sum_{r\in\mathrm{R}} B^i_r\, t^{\mathrm{W}}_r\right)/\,Q_i\,. \tag{3.14}$$

$$t^{\mathrm{R}}_{\mathrm{R}i} = \left(\sum_{r\in\mathrm{R}} T^i_r / v_r\right)/\,Q_i\,. \tag{3.15}$$

$$t^{\mathrm{T}}_{\mathrm{RR}\,i} = \left(\sum_{r,s\in\mathrm{R}} \hat{M}^i_{rs}\, t^{\mathrm{T}}_{rs,i}\right)/\,Q_i\,. \tag{3.16}$$

Then, the demand function D_i relates demand volume Q_i to g_i:

(Demand-2) $$Q_i = \mathrm{D}_i(g_i)\,. \tag{3.17}$$

The Assignment ratios $x_{ri}^{\mathrm{X}} \in \left\{ I/Q, O/Q, B/Q, \hat{M}/Q, T/Q \right\}$ are taken as exogenous for each demand segment. So, the dependence of the flow quantities $F_{ri}^{\mathrm{X}} \in \left\{ I, O, B, \hat{M}, T \right\}$ on Q_i and the respective ratios x_{ri}^{X} amounts to an assignment function that can be denoted in a generic way as follows:

(Demand-3) $$F_{ri}^{\mathrm{X}} = x_{ri}^{\mathrm{X}}.Q_i\,. \tag{3.18}$$

Figure 4 depicts the sub-network of causal influences for the demand side.

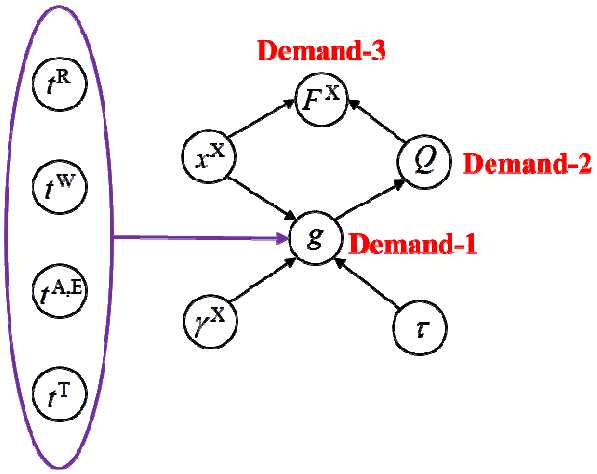

Fig. 4. Causal relationships concerning Demand.

3.4 From Usage to Traffic conditions

Let us now turn to the features of traffic that stem from the aggregation of individual trips into local flows on network elements. Such flows meet service capacity of two kinds: first, at station dwelling of transit vehicles, second aboard these vehicles.

On average over lines and stations in a component during the period under study, the dwelling time spent by a vehicle at a station depends on the numbers of boarding and alighting passengers, respectively, and also on the exchange capacity of the vehicle, denoted k_r^{AB} and counted as the number of passenger channels along a vehicle side (i.e. number of doors times their average number of passenger streams).

As there are $\sigma_r \varphi_r$ such dwelling stops per unit time and traffic direction, the average number of boarding people per station and line and direction amounts to the following:

$$b_r = B_r^{\mathrm{I}} / (\sigma_r \varphi_r \mathrm{H}), \text{ wherein } B_r^{\mathrm{I}} \equiv \textstyle\sum_{i \in \mathrm{I}} B_r^i\,. \tag{3.19}$$

Similarly, the average number of alighting people per stop amounts to:

$$a_r = A_r^{\mathrm{I}} / (\sigma_r \varphi_r \mathrm{H}), \text{ wherein } A_r^{\mathrm{I}} \equiv \textstyle\sum_{i \in \mathrm{I}} A_r^i\,. \tag{3.20}$$

Assuming individual passage time of p_r^a or p_r^b on average for alighting and boarding, respectively, the overall passenger time spent at exchange adds up to $a_r p_r^a + b_r p_r^b$ and is split between the passenger channels of the vehicle, yielding vehicle dwell time as follows:

(Traffic-1) $$\omega_r = \omega_r^0 + \frac{a_r p_r^a + b_r p_r^b}{k_r^{AB}}, \tag{3.21}$$

Wherein ω_r^0 is a fixed, minimum time that accounts for reduced speed at deceleration then acceleration as well as for door opening and closing.

Aboard the vehicle, the average passenger flow determines both seat availability and the standing area available on average per standing passenger. Let us denote:

+ y_r the average passenger load per vehicle in component r, which is split into $\underline{y}_r$ seated passengers and $\bar{y}_r$ standing passengers,

+ $\underline{k}_r$ the seat capacity aboard a vehicle in component r and $\bar{k}_r$ the standing capacity,

In the component during the study period of length H, the vehicles travel a total distance of $N_r v_r \mathrm{H}$ while users from segment i travel a total distance of T_r^i, so it comes out that

$$y_r = \frac{T_r^{\mathrm{I}}}{N_r v_r \mathrm{H}}, \text{ with } T_r^{\mathrm{I}} = \sum_{i \in \mathrm{I}} T_r^i \,. \tag{3.22}$$

From this stems the seated density $\underline{\rho}_r$ and the standing density $\bar{\rho}_r$, respectively:

$$\underline{\rho}_r = \min\left\{U_r, \frac{y_r}{\underline{k}_r}\right\}, \tag{3.23}$$

$$\bar{\rho}_r = \frac{(y_r - \underline{k}_r \underline{\rho}_r)^+}{\bar{k}_r} \,. \tag{3.24}$$

In these formulas, coefficient U_r would have value 1 were the passengers evenly distributed among the vehicle runs and between the stations and along the vehicles. A value less than 1 can be given to take unevenness into account.

It thus remains to account for local densities in passenger discomfort, by way of specific cost functions $\underline{\gamma}_r^{\mathrm{R}}$ and $\bar{\gamma}_r^{\mathrm{R}}$:

(Traffic-2)
$$\gamma_r^{\mathrm{R}} = \frac{\underline{k}_r \underline{\rho}_r \underline{\gamma}_r^{\mathrm{R}}(\underline{\rho}_r) + (y_r - \underline{k}_r \underline{\rho}_r)\bar{\gamma}_r^{\mathrm{R}}(\bar{\rho}_r)}{y_r} \,. \tag{3.25}$$

Another capacity phenomenon may be of interest: that of track capacity, especially for railway modes. A relevant indicator is track reservation η_r due to both track occupancy at station dwelling and safety margins: on average, denoting by ϖ_r a minimum headway time taken as safety margin between successive vehicles, it holds that.

$$\eta_r = \varphi_r . (\omega_r + \varpi_r) \,. \tag{3.26}$$

This is bound to influence vehicle running speed, through a specific relationship as follows:

(Traffic-3)
$$\mathrm{V}_r = \hat{\mathrm{V}}_r(\eta_r) \,. \tag{3.27}$$

Figure 5 illustrates the sub-network of causal relationships for traffic phenomena. Additional traffic phenomena could be modelled, for instance the increase of passenger wait time for vehicle boarding when the ratio of boarding flow to the product of exchange capacity by the system time allowed for boarding approaches 1.

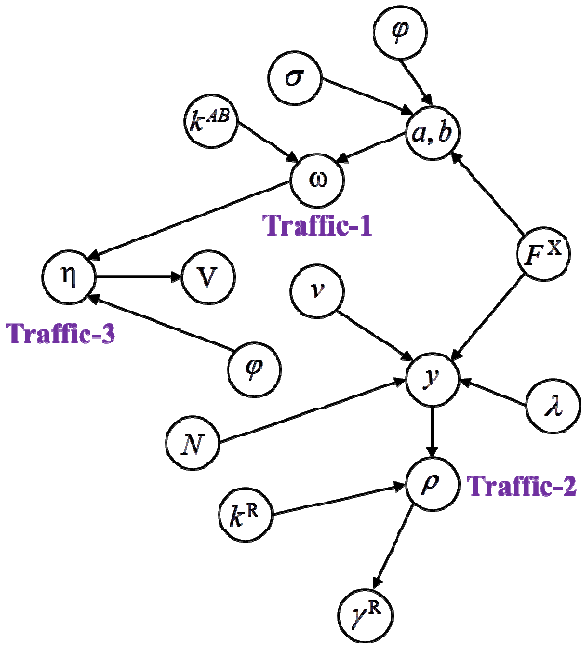

Fig. 5. Causal relationships concerning Usage and Traffic.

3.5 Model architecture

By subsystem, we have identified the model elements and the causal relationships that put them in straightforward interaction; a graphical synthesis has been provided. We can now bring together the subsystems: this is done graphically, too, by the diagram in Figure 6 (which provides justification for the pre-set positions in the previous diagrams by subsystem). The overall diagram displays the model architecture that mimics the system's logical structure. The diagram also demonstrates the system complexity and the pragmatic need to model it in order to manage it.

The model architecture constitutes an important outcome of our systemic analysis. We have made the causal influences explicit in it so as to respect the "complexity of meaning" inherent to the system. The systemic diagram of Van Nes and Bovy (2000) was a useful reference: however it does not include the technical relationships and it represents the supply side in terms of "network spatial properties" such as spatial accessibility, instead of action levers such as fleet size.

Complementarily, we have endeavoured to reduce the "complexity of abundance" by taking a statistical, generic approach to detailed features such as point location, modal elements, individual passengers and their trips. This approach stems straightforwardly from previous analytical studies at Delft Technical University in the 1990s and 2000s (Van Nes, 2002).

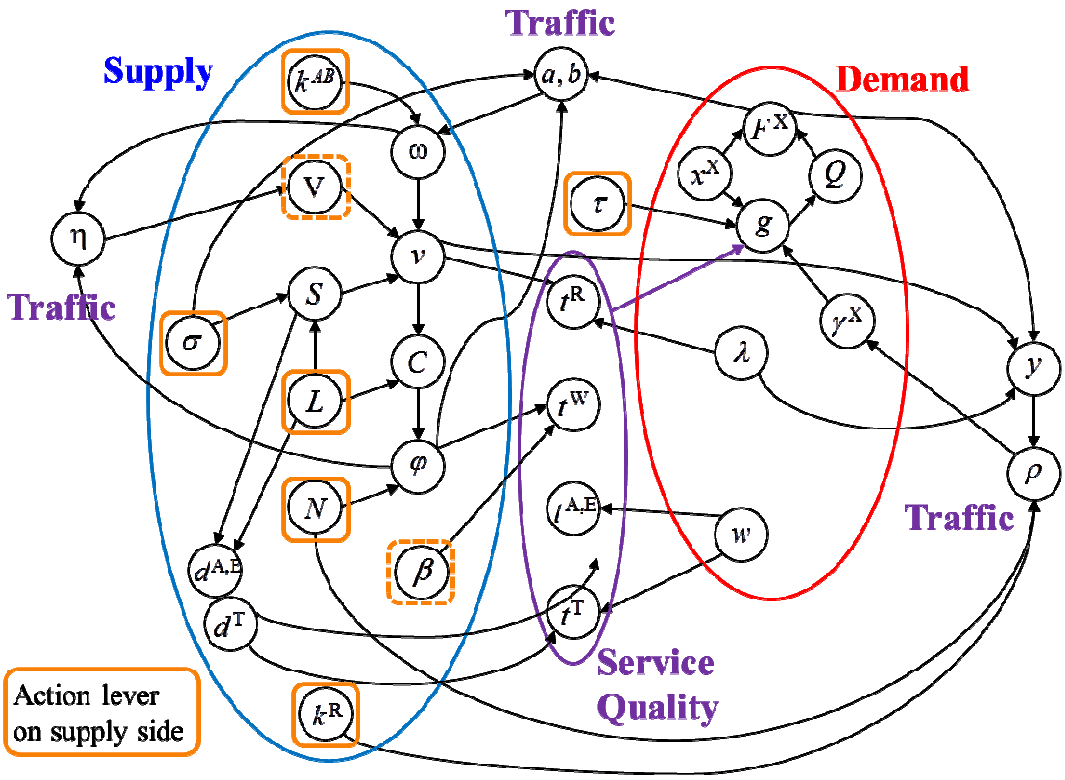

Fig. 6. Causal diagram of the model.

4. Economic analysis

At that stage, the causal model is above all a technical one as it describes the system elements and their straightforward interactions. It has also some economic significance on the demand side since users' individual generalized costs are modelled and determine demand volumes. In this Section, we set up the economic rationale for network management by turning to economic impacts and agents (§ 4.1) and identifying values (§ 4.2) and costs (§ 4.3) to them. We then build up the function of system cost that synthesizes the system technical and economic performance (§ 4.4).

4.1 System impacts as stakes for economic agents

Two kinds of economic agents are primarily involved in system operations and performance: on the demand side the individual users can make use of services and benefit from their availability, whereas on the supply side there are service providers (say operators) that combine the technical means (infrastructure, vehicles and operations) in order to deliver the commercial services.

In fact, the system impacts are not limited to its technical and commercial operations. We have identified the issues of tariffs that cost to the users but yield commercial revenues to the suppliers, of service quality that benefits to the users yet involves technical means hence impose costs to the operators, of passenger traffic that interplays with productive operations.

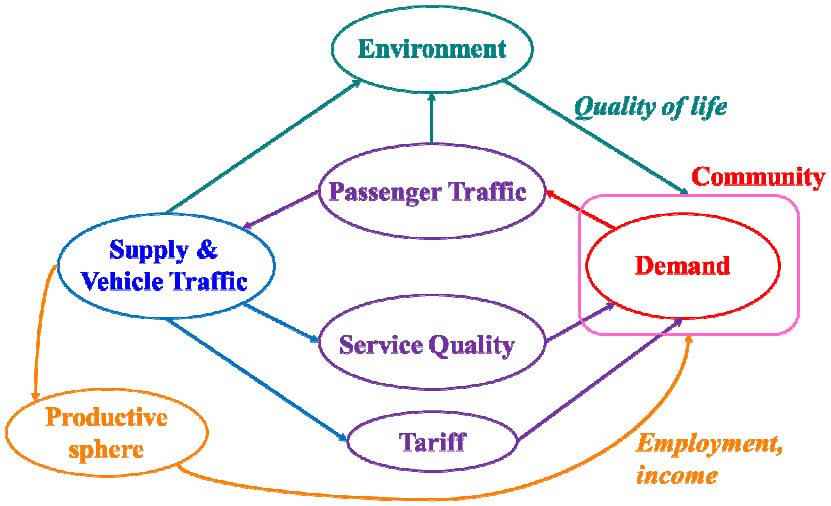

Fig. 7. Synthesis of system impacts and agents.

Fig. 7 summarizes the model architecture in these respects and introduces two additional categories of economic agents, namely the Environment and the Productive sphere. As is well-known (e.g. Combes and Leurent, 2016), transportation – especially traffic operations – exerts a series of impacts on the environment: at the local level, there are pollutant emissions from thermic motors and noise emissions, the consumption of energy and matters, whereas at the global level there is the contribution to greenhouse gases together with other broad consequences of resource depletion and pollution of natural resources (soil, water, air). These impacts are external to the interaction between supply and demand: they concern the Community as a whole, as the environmental quality is important in the Quality of life and its sustainability. Although the Community of course includes the service users, it may be better thought of as the set of residents that avail themselves of the environmental conditions locally and benefit from Quality of life.

As concerns the Productive sphere, it should be kept in mind that the money flows that stem from operators' expenditure hence as costs to them go to other economic agents as commercial revenues, hence as value streams (e.g. Leurent and Windisch, 2015). The relevant set of agents includes not only the resource providers (infrastructure builders, vehicle makers, maintenance, cleaning, energy provision…) but also the employees of the transportation firms that belong to the Community and get wages as income to them. The interaction between the transportation system and the Productive sphere is a system externality of socio-economic kind.

For the sake of simplicity we shall hereafter restrict our scope to Supply and Demand in interaction. However the two broad kinds of externalities, environmental versus socioeconomic, should be included in further analysis to better compare alternative transit modes in a truly multicriteria framework.

4.2 Agents' Values

By demand segment i.e. here by O-D pair i, the generalized cost to an individual user may be expressed in a generic way as follows:

$$g_i = \tau_i + \sum_X \gamma^X t^X x^X \,, \tag{4.1}$$

In which the summation over X denotes the different stages along a trip (Access / Egress, Wait, Run, eventual Transfers). The cost of each stage is basically a time cost γ^X multiplied

by a physical time $t^{\mathrm{X}}x^{\mathrm{X}}$ in which x^{X} is an assignment factor. Index i as well as the location details (components, transfers within / between components) have been omitted in the summation so as to preserve legibility. The money part of the cost, τ_i, might also be decomposed along the trip sequence, were the tariff based on the effective usage.

In transport economics, it is fairly standard to take the generalized cost as the argument of the demand function, which is the basic tool to evaluate consumer surplus.

As an elemental demand δQ just above level $Q = \mathrm{D}(G)$ would take the service at cost $G - \delta G$ but not at cost G, its own gross surplus belongs to $\left[\mathrm{D}^{(-1)}(Q+\delta Q), \mathrm{D}^{(-1)}(Q)\right]$ - say it amounts to $\mathrm{D}^{(-1)}(Q)$, while the net surplus at cost G is null. As the full demand of the segment is made up of all elements with disaggregate gross surplus higher than G, its aggregate gross surplus is the following:

$$\mathrm{GUS}_i = \int_0^{Q_i} \mathrm{D}_i^{(-1)}(q)\,\mathrm{d}q\,, \tag{4.2}$$

with associated aggregate net surplus of:

$$\mathrm{NUS}_i \equiv \mathrm{GUS}_i - Q_i.G_i = \int_{G_i}^{\infty} \mathrm{D}_i(g)\,\mathrm{d}g\,. \tag{4.3}$$

By adding up the segment gross (resp. net) surplus over the demand segments $i \in \mathrm{I}$, we obtain the global demand gross (resp. net) surplus.

As concerns the operators, the primary source of value for them consists in the commercial revenues from paid fees, i.e. $\sum_{i\in\mathrm{I}} Q_i\tau_i$. This is often supplemented by government subsidies and always diminished by taxes.

Let us take a net-of-tax approach: as subsidies are often related to usage, they constitute additional revenues analogous to those from tariff fees. The combined value of commercial fare τ_i and the "related" subsidy $\Delta\tau_i$ may be denoted τ'_i, so that the net-of-tax operators' revenue amounts to

$$\mathrm{NOR} = \sum_{i\in\mathrm{I}} Q_i\,\tau'_i\,. \tag{4.4}$$

4.3 Agents' Costs

The subsidizing authority bears a cost of $\Delta\tau_i$ per trip on segment i, so it bears an aggregate cost of

$$\mathrm{AC} = \sum_{i\in\mathrm{I}} Q_i\,\Delta\tau_i\,. \tag{4.5}$$

Each user according to his segment i incurs a generalized cost of g_i. By aggregation, users' generalized costs amount to

$$\mathrm{GUC} = \sum_{i\in\mathrm{I}} Q_i\,g_i\,. \tag{4.6}$$

The tariff part of that can be called also the money costs to users,

$$\mathrm{MUC} = \sum_{i\in\mathrm{I}} Q_i\,\tau_i\,. \tag{4.7}$$

Net of tariffs, the users' costs are time-based costs:

$$\mathrm{TUC} = \sum_{i\in\mathrm{I}} Q_i\,(g_i - \tau_i) = \sum_{i\in\mathrm{I}} Q_i \sum_{r\in\mathrm{R,X}} \gamma_{ri}^{\mathrm{X}}\,x_{ri}^{\mathrm{X}}\,t_{ri}^{\mathrm{X}}\,. \tag{4.8}$$

On their side, operators bear production costs that include holding costs for the purchase or hiring and the maintenance of their production means, together with operational costs for the

means' utilization, which involves the consumption of consumable goods such as energy. We shall analyse the operators' costs primarily with respect to cost-inductors i.e. by kind of means either linear infrastructure or station or vehicle.

The associated cost inductors are the supply-related variables, L_r, σ_r and N_r, respectively. The distinction of the component enables us to specify the mode and also the region in the urban area, in order to emphasize the cost of ground space which is higher in more central places. At this stage, we associate unit costs denoted c_{Lr}, $c_{\sigma r}$ and c_{Nr} to each basic supply variable taken by period of length H. Every such unit cost involves holding (capital) as well as utilization (operations) costs. The resulting operators' costs can be evaluated by component,

$$OC_r = c_{Lr}.L_r + c_{\sigma r}.\sigma_r + c_{Nr}.N_r, \tag{4.9}$$

and aggregated other all transit modes and the whole urban area as follows:

$$OC = \sum_{r\in R} OC_r = \sum_{r\in R} c_{Lr}.L_r + c_{\sigma r}.\sigma_r + c_{Nr}.N_r . \tag{4.10}$$

4.4 System costs

Let us now sum up the costs minus the values to all of the agents in the system – authority, users and operators, under the limitations mentioned earlier about which impacts are here considered.

The authority bears costs AC as stated in (4.5), neglecting both environmental costs and the values yielded to the community by taxes, impositions, accessibility advantages and effects on the productive sphere. The users benefit from gross surplus GUS that aggregates those of the demand segments as stated in (4.2),

$$GUS = \sum_{i\in I} GUS_i = \sum_{i\in I} \int_0^{Q_i} D_i^{(-1)}(q)\,dq , \tag{4.11}$$

By subtracting users' costs GUC given in (4.6), the net users' surplus amounts to:

$$NUS = \sum_{i\in I} NUS_i = \sum_{i\in I} \int_{G_i}^{\infty} D_i(g)\,dg . \tag{4.12}$$

The operators get commercial and subsidized revenues of NOR and bear costs OC, so their net profit amounts to

$$OP = NOR - OC . \tag{4.13}$$

The overall system cost is defined as the sum of its agents' costs net from the surpluses and revenues, i.e.:

$$\begin{aligned} SC &= AC - NUS - OP , \\ SC &= AC + GUC - GUS - N - OR + OC , \\ SC &= TUC + OC - GUS , \end{aligned} \tag{4.14}$$

Since OR is compensated in AC and the money costs to the users, MUC = GUC – TUC.

Function SC encompasses the overall net costs of the transit system under the exogenous structure of demand and the supply set-ups (the basic variables L_r, σ_r, N_r, τ_i and the derived variables). Thus it is taken as the objective function to be minimized in order to optimize system planning.

5. User equilibrium and system optimization

Having modelled economic agents with the values and costs to them, we are now ready to study their joint behaviour and the scope for system management. To do that, we shall first synthesize the model even further owing to some mathematical formalization (§ 5.1); this enables us to deal with the determination of the endogenous system state on the basis of an exogenous supply plan (§ 5.2). Then we address the issue of system optimization by providing an analytical scheme to determine an optimal supply plan (§ 5.3).

5.1 Mathematical synthesis

Whereas the causal model in fact a synthesis of the technical features of the system, the unit costs constitute a synthesis of many economic features. Figure 8 encompasses the technical and economic features so far studied. The block diagram in it characterizes the straightforward influences between the system characteristics. By chaining influences along sequences, e.g. from supply variables S_r of station spacing, to users' waiting times t^{W}_{ri}, passing by commercial speed v_r, cycle time C_r and service frequency φ_r, we can identify *influence paths*. In other words, figure 8 depicts the topological structure of influences within the system.

The vector of supply variables, $\mathbf{S}_{\mathrm{X}} = \left[(L_r, \sigma_r, N_r, \beta_r, k_r^{\mathrm{AB}}, \bar{k}_r, \underline{k}_r)_{r\in\mathrm{R}}, (\tau_i)_{i\in\mathrm{I}}\right]$, is controlled by the operator: hence the X subscript, standing for Exogenous. The rest of the supply variables constitute the vector of Endogenous supply characteristics, denoted $\mathbf{S}_{\mathrm{N}}$ with subscript N:

$\mathbf{S}_{\mathrm{N}} = \left[(S_r, C_r, \varphi_r, \omega_r, v_r, d_r^{\mathrm{A}})_{r\in\mathrm{R}}, (d_{rs}^{\mathrm{T}})_{r,s\in\mathrm{R}}\right]$: it depends on $\mathbf{S}_{\mathrm{X}}$ and also on the vector of traffic variables, $\mathbf{T} = \left[(\eta_r, a_r, b_r, y_r, \underline{y}_r, \bar{y}_r, \underline{\rho}_r, \bar{\rho}_r)_{r\in\mathrm{R}}\right]$.

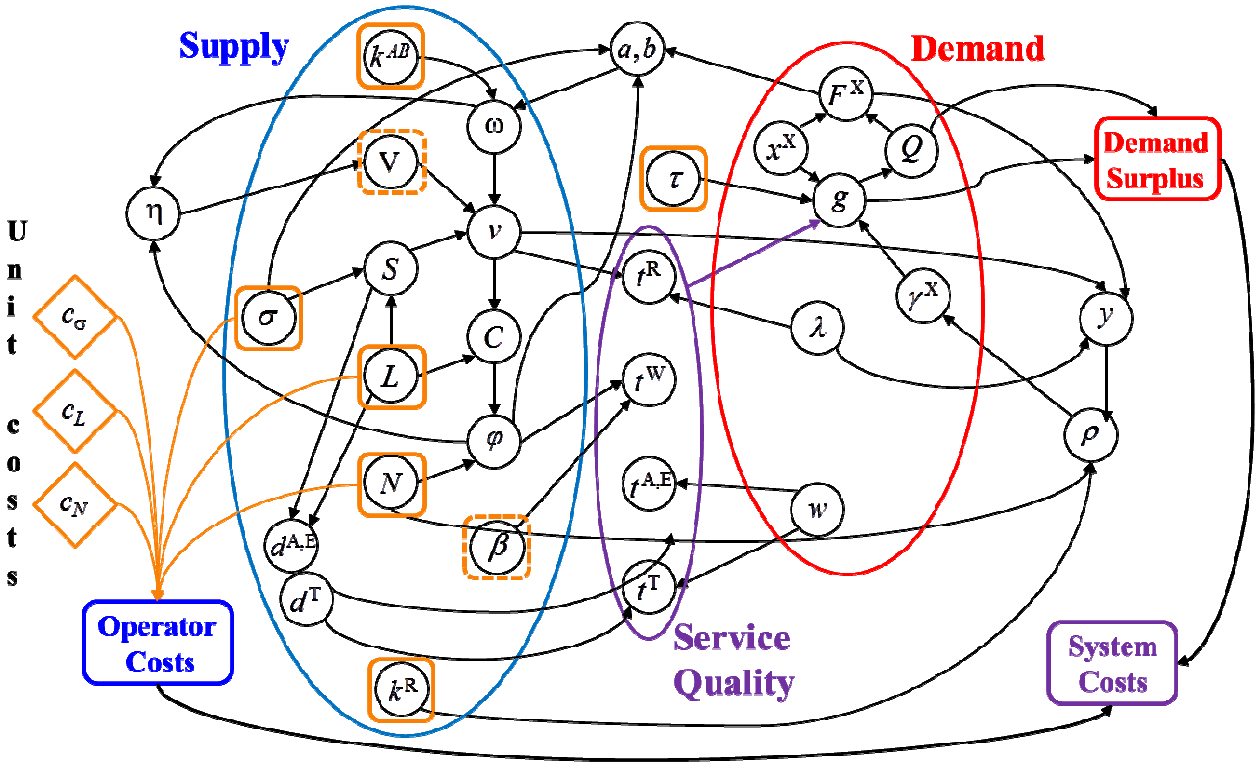

Fig. 8. Diagrammatic model of technical and economic system.

Let us also define:

$$\mathbf{t} = \left[(t^{\mathrm{A}}_{ri}, t^{\mathrm{E}}_{ri}, t^{\mathrm{R}}_{ri}, t^{\mathrm{W}}_{ri})_{r\in\mathrm{R},i\in I}, (t^{\mathrm{T}}_{rs,i})_{r,s\in\mathrm{R},i\in I} \right]$$

$$\boldsymbol{\Gamma} = \left[(\gamma^{\mathrm{A}}_{ri}, \gamma^{\mathrm{E}}_{ri}, \gamma^{\mathrm{R}}_{ri}, \gamma^{\mathrm{W}}_{ri})_{r\in\mathrm{R},i\in I}, (\gamma^{\mathrm{T}}_{rs,i})_{r,s\in\mathrm{R},i\in I} \right]$$

$$\mathbf{g} = \left[(g_i)_{i\in I} \right]$$

$$\mathbf{Q} = \left[(Q_i)_{i\in I} \right]$$

The combined vector $\mathbf{V} = [\mathbf{S}_{\mathrm{X}}, \mathbf{S}_{\mathrm{N}}, \mathbf{T}, \mathbf{t}, \mathbf{g}, \mathbf{Q}, \Gamma]$ constitutes the state vector of the system. $\mathbf{S}_{\mathrm{X}}$ is its exogenous part, whereas the rest of the vector, denoted as $\mathbf{V}_{\mathrm{N}}$, is endogenous.

Let us now use vector notation and mathematical functions to described the modeled influences in a still more synthetic way: denoting $\mathbf{S} = [\mathbf{S}_{\mathrm{X}}, \mathbf{S}_{\mathrm{N}}]$,

$\mathbf{S}_{\mathrm{N}} = \mathbf{F}_{\mathrm{SN}}(\mathbf{S}_{\mathrm{X}}, \mathbf{T})$ is the supply function, (5.1a)

$\mathbf{t} = \mathbf{F}_{\mathrm{t}}(\mathbf{S})$ is the user time function, (5.1b)

$\mathbf{T} = \mathbf{F}_{\mathrm{T}}(\mathbf{S}, \mathbf{Q})$ is the traffic function, (5.1c)

$\boldsymbol{\Gamma} = \mathbf{F}_{\Gamma}(\mathbf{T})$ is the comfort function, (5.1d)

$\mathbf{Q} = \mathbf{F}_{\mathrm{Q}}(\mathbf{g})$ is the demand function, (5.1e)

$\mathbf{g} = \mathbf{F}_{\mathrm{g}}(\boldsymbol{\tau}, \mathbf{t}, \Gamma)$ the user cost function, (5.1f)

it involves $\boldsymbol{\tau}$ that belongs to $\mathbf{S}_{\mathrm{X}}$ while the rest of $\mathbf{S}$ exert indirect influences via $\mathbf{t}$.

The system of equations (5.1) is illustrated in figure 9 and may be put as a unified function

$$\mathbf{V}_{\mathrm{N}} = \mathbf{F}_{\mathrm{VN}}(\mathbf{S}_{\mathrm{X}}, \mathbf{V}_{\mathrm{N}}) . \quad (5.2)$$

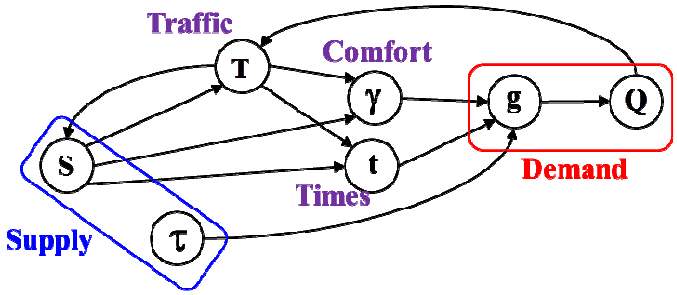

Fig. 9. Synthetic diagram of influences.

5.2 On User and Traffic Equilibrium

Eqn (5.2) synthesizes in a formal way the mutual interaction between the subsystems. It constitutes a Fixed Point Problem (FPP) with respect to endogenous state vector $\mathbf{V}_{\mathrm{N}}$. The FPP characterizes an equilibrium state between supply, traffic and demand, which is akin to the notion of Traffic equilibrium or User equilibrium in the theory of network traffic assignment. Thus, the system state appears as a solution of an FPP of User and Traffic Equilibrium (UTE). Despite an important part of the supply state is endogenous, we avoid to call such a solution state a supply-demand equilibrium because it captures the technical determination of supply but not its economic behaviour.

The determination of an UTE raises the issues of the existence and uniqueness of solutions to the FPP. Loosely speaking, a mathematical proof of existence involves some smoothness property of mapping $\mathbf{F}_{VN}$; this is likely to hold since the constituent elementary functions are continuous almost everywhere. A proof of uniqueness could be based on a property of contraction for the mapping. Both issues, however, are left aside for further research with more mathematical orientation. Hereafter we shall take the two issues as granted.

An important case where this holds obviously is the Uncongested case such that the variations in traffic conditions $\mathbf{T}$ entail variations neither in $\mathbf{\Gamma}$ nor in $\mathbf{S}$. Then, the state determination involves no feedback loops (Cf. figure 10) and is thus straightforward.

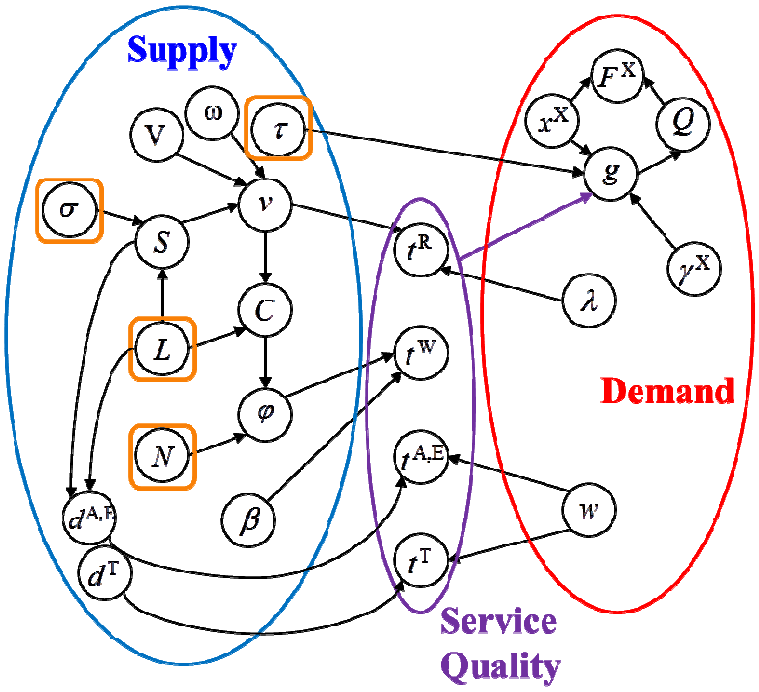

Fig. 10. Synthetic diagram of the uncongested model.

5.3 On System Optimization

The mathematical synthesis enables us to study the issue of system optimization in an analytical way.

System Optimum Problem (SOP): minimize System costs SC with respect to "exogenous supply" state vector $\mathbf{S}_X$, subjected to technical and economic relationships in mapping $\mathbf{F}_{VN}$:

$$\text{Min } SC(\mathbf{S}_X, \mathbf{V}_N) \quad \text{submitted to} \quad \mathbf{V}_N = \mathbf{F}_{VN}(\mathbf{S}_X, \mathbf{V}_N). \tag{5.3}$$

In other words, by solving the mathematical program (5.3) we obtain the optimum levels of supply provision either basic (infrastructure length, station number, fleet size, capacities, tariffs) or derived (station spacing, access distances, commercial speed, service frequency).

A robust approach to solve this kind of problem is to relax the constraints by associating "dual" variables $\boldsymbol{\mu} \equiv [\mu_{SN}, \mu_T, \mu_t, \mu_t, \mu_\Gamma, \mu_g, \mu_Q]$ to each part of them. Assuming that mapping $\mathbf{F}_{VN}$ is differentiable, system optimization requires the following "primal-dual optimality conditions" to hold:

$$\begin{aligned} &\nabla \mathrm{SC} + \boldsymbol{\mu}.(\nabla \mathbf{U} - \nabla \mathbf{F}_{\mathrm{VN}}) = 0 \\ &\mathbf{V}_{\mathrm{N}} = \mathbf{F}_{\mathrm{VN}}(\mathbf{S}_{\mathrm{X}}, \mathbf{V}_{\mathrm{N}}) \end{aligned} \quad . \tag{5.4}$$

Wherein $\mathbf{U} : \mathbf{V}_{\mathrm{N}} \mapsto \mathbf{V}_{\mathrm{N}}$, $\nabla \mathrm{SC}$ and $\nabla \mathbf{F}_{\mathrm{VN}}$ are taken with respect to $\mathbf{V}$, $\nabla \mathbf{U}$ w.r.t $\mathbf{V}_{\mathrm{N}}$.

The primal-dual optimality conditions yield the advantages of a smooth enough state vector and differentiable functions $\mathbf{F}_{\mathrm{VN}}$ and SC. This is an important advantage over the discrete approach to network design, on both analytical and numerical grounds. Of course the discrete approach is superior when it comes to providing detailed features such as the particular location of lines or stations.

The mathematical treatment here is somewhat loose and mostly indicative. We are more interested in the model relevance for planning policies. In fact, policy considerations are likely to impose additional constraints, such as

$(L_r, \sigma_r) \geq (L_r^0, \sigma_r^0)$ To capitalize over system history,

$\varphi_r \geq \varphi_r^0$ To enhance the local level of service,

$d_{Ar} \leq d_{Ar}^{\max}$ To satisfy a requirement of local accessibility in space.

Along with specific political requirements, the major planning issue comes to which action levers are available: for instance, is it feasible to build additional lines, or to vary the tariffs?

6. The uncongested model: economic theory

Our aim here is to address system optimization of the uncongested model and establish some theoretical properties linking the variables (hence the system features in quantitative terms) of interest. These properties stem from optimality conditions by suitable reinterpretation with respect to the variables "under control" on the supply side. A companion objective is to establish the existence and uniqueness of an optimal system state, together with an efficient and straightforward solution scheme.

After stating the economic program of system optimization and its primal-dual conditions of optimality (§ 6.1), we focus on demand flows and tariff optimization (§ 6.2). Then, we deal with the quality of service characteristics that play a crucial role and enable us to separate the supply side by network component (§ 6.3). Next, we address the analytical determination of, successively frequency and fleet size (§ 6.4), the number and spacing of stations (§ 6.5), commercial speed and line length (§ 6.6). To each range of issues is associated a specific relationship of theoretical interest. We then put together the three relationships and emphasize their economic meaning (§ 6.7). Lastly, we establish theoretical properties of existence and uniqueness for a System Optimum State and outline the solution scheme (§ 6.8): the detailed material is given in a specific appendix (cf. Section 9).

6.1 Economic program with primal-dual optimality conditions

The economic program for system optimization involves an objective function SC and a set of constraints as follows:

$$\mathrm{SC} \equiv \sum_{r \in \mathrm{R}} \mathrm{c}_{Lr} . L_r + \mathrm{c}_{\sigma r} . \sigma_r + \mathrm{c}_{Nr} . N_r + \sum_{i \in \mathrm{I}} Q_i (g_i - \tau_i) - \int_0^{Q_i} \mathrm{D}_i^{(-1)}(q) \, \mathrm{d}q \,, \tag{6.1}$$

$$S_r = \frac{L_r}{\sigma_r} \quad \text{with associated multiplier } \mu_{Sr},\ \forall r \in \mathrm{R}, \tag{6.2}$$

$$\frac{1}{v_r} = \frac{1}{\mathrm{V}_r} + \frac{\omega_r}{S_r} \quad \text{with associated multiplier } \mu_{vr},\ \forall r \in \mathrm{R}, \tag{6.3}$$

$$\varphi_r = \frac{N_r v_r}{2L_r} \quad \text{with associated multiplier } \mu_{\varphi r},\ \forall r \in \mathrm{R}, \tag{6.4}$$

$$d_r^{\mathrm{A}} = \tfrac{1}{4}\left(S_r + \frac{\chi_r \mathrm{A}_r}{L_r}\right) \quad \text{with associated multiplier } \mu_{Ar},\ \forall r \in \mathrm{R}, \tag{6.5}$$

$$g_i = \tau_i + \textstyle\sum_{r \in \mathrm{R},\mathrm{X}} \gamma_{ri}^{\mathrm{X}}\, x_{ri}^{\mathrm{X}}\, t_{ri}^{\mathrm{X}},\ \forall i \in \mathrm{I}. \tag{6.6}$$

$$Q_i = \mathrm{D}_i(g_i),\ \forall i \in \mathrm{I}. \tag{6.7}$$

We shall deal with user costs g_i and demand flows Q_i in a straightforward way, without associating specific multipliers to them. In the uncongested case, we take the discomfort costs γ_{ri}^{X} as constant, as well as dwell times ω_r and reference speeds V_r.

We also include a budgetary constraint to compel the system users to contribute to its funding at a floor level pre-set to SR (for Supply Revenue),

$$\textstyle\sum_{i \in \mathrm{I}} Q_i \tau_i \geq \mathrm{SR}, \quad \text{with associated multiplier } \mu_{\mathrm{B}} \geq 0. \tag{6.8}$$

Precisely, the economic program consists in:

$$\text{Min SC with respect to } \boldsymbol{\tau}_{\mathrm{I}} \equiv (\tau_i)_{i \in \mathrm{I}},\ \mathbf{L}_{\mathrm{R}} \equiv (L_r)_{r \in \mathrm{R}},\ \boldsymbol{\sigma}_{\mathrm{R}},\ \mathbf{N}_{\mathrm{R}},\ \mathbf{S}_{\mathrm{R}},\ \varphi_{\mathrm{R}},\ \mathbf{v}_{\mathrm{R}} \text{ and } \mathbf{d}_{\mathrm{R}}^{\mathrm{A}}, \tag{6.9}$$
$$\text{under constraints (6.2-8).}$$

To relax the constraints, we consider the Lagrangian function of the optimization program, namely

$$\begin{aligned} &£(\boldsymbol{\tau}_{\mathrm{I}}, \mathbf{L}_{\mathrm{R}}, \boldsymbol{\sigma}_{\mathrm{R}}, \mathbf{N}_{\mathrm{R}}, \mathbf{S}_{\mathrm{R}}, \varphi_{\mathrm{R}}, \mathbf{v}_{\mathrm{R}}, \mathbf{d}_{\mathrm{R}}^{\mathrm{A}}, \boldsymbol{\mu}_{S\mathrm{R}}, \boldsymbol{\mu}_{\varphi\mathrm{R}}, \boldsymbol{\mu}_{v\mathrm{R}}, \boldsymbol{\mu}_{A\mathrm{R}}, \mu_{\mathrm{B}}) \\ &\equiv \mathrm{SC} - \mu_{\mathrm{B}}\left(\sum_{i \in \mathrm{I}} \mathrm{D}_i(g_i)\tau_i - \mathrm{SR}\right) \\ &+ \sum_{r \in \mathrm{R}} \left\{ \mu_{Sr}\left(S_r - \tfrac{L_r}{\sigma_r}\right) + \mu_{\varphi r}\left(\varphi_r - \tfrac{N_r v_r}{2L_r}\right) + \mu_{vr}\left(\tfrac{1}{v_r} - \tfrac{1}{\mathrm{V}_r} - \tfrac{\omega_r}{S_r}\right) + \mu_{Ar}\left(d_r^{\mathrm{A}} - \tfrac{1}{4}\left(S_r + \tfrac{\chi_r \mathrm{A}_r}{L_r}\right)\right) \right\} \end{aligned} \tag{6.10}$$

The primal-dual conditions for optimality basically consist in equating to zero the first-order derivatives of the Lagrangian function – with a little more sophisticated treatment for the inequality constraint (6.8):

$$\frac{\partial £}{\partial x} = 0 \quad \text{for any primal variable } x, \tag{6.11a}$$

$$\frac{\partial £}{\partial \mu} = 0 \quad \text{for any dual variable } \mu \text{ other than } \mu_{\mathrm{B}}, \tag{6.11b}$$

$$\mu_{\mathrm{B}} \geq 0,\ \frac{\partial £}{\partial \mu_{\mathrm{B}}} \leq 0 \quad \text{and} \quad \mu_{\mathrm{B}} . \frac{\partial £}{\partial \mu_{\mathrm{B}}} = 0. \tag{6.11c}$$

6.2 Demand flows and tariff setting

In a formal way, let us calculate

$$\frac{\partial £}{\partial g_i} = \frac{\partial}{\partial g_i}(-\mathrm{NUS}_i - Q_i\tau_i - \mu_B Q_i\tau_i) \text{ with } \mathrm{NUS}_i = \int_{g_i}^{\infty} \mathrm{D}_i(g)\,\mathrm{d}g$$
$$= Q_i - \tau_i(1+\mu_B)\dot{Q}_i \text{ since } \mathrm{D}_i(g_i) = Q_i \text{ and denoting } \mathrm{D}_i' \text{ by } \dot{Q}_i.$$

Now, as τ_i contributes to the Lagrangian function both per se and via g_i, we derive that

$$\frac{\mathrm{d}£}{\mathrm{d}\tau_i} = \frac{\partial £}{\partial \tau_i} + \frac{\partial £}{\partial g_i}.\frac{\partial g_i}{\partial \tau_i}$$
$$= -Q_i - \mu_B Q_i + Q_i - \tau_i(1+\mu_B)\dot{Q}_i$$
$$= -\mu_B Q_i - \tau_i(1+\mu_B)\dot{Q}_i$$

System optimization with respect to tariffs requires that $\mathrm{d}£/\mathrm{d}\tau_i = 0$, hence that

$$\tau_i = -\frac{\mu_B}{1+\mu_B}\frac{Q_i}{\dot{Q}_i}. \tag{6.12}$$

which is compatible with non-negative tariffs since $\mu_B \geq 0$ and $\dot{Q}_i$ can be expected to be negative. Denoting by $\varepsilon_i \equiv \mathrm{d}\ln Q_i / \mathrm{d}\ln g_i$ the elasticity of demand volume to individual cost, then $\dot{Q}_i / Q_i = \varepsilon_i / g_i$ so the previous condition can be put as

$$\tau_i = -\frac{\mu_B}{1+\mu_B}\frac{g_i}{\varepsilon_i}. \tag{6.13}$$

The main implication is that $(1+\mu_B)\tau_i\dot{Q}_i = -\mu_B Q_i$, so that

$$\frac{\partial £}{\partial g_i} = Q_i + \mu_B Q_i = \bar{\mu} Q_i,$$

Wherein $\bar{\mu} \equiv 1 + \mu_B$.

This enables us to deal with both elastic demand and variable demand in a unified framework, since under fixed demand D_i is constant so that $\partial £ / \partial g_i = Q_i$, which is compatible with (6.13) by setting $\bar{\mu}$ to value 1. Then, the relationship $\bar{\mu} = 1 + \mu_B$ induces that μ_B is set to zero, in due consistency with the irrelevance of the tariffs as drivers of the system state and the neutralization of the budget requirement.

Under variable demand, the relationship between τ_i and g_i can be restated as follows, letting $\theta_i \equiv g_i - \tau_i$ denote the time-based part in the generalized cost:

$$\bar{\mu}\tau_i = (1-\bar{\mu}).(\tau_i + \theta_i)/\varepsilon_i,$$

Hence,
$$(-\varepsilon_i \frac{\bar{\mu}}{\bar{\mu}-1} - 1)\,\tau_i = \theta_i. \tag{6.14}$$

If demand elasticity is constant and uniform among demand segments, then aggregation over them yields a linear relationship between MUC and TUC, namely:

$$\mathrm{TUC} = (-\varepsilon \frac{\bar{\mu}}{\bar{\mu}-1} - 1)\,\mathrm{MUC}, \text{ with } \mathrm{TUC} \equiv \sum_{i\in \mathrm{I}} Q_i\theta_i \text{ and } \mathrm{MUC} \equiv \sum_{i\in \mathrm{I}} Q_i\tau_i.$$

Under fairly common value of $\varepsilon = -.6$, we can expect that TUC and MUC would be of similar order of magnitude at system optimum.

Another consequence of (6.14) concerns the value of the dual parameter $\overline{\mu}$: as both τ_i and θ_i are expected to be positive, then so is their ratio, leading to $-\varepsilon_i\overline{\mu} - \overline{\mu} + 1 > 0$ for every $i \in \mathrm{I}$, hence to

$$\overline{\mu} < \frac{1}{1+\varepsilon_\mathrm{M}}, \text{ with } \varepsilon_\mathrm{M} = \inf\{\varepsilon_i : i \in \mathrm{I}\}.$$

Value $\varepsilon_\mathrm{M} = -.6$ yields an upper bound of $\mu_\mathrm{U} = 2.5$ on $\overline{\mu}$.

Table 2. First-order derivatives of Lagrangian function.

$$\frac{\partial £}{\partial N_r} = \mathrm{c}_{Nr} - \mu_{\varphi r}\frac{v_r}{2L_r} \text{ with } \mu_{\varphi r} = \mathrm{c}_{Nr}\frac{2L_r}{v_r}$$

$$\frac{\partial £}{\partial \varphi_r} = -\frac{\overline{\mu}E_r^\mathrm{W}\beta_r}{2\varphi_r^2} + \mu_{\varphi r},$$

$$\frac{\partial £}{\partial \sigma_r} = \mathrm{c}_{\sigma r} + \mu_{Sr}\frac{L_r}{\sigma_r^2} \text{ with } \mu_{Sr} = -\mathrm{c}_{\sigma r}\frac{\sigma_r^2}{L_r}$$

$$\frac{\partial £}{\partial S_r} = \mu_{Sr} + \mu_{vr}\frac{\omega_r}{S_r^2} - \tfrac{1}{4}\mu_{Ar} \text{ with } \mu_{Ar} = -\overline{\mu}\,E_r^\mathrm{A}$$

$$\frac{\partial £}{\partial d_{Ar}} = \mu_{Ar} + \overline{\mu}\,E_r^\mathrm{A}$$

$$\frac{\partial £}{\partial v_r} = -\frac{\overline{\mu}E_r^\mathrm{R} + \mu_{vr}}{v_r^2} - \mu_{\varphi r}\frac{N_r}{2L_r} \text{ with } \mu_{vr} = -\mu_{\varphi r}\frac{N_r}{2L_r}v_r^2 - \overline{\mu}E_r^\mathrm{R}$$

$$\frac{\partial £}{\partial L_r} = \mathrm{c}_{Lr} - \frac{\mu_{Sr}}{\sigma_r} + \frac{\mu_{Ar}}{4L_r^2}\chi_r\mathrm{A}_r + \mu_{\varphi r}\frac{N_r v_r}{2L_r^2}$$

6.3 Quality of service in quantitative terms

Quality of service to system users is modelled by the time spent in a series of stages, X in {A, W, R, E, T} where letter A stands for Access, E for Egress, R for Run, W for Wait and T for Transfer. We shall omit Transfer times and focus on X in {A, E, R, W} and the associated characteristic times t_r^X spent in a given component r. Characteristic times are useful as intermediary variables between the supply-side factors and the objective function as well as the associated Lagrangian function. Let us then calculate the partial derivatives of the modified objective function $\mathrm{SC}^\# \equiv \mathrm{SC} - \mu_\mathrm{B}(\sum_{i\in\mathrm{I}}\mathrm{D}_i(g_i)\tau_i - \mathrm{SR})$ with respect to characteristic times.

Notionally,

$$\frac{\partial \mathrm{SC}^\#}{\partial t_r^\mathrm{X}} = \sum_{i\in\mathrm{I}}\frac{\partial \mathrm{SC}^\#}{\partial g_i}\frac{\partial g_i}{\partial t_r^\mathrm{X}} = \sum_{i\in\mathrm{I}}\overline{\mu}\,Q_i\,x_{ri}^\mathrm{X}\,\gamma_{ri}^\mathrm{X}, \tag{6.15}$$

In which the x_{ri}^{X} are the assignment factors and each γ_{ri}^{X} represents discomfort cost per time unit. So the Right Hand Side in (6.15) is a value of time aggregated over all users in component r.

As for wait time,

$$\frac{\partial \mathrm{SC}^{\#}}{\partial t_r^{\mathrm{W}}} = \overline{\mu} \sum_{i \in \mathrm{I}} Q_i \, x_{ri}^{\mathrm{W}} \, \gamma_{ri}^{\mathrm{W}} = \overline{\mu} \, E_r^{\mathrm{W}} , \tag{6.16}$$

Wherein $E_r^{\mathrm{W}} \equiv \sum_{i \in \mathrm{I}} \gamma_i^{\mathrm{W}} B_r^i$. We shall refer to E_r^{W} as the Effect of waiting time in the component.

Concerning run time per unit distance, the characteristic variable is $t_r^{\mathrm{R}} \equiv 1/v_r$ so that

$$\frac{\partial \mathrm{SC}^{\#}}{\partial t_r^{\mathrm{R}}} = \overline{\mu} \sum_{i \in \mathrm{I}} \gamma_{ri}^{\mathrm{R}} \, Q_i \, x_{ri}^{\mathrm{R}} \, \lambda_r^i = \overline{\mu} \, E_r^{\mathrm{R}} , \tag{6.17}$$

Wherein $E_r^{\mathrm{R}} \equiv \sum_{i \in \mathrm{I}} \gamma_i^{\mathrm{R}} T_r^i$.

As for Access and Egress times, the demand segments may exhibit different speeds, so the characteristic variable of major interest is the access distance is d_{Ar} by component r. It holds that by demand segment, letting $\mathrm{SC}_{\mathrm{i}}^{\#} \equiv Q_i(g_i - \tau_i) - \int_0^{Q_i} \mathrm{D}_i^{(-1)}(q)\,\mathrm{d}q - \mu_{\mathrm{B}} Q_i \tau_i$ (still under $Q_i = \mathrm{D}_i(g_i)$),

$$\frac{\partial \mathrm{SC}_i^{\#}}{\partial t_{ri}^{\mathrm{A}}} = \frac{\partial \mathrm{SC}_i^{\#}}{\partial g_i} \frac{\partial g_i}{\partial t_{ri}^{\mathrm{A}}} = \overline{\mu} \, Q_i \, x_{ri}^{\mathrm{A}} \, \gamma_{ri}^{\mathrm{A}} = \overline{\mu} \, \gamma_{ri}^{\mathrm{A}} \, I_r^i ,$$

$$\frac{\partial \mathrm{SC}_i^{\#}}{\partial t_{ri}^{\mathrm{E}}} = \frac{\partial \mathrm{SC}_i^{\#}}{\partial g_i} \frac{\partial g_i}{\partial t_{ri}^{\mathrm{E}}} = \overline{\mu} \, Q_i \, x_{ri}^{\mathrm{E}} \, \gamma_{ri}^{\mathrm{E}} = \overline{\mu} \, \gamma_{ri}^{\mathrm{E}} \, O_r^i .$$

Thus, with respect to the access distance, as $\mathrm{SC}^{\#} = \sum_{i \in \mathrm{I}} \mathrm{SC}_{\mathrm{i}}^{\#} + \mu_{\mathrm{B}} \mathrm{SR}$,

$$\frac{\partial \mathrm{SC}^{\#}}{\partial d_{Ar}} = \sum_{i \in \mathrm{I}} \frac{\partial \mathrm{SC}_i^{\#}}{\partial t_{ri}^{\mathrm{A}}} \frac{\partial t_{ri}^{\mathrm{A}}}{\partial d_{Ar}} + \frac{\partial \mathrm{SC}_i^{\#}}{\partial t_{ri}^{\mathrm{E}}} \frac{\partial t_{ri}^{\mathrm{E}}}{\partial d_{Ar}} \quad \text{with} \quad \frac{\partial t_{ri}^{\mathrm{A}}}{\partial d_{Ar}} = \frac{1}{w_i} = \frac{\partial t_{ri}^{\mathrm{E}}}{\partial d_{Ar}} ,$$

So that

$$\frac{\partial \mathrm{SC}^{\#}}{\partial d_{Ar}} = \overline{\mu} \, E_r^{\mathrm{A}} , \tag{6.18}$$

wherein $E_r^{\mathrm{A}} \equiv \sum_{i \in \mathrm{I}} (\gamma_i^{\mathrm{A}} I_r^i + \gamma_i^{\mathrm{E}} O_r^i)/w_i$ is the indicator of effect for terminal distance onto system cost.

To sum up, each indicator of effect is an economic value derived by aggregation over the demand segments.

6.4 Service frequency and fleet size

About fleet size N_r, we have that

$$\frac{\partial £}{\partial N_r} \equiv \mathrm{c}_{Nr} - \mu_{\varphi r} \frac{v_r}{2L_r} . \tag{6.19}$$

So the related optimality condition, $\partial £ / \partial N_r = 0$, yields that

$$\mu_{\varphi r} = \mathrm{c}_{Nr} \frac{2L_r}{v_r} \quad (= \mathrm{c}_{Nr} C_r), \tag{6.20}$$

In which we recognize the cycle time C_r.

As for service frequency φ_r, as

$$\frac{\partial \pounds}{\partial \varphi_r} \equiv \frac{\partial \pounds}{\partial t_r^{\mathrm{W}}} \frac{\partial t_r^{\mathrm{W}}}{\partial \varphi_r} + \mu_{\varphi r} = -\frac{\overline{\mu} E_r^{\mathrm{W}} \beta_r}{2\varphi_r^2} + \mu_{\varphi r}, \tag{6.21}$$

The optimality condition $\partial \pounds / \partial \varphi_r = 0$ amounts to

$$\mu_{\varphi r} = \frac{\overline{\mu} E_r^{\mathrm{W}} \beta_r}{2\varphi_r^2}, \tag{6.22}$$

Or equivalently to

$$\varphi_r^2 = \frac{\overline{\mu} E_r^{\mathrm{W}} \beta_r}{2\mu_{\varphi r}}. \tag{6.23}$$

The quadratic dependence of frequency onto a demand aggregate under system optimization is a well-known property for the operations of transit lines (Small and Verhoef, 2007). Combining (6.20) and (6.23) to the dual optimality condition with respect to $\mu_{\varphi r}$, namely $\partial \pounds / \partial \mu_{\varphi r} = 0$ wherein $\partial \pounds / \partial \mu_{\varphi r} \equiv \varphi_r - N_r v_r / 2L_r$, we get that

$$\varphi_r^2 = \frac{\overline{\mu} E_r^{\mathrm{W}} \beta_r}{2\,\mathrm{c}_{Nr}} \frac{v_r}{2L_r} = \frac{\overline{\mu} E_r^{\mathrm{W}} \beta_r}{2\,\mathrm{c}_{Nr} N_r} \varphi_r,$$

Hence that

$$\varphi_r = \frac{\overline{\mu} E_r^{\mathrm{W}} \beta_r}{2\,\mathrm{c}_{Nr} N_r}. \tag{6.24}$$

Which links the optimum frequency to the operator cost of the rolling stock, $\mathrm{c}_{Nr} N_r$.

Indeed this is a remarkable economic property, as is evidenced by the following equivalent form,

$$\mathrm{c}_{Nr} N_r = \frac{\overline{\mu} E_r^{\mathrm{W}} \beta_r}{2\varphi_r} = \overline{\mu}\, \mathrm{UC}_r^{\mathrm{W}}.$$

Put in words, under system optimization the operator cost of the rolling stock, $\mathrm{c}_{Nr} N_r$, is balanced by the user cost of waiting, denoted $\mathrm{UC}_r^{\mathrm{W}}$, since E_r^{W} is the aggregate time value of a unit waiting time while $\beta_r / 2\varphi_r$ is the average individual waiting time. The balancing is exact under fixed demand ($\overline{\mu} = 1$), whereas under variable demand it involves $\overline{\mu} \geq 1$ as a lever arm on the demand side.

In the solution scheme, eqn. (6.24) will be used as the following relationship between L_r, v_r and N_r:

$$L_r = \frac{\mathrm{c}_{Nr} v_r N_r^2}{\overline{\mu} \beta_r E_r^{\mathrm{W}}}. \tag{6.25}$$

6.5 On station number and spacing

Coming to the number of stations σ_r in the network component, we have that

$$\frac{\partial £}{\partial \sigma_r} \equiv c_{\sigma r} + \mu_{Sr} \frac{L_r}{\sigma_r^2}, \tag{6.26}$$

So the optimality condition associated to σ_r, namely $\partial £ / \partial \sigma_r = 0$, yields that

$$\mu_{Sr} = -c_{\sigma r} \frac{\sigma_r^2}{L_r}. \tag{6.27}$$

The partial derivative of the Lagrangian function with respect to station spacing S_r is

$$\frac{\partial £}{\partial S_r} \equiv \mu_{Sr} + \mu_{vr} \frac{\omega_r}{S_r^2} - \tfrac{1}{4} \mu_{Ar}. \tag{6.28}$$

Thus the related optimality condition, $\partial £ / \partial S_r = 0$, relates the dual variables and S_r. We will use it in the following form, wherein μ_{Sr} is replaced by $-c_{\sigma r} \sigma_r^2 / L_r$ using (6.27):

$$c_{\sigma r} \frac{\sigma_r^2}{L_r} = \mu_{vr} \frac{\omega_r}{S_r^2} - \tfrac{1}{4} \mu_{Ar}. \tag{6.29}$$

As regards the terminal distance, d_{Ar}, the partial derivative of the Lagrangian function with respect to it is

$$\frac{\partial £}{\partial d_{Ar}} \equiv \mu_{Ar} + \frac{\partial \mathrm{UC}}{\partial d_{Ar}} = \mu_{Ar} + \overline{\mu} E_r^{\mathrm{A}}. \tag{6.30}$$

So the related optimality condition, $\partial £ / \partial d_{Ar} = 0$, states that

$$\mu_{Ar} = -\overline{\mu} E_r^{\mathrm{A}}. \tag{6.31}$$

Using the dual condition, $\partial £ / \partial \mu_{Sr} = 0$, under system optimization it holds that $S_r = L_r / \sigma_r$, so that (6.28) is equivalent to the following

$$c_{\sigma r} \sigma_r = \mu_{vr} \frac{\omega_r}{S_r} + \overline{\mu} E_r^{\mathrm{A}} \frac{S_r}{4}. \tag{6.32}$$

This is indeed a second remarkable economic property, which states that under system optimization the operator's cost of stations balances the "longitudinal" part of the users' costs of walking, up to the partial effect of station spacing onto commercial speed and a lever arm of $\overline{\mu} \geq 1$ on the demand side. The equality $d_r^{\mathrm{A}} = \frac{1}{4}(S_r + \chi_r \mathrm{A}_r / L_r)$ is ensured by the dual optimality condition associated to the dual variable $\mu_{\mathrm{A}r}$, $\partial £ / \partial \mu_{\mathrm{A}r} = 0$.

6.6 Commercial speed and infrastructure length

Concerning commercial speed v_r, the partial derivative of the Lagrangian function is

$$\begin{aligned} \frac{\partial £}{\partial v_r} &\equiv \frac{\partial \mathrm{UC}}{\partial t_r^{\mathrm{R}}} \frac{\partial t_r^{\mathrm{R}}}{\partial v_r} - \frac{\mu_{vr}}{v_r^2} - \mu_{\varphi r} \frac{N_r}{2L_r} \\ &= -\frac{\overline{\mu} E_r^{\mathrm{R}} + \mu_{vr}}{v_r^2} - \mu_{\varphi r} \frac{N_r}{2L_r} \end{aligned}. \tag{6.33}$$

On replacing $\mu_{\varphi r}$ by $\overline{\mu} E_r^{\mathrm{W}} \beta_r / 2\varphi_r^2$ and φ_r by $N_r v_r / 2L_r$, the optimality condition related to v_r, $\partial £ / \partial v_r = 0$, yields that

$$\frac{\overline{\mu} E_r^{\mathrm{R}} + \mu_{vr}}{v_r^2} = -\mu_{\varphi r} \frac{N_r}{2L_r} = -\overline{\mu} E_r^{\mathrm{W}} \beta_r \frac{4L_r^2}{2N_r^2 v_r^2} \frac{N_r}{2L_r}$$

$$= -\frac{\mathrm{c}_{Nr} N_r}{v_r} \text{ owing to (6.25)}$$

So that: $v_r = \dfrac{-\mu_{vr} - \overline{\mu} E_r^{\mathrm{R}}}{\mathrm{c}_{Nr} N_r}$.

For this condition to hold true it requires that $\mu_{vr} < 0$.

From (6.32): $$\mathrm{c}_{\sigma r} \sigma_r + (\mathrm{c}_{Nr} N_r v_r + \overline{\mu} E_r^{\mathrm{R}}) \frac{\omega_r}{S_r} = \overline{\mu}\, E_r^{\mathrm{A}} \frac{S_r}{4}. \quad (6.34)$$

Gathering (6.32) and (6.34), we obtain that

$$\mathrm{c}_{\sigma r} \sigma_r + (\mathrm{c}_{Nr} N_r v_r + \overline{\mu} E_r^{\mathrm{R}}) \frac{\omega_r}{S_r} = \overline{\mu}\, E_r^{\mathrm{A}} \frac{S_r}{4}, \quad (6.35)$$

Which is the operational form of the second property in the solution scheme.

As for infrastructure length L_r, the related partial derivative of the Lagrangian function is

$$\frac{\partial £}{\partial L_r} \equiv \mathrm{c}_{Lr} - \frac{\mu_{Sr}}{\sigma_r} + \frac{\mu_{Ar}}{4L_r^2} \chi_r \mathrm{A}_r + \mu_{\varphi r} \frac{N_r v_r}{2L_r^2}. \quad (6.36)$$

In the related optimality condition, $\partial £ / \partial L_r = 0$, we shall replace μ_{Sr} by $-\mathrm{c}_{\sigma r} \sigma_r^2 L_r$, $\mu_{\varphi r}$ by $\mathrm{c}_{Nr}\, 2L_r / v_r$ and μ_{Ar} by $-\overline{\mu} E_r^{\mathrm{A}}$ to get that

$$\mathrm{c}_{Lr} + \mathrm{c}_{\sigma r} \frac{\sigma_r}{L_r} + \frac{\mathrm{c}_{Nr}}{L_r} N_r = \overline{\mu}\, E_r^{\mathrm{A}} \frac{\chi_r \mathrm{A}_r}{4L_r^2}.$$

Or equivalently, $$\mathrm{c}_{Lr} L_r + \mathrm{c}_{\sigma r} \sigma_r + \mathrm{c}_{Nr} N_r = \overline{\mu}\, E_r^{\mathrm{A}} \frac{\chi_r \mathrm{A}_r}{4L_r}. \quad (6.36)$$

This constitutes the third remarkable economic property under system optimization: the full operator costs (Left Hand Side in the equality) balances the transversal part of the users' costs of walking, up to a lever arm of $\overline{\mu} \geq 1$ on the demand side. Under fixed demand it holds that $\overline{\mu} = 1$ hence the matching is perfect and it means that the Time User Costs are quite greater than the Operator Costs in the full System Cost.

Consequently, a fare policy requiring the users to cover the full costs of supply might be quite tolerable to them, since the related fee would only constitute a minor part in the individual generalized cost.

6.7 A synthesis of conditions for system equilibrium

To sum up, the primal-dual optimality conditions for the modelled transit system with no congestion amount to a net of 3R+1 relationships between three basic supply variables by network component, (L_r, σ_r, N_r), and the dual variable $\overline{\mu}$ associated to a constraint of cost recovery. Let us gather them as a reduced set of conditions:

$$\bar{\mu} \geq 1, \quad \frac{\partial £}{\partial \bar{\mu}} \equiv \mathrm{SR} - \sum_{i \in \mathrm{I}} \tau_i Q_i \leq 0, \quad (\bar{\mu} - 1).\frac{\partial £}{\partial \bar{\mu}} = 0. \tag{6.38a}$$

From (6.25): $L_r = \dfrac{c_{Nr} v_r N_r^2}{\bar{\mu} \beta_r E_r^{\mathrm{W}}}$ wherein $\dfrac{1}{v_r} = \dfrac{1}{V_r} + \dfrac{\omega}{S_r}$ and $S_r = \dfrac{L_r}{\sigma_r}$. (6.38b)

From (6.35): $c_{\sigma r}\sigma_r + (c_{Nr} N_r v_r + \bar{\mu} E_r^{\mathrm{R}}) \dfrac{\omega_r}{S_r} = \bar{\mu}\, E_r^{\mathrm{A}} \dfrac{S_r}{4}$. (6.38c)

From (6.37): $c_{Lr} L_r + c_{\sigma r}\sigma_r + c_{Nr} N_r = \bar{\mu}\, E_r^{\mathrm{A}} \dfrac{\chi_r A_r}{4 L_r}$. (6.38d)

A basic vector $\left[(L_r, \sigma_r, N_r)_{r \in \mathrm{R}}, \bar{\mu}\right]$ that solves the previous system enables one to derive all characteristics of a system state under system optimization.

A couple of comments are in order here. First, under fixed demand we have $\bar{\mu} = 1$ so the reduced problem breaks into R problems each with three equations in three unknowns: then, system optimization is decentralized by network component.

Second, under variable demand the dual variable $\bar{\mu}$ summarizes the coupling between the components, through the determination of the demand which puts the components in interaction (in a "weak form" of interaction, actually).

Third, the integration of several components in a unified framework of system optimization induces inter-regional and inter-modal analogies: between any two components r and s, a number of characteristic features must be proportionate, notably so:

From (6.25), $\dfrac{c_{Nr} v_r N_r^2}{L_r \beta_r E_r^{\mathrm{W}}} = \bar{\mu}$,

From (6.35), $\dfrac{c_{\sigma r}\sigma_r + c_{Nr} N_r v_r \omega_r / S_r}{E_r^{\mathrm{A}} S_r / 4 - E_r^{\mathrm{R}} \omega_r / S_r} = \bar{\mu}$,

From (6.37), $\dfrac{c_{Lr} L_r + c_{\sigma r}\sigma_r + c_{Nr} N_r}{E_r^{\mathrm{A}} \chi_r A_r / 4 L_r} = \bar{\mu}$.

Furthermore, all of these quantities are equal to $\bar{\mu}$ hence to one another in a given component as well as between components.

Fourth and last, the equality between operator costs and the user cost of "transversal" access indeed constitutes a "golden rule" for transit networks, as it matches the cost of supply to the cost of transversal access, that for all trips in either a given component, or a specific area, or the multimodal network in the full urban area. So its spatial outreach goes far beyond the market area of a single line of transit.

Let us summarize the three remarkable economic properties and emphasize their respective economic significance:

[1] $c_{Nr} N_r = \bar{\mu}\, \mathrm{UC}^{\mathrm{W}}$ i.e. the cost of the rolling stock balances that of waiting,

[2] $c_{\sigma r}\sigma_r + \bar{\mu}.(\mathrm{UC}^{\mathrm{W}} + \mathrm{UC}^{\mathrm{R}})(1 - \dfrac{v_r}{V_r}) = \bar{\mu}\, \mathrm{UC}^{\mathrm{A}}_{\mathrm{longitudinal}}$ i.e. the users' costs of longitudinal access is balanced by the station costs, plus the dwelling proportion of commercial time multiplied by the cost of in-system time to the users..

[3] $c_{Lr}L_r + c_{\sigma r}\sigma_r + c_{Nr}N_r = \bar{\mu}\,\mathrm{UC}^{\mathrm{A}}_{\mathrm{transversal}}$ i.e. the supply cost balances that of transversal access.

The second relationship stems from (6.35), by replacing $c_{Nr}N_r v_r$ with $v_r\bar{\mu}\,\mathrm{UC}^{\mathrm{W}}$ owing to [1] and also E_r^{R} with $v_r\,\mathrm{UC}^{\mathrm{R}}$. Then, $c_{Nr}N_r v_r + \bar{\mu}E_r^{\mathrm{R}} = \bar{\mu}.(\mathrm{UC}^{\mathrm{W}} + \mathrm{UC}^{\mathrm{R}})v_r$. Yet it does also hold that $v_r.\omega_r / S_r = 1 - v_r/\mathrm{V}_r$ i.e. the proportion of "running time" that is spent at station dwelling.

By adding up the two last relationships and remarking that $\mathrm{TUC} = \mathrm{UC}^{\mathrm{R}} + \mathrm{UC}^{\mathrm{W}} + \mathrm{UC}^{\mathrm{A}}$ (recall that the transfer costs were omitted), it comes out that

$$\frac{\mathrm{OC} + c_\sigma\sigma}{\bar{\mu}} + (1 - \frac{v}{\mathrm{V}})(\mathrm{TUC} - \mathrm{UC}^{\mathrm{A}}) = \mathrm{UC}^{\mathrm{A}}$$

Hence that $\mathrm{UC}^{\mathrm{A}} = \frac{1}{2 - v/\mathrm{V}}\frac{\mathrm{OC} + c_\sigma\sigma}{\bar{\mu}} + \frac{1 - v/\mathrm{V}}{2 - v/\mathrm{V}}\mathrm{TUC}$:

The total value of users' access time is worth more than half the operator costs plus a fraction of the time users costs. Under variable demand and constant elasticity, $\mathrm{TUC} = \frac{1-(1+\varepsilon)\bar{\mu}}{\bar{\mu}-1}\mathrm{MUC}$, so that

$$\mathrm{UC}^{\mathrm{A}} = \frac{1}{2 - v/\mathrm{V}}\frac{\mathrm{OC} + c_\sigma\sigma}{\bar{\mu}} + \frac{1 - v/\mathrm{V}}{2 - v/\mathrm{V}}.\frac{1 - (1+\varepsilon)\bar{\mu}}{\bar{\mu} - 1}\mathrm{MUC}\,.$$

6.8 Formal properties and solution scheme

In the Appendix (Section 9), we provide a solution scheme, which enabled us to derive formal properties of existence and uniqueness for a System Optimum State, together with some properties of sensitivity analysis. Let us here outline the main outcomes.

Proposition 1: Existence and uniqueness of a system optimum state under Fixed demand. *Under "Fixed demand" defined as given parameter $\bar{\mu}$ and demand flows $\mathbf{Q}_{\mathrm{I}} = (Q_i : i \in \mathrm{I})$, there exists a System Optimal state, which is unique.*

Under "Fixed demand" the System Optimum problem is separable and reduces to a reduced problem of 3 equations linking three reduced unknowns $L' \equiv L/\bar{\mu}$, $\sigma' \equiv \sigma/\bar{\mu}$ and $N' \equiv N/\bar{\mu}$. It turns out that station spacing S is the main factor, with respect to whom the SO sub-problem of the component reduces in a single equation in S only.

The single equation in S amounts to match a function of S that increases from zero to infinity, with a given value that is non-negative: thus the equation has one and one only solution. The solution depends on $\bar{\mu}$ and the demand flows $\mathbf{Q}_{\mathrm{I}}$ (via the Effect factors) as parameters in a simple way, which enables us to derive the following property of sensitivity for the reduced optimal state $(\hat{L}'_S,\ \hat{N}'_S,\ S)$ and the reduced operator cost function, $\mathrm{OC}'_S \equiv c_{\mathrm{L}}L' + c_\sigma\sigma' + c_{\mathrm{N}}N'$:

Proposition 2: Sensitivity analysis of optimum system state (under fixed demand).
[i] The reduced system optimum solution $(\hat{L}'_S,\ \hat{N}'_S,\ S)$ decreases with $\bar{\mu}$ and so does the reduced operator cost, OC'_S. [ii] With respect to region area A, the system optimum solution $(\hat{L}'_S,\ \hat{N}'_S,\ S)$ increases and so does the reduced operator cost, OC'_S.

Thus, under "Fixed Demand" it is easy to solve the problem of System Optimization. Under Variable demand, both $\overline{\mu}$ and the demand flows $\mathbf{Q}_I$ are endogenous as well as the supply side variables $\left[(L_r, \sigma_r, N_r) : r \in \mathrm{R}\right]$. We address the SO problem under Variable Demand as a three-level optimization program, with upper level to deal with $\overline{\mu}$, intermediary level to solve for $\mathbf{Q}_I$ and lower level to solve for the supply side variables on the basis of $\overline{\mu}$ and $\mathbf{Q}_I$. Of course the lower level consists in a Fixed Demand SO problem.

The following Proposition holds true:

Proposition 3: Existence and uniqueness of System Optimum State under Variable demand. *Assuming that the target revenue SR is feasible, there exists a state of System Optimum for the transit system under variable demand. This state is unique.*

Our solution scheme to address the upper level problem consists in a dichotomy search on $\overline{\mu}$ only so as to meet the condition of minimum commercial revenue. From each candidate value of $\overline{\mu}$, we derive the demand flows $\mathbf{Q}_I$ and the supply side variables that jointly satisfy all of the other conditions for optimality, by solving a Fixed Point Problem in $\mathbf{Q}_I$ which uses the Fixed Demand SO problem as a sub-program. Conditionally to $\overline{\mu}$, this Fixed Point Problem makes up the intermediary level in the overall, three-level scheme.

7. Conclusion

This paper owes much to previous contributions to the theory of transit network design. Our model of system elements and relationships essentially builds upon Van Nes (2002), whereas our synthetic diagram is both more detailed (adding up causal relationships and action levers) and more structured than the initial version (Van Nes and Bovy, 2000). We extended the model to deal with spatial heterogeneity (yet in a coarse way): this involves the distinction of regions within the study area and the identification of origin-destination pairs. Another theoretical extension pertains to traffic phenomena and their feedbacks on supply performance as well as on quality of service. Having stated the two related problems of, first, system state determination under exogenous supply set-ups and, second, state optimisation by acting on supply-side levers, we provided a theoretical study of system optimization. Three "golden rules" for the optimal design and management of urban transit networks have been established in a fairly generic framework (yet in the uncongested case). A fourth rule pertains to second best optimal tariffs under variable demand. Solution algorithms have been provided for the fixed demand case then for variable demand under a threshold constraint on commercial revenues.

All in all, the strategic model is both technical and economic; it lies halfway of, on one hand, network traffic assignment models which feature out technical relationships and the supply-demand interactions and, on the other hand, theoretical models in transport economics at the local or line or aggregate level. The generic framework allows for theoretical analysis. However it can support neither the spatial nor the temporal details that are required in the design of particular projects.

Beyond system understanding, we believe that the model is suitable for practical applications, ranging from synthetic statistical description of an existing system, to the assessment of observed or hypothetical system states against a theoretically optimal state, passing by comparisons between network components within the system and also between the transit systems of different urban areas.

Acknowledgement. This work has been supported by two Academic Chairs: first, the ENPC Chair on the "Socioeconomics and Modeling of Urban Passenger Transit" sponsored by STIF (the Mobility Organizing Authority in greater Paris), second, the ParisTech Chair on the "Eco-design of Building Sets and Infrastructure" sponsored by Vinci Group. We are grateful to both STIF and Vinci for their continued support and inspiration.

8. References

Andreasson, I., Leurent, F., Corman, F. and dell'Olio, L. (2016) Modelling the Diversity and Integration of Transit Modes. In Gentile, G. & Noeckel, K. (eds): *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, Springer Tracts on Transportation and Traffic, Vol. 10: 485-520.

Beckmann M. McGuire C.B. and Winsten C.B. (1956) Studies in the Economics of Transportation. Yale University Press.

Combes F. and Leurent F. (2016) Urban economics and passenger transportation: concentrate flows, design lines. Chapter 5 in Peuportier B., Leurent F. and Roger-Estrade J. (eds), *Eco-design of Building Sets and Infrastructure*. CRC Press, pp. 66-84.

De Rus G. (2010) Introduction to Cost-Benefit Analysis; Looking for reasonable shortcuts. Edward Elgar, 249 pages.

Dupuit J. (1844) De la Mesure de l'utilité des Travaux Publics. Annales des Ponts et Chaussées : Mémoires et Documents, 2nd series 8(2): 332-375. Trans. R.M. Barback, "On the Measurement of Utility for Public Works". International Economic Papers, 1952, 2: 83-110.

Leurent F., Windisch E. (2015) Benefits and costs of electric vehicles for the public finances: an integrated valuation model based on input-output analysis, with application to France. Research in Transportation Economics 415:51-62.

Ortuzar J. de Dios, Willumsen L. (2004) Modeling Transport. Wiley.

Quinet E. and Vickerman R. (2004) Principles of Transport Economics. Edward Elgar.

Small K. and Verhoef E. (2007) The Economics of Urban Transportation. 2nd Edition. Routledge, London.

Van Nes, R. (2002) Design of multimodal transport networks: a Hierarchical Approach. T2002/5, TRAIL Thesis Series, Delft University Press, 286 pages.

Van Nes R. and Bovy P.H.L. (2000), The importance of objectives in urban transit network design, Transportation Research Record 1735: 25-34.

LAET + May / MARS

IMPACT

Papageorgiou

Strategic models of traffic

Combinatorial models of optimum design

9. Appendix: solution scheme and formal properties

In order to solve the problem of System Optimization, we designed a computation scheme that decomposes the problem in three layers: from bottom-up, the determination of supply-side variables conditionally to dual parameter $\bar{\mu}$ and demand flows $\mathbf{Q}_\mathrm{I} = (Q_i : i \in \mathrm{I})$, then the determination of demand flows $\mathbf{Q}_\mathrm{I}$ conditionally to $\bar{\mu}$, and lastly the determination of $\bar{\mu}$ itself. The lower layer constitutes the Fixed demand problem of System Optimization and it turns out to be separable by network component.

The building-up of the solution scheme enables us to establish formal properties of existence and uniqueness of an SO state. These are obtained at first under Fixed demand, then under Variable demand. So the Section body is in two parts. Subsection 9.1 deals with the Fixed demand problem, its specific solution scheme and formal properties. Then, Subsection 9.2 addresses Variable demand and the determination of the dual parameter $\bar{\mu}$ that has a key role.

9.1 Algorithm for fixed demand and given dual parameter

Here the dual variable $\bar{\mu}$ is taken as a parameter and the $\mathbf{Q}_\mathrm{I} \equiv (Q_i)_{i\in\mathrm{I}}$ are fixed. This approach includes the fixed demand case, in which $\bar{\mu}$ is set to 1 and the flows are fixed. In a more general setting, let us assume that $\bar{\mu}$ is fixed and also the demand volumes, hence the effect factors E_r^X for all components r in R.

Thus, the reduced problem of system optimization (6.38)[1]-[3] is separable by component and consists in three non-linear equations in three unknowns (we drop index r for simplicity):

[1'] $$L' = \frac{\mathrm{c_N}\, v\, N'^2}{\beta E^\mathrm{W}} \quad \text{wherein} \quad \frac{1}{v} = \frac{1}{\mathrm{V}} + \frac{\omega}{S} \quad \text{and} \quad S = \frac{L'}{\sigma'}.$$

[2'] $$\mathrm{c}_\sigma \sigma' + (\mathrm{c_N} N' v + E^\mathrm{R}) \frac{\omega}{S} = \frac{1}{4} E^\mathrm{A} S.$$

[3'] $$\mathrm{c_L} L' + \mathrm{c}_\sigma \sigma' + \mathrm{c_N} N' = \frac{1}{4} \frac{E^\mathrm{A}}{\bar{\mu}} \frac{\chi \mathrm{A}}{L'}.$$

In this re-statement, we have taken reduced variables $L' \equiv L/\bar{\mu}$, $\sigma' \equiv \sigma/\bar{\mu}$ and $N' \equiv N/\bar{\mu}$ with respect to which there only remains one dependency on $\bar{\mu}$ viz. in the right hand side of [3']. Furthermore, $S' = L'/\sigma' = S$ and $v' = v$ so we shall focus on S as the main variable, together with L' and N'.

From [1'], L' is fully determined by N' and S (through v), so it can be denoted $\hat{L}'_{N',S}$.

By substituting $\mathrm{c_N}\, v_S N'^2 / \beta E^\mathrm{W}$ to L' in [2'] via $\sigma' = L'/S$, and multiplying by S, we get that

$\frac{\mathrm{c}_\sigma\, \mathrm{c_N}\, v_S}{\beta E^\mathrm{W}} N'^2 + \omega\, \mathrm{c_N} v_S N' = \frac{1}{4} E^\mathrm{A} S^2 - \omega E^\mathrm{R}$, or equivalently

$$N'^2 + \frac{\omega \beta E^\mathrm{W}}{\mathrm{c}_\sigma} N' - \frac{\omega \beta E^\mathrm{W}}{\mathrm{c}_\sigma\, \mathrm{c_N}\, v_S} \left(\tfrac{1}{4\omega} E^\mathrm{A} S^2 - E^\mathrm{R}\right) = 0. \tag{9.1}$$

Which merely amounts to a second-order equation in N' that is parameterized by S.

Denoting $\xi \equiv \frac{\omega\,\beta\,E^{\mathrm{W}}}{\mathrm{c}_\sigma}$ and $\zeta_S \equiv \frac{\xi}{\mathrm{c}_{\mathrm{N}}\,v_S}(\frac{1}{4\omega}E^{\mathrm{A}}S^2 - E^{\mathrm{R}})$, the equation reduces to

$$N'^2 + \xi N' - \zeta_S = 0\,. \tag{9.2}$$

The discriminant amounts to

$$\Delta = \xi^2 + 4\zeta_S\,, \tag{9.3}$$

Which is non-negative iff $\frac{1}{\omega}E^{\mathrm{A}}S^2 - 4E^{\mathrm{R}} \geq -\xi\,\mathrm{c}_N v_S$, or $S^2 \geq \frac{\omega}{E^{\mathrm{A}}}(4E^{\mathrm{R}} - \xi\,\mathrm{c}_N v_S)$.

We shall satisfy ourselves with the stricter requirement that $\zeta_S \geq 0$ i.e. $S^2 \geq 4\omega E^{\mathrm{R}}/E^{\mathrm{A}}$ or

$$S \geq S_0 \equiv 2\sqrt{\omega\,E^{\mathrm{R}}/E^{\mathrm{A}}}\,. \tag{9.4}$$

Under this requirement, $\Delta \geq \xi^2$ hence the second-order equation in N' admits one nonnegative solution only, namely

$$\hat{N}'_S = \frac{\sqrt{\Delta}-\xi}{2} = \frac{\xi}{2}\psi_S \quad \text{with} \quad \psi_S \equiv \sqrt{1 + \frac{E^{\mathrm{A}}}{\xi\,\mathrm{c}_{\mathrm{N}}\,\omega\,v_S}(S^2 - S_0^2)} - 1\,. \tag{9.5}$$

Replacing N' by $\frac{1}{2}\xi\psi_S$ in [1'], we obtain L' as a function of S only,

$$\hat{L}'_S = \frac{\mathrm{c}_{\mathrm{N}}\,\xi^2\,v_S}{4\beta\,E^{\mathrm{W}}}\psi_S^2\,. \tag{9.6}$$

In turn, $\sigma' = L'/S$ is a function of S only,

$$\hat{\sigma}'_S = \frac{\mathrm{c}_{\mathrm{N}}\,\xi^2}{4\beta\,E^{\mathrm{W}}}v_S\frac{\psi_S^2}{S}\,. \tag{9.7}$$

By replacing in [3'] variables L', σ' and N' by $\hat{L}'_S$, $\hat{\sigma}'_S$ and $\hat{N}'_S$, respectively, we obtain a single equation S in only:

$$\frac{\mathrm{c}_{\mathrm{N}}\,v_S\xi^2}{4\beta\,E^{\mathrm{W}}}\psi_S^2\left(\mathrm{c}_{\mathrm{L}}\frac{\mathrm{c}_{\mathrm{N}}\,v_S\xi^2}{4\beta\,E^{\mathrm{W}}}\psi_S^2 + \mathrm{c}_\sigma\frac{\mathrm{c}_{\mathrm{N}}\,\xi^2}{4\beta\,E^{\mathrm{W}}}v_S\frac{\psi_S^2}{S} + \mathrm{c}_{\mathrm{N}}\frac{\xi}{2}\psi_S\right) = \frac{E^{\mathrm{A}}}{4\overline{\mu}}\chi\mathrm{A}\,. \tag{9.8}$$

Notice that the Right Hand Side does not depend on S, while the Left Hand Side does not involve $\overline{\mu}$. We shall deal with (9.8) by considering the LHS as a function of S, more precisely a product function that multiplies $\hat{L}'_S$ by $\mathrm{OC}'_S \equiv \mathrm{c}_{\mathrm{L}}\hat{L}'_S + \mathrm{c}_\sigma\hat{\sigma}'_S + \mathrm{c}_{\mathrm{N}}\hat{N}'_S$.

Lemma. *Functions* $\hat{N}'_S$, $\hat{L}'_S$ *and* OC'_S *are increasing with* $S \geq S_0$.

Proof. Function $S \mapsto S^2 - S_0^2$ is non-negative on $[S_0, \infty)$, thus so does ψ_S hence $\hat{N}'_S$, too. As for $v_S = 1/(\frac{1}{\mathrm{V}} + \frac{\omega}{S})$, it is positive and has derivative $v'_S = \omega v_S^2/S^2$ which also is positive. Thus function $S \mapsto v_S$ is both positive and strictly increasing with S on $[S_0, \infty)$. Coming back to ψ_S,

$$\begin{aligned}\psi'_S &= \frac{1}{2(\psi_S+1)}\frac{E^{\mathrm{A}}}{\xi\, \mathrm{c}_N\, \omega}\left(\frac{2S}{v_S}-(S^2-S_0^2)\frac{v'_S}{v_S^2}\right)\\ &= \frac{1}{2(\psi_S+1)}\frac{E^{\mathrm{A}}}{\xi\, \mathrm{c}_N\, \omega}\left(\frac{2S}{v_S}-(S^2-S_0^2)\frac{\omega}{S^2}\right) = \frac{1}{2(\psi_S+1)}\frac{E^{\mathrm{A}}}{\xi\, \mathrm{c}_N\, \omega}\left(\frac{2S}{\mathrm{V}}+2\omega-\omega+\omega\left(\frac{S_0}{S}\right)^2\right)\\ &= \frac{1}{2(\psi_S+1)}\frac{E^{\mathrm{A}}}{\xi\, \mathrm{c}_N\, \omega}\left(\frac{2S}{\mathrm{V}}+\omega+\omega\left(\frac{S_0}{S}\right)^2\right)\\ &\geq 0\end{aligned}$$

So ψ_S is increasing as well as non-negative on $[S_0,\infty)$. The same applies to $S\mapsto\psi_S^2$. Then, from its definition the function $S\mapsto \hat{L}'_S \equiv \frac{\mathrm{c_N}\xi^2}{4\beta E^{\mathrm{W}}} v_S\psi_S^2$ is the product of two non-negative and increasing functions on $[S_0,\infty)$, which makes it non-negative and increasing on that domain. Concerning OC'_S,

$$\begin{aligned}\mathrm{c_L}\hat{L}'_S+\mathrm{c}_\sigma\hat{\sigma}'_S+\mathrm{c_N}\hat{N}'_S &= \tfrac{1}{2}\xi\,\mathrm{c_N}\,\psi_S\left[\mathrm{c_L}\frac{v_S\xi}{2\beta E^{\mathrm{W}}}\psi_S+\mathrm{c}_\sigma\frac{v_S\xi}{2\beta E^{\mathrm{W}}}\frac{1}{S}\psi_S+1\right]\\ &= \tfrac{1}{2}\xi\,\mathrm{c_N}\,\psi_S\left[1+\tfrac{1}{2}\omega v_S\psi_S\left(\frac{\mathrm{c_L}}{\mathrm{c}_\sigma}+\frac{1}{S}\right)\right] \text{ since } \xi=\omega\beta E^{\mathrm{W}}/\mathrm{c}_\sigma\end{aligned}$$

As we already know that both ψ_S and $\psi_S^2 v_S$ are increasing with respect to S on $[S_0,\infty)$, we only have to check that $S\mapsto\psi_S^2 v_S/S$ is increasing on that domain, in order to demonstrate that so is OC'_S (postulating non-negative unit costs). By derivation,

$$\frac{d}{dS}\frac{\psi_S^2 v_S}{S} = \frac{1}{S}\frac{d}{dS}\left[\psi_S^2 v_S\right]-\frac{\psi_S^2 v_S}{S^2}$$

$$\begin{aligned}\frac{d}{dS}\frac{\psi_S^2 v_S}{S} &= \frac{1}{S}\left[2v_S\psi_S\psi'_S+\psi_S^2 v'_S\right]-\frac{\psi_S^2 v_S}{S^2} = \frac{v_S\psi_S}{S^2}\left[2S\psi'_S+\frac{\omega\psi_S v_S}{S}-\psi_S\right]\\ &= \frac{v_S\psi_S}{S^2}\left[2S\psi'_S+v_S\psi_S\left(\frac{\omega}{S}-\frac{1}{v_S}\right)\right] = \frac{v_S\psi_S}{S^2}\left[2S\psi'_S-\frac{v_S\psi_S}{\mathrm{V}}\right]\\ &= \frac{v_S\psi_S}{S^2}\left[\frac{2S}{2(\psi_S+1)}\alpha\left(\frac{2S}{\mathrm{V}}+\omega+\omega\left(\frac{S_0}{S}\right)^2\right)-\frac{v_S\psi_S}{\mathrm{V}}\right] \text{ with } \alpha\equiv\frac{E^{\mathrm{A}}}{\xi\,\mathrm{c_N}\,\omega}\\ &= \frac{v_S\psi_S}{S^2(\psi_S+1)}\left[\alpha S\left(\frac{2S}{\mathrm{V}}+\omega+\omega\left(\frac{S_0}{S}\right)^2\right)-\frac{v_S\psi_S}{\mathrm{V}}(\psi_S+1)\right]\\ &= \frac{v_S\psi_S}{S^2(\psi_S+1)}\left[\alpha S\left(\frac{2S}{\mathrm{V}}+\omega\left(1+\left(\frac{S_0}{S}\right)^2\right)\right)-\frac{v_S}{\mathrm{V}}\left[(\psi_S+1)^2-\psi_S-1\right]\right]\\ &= \frac{v_S\psi_S}{S^2(\psi_S+1)}\left[\alpha\left(\frac{2S^2}{\mathrm{V}}+\omega\left(S+\frac{S_0^2}{S}\right)-\frac{v_S}{\mathrm{V}}\frac{S^2-S_0^2}{v_S}\right)+\psi_S\frac{v_S}{\mathrm{V}}\right] \text{ as } (\psi_S+1)^2-1=\alpha\frac{S^2-S_0^2}{v_S}\\ &= \frac{v_S\psi_S}{S^2(\psi_S+1)}\left[\frac{E^{\mathrm{A}}}{\xi\,\mathrm{c_N}\,\omega}\left(\frac{S^2}{\mathrm{V}}+\omega\left(S+\frac{S_0^2}{S}\right)+\frac{S_0^2}{\mathrm{V}}\right)+\psi_S\frac{v_S}{\mathrm{V}}\right]\end{aligned}$$

which is indeed non-negative, implying that $S\mapsto\psi_S^2 v_S/S$ is increasing as well as non-negative.

Proposition 1: Existence and uniqueness of a system optimum state under fixed demand.
[i] Function $S \mapsto \hat{L}'_S . \mathrm{OC}'_S$ *is increasing with* S *on* $[S_0, \infty)$ *and varies from 0 to* $+\infty$.
[ii] Therefore equation [3'] in S *admits a solution, which is unique.*

Proof. On domain $]S_0, \infty[$ both functions $\hat{L}'_S$ and OC'_S are positive and strictly increasing since their derivatives are strictly positive there (postulating positive unit costs). Therefore their product is positive and strictly increasing, too. From its definition, $\psi_S = 0$ at S_0 and tends to infinity as S increases indefinitely. Thus so do $\hat{N}'_S$ and $\hat{L}'_S$ since $v_S \to \mathrm{V} > 0$ when $S \to +\infty$. This applies in turn to OC'_S and $\hat{L}'_S . \mathrm{OC}'_S$, which establishes [i].

[ii] From the previous point, each value $y \in [0, \infty)$ is the image of one and one only $S_y \in [S_0, \infty)$. This applies notably to $y = \frac{1}{4} E^{\mathrm{A}} \chi \mathrm{A} / \overline{\mu}$. Thus Equation [3'] admits one and one only solution in S. This solution induces $\hat{L}'_S$, $\hat{\sigma}'_S$ and $\hat{N}'_S$ that solve the reduced problem of system optimization, meaning that there exists an optimal system state, which is unique.

Proposition 2: Sensitivity analysis of optimum system state (under fixed demand).
[i] The reduced system optimum solution ($\hat{L}'_S$, $\hat{N}'_S$, S*) decreases with* $\overline{\mu}$ *and so does the reduced operator cost,* OC'_S.

[ii] With respect to region area A, the system optimum solution ($\hat{L}'_S$, $\hat{N}'_S$, S*) increases and so does the reduced operator cost,* OC'_S.

Proof. [i] As $\hat{L}'_S$ and $\hat{N}'_S$ increase with S, the chain functions $\hat{L}'_{S(y)}$ and $\hat{N}'_{S(y)}$ increase with y. As $\overline{\mu}$ increases, $y = \frac{1}{4} E^{\mathrm{A}} \chi \mathrm{A} / \overline{\mu}$ decreases and so does S_y that solves the associated equation. By chaining, both functions $\hat{L}'_{S \circ y(\overline{\mu})}$ and $\hat{N}'_{S \circ y(\overline{\mu})}$ decrease with $\overline{\mu}$. The same reasoning applies to function OC'_S.

[ii] With respect to A, y increases so the same line of argumentation applies to the optimal solution, yet in the opposite direction of variation.

Then, in the fixed demand case, the wider the area to serve, the more infrastructure length and service vehicles are required for optimal service provision, yet also with increased station spacing and with larger operator costs.

We may tentatively apply the same line of reasoning to the shape parameter, χ: larger values (e.g. 2 for grid networks) would yield larger solution triplets and operator costs than smaller ones (e.g. 1 for directional parallel lines). But this may well exceed the outreach of the model.

9.2 Master program to determine dual parameter and demand flows

In addition to the fixed demand problem, the variable demand problem involves the determination of the dual parameter $\overline{\mu}$ that acts as a scaling parameter, together with that of the demand flows: both are taken as exogenous factors in the fixed demand problem.

So, in the master program dealing with system optimization under variable demand, let us specify the endogenous vector as follows: $[\overline{\mu}, (Q_i)_{i \in \mathrm{I}}]$. From it stem all of the other variables in the original system optimization problem, through the following algorithm:

"Derivation chain":

(i) By network component r :

(a) Derive Effect factors E_r^{X} by status X in {A/E, R, W} from current demand flows.

(b) Solve the "fixed demand" problem to yield S_r, L_r and N_r.

(c) Derive the related variables σ_r, v_r, φ_r and d_{Ar}.

(ii) By demand segment i :

(a) Derive the user times by stage X in {A/E, R, W} from the related supply variables, $t_i^{\mathrm{X}} = \sum_{r\in \mathrm{R}} x_{ri}^{\mathrm{X}} t_r^{\mathrm{X}}$.

(b) Derive the individual time costs, $\theta_i := \gamma_i^{\mathrm{A}} t_i^{\mathrm{A}} + \gamma_i^{\mathrm{E}} t_i^{\mathrm{E}} + \gamma_i^{\mathrm{R}} t_i^{\mathrm{R}} + \gamma_i^{\mathrm{W}} t_i^{\mathrm{W}}$.

(c) Derive the optimal tariffs, $\tau_i := -\frac{\overline{\mu}-1}{\overline{\mu}} Q_i / \dot{Q}_i$ by solving the optimality condition about τ_i as a fixed point problem, $\tau_i := -\frac{\overline{\mu}-1}{\overline{\mu}} \mathrm{D}_i(\tau_i + \theta_i) / \dot{\mathrm{D}}_i(\tau_i + \theta_i)$.

(d) Let $g_i := \tau_i + \theta_i$.

Thus we can avail ourselves of "Derivation chain" as a subprogram. It exerts obvious feedback on the current state vector, since its $g_i^{(k)}$ outcomes as of iteration k enable us to infer "new" demand flows through the demand functions: $\hat{Q}_i^{(k)} := \mathrm{D}_i(g_i^{(k)})$.

Yet it is more robust to update the demand flows in a progressive way, by a rule of convex combinations:

$$Q_i^{(k+1)} := u_k \, \mathrm{D}_i(g_i^{(k)}) + (1-u_k) Q_i^{(k)}, \tag{9.9}$$

Wherein $(u_k)_{k\geq 0}$ is a sequence of non-negative numbers that decrease to zero.

A convergence criterion as follows:

$$\mathrm{CC}^{(k)} := \sum_{i\in \mathrm{I}} \left[\mathrm{D}_i(g_i^{(k)}) - Q_i^{(k)} \right]^2, \tag{9.10}$$

is useful to check the convergence of the computational scheme with respect to the demand flows.

Assuming that this convergence has been achieved satisfactorily, the last issue to address consists in making the dual parameter $\overline{\mu}$ satisfy its own condition of optimality.

As a higher $\overline{\mu}$ puts more emphasis on the amount of commercial revenues in the Lagrangian function, the solution optimized with respect to all of the other variables, taken as a function of $\overline{\mu}$ only, will yield larger commercial revenues as $\overline{\mu}$ increases: denoting $\boldsymbol{\tau}\mathbf{Q}(\overline{\mu}) \equiv \sum_{i\in \mathrm{I}} \tau_i^{(\overline{\mu})} Q_i^{(\overline{\mu})}$, function $\overline{\mu} \mapsto \boldsymbol{\tau}\mathbf{Q}(\overline{\mu})$ is an increasing one.

If $\boldsymbol{\tau}\mathbf{Q}(1) \geq \mathrm{SR}$ then the optimality condition is satisfied, so $\overline{\mu} = 1$ together with the derived outcomes provides an optimal system state.

In most cases, however, we can expect that $\boldsymbol{\tau}\mathbf{Q}(1) < \mathrm{SR}$, so that value 1 for $\overline{\mu}$ is only a lower bound and we have to look for a higher value in order to satisfy the budget constraint. To do that, we put forward a dichotomy strategy, by progressively halving an interval $[\mu_{\mathrm{L}}, \mu_{\mathrm{U}}]$ between a lower bound μ_{L} and an upper bound μ_{U}, of which either one is updated in each iteration of the dichotomy algorithm that follows.

Dichotomy scheme:

(i) Set initial interval bounds $\left[\mu_{\mathrm{L}}^{(1)}, \mu_{\mathrm{U}}^{(1)}\right]$ *and iteration counter* $k \coloneqq 1$.

(ii) Let $\hat{\mu} \coloneqq \frac{1}{2}(\mu_{\mathrm{L}}^{(k)} + \mu_{\mathrm{U}}^{(k)})$ *and derive the optimal system state conditionally to* $\hat{\mu}$.

(iii) If $\left|\boldsymbol{\tau}\mathbf{Q}(\hat{\mu}) - \mathrm{SR}\right|$ *is smaller than a given preset tolerance then Terminate, else continue.*

(iv) If $\boldsymbol{\tau}\mathbf{Q}(\hat{\mu}) < \mathrm{SR}$ *then {* $\mu_{\mathrm{L}}^{(k+1)} \coloneqq \hat{\mu}$ and $\mu_{\mathrm{U}}^{(k+1)} \coloneqq \mu_{\mathrm{U}}^{(k)}$ *} else {* $\mu_{\mathrm{L}}^{(k+1)} \coloneqq \mu_{\mathrm{L}}^{(k)}$ and $\mu_{\mathrm{U}}^{(k+1)} \coloneqq \hat{\mu}$ *}.*

(v) Update $k \coloneqq k+1$ *and go to Step (ii).*

Assuming "reasonable" demand functions with elasticity that is bounded and greater than -1, a natural candidate as initial upper bound is (Cf. eqn. (6.14)):

$$\mu_{\mathrm{M}} \coloneqq \frac{1}{1+\varepsilon_{\mathrm{M}}} \quad \text{wherein } \varepsilon_{\mathrm{M}} \coloneqq \min\{\varepsilon_i : i \in \mathrm{I}\}. \tag{9.10}$$

Effectively, if $\boldsymbol{\tau}\mathbf{Q}(\mu_{\mathrm{M}}) > \mathrm{SR}$ then we can use it as the initial upper bound in the Dichotomy scheme. Otherwise, if $\boldsymbol{\tau}\mathbf{Q}(\mu_{\mathrm{M}}) < \mathrm{SR}$ then we can use μ_{M} to update μ_{L} and test $2\mu_{\mathrm{M}}$ as candidate upper bound, and repeat that until the inequality constraint is satisfied.

For the scheme to succeed, it requires that the target revenue SR can be achieved. Taking this condition as granted, then the dichotomy scheme will yield a value of $\overline{\mu}$ that meets the constraint, and this value will be the unique solution since function $\boldsymbol{\tau}\mathbf{Q}$ increases with $\overline{\mu}$. In other words:

Proposition 3: Existence and uniqueness of a system optimum state under Variable demand. *Assuming that the target revenue SR is feasible, there exists a state of System Optimum for the transit system under variable demand. This state is unique.*