Optimal resources allocation to elementary networks

G. Bellei, G. Gentile, N. Papola
Dipartimento di Idraulica, Trasporti e Strade
University or Rome "La Sapienza",
Via Eudossiana 18, 00184 Rome, Italy
Email: gbellei@dits.ing.uniroma1.it

Abstract

A wide range of transport bilevel problems are investigated referring to an elementary network consisting of one Origin Destination (OD) pair, with a given demand, connected by two links. In this context it is shown that these problems, generally non-convex, exhibit several local minima. Most results are supplied in a graphical form and analytical proofs are developed for the NDP with linear investment functions.

1 Introduction

Since the work of Tobin and Friesz[6] has opened the field of assignment sensitivity analysis, several contributions (for example Friesz et al.[3], Yang[8], Davis[2]), aiming at devising solution algorithms for classical transport bilevel optimisation problems (Network Design, Traffic Control Problem, OD Matrix Estimation etc.), exploited their approach to achieve lower level problem differentiation. These algorithms should be considered as heuristics, no exact and general solution to bilevel optimisation problems being available up to now.

In this paper, we formulate and analyse a specific Network Design Problem (NDP), referring to the most elementary conceivable network.

The essentiality of network topology will make the NPD solution a trivial task. On the other hand, this permits us to investigate rather deeply the structure of the problem: thanks to the extreme simplicity of the network, only few variables will be involved, so that we are in a position to see literally what is going to happen as the parameters of the problem vary. Concretely, we will consider a single OD pair, with a given transport demand q, connected by two links/paths a and b, whose average travel costs c_p , $p=a,\ b$, are known functions of the traffic flow f_p on the link and of some kind of resource k_p allocated to it. We then investigate problem (1), in the following referred to as Basic Problem (BP)

$$\begin{aligned}
&\underset{k_a,k_b}{\mathrm{Min}} \quad C = c_a\left(f_a,k_a\right)\cdot f_a + c_b\left(f_b,k_b\right)\cdot f_b \\
&\text{s. to}: \\
&i_a\left(k_a\right)+i_b\left(k_b\right)=I \quad k_a,k_b \geq 0 \\
&f_a + f_b = q \quad f_a,f_b \geq 0 \\
&\left(c_a-\xi\right)\cdot f_a = \left(c_b-\xi\right)\cdot f_b = 0 \quad c_a,c_b \geq \xi
\end{aligned} \tag{1}$$

where:
. the total user cost C is assumed as objective function;
. the whole available budget I is supposed to be spent;
. the equilibrium flow pattern is determined according to the deterministic user equilibrium (DUE);
. investment functions i_p , $p=a,\ b$, are assumed to be linear.

Some modification of the BP will be then introduced and analysed through numerical examples.

2 Link cost functions

Most of the frequently used link congestion cost functions are of the multiplicative type, namely $c=u\cdot[1+d(f)]$, where u is a constant and d is an increasing function of the link flow f. Such a formalisation is not satisfactory when the construction of new infrastructures is involved. In order not to ignore, as it often happens, this aspect of the problem, we use the following functional form:

$$\begin{aligned}
&c_p\left(f_p,k_p\right)=u_p\left(k_p\right)\cdot\left[1+d_p\left(x_p\right)\right] \quad , \quad x_p=\frac{f_p}{k_p} \\
&u_p{}'\leq 0 \quad , \quad d_p\left(0\right)=0 \quad , \quad d_p{}'\geq 0 \qquad p=a,b
\end{aligned} \tag{2}$$

where $u(k)$ is assumed to be a decreasing function of the capacity k and where the $c(f,k)$ functions are assumed to be continuous, strictly increasing with respect to f and strictly decreasing with respect to k. The (2) becomes of the B.P.R. type when $d_p=(1+\alpha_p \cdot x_p\hat{}\beta_p)$, and of the Davidson modified type when $d_p=[1+J_p \cdot x_p/(1-x_p)]$ if $x_p<\mu_p$, $d_p=[1+J_p \cdot (x_p-\mu_p^2)/(1-\mu_p)^2)]$ otherwise. Moreover, in order to avoid optimal solutions that imply the allocation of too small resource quantities, which have no physical meaning, and still contemplate the case of "no investment at all" for some planned link, instead of formulating the problem as a mixed integer program, zero capacity links are associated to an infinitely high uncongested travel time.

In formalising NDP (1) we use the concept of resource without any other specification; this approach allows us to place NDP and Traffic Control Problem (TCP) (Yang & Yagar[7]) in the same framework, when dealing with our elementary network. On this regard, we remember that a detailed representation of traffic lights operation or a different traffic control / route choice problem formalisation (Smith[5]) lead to completely different approaches, while a broad range of traffic management tools may be represented, in aggregate form, as a (road) space and (green) time resources allocation problem. Moreover, while it has been shown (Papola[4]) that, in practical cases, the minimum cycle can be regarded as a good approximation of the optimal cycle, as the saturation degree becomes only moderately high, the corresponding average delay can be conveniently used as a cost function in the context of TCP since it can be reduced to the form $c=u \cdot (1+d)$. In this case resources and investment are the same and represent green time values to be allocated on two conflicting approaches belonging to the same OD pair.

3 Analysing the Basic Problem

3.1 Link/Path utilisation pattern

Budget constrains can be eliminated from the problem (1), by expressing the capacities k_p as a function of the budget share θ, utilising the inverse $k_p(i_p)$, so assuming θ as design variable: $i_a=\theta \cdot I$, $i_b=(1-\theta)\cdot I$, $\theta\in[0,1]$.

We will first deal with the case $u'\neq 0$, where the limit of $u(k)$ as k approaches zero is $+\infty$. The assumption that to zero capacity links corresponds an infinitely high uncongested travel cost gives places to the existence of three ranges of θ: $\theta\in[0,\theta_A)$, only link b is utilised;

$\theta\in(\theta_A, \theta_B)$, both links are utilised; $\theta\in(\theta_B,1]$, only link a is utilised. Points A and B in figure 1 represent limit situations where respectively $c_a(0,\theta)=c_b(q,\theta)$ and $c_a(q,\theta)=c_b(0,\theta)$.

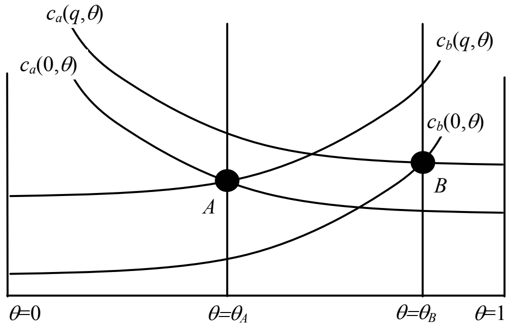

Figure 1: determination of the path utilisation ranges through points A and B.

The existence and the uniqueness of points A and B can be easily proved. In fact, owing to the properties of the $c(f,\theta)$ functions, it is $c_a(0,0)>c_b(q,0)$ and $c_a(0,1)<c_b(q,1)$. The assertion follows. The same holds for point B. The properties of the functions also imply that $\theta_A<\theta_B$, as can be seen in figure 1.

3.2 Analysis of total cost function

We will analyse here the shape of the total cost C when a monodimensional function of the budget share is taken into consideration, by determining f_a and f_b, for any θ, according to DUE.

For values of $\theta\in[0,\theta_A)$, total cost C is an increasing function. Investment is in fact subtracted to the utilised path and it is given to the unutilised one, which, in practical terms, means a loss of capacity. Symmetrically, for values of $\theta\in(\theta_B,1]$, total cost C is a decreasing function (see figure 1).

Let us consider now the range $\theta\in(\theta_A,\theta_B)$ where both paths are utilised. Within this range the variation of path cost, following an investment reallocation $\mathrm{d}\theta$, is equal on both paths ($\mathrm{d}c_a=\mathrm{d}c_b$) by definition. Then the necessary and sufficient condition for the total cost $C=c_a{\cdot}f_a+c_b{\cdot}f_b=c_a{\cdot}q$ to decrease ($\mathrm{d}C<0$) as a consequence of this investment reallocation is that $\mathrm{d}c_a=\mathrm{d}c_b<0$. Considering again investment functions, it is: $\mathrm{d}i_p=i_p^k{\cdot}\mathrm{d}k_p$, which implies $\mathrm{d}k_p=\mathrm{d}i_p/i_p^k$, $p=a,\ b$; so that, expressing cost as a function of capacity as in eqn. (2):

$$\mathrm{d}c_p = c_p^f \cdot \mathrm{d}f_p + c_p^k \cdot \mathrm{d}k_p = c_p^f \cdot \mathrm{d}f_p + c_p^k \cdot \frac{\mathrm{d}i_p}{i_p^k} \qquad (3)$$

where the superscripts f and k denote derivatives with respect to the corresponding variables. The DUE implies:

$$c_a^f \cdot \mathrm{d} f_a + c_a^k \cdot \frac{\mathrm{d} i_a}{i_a^k} = c_b^f \cdot \mathrm{d} f_b + c_b^k \cdot \frac{\mathrm{d} i_b}{i_b^k} \tag{4}$$

so that, being, obviously, $\mathrm{d}f_b = -\mathrm{d}f_a$ and $\mathrm{d}i_b = -\mathrm{d}i_a$, the derivative of the equilibrium flow with respect to investment (on link a) is:

$$\frac{\mathrm{d} f_a}{\mathrm{d} i_a} = -\frac{c_a^k / i_a^k + c_b^k / i_b^k}{c_a^f + c_b^f} \tag{5}$$

while the derivative of total cost results to be:

$$\frac{\mathrm{d} C}{\mathrm{d} i_a} = q \cdot \frac{\mathrm{d} c_a}{\mathrm{d} i_a} = q \cdot \frac{c_b^f \cdot c_a^k / i_a^k - c_a^f \cdot c_b^k / i_b^k}{c_a^f + c_b^f} \tag{6}$$

When considering cost functions of the less general but extensively used form (2), partial derivatives are:

$$c_p^k = u_p' \cdot (1 + d_p) - u_p \cdot d_p' \cdot \frac{f_p}{k_p^{\,2}} \quad , \quad c_p^f = d_p' \cdot \frac{u_p}{k_p} \tag{7}$$

and derivative of equilibrium cost c_a is:

$$\frac{\mathrm{d}c_a}{\mathrm{d}i_a} = \frac{\dfrac{d_b' \cdot u_b}{i_a^k \cdot k_b} \cdot \left[u_a' \cdot (1+d_a) - u_a \cdot d_a' \cdot \dfrac{f_a}{k_a^{\,2}} \right] - \dfrac{d_a' \cdot u_a}{i_b^k \cdot k_a} \cdot \left[u_b' \cdot (1+d_b) - u_b \cdot d_b' \cdot \dfrac{f_b}{k_b^{\,2}} \right]}{\dfrac{d_b' \cdot u_b}{k_b} + \dfrac{d_a' \cdot u_a}{k_a}} \tag{8}$$

Equation (8) in points A and B, where there is no congestion on the empty link ($d(0)=0$), simplifies considerably. When the additional condition $d'(0)=0$ is assumed, as it happens when using BPR functions, remembering that $i_a=\theta \cdot I$, from the (6) and the (8) we get:

$$\lim_{\theta \to \theta_A^{+}} \frac{\mathrm{d} C}{\mathrm{d} \theta} = \frac{u_a'}{i_a^k} \cdot q \cdot I \le 0 \quad , \quad \lim_{\theta \to \theta_B^{-}} \frac{\mathrm{d} C}{\mathrm{d} \theta} = -\frac{u_b'}{i_b^k} \cdot q \cdot I \ge 0 \tag{9}$$

In conclusion, when $u' \neq 0$ and $d'(0)=0$, the total cost function $C(\theta)$ is increasing on the left of points A and B and decreasing on the right, which means that these points are two local maxima. Moreover, because the derivative on the left is not equal to that on the right, in points A and B $C(\theta)$ has a cusp (see figure 2).

In the numerical examples we use BPR functions with parameters $\alpha_a=\alpha_b=1{,}5$, $\beta_a=\beta_b=4$, assuming $u_p(k_p)=l_p/V^0_p(k_p)$. As far as the u functions parameters is concerned, the following values have been used:

l_a=50 km, l_b=45 km (l_b=35 km when u'=0); when $u' \neq 0$ $V^0_p(k_p) = \rho \cdot k_p{}^{\sigma}$, ρ=0,3 , σ=0,7; when u'=0 $V^0_p(k_p)$=50 km/h. In order to deal with more appreciable values, in the figures, instead of the total cost C, the average total cost C/q has been depicted.

It can be observed that, since in the central range total cost $C(\theta)$ is decreasing on the left hand side and increasing on the right hand side, the continuity of the function and of its derivative ensures the existence of at least one relative minimum. As it is shown in figure 2 the actual shape of the function depends on the demand / budget ratio: if it is sufficiently low, then only one relative minimum arises; vice versa high values of the ratio exhibits two relative minima and one relative maximum.

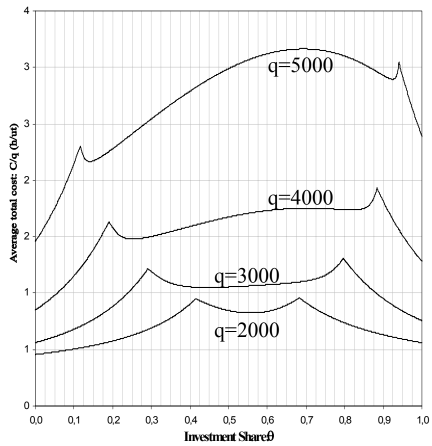

Figure 2: BP average total cost functions vs. θ when $u' \neq 0$.

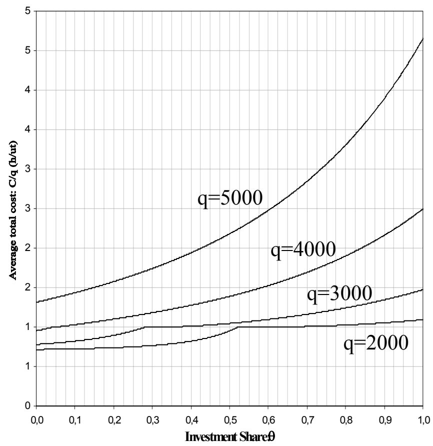

Figure 3: BP average total cost functions vs. θ when u'=0.

Let us consider now the case where u does not depend upon k (u'=0). With no loss of generality let $u_a > u_b$, so that path b is always utilised. Naturally enough point B then disappears, while, as it can be shown, the range where path b alone is utilised appears only if $k_b{}^{max} > k_b{}^{lim}$, being $k_b{}^{max}$ the maximum capacity corresponding to the whole budget I and $k_b{}^{lim}$ a specific capacity value defined by the following equation: $d_b(q/k_b{}^{lim}) = u_a/u_b - 1$. Considering that u'=0, when point A exists, the derivative of the total cost function $C(\theta)$ is positive on its left and zero on its right (see eqn. (9)), so that this is no more a maximum; though it is clearly a cusp (see figure 3).

A central hypothesis in achieving the above results was $d'(0)$=0. As eqn. (8) shows, in cases where $u' \neq 0$ but $d'(0)>0$ the sign of dC/dθ depends on the cost function parameters.

3.3 Global minimum

In the present case, as expected, the numerical results show that the global minimum is achieved when the whole budget is allocated to one of the two paths, which of them depending on the parameters values.

The assertion can be easily proved for the particular case of a symmetrical network with $u'\neq 0$. In this case, in fact, for $\theta\in(\theta_A,\theta_B)$ the stationary point must fall, for a matter of symmetry, exactly in the middle. Naming K the total capacity allowed by available budget I, in correspondence of this point, saturation degree is the same for both links and it is equal to $(q/2)/(K/2)=q/K$. This is also the saturation degree experienced in the two external solutions and congestion cost is thus the same in these three points. Uncongested travel cost, however, is decreasing with capacity; thus, since $C(0)=C(1)=q\cdot u(K)\cdot[1+d(q/K)] < C(1/2)=q\cdot u(K/2)\cdot[1+d(q/K)]$, it is evident that external solutions are better than the central one.

4 Modifying the Basic Problem

4.1 Different route choice models

In figure 4 are depicted the average total cost curves C/q, for $u'\neq 0$, when considering a deterministic system equilibrium DSE flow assignment; the corresponding DUE curves are also depicted for comparison. It is worth noting that DSE curves are, by definition, regularly non-above the corresponding DUE ones.

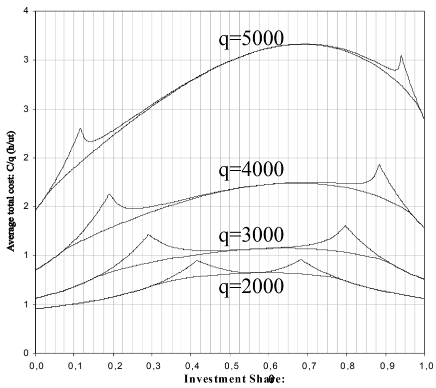

Figure 4: DSE and DUE average total cost functions vs. θ when $u'\neq 0$.

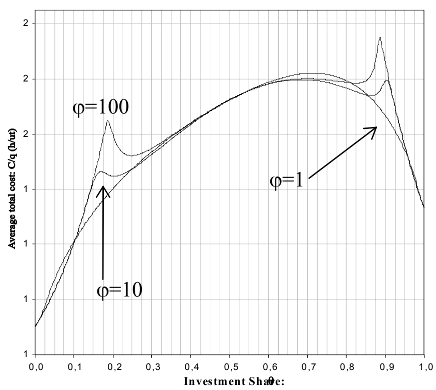

Figure 5: SUE average total cost functions vs. θ when $u'\neq 0$.

In figure 5 are depicted the average total cost curves C/q when stochastic user equilibrium (SUE) is considered, adopting a logit model with different values of the parameter φ, while q is kept equal to 4000 veic/h. The case $u'\neq 0$ is considered. Since now are always utilised both paths, it does not make sense considering different ranges of utilisation and cusps do not exist any more; yet for sufficiently high values of the logit parameter φ, (low utility variance) a shape similar, a part from the cusps, to the ones already met in the BP can be easily recognised. Clearly non-convexity, being an intrinsic feature of the problem, does not disappear by switching to the stochastic approach, (see figure 5).

4.2 Scale economies and diseconomies

Investment functions i_p take into account physical features of the links as well as environmental costs associated to resource allocation such as traffic pollution, visual impact, contiguity with environmentally sensitive or densely populated areas and land separation effects.

In the present work, to get numerical examples, the following simple form has been considered: $i_p = l_p \cdot \chi_p \cdot k_p \hat{}\, \gamma_p$, being l_p the length of the link, χ_p and γ_p two calibration parameters. Throughout all examples we assume χ_p to be equal to 2.500.000 for both paths, while budget I was taken equal to 500 billions of It.Lires. Scale economies and diseconomies are then simulated by considering respectively a value for γ_p less than one (0,7 in the examples) and higher than one (1,4 in the examples).

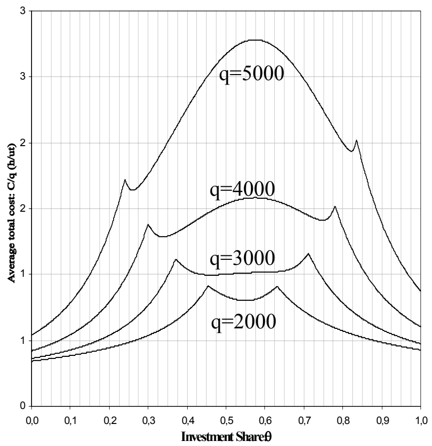

Figure 6: scale economies average total cost functions vs. θ when $u'\neq 0$.

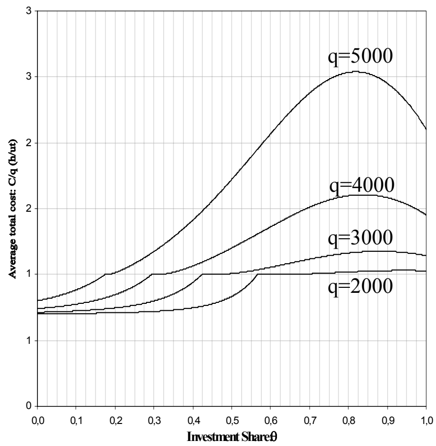

Figure 7: scale economies average total cost functions vs. θ when $u'=0$.

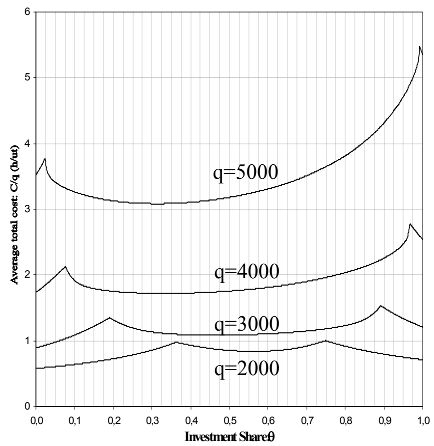

Figure 8: scale diseconomies average total cost functions vs. θ when $u' \neq 0$.

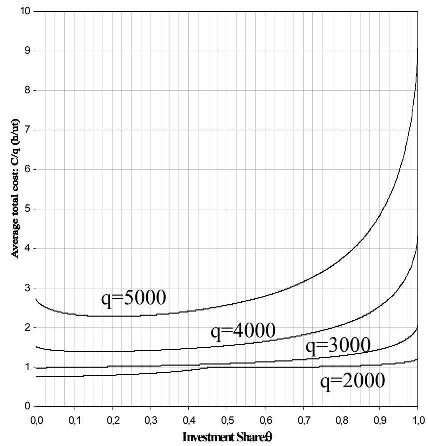

Figure 9: scale diseconomies average total cost functions vs. θ when $u'=0$.

Scale economies can be justified in connection with new roar construction, scale diseconomies in connection with the above mentioned environmental costs. The effects of scale economies and scale diseconomies on the total cost function shape, for different values of q, are depicted in figures 6 and 7 and in figures 8 and 9, respectively; where figures 6 and 8 refer to $u' \neq 0$ and figures 7 and 8 to $u'=0$. It is well worth noting that in the case of scale diseconomies internal solutions tend to be better so that the global minimum may correspond to points where resources are allocated to both links. Scale economies still imply an “only to the best link” solution.

4.3 Relaxing the budget constraint

We consider now the following problem:

$$\begin{aligned}
&\underset{k_a,k_b}{\text{Min}} \quad F = \lambda \cdot \left[c_a\left(f_a,k_a\right)\cdot f_a + c_b\left(f_b,k_b\right)\cdot f_b\right] + \psi \cdot \left[i_a\left(k_a\right) + i_b\left(k_b\right)\right] \\
&\text{s. to:} \\
&f_a + f_b = q \quad f_a, f_b \geq 0 \quad k_a, k_b \geq 0 \\
&\left(c_a - \xi\right)\cdot f_a = \left(c_b - \xi\right)\cdot f_b = 0 \quad c_a, c_b \geq \xi
\end{aligned} \tag{10}$$

where, contrariwise to problem (1), the budget constraint is removed and the cost of the investment is added to the total user cost in the objective function and where λ and ν are assumed equal to 43.800.000 and 0,1 respectively.

In figures 10 and 11 are depicted the $F(k_a,k_b)$ surfaces by means of sections F=constant where no scale economies or diseconomies are present. By means of vertical sections, I=constant, we can obtain, instead, two dimensional curves identical, as far as the shape is concerned, to the above corresponding $C(\theta)$ curves. In particular, figure 10 refers to the case $u'\neq 0$, while figure 11 refers to $u'=0$. In both cases, as expected, global minimum lays in points where it is $k_a=0$ or, alternatively, $k_b=0$.

The same problem, formulated according to a DSE gives place to a saddle point as depicted in figure 12.

6 Conclusions

A number of different conceivable NDP nave been examined. Namely, from the supply side (upper problem): total user costs subject to budget constraint, with scale economies and diseconomies; total user cost plus investment cost. From the demand side (traveller's' response): deterministic user equilibrium; stochastic user equilibrium; deterministic system equilibrium.

Despite the extreme simplicity of the network, and really thanks to that, we were able to fully investigate the behaviour of the system and to carry out some general results. The non-convexity appears to be an intrinsic characteristic of bilevel problems: in fact, whatever case we consider, unless we limit ourselves to assume the uncongested travel time not dependent upon the investment, several local minima are present. Moreover, on condition that the parameters are opportunely specified, the system exhibits similar behaviours, apart from some specificity like the existence of the cusps within the deterministic framework. Finally it is worth noting that slight variations of the parameter values modify substantially the behaviour of the system.

The results, far from being conclusive, represent an interesting starting point in the direction of investigating the general behaviour of the problem when real networks are considered.

References

[1] Bell, M.G.H. & Iida, Y., *Transportation Network Analysis*, John Wiley & Sons, Chichester, 1997.

[2] Davis, G., Exact local solutions to the continuous network design problem via stochastic user equilibrium assignment, *Transpn Res*,. **27B**, pp. 61-75, 1993.
[3] Friesz, T.L., Tobin, R.L., Cho, H.J. & Mehta, N.J., Sensitivity analysis based algorithms for mathematical programs with variation inequality constraints, *Math. Prog*,. **48**, pp. 265-284, 1990.
[4] Papola, N., Teoria del deflusso veicolare e suo impiego nella progettazione e nella regolazione, *Atti di Trasporti*, **10**, 1988.
[5] Smith, M.J. & Van Vuren, T., Traffic Equilibrium with Responsive Traffic control, *Transpn Sci.*, **27**, pp. 118-132, 1993.
[7] Tobin, R.L. & Friesz, T.L, Sensitivity analysis for equilibrium network flows, *Transpn Sci.*, **22**, pp. 242-250, 1988.
[8] Yang, H. & Yagar, S., Traffic assignment and signal control in saturated road networks, *Transpn Res.*, **29A**, pp. 125-139, 1995.
[9] Yang, H., Heuristic algorithm for the bilevel origin-destination matrix estimation problem, *Transpn Res*. **29B**, pp. 231-242, 1995.

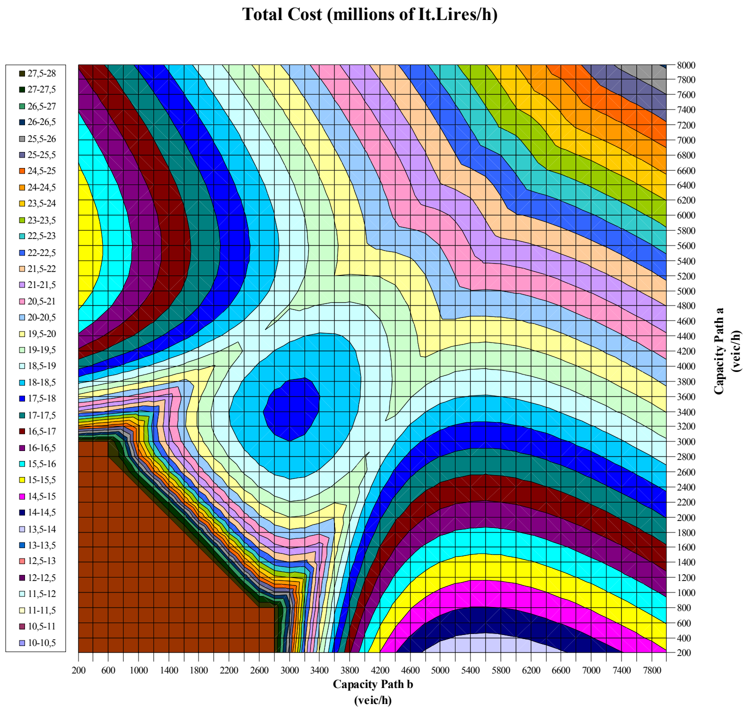

Figure 10: $F(k_b,k_a)$ surface depicted by means of sections F=constant where no scale economies or diseconomies are present and $u'\neq 0$.

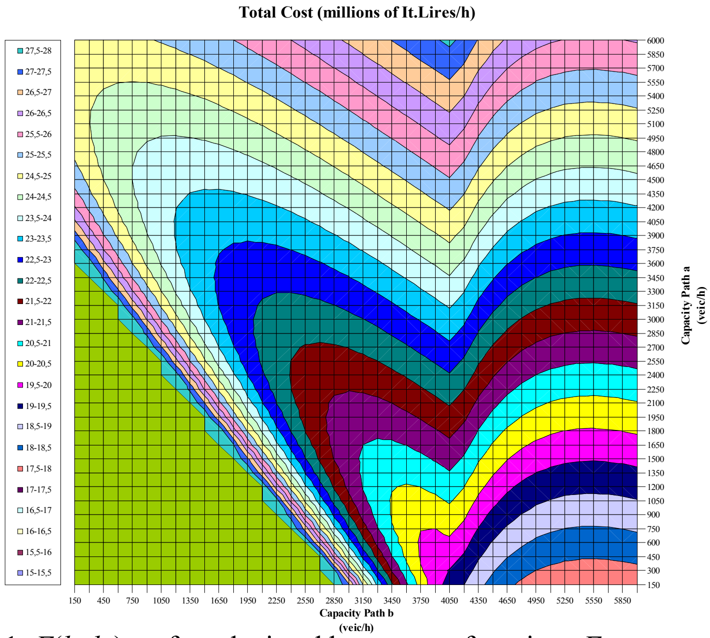

Figure 11: $F(k_b,k_a)$ surface depicted by means of sections F=constant where no scale economies or diseconomies are present and $u'=0$.

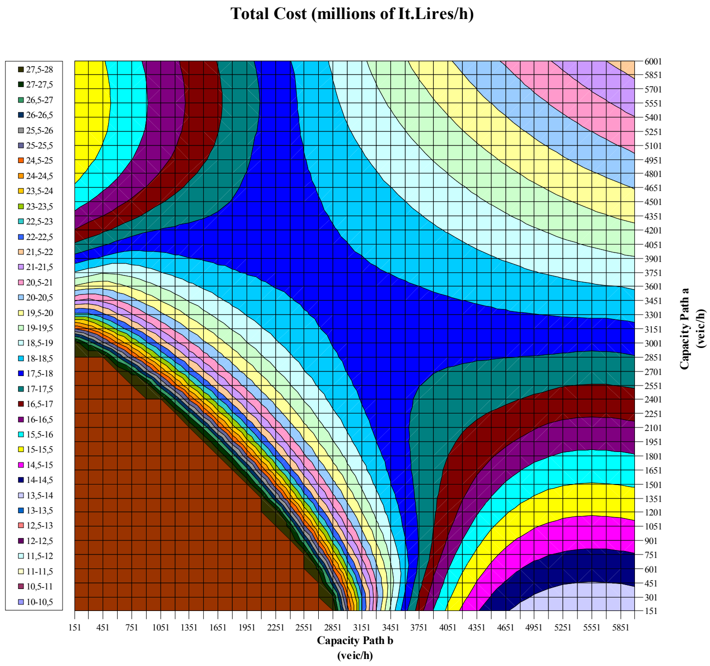

Figure 12: $F(k_b,k_a)$ surface when the problem is the same as (10) but formulated according to a DSE and $u'\neq 0$.