Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 23 (2017) 700–719

Transportation Research Procedia

www.elsevier.com/locate/procedia

22nd International Symposium on Transportation and Traffic Theory

On characterizing the relationship between route choice behavior and optimal traffic control solution space

Marco Rinaldi[a]*, Chris M.J.Tampère[b], Francesco Viti[a]

[a]*University of Luxembourg, FSTC, 6 rue Richard Coudenhove-Kalergi, L-1359 Luxembourg, Luxembourg*
[b]*KU Leuven, L-Mob Leuven Mobility Research Centre, Celestijnenlaan 300A, 3001 Leuven, Belgium*

Abstract

Explicitly including the dynamics of users' route choice behaviour in optimal traffic control applications has been of interest for researchers in the last five decades. This has been recognized as a very challenging problem, due to the added layer of complexity and the considerable non-convexity of the resulting problem, even when dealing with simple static assignment and analytical link cost functions. In this work we establish a direct behavioural connection between the different shapes and structures emerging in the solution space of such problems and the underlying route choice behaviour. We specifically investigate how changes in the active equilibrium route set exert direct influence on the solution space's structure and behaviour. Based on this result, we then formulate and validate a constrained version of the original problem, yielding desirable properties in terms of solution space regularity.

Peer-review under responsibility of the scientific committee of the 22nd International Symposium on Transportation and Traffic Theory.

Keywords: Traffic Control; Route Choice; Optimization

* Corresponding author. Tel.: +352/46 66 44-6281; fax: +352/46 66 44-36281.
E-mail address: marco.rinaldi@uni.lu

Peer-review under responsibility of the scientific committee of the 22nd International Symposium on Transportation and Traffic Theory.
10.1016/j.trpro.2017.05.039

1. Introduction

When considering highly saturated road networks, it has been widely recognized that capturing user's route choice alongside vehicle propagation and congestion dynamics is essential to the development of advanced mobility management strategies. As congestion increases, users will naturally explore different options for reaching their destination, first and foremost in terms of rerouting. From a theoretical perspective, this concept has been historically defined under the name of User Equilibrium (Wardrop, 1952), and essentially captures the natural tendency of road users to seek minimum cost connections from their origin to their desired destinations.

Traffic models have gained popularity over the years, especially in connection with practices such as traffic planning and management. The reason behind this is that as models evolved towards higher accuracy, lower computational costs and improved behavioural richness, they became very attractive tools to predict the effects of changes in either demand management or infrastructure planning. Such an example is the development of model-based optimal traffic control policies, which can be seen as an instance of the Network Design Problem (Johnson et al., 1978). In these applications, an optimization framework aiming at meeting specific policy goals is wrapped around a traffic model, such that the latter can be procedurally used to guide the optimization procedure towards desirable solutions.

Congested traffic behaves in rather complex patterns: as travel times arise, different geographical portions of the network become behaviourally connected, through either spillback dynamics and/or vehicle rerouting. Computing optimal coordinated control in these instances, a problem referred to as the anticipatory traffic control problem ever since the seminal work of (Allsop, 1974), exhibits considerable difficulties. Sensitivity analysis is a necessary tool to understand how equilibrium-constrained control problems behave, but due to unfavourable conditions such as non-uniqueness of path flows and (under lax assumptions) discontinuous behaviour of route selection dynamics, this rather standard tool's performance is strongly jeopardized.

In previous work we empirically assessed how even for very simple scenarios these behavioural concerns strongly influence the shape and structure of standard network control objective functions (Rinaldi and Tampère, 2015). Specifically, we showed how for deterministic static user equilibrium constrained problems, the Total Cost objective function for a very small network with monotonically increasing separable cost functions exhibited strong non-linear behaviour, and how gradient-based optimization techniques cannot generally be guaranteed to reach global optimality. In this novel work, we analyse the underlying causes for these unfavourable conditions, and employ the newly developed insights to improve the tractability of the anticipatory traffic control problem. Our focus is specifically the optimal pricing control problem, subject to static deterministic User Equilibrium.

Specifically, we claim the following contributions for this paper:

- We observe a direct connection between the active route set (expressed through an atomic representation – see Section 3.1) and specific features (later referred to as "regions") in the problem's solution space shape.
- We validate our observation analytically and numerically on a small example.
- We extend the standard bi-level optimization approach used in anticipatory control with a set of constraints capable of capturing this direct connection, and benefiting from it in terms of solution space regularity.
- We showcase the impact of the newly developed constraints for two small networks, analysing optimization performance both in full and partial controllability instances.

To achieve these objectives, we choose a specific way of describing route sets and route choice behaviour, introduced by (Bar-Gera, 2010, 2006). As we will detail in later sections, this powerful tool allows to atomically capture route choice and route sets obviating the need of exhaustive enumeration. To the best of our knowledge, this is the first exploration of how to connect a specific route choice description to bi-level optimization with equilibrium constraints and to take explicit advantage of the system optimal route set structure in optimizing 2nd best network pricing.

The remainder of this paper is structured as follows: Section 2 provides a literature review of equilibrium problems, their sensitivity and their impact in constrained optimization. In Section 3 we explore the relationship between route choice and solution space of bi-level NDP formulations, and discuss the arising of regions and their properties, focusing on how this concept can be exploited to develop better optimization techniques. After these techniques are introduced, in Section 4 we explore how these behave in two small sized networks, highlighting the difference between our novel approach and standard gradient-based optimization. Finally, we provide conclusions and remarks related to future research in Section 5.

2. Literature Review

Deriving mathematical descriptions of user behaviour, specifically in terms of route choice, in transportation networks has been of interest for researchers since the mid-1900s, a notable example being the seminal work of (Wardrop, 1952). In his "first principle", Wardrop introduced the concept of static User Equilibrium (UE) as a macroscopic description of route choice behaviour, in which he postulated that, ideally, users distribute over a network such that "no individual could independently change his/her choice without incurring in an increase in his/her own travel cost". Although UE is a simplification of individual choices in a heterogeneous population (as it assumes homogeneous behaviour, and solely considers the cost of travel as the decisive variable), it has been widely adopted by researchers, for its immediacy and, under assumptions, greater mathematical tractability (Beckmann et al., 1956). A considerable research track has been in the meantime involved in developing finer, more comprehensive models for route choice; while these sophisticated models are rather outside of this paper's scope, we refer the interested reader to the review paper of (Prato, 2009).

In parallel to these developments, higher level problems connected with design choices, such as infrastructure planning and operations, demand estimation and traffic control were extensively researched under the common name of Network Design Problem (NDP) (Abdulaal and LeBlanc, 1979; Johnson et al., 1978). These problems are often formulated in terms of optimization, where the main objective function (e.g. determining optimal pricing, minimizing errors between estimated and measured flows in demand estimation,…) is subject to User Equilibrium constraints.

In this work we focus specifically on the problem of determining optimal traffic control policies subject to UE constraints, a practice often referred to as the anticipatory traffic control problem, which originated in the work of (Allsop, 1974). This problem has received notable attention over the years, both in terms of pricing controllers (Yan and Lam, 1996; Yang and Bell, 1997, 1997; Zhang and Yang, 2004) and optimal intersection signalling (Cantarella et al., 1991; Smith and Ghali, 1990; Yang and Yagar, 1995). More recently, these problems have been formulated and applied to dynamic UE, e.g. in the works of (Rinaldi et al., 2016; Taale and Hoogendoorn, 2012; Ukkusuri et al., 2013). Including UE constraints in NDP problems introduces however unfavourable hardships, especially in terms of optimization complexity. Sensitivity analysis of equilibrium-constrained problems has therefore received major attention by researchers, as its nature and properties pose considerable limitations on how well widespread optimization schemes might converge towards the global minimum. Works such as those of (Friesz et al., 1990; Tobin and Friesz, 1988; Yang and Bell, 2007; Ying and Miyagi, 2001) have been exploring the properties of UE constraints sensitivity. A common conclusion is that the introduction of UE constraints considerably influences the shape and structure of sensitivity, and thus the overall performance of objective function minimization procedures. Indeed, UE constraints introduce non-convexity to the problem and, especially, non-smoothness. In his paper, Patriksson (2004) further analysed the origin of non-smoothness and proposed an exact scheme for determining sensitivity and descent directions of UE constrained problems under specific conditions, employing sub-gradients as a way to directly deal with the problem's unattractive properties.

In this paper we seek an alternative, simpler solution to this issue, with the main objective of enabling the most general purpose class of optimization schemes, gradient-based optimization, to reach high levels of performance. We achieve this by first isolating the cause of non-smoothness and non-linearity in UE constrained problems, and then by exploiting this knowledge to modify the overall problem, based on specific prior knowledge, to improve its tractability. In order to achieve our objectives, we base ourselves on observations related to how UE problems have been solved algorithmically. One of the key complexities related to UE problems is the non-uniqueness and combinatorial characteristics of route set enumeration. In the last decade, few authors have introduced UE algorithms who obviate this enumeration while improving both on overall computational time and, more importantly, precision. In their works, authors such as Dial (2006) and Gentile et al. (2004) have elaborated on decomposing the problem of User Equilibrium by considering separated sub-networks, either origin-rooted or destination-rooted, and have shown that overall network equilibrium can be reached by independently equilibrating these sub-networks. In his works, Bar-Gera reached similar conclusions (Bar-Gera, 2010, 2002), although his chosen sub-network representation is considerably more atomic. Throughout the rest of the paper we indeed adopt Bar-Gera's chosen description of route sets and route choice. As we'll discuss in the next Section, we believe this description to be sufficient, under specific hypotheses, to fully represent route choice behaviour. We then establish a direct connection between this route choice representation and the shape and structure of sensitivity of UE constrained problems.

3. Methodology

3.1. Pairs of Alternate Segments (PASs) as a description of route choice

We begin this section by providing formal definitions of “Paired Alternative Routes” and “Paired Alternative Segments” (Bar-Gera, 2006):

Def: A dependent set of paired alternative routes R^0 is a set of n pairs of routes, possibly with repetitions, connecting all ODs in the network. The set is defined as $R^0 = \{[r_1, r'_1, r_2, r'_2\, r_3, r'_3, ...]\}$ such that the two following conditions are satisfied: (a) each pair of routes $[r_i, r'_i]$, named “primary” and “secondary”, consists of two alternatives for the same OD pair pq, and (b) the total number of times a link is used by the primary routes is the same as the amount of times it is used by secondary routes.

This definition stems from the author’s objectives of studying the composition of route sets from the perspective of solving the Maximum Entropy User Equilibrium problem. The author shows that in solutions in which “primary” routes appear as part of the minimum-cost equilibrium route set R^*, “secondary” routes defined as above are always part of the set itself and vice-versa, that is $r_i \in R^* \leftrightarrow r'_i \in R^*$. This implies that describing the route choice problem in terms of PASs is highly beneficial in terms of forming a consistent (if not correct) approximation of the equilibrium route set.

Paired Alternative Segments (PASs) are successively defined as a set of *basic route pairs* T^0, composed such that all routes forming the original set of paired alternative routes R^0 can be equivalently obtained through unions of these basic route pairs on *all* OD pairs pq:

$$\begin{aligned} & T^0 = \cup_{pq} T^0_{pq},\ T^0_{pq} \subseteq \{[r, r'] : r, r' \in R^0_{pq}\} \\ & r = \{s_0, s_1, ..., s_m\} \in R^0_{pq} : [s_{i-1}, s_i] \in T^0_{pq}\ \forall 1 \le i \le m \\ & r' = \{s'_0, s'_1, ..., s'_n\} \in R^0_{pq} : [s'_{i-1}, s'_i] \in T^0_{pq}\ \forall 1 \le i \le n \\ & R^0 = \cup_{pq} R^0_{pq}, \end{aligned} \tag{1}$$

where $s_0, ..., s_m$ are route segments composing the original routes $r \in R^0_{pq} \subseteq R^0$. PASs can thus be regarded as basic building blocks for representing route sets and, consequentially, route choice behaviour.

As mentioned earlier, path-based algorithms for static UE assignment exploiting (directly or indirectly) the concept of PASs have been introduced in literature (Bar-Gera, 2010; Dial, 2006; Nie, 2010). These algorithms yield, upon convergence, not only equilibrium route/link flows but also the set of PASs T^0 that have been considered throughout the procedure of convergence, together with their final “state”. These states can be summarized, OD pair-wise, as follows:

- **a) Inactive PAS:** These are all PASs which can be theoretically constructed considering all couples $[s_i, s'_i]$ where all flow is on the free-flow shortest segment i, while segment i' has no flow. Although very high in numbers on general network, due to combinatorial explosion, knowledge on these PASs offers no information as to how the network is behaving.
- **b) Active PAS:** Both branches of the PAS are being utilized, which leads the flow distribution of the PAS to be equilibrated (thus, the perceived cost difference between its branches is nil).
- **c) Overly Active PAS:** The set of PASs whose flow has entirely shifted from the uncongested free-flow branch i towards its alternative i'. These PASs represent situations in which the changes of costs due to demand flows (generated by other users in the network traveling on different PASs sharing a few links with the one currently under consideration) has strongly influenced the network, away from its natural, uncongested behaviour.

A simple sketch of this idea is shown in Figure 1, where we assume free flow costs $c_0(s_0) < c_0(s_0')$, $c_0(s_1) < c_0(s_1')$ and $c_0(s_2) < c_0(s_2')$.

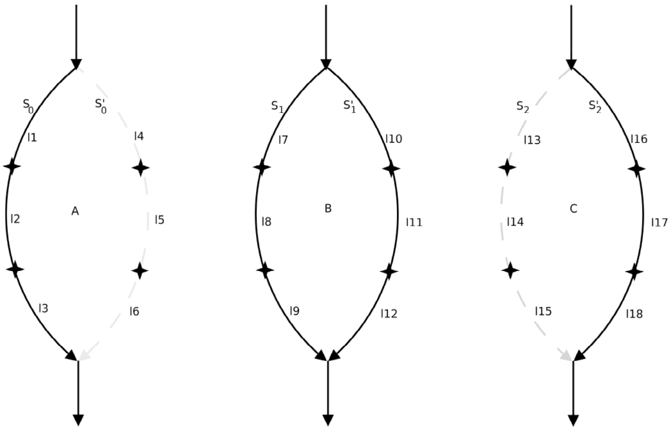


Fig. 1: Three PAS example. Left: Inactive PAS, Centre: Active PAS, Right: Overly Active PAS

In this paper we choose to characterize route sets and route choice behavior by considering the sets b) and c). This isn't a unique representation, as can be simply seen by considering the example of Figure 2.

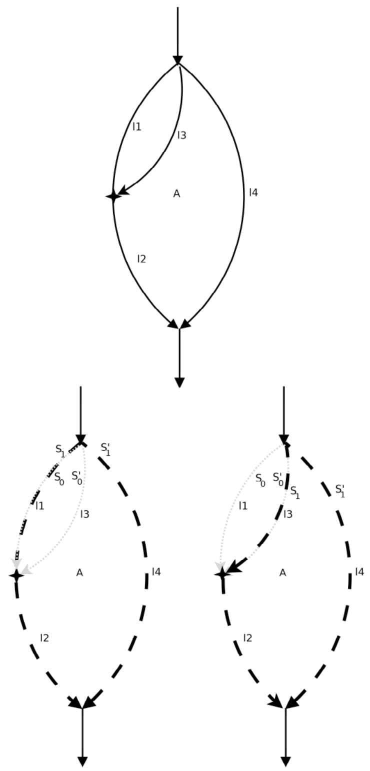
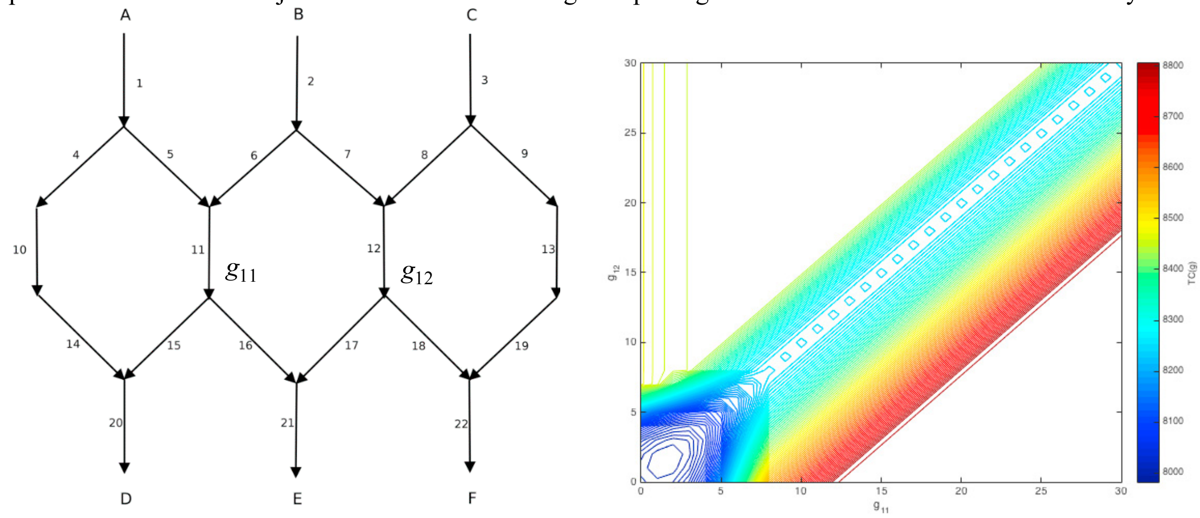


Fig. 2: Multiple PAS representations of the same network

The simple network, composed of four links, can be represented by the two PAS sets $[s_0, s_0']$, shown as bright grey dotted lines, and $[s_1, s_1']$ shown in thicker black lines. The two representations at the bottom of Figure 2, while different

from a visual perspective, are both valid descriptors of the full route choice in this simple three-route example. In fact, PAS structures are sufficient to capture a complete snapshot of the underlying network, under stationary conditions, as long as equilibrium can be shown to be unique. Moreover, as cost or demand changes over the network, we will empirically show how these two sets of PASs will be the most likely affected, while the inactive set a) will only react to major variations. This implies that the completeness of our chosen representation is rather robust when performing sensitivity analysis, a trait that will be exploited in the second portion of this Section.
As discussed earlier in the introduction, our goal is to investigate the origin of strong irregularities in sensitivity of UE constrained problems. This stems from our own previous research, in which we showed how irregular the solution space of the Total Cost objective function is for changes in pricing controllers. We now introduce this very example.



(a) (b)

Fig. 3: Simple three-PAS network (a) and respective total cost solution space shape (b)

For the small network of Figure 3(a) we performed an extensive exploration of the Total Cost objective function's solution space. The example features BPR-like cost functions on links 10-13 and two pricing controllers, g_{11} and g_{12} , equipped on the central links. Demand flows vertically from origins A, B and C towards destinations D, E and F respectively. Each OD couple only has two route alternatives (left or right), which conveniently correspond with the only possible PAS-based representation for this network.
What can be intuitively concluded by interpreting the resulting shape, shown in Figure 3(b), is that different behavioural "regions" emerge in this space: when both g_{11} and g_{12} are below values of 10-12, a fairly regular, convex pattern can be seen, and a clear minimum is found at $g = [2.5, 2.5]$. Outside these boundaries, a diagonal "valley" arises, surrounded by two linearly sloped regions. As individual controllers keep increasing in value, the solution space becomes flat, insensitive to further changes. Finally, a distorted area appears at the top-left of Figure 3(b), in correspondence of values of g_{11} in the interval $[0-3]$ and g_{12} above a value of 12. The multiplicity of areas with different behaviours and, more importantly, different local minima motivates the research performed in this work, in which we try to establish first why these regions exists and further how to efficiently deal with them.

Our key hypothesis is that these regions are due to route activation/deactivation dynamics, i.e. due to changes in the *active route set.* As previously utilized routes become unattractive to flow and/or new routes activate, the overall response of the network changes considerably.

We can show this intuition to hold analytically, by considering a simplified example of the network of Figure 3(a), where linear cost functions are equipped in place of BPR cost curves.

Table 1. Linear cost parameters for the Three-bush network

Link	Parameters
Link 10	$a_{10}=1, b_{10}=10$
Link 11	$a_{11}=2, b_{11}=4$
Link 12	$a_{12}=1, b_{12}=4$
Link 13	$a_{13}=1, b_{13}=10$

Given the network parameters of Table 1, the analytical formulation of UE for this simple case can be setup as follows, assuming all routes being active in the User Equilibrium solution (note that in case of inactive or overly active route pairs, the equalities in Eq (4) need to be modified accordingly into inequalities):

$$\begin{cases} c_{r_1} = f_{10} + 10 \\ c_{r_2} = 2 \cdot f_{11} + 4 \\ c_{r_3} = 2 \cdot f_{11} + 4 \\ c_{r_4} = f_{12} + 4 \\ c_{r_5} = f_{12} + 4 \\ c_{r_6} = f_{13} + 10 \end{cases} \tag{2}$$

$$\begin{cases} f_{10} = f_{r_1} \\ f_{11} = f_{r_2} + f_{r_3} \\ f_{12} = f_{r_4} + f_{r_5} \\ f_{13} = f_{r_6} \end{cases} \tag{3}$$

$$\begin{cases} c_{r_1} = c_{r_2} + g_{11} \\ c_{r_3} + g_{11} = c_{r_4} + g_{12} \\ c_{r_5} + g_{12} = c_{r_6} \end{cases} \tag{4}$$

Considering demand service constraints, a finite number of active/inactive route combinations exists. Specifically, this is equal to the number of possible state combinations (27 in total) of the three PASs composing the network, **I**nactive, **A**ctive and **O**verly-active. We represent each combination with three letters, capturing the current state of the three corresponding PASs (in topological order from left to right). So, for example, **AAA** represents the situation in which all three PASs are active and equilibrated, while for instance **AIO** is the situation in which PAS 1 is active ($c_{r_1} = c_{r_2}$), PAS 2 is inactive ($c_{r_3} < c_{r_4}$) and PAS 3 is overly active ($c_{r_5} > c_{r_6}$).

Different route set activation combinations will reflect in which route flows are included explicitly in equation (3), as well as in which equalities/inequalities will appear in equation (4). In the basic formulation, all six route flows are included and active, which corresponds to combination **AAA**. Solving the resulting system of linear equations allows us to determine analytical values for equilibrium route and link flows, from which we can then straightforwardly derive the sensitivity Jacobian of the UE flows with respect to changes in the two control variables $[g_{11}, g_{12}]$, shown in Equation (5).

$$J_f(g) = \begin{pmatrix} 1/7 & 2/7 \\ -3/7 & 1/7 \\ 1/7 & -5/7 \\ 1/7 & 2/7 \end{pmatrix} \tag{5}$$

Sufficient steering from the three controllers will though transition the system towards different states, e.g. a very high value for controller g_{12} will trigger combination **AIO**, in which only the left PAS is Active while the central one is Inactive and the right is Overly-active. The corresponding equilibrium problem (3) can be rewritten as follows, removing route flows f_{r_4}, f_{r_5}. It's immediate to see from Equation (7) that the system's sensitivity is notably different as these routes deactivate.

$$\begin{cases} f_{10} = f_{r_1} \\ f_{11} = f_{r_2} + f_{r_3} \\ f_{12} = 0 \\ f_{13} = f_{r_6} \end{cases} \tag{6}$$

$$J_f(g) = \begin{pmatrix} 0 & 1 \\ -1/2 & 1/2 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{7}$$

We believe this difference in sensitivity to be the key driver behind the unattractive behaviour of the Total Cost solution space. In general networks, considering enumerative or implicit extended route set descriptions, keeping track of these changes would prove simply unfeasible in terms of computational effort. By tracking PASs and their state, however, we can considerably reduce this burden, as PASs will anyhow arise throughout the numerical UE algorithm of choice.
The next set of tests is devised to validate our hypothesis on an analytically unfriendly case. By assigning a unique numerical ID to each new, unexplored combination of PAS states arising throughout the solution space exploration, we can easily keep track of where in the solution space these regions are, and where they border with other regions. The results shown in Figure 4 indeed further confirm our intuition: the algorithmically drawn borders do indeed fit exactly our former qualitative analysis.

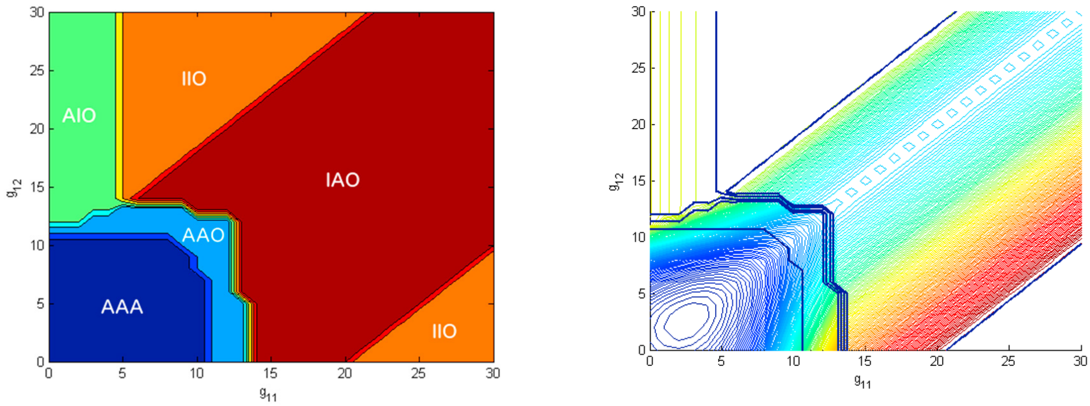


Fig. 4: Highlighted regions and borders in solution space. Each region is associated with three letters describing the state of the three PAS's composing the network (**AAA** = all active, **AAO** = **A**ctive, **A**ctive, **O**verly active, …).

We moreover assess that the nature, location and shape of these regions shows strong dependency upon the network's supply characteristics. The analytical flow-control Jacobian formulation, if considered in its parametric form (which can be found for the two former cases in Appendix A), clearly shows how its values solely depend upon the balance of the different links' a_l parameter values. Based upon the cost balances between the three intertwined PASs, different borders will therefore ensue in different locations at a macroscopic level.
In Figure 5 we show this to hold for non-linear examples, where we increased the β parameter of link 10 fivefold (the original values are presented in Table 2). While the overall shape of the objective function is considerably distorted, regions and borders still arise entirely following our intuition.

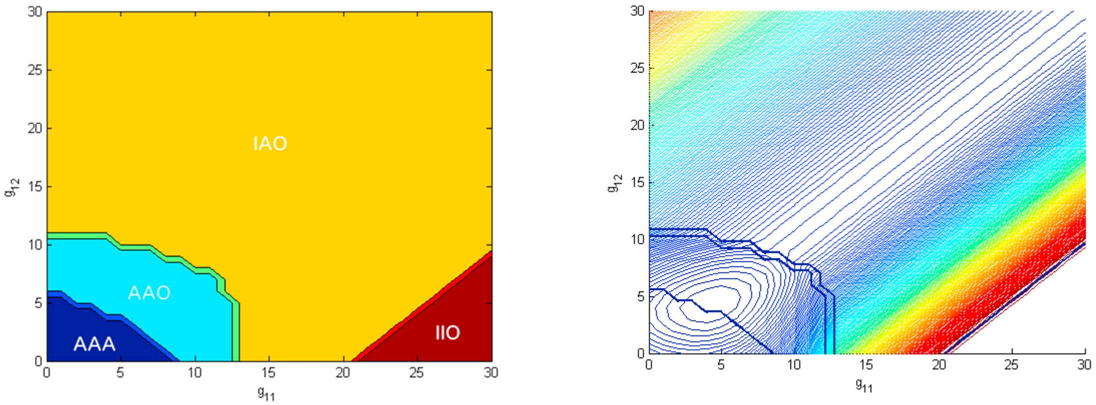


Fig. 5: Highlighted regions and borders in solution space. Each region is associated with three letters describing the state of the three PAS's composing the network (**AAA** = all active, **AAO** = **A**ctive, **A**ctive, **O**verly active, …).

While such a graphical exploration quickly becomes unfeasible with size of network and, especially, number of controllers, it is our belief that this region-based behaviour holds for general networks, under the assumption of static deterministic user equilibrium and variable, cost-dependent route set.
In the next Subsection, we exploit this added knowledge to develop ad-hoc optimization approaches that explicitly consider the existence and nature of regions, yielding modified gradient-based algorithms that – even though global optimality cannot be guaranteed – have significantly more favourable characteristics than the original versions.

3.2. PAS-constrained anticipatory traffic control optimization

As discussed earlier, tracking route activation/deactivation dynamics appears to be key in understanding where and how the landscape of the Total Cost objective function presents irregularities and sudden behavioural changes.
By feeding the optimization procedure with a-priori information on what to expect in terms of sudden changes in sensitivity, we exploit this concept in order to improve its performance and optimality. However, obtaining a-priori knowledge on all regions and all borders requires extensive solution space exploration, which of course isn't feasible. We instead introduce PAS related *constraints* to the Total Cost minimization procedure, by determining a "target" region, which would be considered optimal in terms of active route set.
We determine such a region by computing a PAS-wise description of System Optimum assignment, which, according to Wardrop's "second principle", represents the best possible configuration of the road network's supply. Learning which PASs are active and overly-active in the target solution allows to derive explicit constraints to how costs should ideally be distributed in portions of the network. Imposing these constraints will help determining globally optimal control policies, especially in instances in which not all controllers necessary to achieve System Optimum are equipped on the network.
Obtaining a System Optimum assignment solution, which includes PAS information implies introducing minor modifications to a path-based algorithm of choice, such that the link cost functions include not only the average cost component, but also the marginal cost component. Under the assumption of monotonic, separable link cost functions

the extended cost functions $c_l(f_l)$ this can be formulated as follows:

$$c_{TOT} = c_{AC} + c_{MC} = c_l(f_l) + f_l \cdot \frac{dc_l}{df_l}, \tag{8}$$

where c_{AC} is the average cost component and c_{MC} the marginal cost component.

Once this simple modification is introduced, the algorithm will equalize marginal total system costs, i.e. the flow derivative of the total cost $f_l \cdot c_l(f_l)$ incurred on link l, following 2nd Wardrop equilibrium principle. As the chosen path-based algorithm operates in PASs, the PAS set T_{SO}^0 and states pertaining to System Optimum can also be obtained. This set contains important information also for the calculation of second-best optimal solutions aiming at minimizing total time spent subject to partial control over the network flows. Indeed, compared to the states of the UE PAS set, the SO PAS set shows which PAS branches need to carry less or more flow for total cost to decrease. Even though partial controllability settings might not be sufficient to reach the SO PAS configuration, it still represents a plausible direction in which improvements of system costs can likely be found.

In order to deduce link-wise cost constraints we have to analyze the p PASs composing the set T_{SO}^0, and, depending on their states, either equality or inequality constraints will be derived. We detail this procedure through the simple example shown in Figure 8. In this network two PASs, A and B, compose the full set, where $[s_0, s'_0], [s_1, s'_1]$ are the four basic route segments composing the network. Each segment is further composed of three links: $s_0 = \{l_1, l_2, l_3\}, s'_0 = \{l_4, l_5, l_6\}, s_1 = \{l_7, l_8, l_9\}, s'_1 = \{l_{10}, l_{11}, l_{12}\}$.

As symbolized by the difference in strokes and colours, PAS A is Active and PAS B is Overly-active. Analysing link costs, we can assume that the system of linear equations (9) holds for this simple example.

$$\begin{cases} c_1 + c_2 + c_3 = c_4 + c_5 + c_6 \\ c_7 + c_8 + c_9 > c_{10} + c_{11} + c_{12} \end{cases} \tag{9}$$

Assuming that this situation represents the System Optimal configuration, from system (9) we can easily derive a set of equality and inequality constraints, expressing how the different links composing the two PASs A and B behave in terms of cost when close to SO. For this simple network, the two equality and inequality constraint sets would take the following form:

$$\begin{aligned} &A_{eq} \cdot c = 0 \\ &A_{eq} = \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned} \tag{10}$$

$$\begin{aligned} &A_{ineq} \cdot c < 0 \\ &A_{ineq} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & 1 & 1 & 1 \end{pmatrix} \end{aligned} \tag{11}$$

In general, thus, given a network N composed by $|l|$ links, $l \in L$ whose System Optimal PAS configuration T_{SO}^0 exhibits m PASs in equilibrium and v PASs in overly-active state (with $m + v = p$ the total number of PASs), the resulting constraints can be represented by the two matrices $A_{eq} \in \Re^{m \times |l|}$, $A_{ineq} \in \Re^{v \times |l|}$.

Our objective is to employ these constraints on the bi-level anticipatory traffic control problem, so to improve its convergence towards desired solutions, and furthermore to avoid issues related with insensitive regions.

Considering Total Cost as the objective function of the policymaker and introducing a control variable vector $g = [g_m] \; \forall \in M$ that represents pricing variables on a subset of links $M \subseteq L$, we derive the following constrained optimization formulation:

$$
\begin{aligned}
&\min_g \sum_{l\in L} f_l(c_l(f_l,g))\cdot c_l(f,g) \\
&s.t \begin{cases} (i)\, f \in S_{f^*} \\ (ii)\, S_{f^*} :\{f : f = \arg\min \sum_{l\in L}\int_0^{f_l} c_l(y_l)dy_l, f\in S_f\} \\ (iii)\, A_{eq}\cdot c = 0 \\ (iv)\, A_{ineq}\cdot c < 0 \end{cases}
\end{aligned}
\tag{12}
$$

Where S_{f^*} represents the feasibility set of equilibrium flows and S_f the (larger) set of demand-feasible flows. Constraints (i) and (ii) represent the lower level User Equilibrium problem, while (iii) and (iv) are the equality/inequality cost-wise constraints that the different PASs in the network should respect in order to meet the System Optimal configuration.

The two additional constraint sets derived from PAS configurations will strongly influence the feasible solution space size for the original objective function, therefore reducing the chances of optimization procedures encountering local minima. However, especially when considering simple gradient-based algorithms, as is this paper's objective, these constraints will also impose very strict conditions on which initial point g_0 can be used for optimization. Initial points which push the objective function outside of the feasible boundaries will be rejected, thus creating the necessity of a specific methodology to ensure initial point feasibility. Moreover, there is no guarantee that this feasible space is actually non-empty and hence existence of such an initial point.

In order to avoid these issues, in (13) we introduce a Lagrangian relaxation of (12), in which the additional PAS-based constraints appear in the objective function itself in form of penalty terms.

$$
\begin{aligned}
&\min_{g,\lambda} \sum_{l\in L} f_l(c_l(f_l,g))\cdot c_l(f,g) + \sum_{m\subseteq T_{SO}^0} \lambda_m \cdot |(A_{eq}[m,l]\cdot c)| + \sum_{v\subseteq T_{SO}^0} \lambda_v \cdot (A_{ineq}[v,l]\cdot c - \epsilon) \\
&s.t. \begin{cases} (i)\, f \in S_{f^*} \\ (ii)\, S_{f^*} :\{f : f = \arg\min \sum_{l\in L}\int_0^{f_l} c_l(y_l)dy_l, f\in S_f\} \\ (iii)\, \lambda_m, \lambda_v \ge 0 \end{cases}
\end{aligned}
\tag{13}
$$

where $p = m \cup v \in T_{SO}^0$ are all PASs appearing in the given network's System Optimal configuration, A_{eq}, A_{ineq} the matrices representing the corresponding equilibrium/disequilibrium conditions in the System Optimal configuration and $\lambda = [\lambda_v, \lambda_m]$ the respective Lagrangian multiplication terms. With the notation $A_{eq}[m,l]$ we represent the $m-th$ row $l-th$ column element of matrix A_{eq}.

As will be shown in the experimental results section, under the assumption of full controllability, this obviates the need to find a good or feasible initial point, and instead creates conditions under which any initial choice $[g_0, \lambda_0]$ is guaranteed to lead the system towards an optimum. Moreover, under the same assumption, this relaxation guarantees reaching SO from any initial choice $[g_0, \lambda_0]$. Finally, it has to be noted that throughout the optimization procedure no further information in terms of PASs is necessary: the constraints, either in relaxed or strict form, are solely based upon the current link cost vector c. This formulation is sufficient to fully capture the network's behaviour as long as conditions for equilibrium existence and uniqueness are met.

In the experimental results section this latter condition will be explored, showing how the choice of the initial values λ_0 for the Lagrangian multipliers under conditions in which full controllability cannot be guaranteed influence the behaviour of gradient-based optimization.

4. Experimental results

To validate our approach, we perform tests on two basic networks, specifically devised to (i) keep complexity as low as possible in order to allow full solution space exploration and visualization and (ii) exhibit all essential behavioural difficulties proper of larger network instances. Through both test cases, our main objective is to empirically validate whether explicitly including PAS constraints, specifically in the relaxed form of formulation (13), is beneficial in terms of solution space shape regularity and, thus, optimization. Test Case II aims furthermore at collecting some intuition related to the validity of the proposed approach when dealing with partial controllability instances, that is when not all PAS constraints might be met (a condition akin to second best pricing). Validation on larger networks is left to future research.
Deterministic static user equilibrium and system optimum assignment throughout this section are computed through (variations of) the DIAL B deterministic assignment procedure (Dial, 2006). MATLAB®'s Optimization Toolbox's *fmincon* is employed for all optimizations, the underlying method being gradient descent with quasi-newton BFGS Hessian approximation.

4.1. Test Case I

The first case is that presented throughout the methodology section, based upon the network of Figure 3. As mentioned earlier, links 10, 11, 12 and 13 are equipped with the standard BPR cost function: $c_l(f_l) = 1 + \alpha_l \cdot (f_l / 400)_l^{\beta}$, while all others have a constant cost c_0. Links 11 and 12 are further equipped with pricing controllers. The resulting control vector is $g = [g_{11}, g_{12}]$.
As discussed earlier, depending on the configuration of the parameters α, β, different behaviour will ensue in this network. For the first test case we choose these parameters as follows:

Table 2: Link parameters for Test Case 1

Link	Parameters
Link 10	$\alpha_{10} = 0.5, \beta_{10} = 2$
Link 11	$\alpha_{11} = 0.2, \beta_{11} = 4$
Link 12	$\alpha_{12} = 0.5, \beta_{12} = 4$
Link 13	$\alpha_{13} = 0.4, \beta_{13} = 2$

The objective of this first test case is to show how introducing PAS constraints according to (13) positively influences the objective function's regularity. Figure 6(a) shows the solution space contour plots for the Total Cost function depending on different values for the two controllers g_{11}, g_{12}. As discussed earlier, employing a gradient-based optimization algorithm on such an objective function might be detrimental, an initial point g_0 sufficiently far from the global minimum might lead optimization towards local minima.

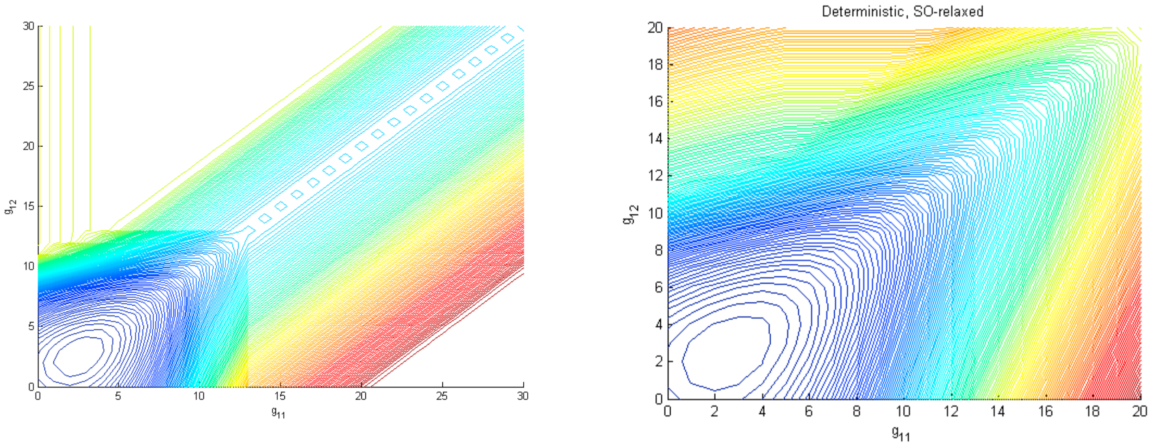


Fig. 6: Solution space for the Total Cost Objective function (a) and its SO-constrained variation (b)

In Figure 6(b) we show instead how the solution space looks like when including constraints derived from the system optimal PAS set T_{SO}^{0} . For this example, a static value for the Lagrangian relaxation terms of $\lambda_m = [0.9]$ has been used.

As can clearly be seen in Figure 6(b), the altered shape of the objective function is considerably more regular, completely convex. This is due to the three System Optimal PAS constraints, which for this rather simple network take the following form (for the sake of readability, only columns 10-13 are shown):

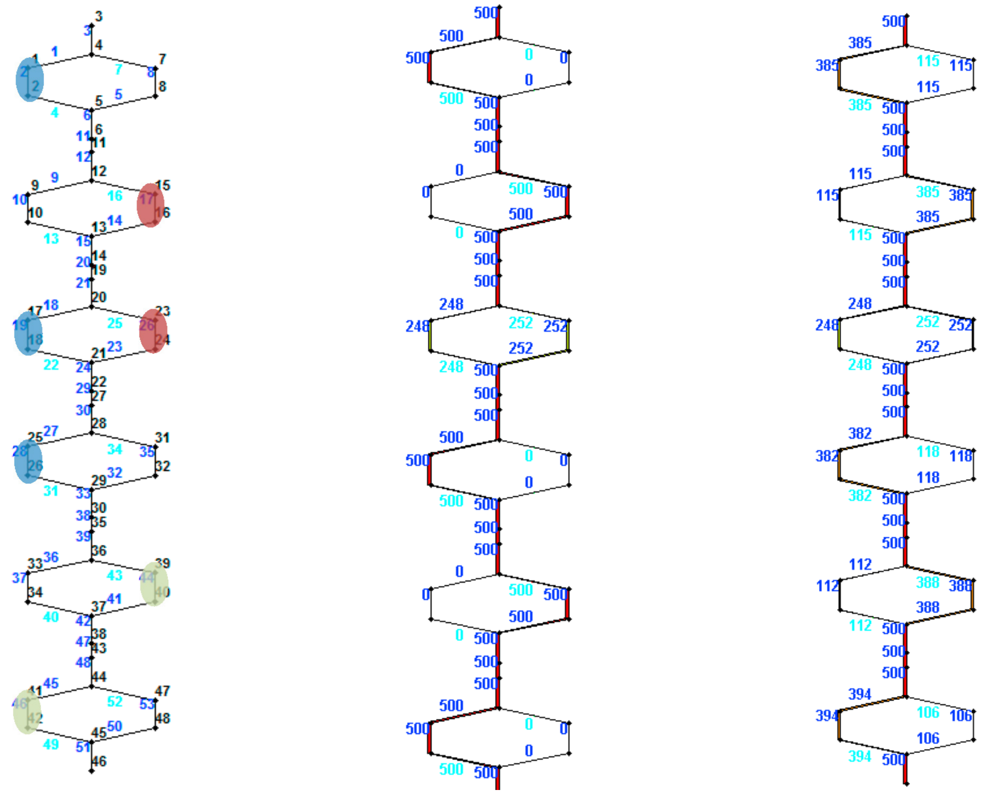
$$A_{eq} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, A_{ineq} = [\ ] \tag{14}$$

Indeed, the optimal configuration is that of equilibrating all three PASs: any control value that would influence route choice towards either selecting the left or right branch alone would be undesirable; in the original solution space of Figure 8(a) these areas are those characterized by the intervals $g_{11} \in [12,20], g_{12} = [12,20]$. By including these constraints as penalty terms in the objective function, points of the solution space that would originally sit in local minima, such as $g = [6,16], g = [16,16]$, are instead penalized due to their relative distance in cost configuration compared to SO conditions. In terms of gradient-based optimization this is highly beneficial, as a global region of attraction towards portions of the solution space exhibiting SO behaviour is introduced by formulation (7).

4.2. Test Case II

In this second example we employ once again a rather simple network, but introduce interdependencies among control variables. Rather than influencing one link at a time, in this example different pricing controllers influence different links at the same time. This allows us to showcase a very important property that networks should exhibit for our formulation (13) to be as successful as possible: full controllability. This network is shown in Figure 7(a).

As indicated by the different coloring scheme, three pricing controllers (red, blue and green) act upon different links on this network. Specifically, controller g_1 imposes pricing on links 2, 19 and 28, controller g_2 on links 17 and 26 and, finally, controller g_3 on links 44 and 46. All links are equipped with BPR cost functions, with a capacity of 400veh/h each, and the different alpha/beta parameters have been carefully tuned such that the User Equilibrium and System Optimum flow configurations would result as shown in Figure 7 (b,c).



(a) (b) (c)

Fig. 7: Network (a), User Equilibrium (b) and System Optimal (c) flow distributions for the Serial PAS network

This specific design has been chosen such that no less than the full action of all three controllers simultaneously would be enough to steer the 6 PASs towards the fully equilibrated nature of the SO solution.

Several tests have been performed on this network, the first set of which shows how different initial points with and without the extra constraints of (13) lead optimization towards entirely different solutions. These results are summarized in Table 3. All tests in this section are performed under deterministic user equilibrium conditions, using the DIAL B algorithm.

Table 3: Optimization test results, full controllability

UE Total Cost	455411.9 veh*h			
SO Total Cost	431117.1 veh*h			
Initial Point	$x_0 = [0.01, 0.01, 0.01]$	$x_0 = [15, 0.01, 10]$	$x_0 = [0.01, 10, 15]$	$x_0 = [30, 30, 30]$
Total Cost	455411.9 veh*h	440276.7 veh*h	455411.9 veh*h	431117.1 veh*h
Total Cost SO-relax, $\lambda_0 = 0.1$	431117.1 veh*h	431117.1 veh*h	431117.1 veh*h	431117.1 veh*h

These results show indeed how our proposed formulation allows descending towards System Optimum from all four randomly chosen initial points, while the standard objective function suffers from considerable locality issues.

As anticipated earlier, in this network all three controllers are necessary to guarantee full optimality. To showcase the impact that this has on the objective function's regularity, and indeed on our relaxed formulation, we have devised one last series of tests in which only a subset of all controllers can be optimized, while the others are fixed to a value of 0. Optimization results for these tests are summarized in Table 4.

Table 4: Optimization test results, partial controllability, $x_0 = [0.01, 0.01, 0.01]$

UE Total Cost	455411.9 veh*h					
SO Total Cost	431117.1 veh*h					
Controller set	1+2	1+3	2+3	1	2	3
Total Cost	455411.9 veh*h	446525.3 veh*h	446525.3 veh*h	455411.9 veh*h	455411.9 veh*h	455411.9 veh*h
Total Cost SO-relax, $\lambda_0 = 0.1$	446276.7 veh*h	446252.3 veh*h	446252.3 veh*h	455411.9 veh*h	455411.9 veh*h	446252.3 veh*h

Thanks to the reduction in dimensionality, we can moreover show the shape of the objective function for these pairs / single controllers. In Figure 8 the solution space for controllers $[g_1, g_2]$ is shown for both the original and SO-relaxed formulation. Figure 9 reports the same results for the controller couple $[g_1, g_3]$.

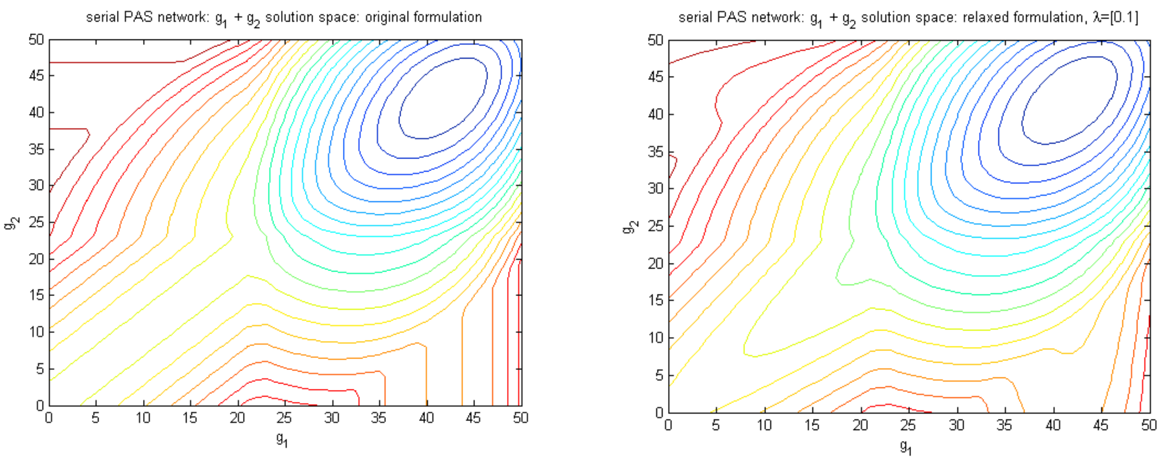


Fig. 8: Total Cost solution space contour plots for the Serial PAS network. $g = [g_1, g_2]$

These figures show how strongly non convex this specific network's objective function is, and how different it is for different controller couples. In the case of Figure 9, even though formulation (7) does introduce some regularity, the local minimum located in $g = [42, 45]$ still has its own clear attraction domain.

Interestingly, the global minimum located in $g = [0, 45]$ does in fact violate SO constraints. This becomes evident if a larger Lagrangian multiplier value is chosen, In Figure 10 we can clearly see how our relaxation would point towards the local minimum rather than the global minimum if a value $\lambda = [0.9]$ were chosen.

This implies that fixing an arbitrary value for λ vector would be problematic for our algorithm in the search for optimality. As the Lagrangian multipliers are though part of the optimization process itself, this eventuality is luckily remote.

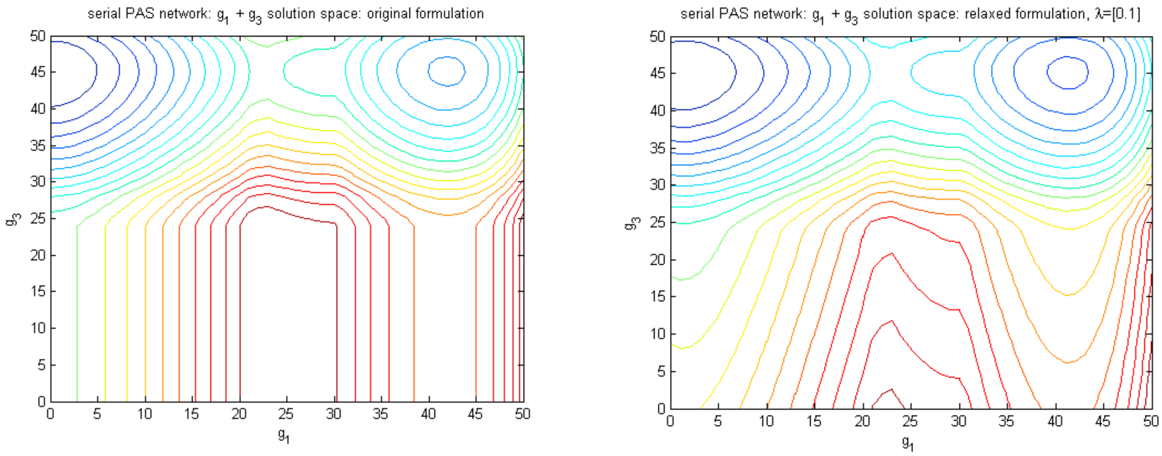


Fig. 9: Total Cost solution space contour plots for the Serial PAS network. $g = [g_1, g_3]$

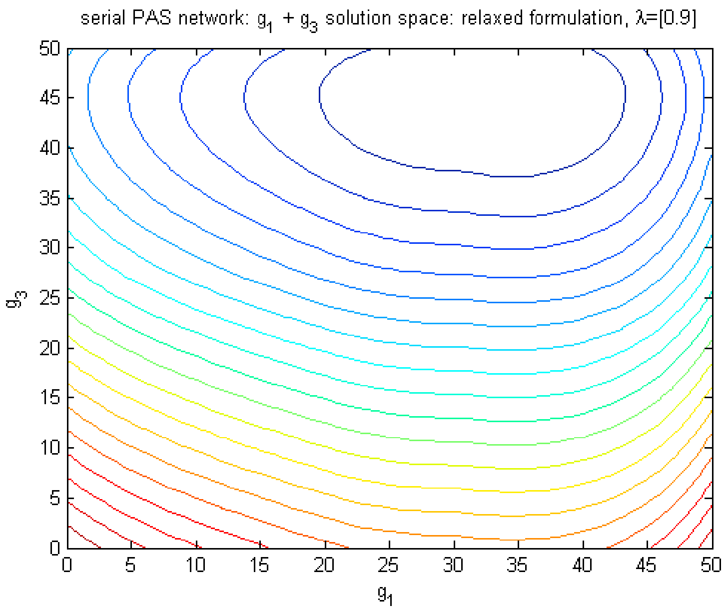


Fig. 10: Total Cost solution space contour plots for the Serial PAS network. $g = [g_1, g_3]$, $\lambda = [0.9]$

In Figure 11 we show the results of our final tests, in which we separately optimize Total Cost when considering only either g_1 or g_3. We superimpose on the mono-dimensional objective function curve the descent steps taken by the optimization algorithm, and indeed show how even when employing an initial value $\lambda_0 = [0.9]$ the optimization process immediately reduces the values of the Lagrangian multipliers, collapsing towards the real objective function, when keeping these high would instead create detrimental results.

On the left hand side of Figure 11, it can clearly be seen that even though the relaxed formulation's initial point is rather higher above the real objective function, due to the additional cost constraints, the optimization procedure considers reducing the Lagrangian multipliers λ a foremost priority, thus leading the optimization towards the globally optimal, although SO constraint violating, solution in $g_1^* = 0$.

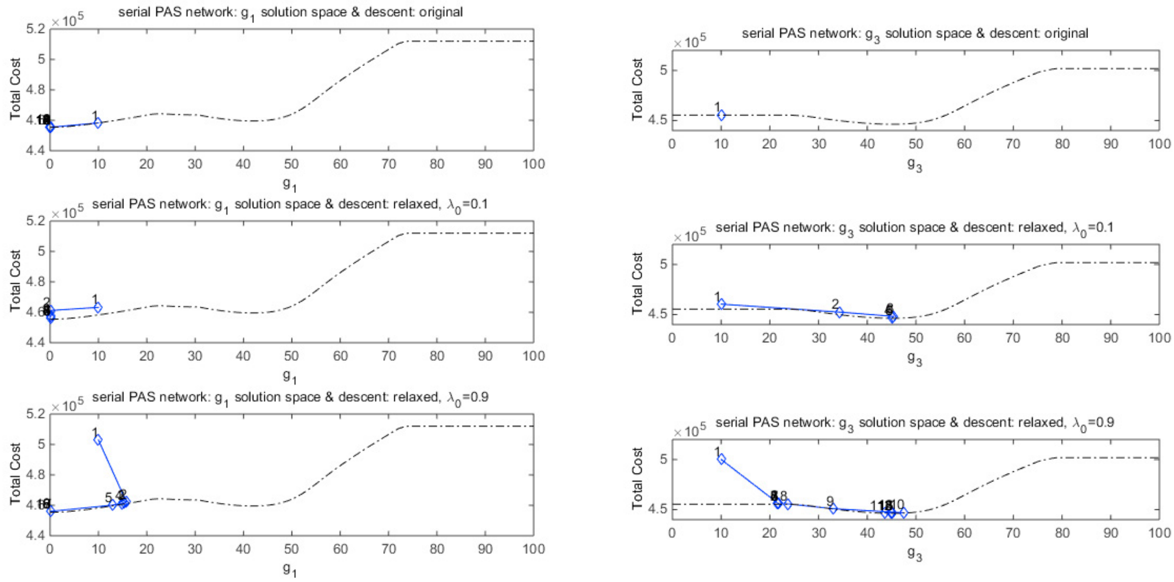


Fig. 11: Solution spaces and descent steps for single-controller instances of the Serial PAS network

Conversely, analysing the right hand side of Figure 14 one can clearly see how without imposing the additional constraints the optimization would stay put in the flat area around the initial point $g_0 = 10$, rather than converge towards the real minimum value $g_3^* = 45$.
Finally, we would like to remark that given a sufficiently unlucky combination of bad initial point x_0 and initial Lagrangian multiplier vector λ_0 our formulation can indeed be led towards the local minimum located in $g_1 = 42$. Statistically this latter is less probable than its opposite, which suggests that employing our relaxation scheme as part of a multi-initial point metaheuristic procedure such as Monte Carlo or Particle Swarm Optimization might prove very useful.

5. Conclusions and future research

In this paper we contributed to the problem of determining optimal pricing control subject to User Equilibrium constraints. Specifically, we investigated the behavioral cause of non-smoothness and non-convexity in the Total Cost objective function and introduced an extension to standard gradient-based descend methods aimed at improving convergence of gradient-based algorithms, especially when dealing with flat, insensitive regions in the solution space.

Our results show that different portions of the Total Cost solution space's landscape are connected to the underlying route set activation/deactivation dynamics. As different routes enter/exit the equilibrium route set, the behavior, shape and locality features of the solution space are severely impacted, effectively giving rise to different "regions".
Based upon these findings, a constrained formulation of the optimal anticipatory traffic control is introduced, in which a "target" region, specifically that pertaining to System Optimum, acts as the ultimate objective.
Our resulting optimization formulations strongly benefit from the additional constraints, and – even though global optimality can still not be guaranteed - yield better tractable problems which converge to significant minima, avoiding instead originally flat regions. Under the assumption of full controllability, the relaxed approach moreover obviates the need to determine a valid initial point from which to begin gradient descent, as the target region constraints act as a global attractor.
Future research includes foremost extending these results to larger, more realistic networks. As the size of networks

and, especially, controllers increase the ability to graphically assess whether or not the proposed framework is successful is lost, demanding extensive empirical testing. Nevertheless, preliminary results show that the findings related to existence and behavior of objective function regions applies even for larger networks, although the quantity and variability of these is considerably greater. Another interesting direction warranting pursuit is the determination of optimal Lagrangian multipliers: due to the high number of regions and constraints in larger networks, explicit optimization of these variables is not a viable practice.

Finally, we will also focus on developing qualitative ways of determining the "target" route set configuration, as algorithmic computation of the System Optimum solution cannot be generalized to instances where link cost functions are no longer separable, as would be the case when considering explicit congestion propagation dynamics.

Appendix A.

For region **AAA**, the network's flow-control Jacobian depends upon the link parameters as follows:

$$J_f(g) = \frac{1}{D}\cdot\begin{pmatrix} a_{12}a_{13} & a_{11}a_{13} \\ -a_{12}a_{13} + a_{10}(a_{12}+a_{13}) & a_{10}a_{13} \\ a_{10}a_{13} & a_{11}a_{13} + a_{10}(a_{11}+a_{13}) \\ a_{10}a_{12} & a_{10}a_{11} \end{pmatrix} \tag{15}$$

$$D = a_{10}a_{11}a_{12} + a_{10}a_{11}a_{13} + a_{10}a_{12}a_{13} + a_{11}a_{12}a_{13}$$

For region **AIO**, the resulting flow-control Jacobian is instead as follows:

$$J_f(g) = \begin{pmatrix} 0 & \frac{1}{a_{10}} \\ -\frac{1}{a_{11}} & \frac{1}{a_{11}} \\ 0 & 0 \\ 0 & \frac{1}{a_{13}} \end{pmatrix} \tag{16}$$

References

Abdulaal, M., LeBlanc, L.J., 1979. Continuous equilibrium network design models. Transp. Res. Part B Methodol. 13, 19–32. doi:10.1016/0191-2615(79)90004-3

Allsop, R.E., 1974. SOME POSSIBILITIES FOR USING TRAFFIC CONTROL TO INFLUENCE TRIP DISTRIBUTION AND ROUTE CHOICE. Presented at the Transportation and Traffic Theory, Proceedings.

Bar-Gera, H., 2010. Traffic assignment by paired alternative segments. Transp. Res. Part B Methodol. 44, 1022–1046. doi:10.1016/j.trb.2009.11.004

Bar-Gera, H., 2006. Primal Method for Determining the Most Likely Route Flows in Large Road Networks. Transp. Sci. 40, 269–286. doi:10.1287/trsc.1050.0142

Bar-Gera, H., 2002. Origin-Based Algorithm for the Traffic Assignment Problem. Transp. Sci. 36, 398–417. doi:10.1287/trsc.36.4.398.549

Beckmann, M., McGuire, C.B., Winsten, C.B., 1956. STUDIES IN THE ECONOMICS OF TRANSPORTATION.

Cantarella, G.E., Improta, G., Sforza, A., 1991. Iterative procedure for equilibrium network traffic signal setting. Transp. Res. Part Gen. 25, 241–249. doi:10.1016/0191-2607(91)90141-C

Dial, R.B., 2006. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. Transp. Res. Part B Methodol. 40, 917–936. doi:10.1016/j.trb.2006.02.008

Friesz, T.L., Tobin, R.L., Cho, H.-J., Mehta, N.J., 1990. Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. Math. Program. 48, 265–284. doi:10.1007/BF01582259

Gentile, G., Meschini, L., Papola, N., 2004. Fast heuristics for continuous dynamic shortest paths and all-or-nothing assignment, in: AIRO 2004.

Johnson, D.S., Lenstra, J.K., Kan, A.H.G.R., 1978. The complexity of the network design problem. Networks 8, 279–285. doi:10.1002/net.3230080402

Nie, Y. (Marco), 2010. A class of bush-based algorithms for the traffic assignment problem. Transp. Res. Part B Methodol. 44, 73–89. doi:10.1016/j.trb.2009.06.005

Patriksson, M., 2004. Sensitivity Analysis of Traffic Equilibria. Transp. Sci. 38, 258–281. doi:10.1287/trsc.1030.0043

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. J. Choice Model. 2, 65–100. doi:10.1016/S1755-5345(13)70005-8

Rinaldi, M., Himpe, W., Tampère, C., 2016. A sensitivity based approach for adaptive decomposition of anticipatory network traffic control. Transp. Res. C Emerg. Technol. doi:10.1016/j.trc.2016.01.005

Rinaldi, M., Tampère, C.M.J., 2015. An extended coordinate descent method for distributed anticipatory network traffic control. Transp. Res. Part B Methodol. 80, 107–131. doi:10.1016/j.trb.2015.06.017

Smith, M.J., Ghali, M., 1990. The dynamics of traffic assignment and traffic control: A theoretical study. Transp. Res. Part B Methodol. 24, 409–422. doi:10.1016/0191-2615(90)90036-X

Taale, H., Hoogendoorn, S., 2012. Network-wide Traffic Management with Integrated Anticipatory Control. Presented at the Transportation Research Board 91st Annual Meeting.

Tobin, R.L., Friesz, T.L., 1988. Sensitivity Analysis for Equilibrium Network Flow. Transp. Sci. 22, 242–250. doi:10.1287/trsc.22.4.242

Ukkusuri, S., Doan, K., Aziz, H.M.A., 2013. A Bi-level Formulation for the Combined Dynamic Equilibrium based Traffic Signal Control. Procedia - Soc. Behav. Sci. 80, 729–752. doi:10.1016/j.sbspro.2013.05.039

Wardrop, J.G., 1952. ROAD PAPER. SOME THEORETICAL ASPECTS OF ROAD TRAFFIC RESEARCH. ICE Proc. Eng. Div. 1, 325–362. doi:10.1680/ipeds.1952.11259

Yan, H., Lam, W.H., 1996. Optimal road tolls under conditions of queueing and congestion. Transp. Res. Part Policy Pract. 30, 319–332.

Yang, H., Bell, M.G., 1997. Traffic restraint, road pricing and network equilibrium. Transp. Res. Part B Methodol. 31, 303–314.

Yang, H., Bell, M.G.H., 2007. Sensitivity Analysis of Network Traffic Equilibrium Revisited: The Corrected Approach. Presented at the 4th IMA International Conference on Mathematics in Transport.

Yang, H., Yagar, S., 1995. Traffic assignment and signal control in saturated road networks. Transp. Res. Part Policy Pract. 29, 125–139.

Ying, J.Q., Miyagi, T., 2001. Sensitivity Analysis for Stochastic User Equilibrium Network Flows—A Dual Approach. Transp. Sci. 35, 124–133. doi:10.1287/trsc.35.2.124.10137

Zhang, X., Yang, H., 2004. The optimal cordon-based network congestion pricing problem. Transp. Res. Part B Methodol. 38, 517–537. doi:10.1016/j.trb.2003.08.001