Optimal control for multi-lane motorways in presence of vehicle automation and communication systems

Claudio Roncoli∗ Markos Papageorgiou∗ Ioannis Papanichail∗

∗ Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania 73100 Greece (e-mail: <croncoli, markos, ipapa>@dssl.tuc.gr).

Abstract: The presence and exploitation of Vehicle Automation and Communication Systems (VACS) while defining optimal control strategies in motorway traffic flow control is addressed in this paper. VACS are supposed to act both as sensors (providing information on traffic conditions) and as actuators, allowing the application of ramp metering, variable speed limit control, and lane changing control. A quadratic programming problem is defined on the basis of a novel first-order traffic flow model for multi-lane motorways. An example is presented in order to illustrate the effectiveness of the proposed optimisation problem.

Keywords: Traffic control, Motorway traffic, Optimal control, Model-based control, Quadratic programming.

1. INTRODUCTION

Traffic congestion is a major issue in modern motorway systems, causing serious infrastructure degradation in and around metropolitan areas. The European Commission estimates that the yearly cost of road traffic congestion in Europe exceeds 120 billion €. Despite the significant advances in the area of Information Technology, it seems that full exploitation of innovative technologies to mitigate motorway traffic congestion is not yet achieved. On the other hand, there has been an enormous interdisciplinary effort, in recent years, by the automotive industry as well as by numerous research institutions around the world to plan, develop, test and start deploying a variety of Vehicle Automation and Communication Systems (VACS) that are expected to revolutionise the features and capabilities of individual vehicles within the next decades. Several potential future scenarios are envisaged, including self-driving vehicles within a completely connected road-vehicle infrastructure.

The notable work by Rao and Varaiya (1994) introduced the concept of highly automated Intelligent Vehicle Highway System (SmartIVHS), describing a possible hierarchical control strategy with the purpose of increasing highway capacity and safety. The authors defined simple policies for prescribing and regulating lane changings and desired speeds in an interconnected control system. Some studies exploited the concept of Automated Highway System (AHS), defining a set of layers and developing control strategies for each one of them. In this context, an interesting work was presented by Li et al. (1997), in which the authors analysed the link layer control problem and proposed a control law for the stabilisation of traffic conditions. It must be highlighted that the concept of platooning (i.e. the organisation of vehicles into closely spaced groups) is often considered as a good approach, capable of increasing the motorway capacity and reducing instability. Another interesting research work was conducted by Hall and Lotspeich (1996); they defined a model, based on linear programming, for assigning traffic to lanes. The AHS was modelled as a static trip-based multi-commodity network, in which the objective was to maximise the total flow subject to predetermined O/D patterns; this model was extended to a dynamic case by Hall and Caliskan (1999). The problem of lane assignment has been also considered in some other research works. Among them, Ramaswamy et al. (1997) described several strategies and formulated two static optimisation problems aiming at the minimisation of the total travel time; the difference between them lies in whether the cost for manoeuvring depends on highway congestion (generating a non-linear problem) or not (generating a linear problem); it was shown that the former problem is definitely more appropriate in case of heavily congested networks.

In any case, a massive use of VACS could pave the way to the application of innovative methodologies for the reduction of traffic congestion. Nevertheless, there should be a considerable effort in studying the complex interrelations between equipped vehicles and traffic conditions. As a matter of fact, none of the described approaches took into account the various aspects characterising traffic dynamics and, hopefully, this work will help to have a first understanding of such a complex problem.

The main contribution of this paper is the definition of an optimal control problem for the flow assignment assuming that many vehicles are equipped with VACS and are able to exchange information with the road infrastructure. Strong emphasis is given to the applicability of the proposed approach in a real motorway, thus, for the formu-
lation of the mathematical problem, some simplifications have been made in order to guarantee the possibility of obtaining a solution in reasonable computation time (e.g. all the model constraints are linear). Section 2 introduces a novel macroscopic traffic flow model, taking into account various original aspects such as lane changings and the capacity drop phenomenon. Based on this model, in Section 3, a quadratic programming formulation is formulated aiming at the minimisation of traffic congestion; the possible control actions are described, and some aspects, related to the quality of the solution, are considered. In Section 4, a motorway stretch is utilised as a test-bed in order to illustrate and try to understand the control actions obtained by solving the optimisation problem. Finally, Section 5 concludes the paper highlighting the main results and introducing some research challenges for the future.

2. A FIRST-ORDER MODEL FOR MULTIPLE-LANE MOTORWAYS

Differently from most of the studied motorway traffic models, the approach presented in this paper considers each lane independently; this is necessary since one of the aims of this model is to capture and define the lane changing behaviour of vehicles. Only few related studies have been carried out in the past and most of them are based on the work by Gazis et al. (1962), in which lane densities on a multi-lane highway were assumed to oscillate around an equilibrium density; this concept was extended by Michalopoulos et al. (1984), who proposed three models for capturing the lane changing behaviour (a single continuum model, a second-order model, and a two-dimensional formulation) and presented some experimental comparisons. In a more recent work, Laval and Daganzo (2006) extended the kinematic wave (KW) theory and proposed a multi-lane KW-based model as a first module of a more complex model; this latter work has some similarities with the approach presented in this section.

A multiple-lane motorway is considered. The index $j = 1, \ldots, J$ is introduced for lanes while $i = 1, \ldots, I$ for segments. Defining a discrete time step $T$, the simulation horizon $K$ is indexed by $k = 1, \ldots, K$ and the simulation time $t$ is $t = kT$. A segment-lane entity is characterised by the following state variables:

- the density $\rho_{i,j}(k)$ [veh/km], i.e. the number of vehicles in the segment $i$, lane $j$ at time step $k$, divided by the segment length $L_i$; and
- the on-ramp queue $w_{i,j}(k)$ [veh], i.e. the number of vehicles queuing at the on-ramp that enters into the motorway at segment $i$, lane $j$.

The control variables are the flows that can enter and exit the segment-lane:

- the longitudinal flow $q_{i,j}(k)$ [veh/h] is the traffic volume leaving segment $i$ and entering segment $i + 1$ during time interval $(k, k + 1]$ (remaining in lane $j$);
- the lateral flow $f_{i,j,j}(k)$ [veh/h] ($j = j \pm 1$) is the traffic volume moving from lane $j$ to lane $j'$ differentially from the model by Laval and Daganzo (2006), it is supposed that vehicles changing lane remain in the same segment during the current time interval; and
- the on-ramp flow $r_{i,j}(k)$ [veh/h].

Assuming that the on-ramp demand $D_{i,j}(k)$ [veh/h] is known, the dynamics equation at on-ramps is:

$$w_{i,j}(k + 1) = w_{i,j}(k) + T[D_{i,j}(k) - r_{i,j}(k)].$$

The next modelling issue to address is the longitudinal flow in absence of control, which will be used as an upper limit for the controlled longitudinal flow. The starting basis for this is the well-known Cell Transmission Model (CTM) (Daganzo, 1994 and Daganzo, 1995) which, however, does not take into account is the capacity drop phenomenon, i.e. the reduction of discharge flow once a congestion is formed. The reasons for this phenomenon are not exactly known, however it seems to be caused by the limited acceleration of vehicles while exiting a congested area. In second-order models, such as METANET (Papageorgiou and Messmer, 1990), the capacity drop is generated by the equations describing the spatiotemporal evolution of speed. This option is not available for first-order LWR models, therefore several attempts have been made in order to overcome this shortcoming. In Hall and Hall (1990) an inverse-lambda fundamental diagram was utilised, whereas Leclercq et al. (2011) proposed an analytical model that extends the classical CTM by incorporating endogenously the capacity drop. In Lebacque (2003), the problem is addressed by imposing an upper bound to the acceleration depending on the traffic phase, distinguishing between LWR and maximum acceleration, whereas in Laval and Daganzo (2006) discrete particles are introduced and treated as moving temporary blockages. A more recent interesting approach was proposed by Srivastava and Geroliminis (2013) who extended the classical LWR model by defining a fundamental diagram with two values of capacity and providing a memory-based methodology to choose the appropriate value in the numerical solution of the problem.

A common aspect of all the aforementioned approaches is the introduction of non-linear terms in their formulation.

Fig. 1. The segment-lane variables used in the model formulation.

The off-ramp flow is determined according to given turning rates $\gamma_{i,j}$ as a percentage of the total flow passing through the segment $q_i(k) = \sum_{j=1}^J q_{i,j}(k)$. An illustration of the segment-lane variables is presented in Fig. 1.
3. A QUADRATIC PROGRAMMING PROBLEM FOR THE OPTIMAL FLOW ASSIGNMENT

As previously mentioned, it is assumed that the control variables are the on-ramps flows, the longitudinal flows, and the lateral flows. The following three control actions are therefore taken into account:

- Ramp Metering (RM) is currently applied on many motorways (see e.g. Papamichail et al., 2010) and does not necessarily require any particular equipment to be performed.
- Mainstream Traffic Flow Control (MTFC) via Variable Speed Limits: it is assumed that the exiting flows (and consequently the speeds) are computed for each segment-lane; thus all equipped vehicles travelling on a segment-lane will receive and apply the respective speed as a speed limit. For a sufficient penetration of equipped vehicles, this will be sufficient to impose the speed limit to non-equipped vehicles as well.
- Lane-Changing Control: the optimal lateral flows are computed for each segment-lane, but the implementation of this control action is more cumbersome and uncertain than the previous two, unless all vehicles are under full guidance by the control center; in this latter case, it is not difficult to implement the control action by sending lane-changing orders to an appropriate number of vehicles. In all other cases, an intermediate algorithm should decide on the number and ID of equipped vehicles that should receive a lane-changing advice, taking into account the compliance rate and the spontaneous lane-changings; the latter may be reduced by involving additional “keep-lane” advice to other vehicles. These issues are currently in course of investigation and development.

Since the main objective is the reduction of traffic congestion, these control actions determine the decision variables of the optimisation problem formulated below. In order to make the problem solvable for large networks, it is formalised as a Quadratic Program (QP), characterised by a convex quadratic cost function and linear constraints. The latter are derived directly from the traffic model described in Section 2. In addition, two important aspects should be taken into account. When applying ramp-metering actions, the size of each queue \( w_{i,j}(k) \) is upper-bounded by a predefined value \( w_{i,j}^{\text{max}} \); since the motorways segment-lanes have also finite storage capacities, a sufficiently high external demand could cause infeasibility of the optimisation problem due to lack of any possibility to accommodate the external demand and still comply with all constraints. To avoid this occurrence, two extra variables are introduced: \( W_{i,j}(k) \) [veh] that represents an extra-queue state variable and \( d_{i,j}(k) \) [veh/h] that represents the demand flow that is capable to enter the constrained queues. The dynamics at on-ramps defined in (2) are thus modified as follows:

\[
\begin{align*}
\dot{w}_{i,j}(k+1) &= w_{i,j}(k) + T [d_{i,j}(k) - r_{i,j}(k)] \\
W_{i,j}(k+1) &= W_{i,j}(k) + T [D_{i,j}(k) - d_{i,j}(k)]
\end{align*}
\]

If a strong penalty factor is associated to variables \( W_{i,j}(k) \), then \( W_{i,j}(k) \equiv 0 \), i.e. \( d_{i,j}(k) \equiv D_{i,j}(k) \) will be forced in the solution, except in case the demand flow is too high, where extra queues may indeed be generated to avoid infeasibility.

The computation of lateral flows is completely delegated to the optimiser; only physical upper bounds and a cost function term are specified.

\[
\begin{align*}
[f_{i,j,j-1}(k) + f_{i,j,j+1}(k)] &\leq \frac{L_i}{T} \rho_{i,j}(k) \\
[f_{i,j-1,j}(k) + f_{i,j+1,j}(k)] &\leq \frac{L_i}{T} [\rho_{i,j}^{\text{jam}} - \rho_{i,j}(k)] \\
f_{i,j,j-1}(k) &\leq f_{i,j,j+1}(k) \leq f_{i,j,j+1}^{\text{max}}
\end{align*}
\]

Equation (5) represents the upper bound for lateral flow determined by the number of vehicles in the current lane-segment; (6) is an upper bound considering the available space for lateral flow entering a segment-lane; and (7) is a hard constraint considered in order to strictly limit lateral movements.

The density is computed according to (1), whereas the following constraints are added for longitudinal flows:

\[
\begin{align*}
q_{i,j}(k) &\leq v_{i,j}^{\text{max}} \rho_{i,j}(k) \\
q_{i,j}(k) &\leq v_{i,j}^{\text{max}} \frac{\rho_{i,j}^{\text{per}} - \rho_{i,j}^{\text{jam}}}{\rho_{i,j}^{\text{per}} - \rho_{i,j}^{\text{jam}}} \rho_{i,j}(k) + \frac{\rho_{i,j}^{\text{per}} - \rho_{i,j}^{\text{jam}}}{\rho_{i,j}^{\text{per}} - \rho_{i,j}^{\text{jam}}} q_{i,j}^{\text{jam}}
\end{align*}
\]
\[ q_{i,j}(k) \leq v_{\text{max}}(i+1,j) \rho_{i+1,j}^{cr} \]
\[ q_{i,j}(k) \leq -v_{\text{max}}(i+1,j) \rho_{i+1,j}^{cr} \rho_{i+1,j} \]

Equations (8) and (9) represent the demand part of the FD, whereas (10) and (11) are the bounds describing the supply part of the FD, reflecting the FD shown in Fig. 2.

In addition, upper-bounds are considered for queues \( w_{i,j}(k) \leq w_{\text{max}}^{\text{cr}} \), entering flows at on-ramps \( (r_{i,j}(k) \leq r_{\text{max}}^{\text{cr}}) \), and exiting flow at off-ramps \( (\gamma_{i,j} \dot{q}_{i,j}(k) \leq q_{i,j}^{\text{cr, max}}) \). Non-negativity constraints are also specified for all the variables.

The cost criterion is defined by the following equations:
\[ \min_{\rho, w, W} Z \]
\[ q, r, f \]
\[ Z = T \sum_{i,j=1}^{I} \sum_{k=1}^{K} [L_{i,j}(k) + w_{i,j}(k)] \]
+ \[ M \sum_{i,j=1}^{I} \sum_{k=1}^{K} W_{i,j}(k) \]
+ \[ \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left[ \beta_{i,j,j-1} f_{i,j,j-1}(k) + \beta_{i,j,j+1} f_{i,j,j+1}(k) \right] \]
+ \[ \lambda^{f} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left[ r_{i,j}(k) - r_{i,j}(k-1) \right]^{2} \]
+ \[ \lambda^{f} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left[ P_{i,j,j-1}(k) + \sum_{j=1}^{J} P_{i,j,j+1}(k) \right] \]
+ \[ \lambda^{xf} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{j=2}^{J} P_{i,j,j}(k) \]
+ \[ \lambda^{xf} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{j=2}^{J} P_{i,j,j}(k) \]

and:
\[ P_{i,j,j-1}(k) = \left[ f_{i,j,j-1}(k) - f_{i,j,j-1}(k-1) \right]^{2} \]
\[ P_{i,j,j+1}(k) = \left[ f_{i,j,j+1}(k) - f_{i,j,j+1}(k-1) \right]^{2} \]
\[ P_{i,j,j}(k) = \left[ q_{i,j}(k) - q_{i,j}(k-1) + v_{\text{max}}^{\text{cr}} \left[ \rho_{i,j}(k-1) + \rho_{i,j}(k) \right] \right]^{2} \]
\[ \left( \rho_{i,j}^{r} \right)^{2} \]
\[ P_{i,j,j}(k) = \left[ q_{i,j}(k) - q_{i-1,j}(k) + v_{\text{max}}^{\text{cr}} \left[ \rho_{i-1,j}(k) + \rho_{i,j}(k) \right] \right]^{2} \]
\[ \left( \rho_{i,j}^{r} \right)^{2} \]

The cost function in (13) is composed of seven different terms, the first three linear and the last four quadratic:

- The first linear term represents the Total Time Spent (TTS); it considers both the time travelled and the time spent queuing at on-ramps; this is the most important factor that indicates the goodness of the solution.
- The second linear term (weighted by \( M \)) is the penalty for the extra-queues generated in the optimal solution. Its penalty weighing factor must be tuned in order to determine that all variables \( W_{i,j}(k) \) are as close as possible to 0 whenever possible.
- The third linear term (weighted by \( \beta_{i,j,j} \)) aims at penalising lateral flows; it has the purpose of reducing the lane changings (thus giving priority to the other control actions).
- The quadratic terms are penalty terms on the variation of control variables from a time step to the next one or between two neighbouring segments. These terms are introduced in order to reduce the space and time fluctuation of the control variables, helping to homogenise their actions and contributing to improve the overall stability of the system. The first one (weighted by \( \lambda^{f} \)) is related to on-ramp flows; the second one (weighted by \( \lambda^{f} \)) penalises the time variation of lateral flows; and the last two (weighted by \( \lambda^{xf} \) and \( \lambda^{xf} \)) have the purpose of penalising the time and space variations of speed values. Since the speed could be defined only by the non-linear relation \( v = q/p \), the proposed approach considers a linear approximation around the values \( p = p^{cr} \) and \( v = v^{max} \); then \( P_{i,j}(k) \) represents the penalisation of speed from one step to the next one, as described in (16), whereas \( P_{i,j}(k) \) is related to the penalisation from one segment to the neighbouring one, as stated in (17).

4. APPLICATION EXAMPLE

4.1 Network description

The motorway stretch, used with the purpose of evaluating and illustrating the concepts described in this paper, is an adapted version of the road section of the motorway A20 from Rotterdam to Gouda, the Netherlands, that was utilised by Schakel and van Arem (2013). This network is particularly interesting because of its infrastructural features (lane-drops, on-ramps and off-ramps) and the potential congestion patterns that could be generated (congestion at the lane-drops, spillback from off-ramps, heavy entering flows at on-ramps). The network, about 12 km in length, is subdivided into 27 segments of 420m–490m in length, as shown in Fig. 3. The time step is set to \( T = 15 \) s; considering a simulation horizon of 1 hour, \( K = 240 \) results. The lanes are numbered 1, . . . , 4 from the inner lane (close to the roadside) to the outer lane (close to the road median); in the remaining paper, lane-drops, on-ramps and off-ramps will be identified by the segment and lane they are related to.

The utilised demand profile is shown in Fig. 4; the reduced entering flows during the last 20 min represent a cool down period that will ensure the disappearing of any congestion at the end of the simulation. In addition, the network is supposed to be empty at \( k = 0 \) (both the densities and ramp queues are initialised...
All links have the same characteristic values: the critical density is set to \( \rho_{i,j}^{cr} = 22 \) veh/km, the jam density is set to \( \rho_{i,j}^{jam} = 180 \) veh/km, the maximum speed is \( v_{i,j}^{max} = 100 \) km/h, and the maximum flow at jam density \( q_{i,j}^{jam} = 1467.4 \) veh/h (this is obtained by setting a slope \( w = w/3 \)). The exit rates at off-ramps are set as follows: \( \gamma_{2,1} = 0.2, \gamma_{8,2} = 0.0085, \gamma_{14,2} = 0.3, \gamma_{19,2} = 0.2 \).

Another significant aspect is represented by the tuning of the cost function weights: once a proper value of \( M \) is determined in order to avoid extra-queues (in this case, \( M = 10 \)), the tuning procedure is mainly focused in keeping virtually the same TTS and, at the same time, trying to obtain reduced lateral flows and smooth control actions. For the linear penalty term related to lateral flow, an important aspect is also represented by the locations of these control actions. In locations where strong lateral flow actions are expected (e.g. at lane-drops and on-ramps), vehicles are encouraged to change lane in the segment immediately upstream by setting the weight \( \beta_{i,j} = 0 \); in all other segments, the values are set to \( \beta_{i,j} = 0.01 \). As a last step, the weigh parameters of quadratic terms were tuned, obtaining the following values: \( \lambda_{f} = 10^{-5}, \lambda_{s} = 10^{-7}, \lambda_{m} = 10^{-5}, \) and \( \lambda_{d} = 10^{-6} \).

### 4.2 Optimisation results

The QP was solved using the barrier method algorithm provided within the optimisation solver Gurobi (Gurobi Optimization, 2013), which takes about 20 seconds to find the optimal solution. In the following, the control actions taken by the optimiser are analysed. As mentioned before, among the three possible control actions only lane-changing is penalised, whereas ramp metering and mainstream flow control have theoretically the same priority. The following different actions could be identified in the optimal control solution:

- In case the overall segment capacity is sufficient to accommodate the flow entering at the on-ramp, the space is created by assigning lateral flow actions (e.g. at the on-ramp in segment 6 lane 2). These actions are shown in Fig. 5: it is interesting to highlight that they lead to achieve the maximum flow at segment 6 (that includes also the flow entering at the on-ramp). Since the lateral movements are performed in the upstream segment, it is also interesting to point out that the phenomenon of having vehicles entering the motorway and changing lane in the same segment is avoided.
- In the lane-drop area of segment 14, the space for vehicles changing lane is created through some MTFC actions in the upstream segments, thus avoiding an excessive increase of density and excessive vehicle movements at the segment of the lane-drop.
- In the area of the on-ramp in segment 16, the main control strategy applied is ramp metering (see Fig. 6); however, the queue reaches its maximum length and the flow is released: for this reason, strong MTFC actions are subsequently performed in the upstream segments. It must be highlighted that MTFC actions are performed only on the inner lane, whereas in neighbouring lanes the flow (and consequently the speed) remains constantly at the maximum value.
- At the on-ramp in segment 22 the situation is similar to the previous case, however the MTFC actions are performed applying variable speed limits upstream as illustrated in Fig. 7 despite the maximum queue length is not reached (Fig. 6).

5. CONCLUSIONS

This paper has introduced an innovative optimal control approach for improving traffic condition in multilane motorways where a sufficient percentage of vehicles are equipped with VACS; a first-order model is designed to
compute lateral movements of vehicles and to reproduce traffic congestion phenomena due to various reasons, including capacity drop. Nevertheless, the restriction of having linear constraints brought to the definition of strong assumptions that must be more thoroughly analysed and verified. As every new model, calibration and validation procedures using real data (or alternatively data computed through a microscopic traffic simulator) are required and they are the subject of ongoing research activities.

In addition, the optimisation problem has produced important and useful results, showing that the use of VACS could generate strong benefits to traffic conditions, alleviate congestion and improve safety. The results of this optimal control problem could be therefore utilised as a starting point to better understand the strategies that must be exploited when defining control actions in presence of VACS. Of course, the direct application of these results is not possible before a sufficient number of vehicles will be equipped with the necessary devices. For this reason, currently the only opportunity to test this methodology is by using a microscopic simulation software; applying a receding horizon framework, several scenarios could be designed in order to investigate and verify some properties (e.g. delay in application of control decision, penetration rate of VACS, etc.).

ACKNOWLEDGEMENTS

The research leading to these results has been conducted in the frame of the project TRAMAN21, which has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 321132.

REFERENCES

Daganzo, C.F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent

with the hydrodynamic theory. Transportation Research Part B: Methodological, 28(4), 269–287.


