Nonlinear Optimal Control Applied to Coordinated Ramp Metering

Apostolos Kotsialos and Markos Papageorgiou, *Fellow, IEEE*

Abstract—The goal of this paper is to describe a generic approach to the problem of optimal coordinated ramp metering control in large-scale motorway networks. In this approach, the traffic flow process is macroscopically modeled by use of a second-order macroscopic traffic flow model. The overall problem of coordinated ramp metering is formulated as a constrained discrete-time nonlinear optimal control problem, and a feasible-direction nonlinear optimization algorithm is employed for its numerical solution. The control strategy's efficiency is demonstrated through its application to the 32-km Amsterdam ring road. A number of adequately chosen scenarios along with a thorough analysis, interpretation, and suitable visualization of the obtained results provides a basis for the better understanding of some complex interrelationships of partially conflicting performance criteria. More precisely, the strategy's efficiency and equitey properties as well as their tradeoff are studied and their partially competitive behavior is discussed. The results of the presented approach are very promising and demonstrate the efficiency of the optimal control methodology for motorway traffic control problems.

Index Terms—optimal control, nonlinear systems, ramp metering, traffic control (transportation).

I. INTRODUCTION

A NUMBER of approaches have been developed in the past for the design of control strategies for motorway networks that involve control measures such as route recommendation with the use of variable message signs (VMS) or equipped vehicles, ramp metering, motorway-to-motorway (mtm) control, etc. In this paper, the optimal control approach (discrete-time formulation) is applied to the design of large-scale optimal coordinated ramp metering control strategies.

Early applications of nonlinear constrained optimal control to coordinated ramp metering were reported in [1], [2], and [3]. In [4] and [5], a similar approach was applied to the Boulevard Peripherique in Paris. In [6]–[8] and [9], the problem of ramp metering as an optimal control problem was considered. In [10] and [11], results related to the integrated control of motorway networks, considering both route recommendation and ramp metering, were reported for a hypothetical network based on the approach described in this paper and the control software tool advanced motorway optimal control (AMOC). AMOC determines optimal control trajectories for arbitrary-topology networks with arbitrary geographically distributed control measures (including, besides ramp metering, mtm control and route recommendation). AMOC considers all the available control measures simultaneously so as to maximize their synergistic effect thereby avoiding conflicting control actions. Results for coordinated ramp metering based on AMOC are reported in [12]. A control strategy based on an optimal control problem formulation integrating ramp metering and variable speed limits was reported in [13]. For a more detailed and systematic overview of ramp metering control strategies see [14].

This paper first provides a thorough description of the optimal control approach to coordinated ramp metering along with a suitable numerical solution algorithm. Based on application of the proposed methodology to the Amsterdam ring road, previously obtained results are enhanced and extended with regard to the computational efficiency, which is a prerequisite for the real-time application of the method, as well as with regard to the operational performance (efficiency versus equity) and related tradeoffs. Both issues (computational effort and operational performance) are of major importance for the practical deployment of the proposed methodology.

The rest of this paper is structured as follows. Section II briefly presents the macroscopic traffic flow model which is used for the design of the control strategy. Section III formulates the problem of coordinated ramp metering as a constrained nonlinear discrete-time optimal control problem. Section IV presents the numerical solution algorithm for the formulated problem. Section V presents the application results for the Amsterdam ring road. The main conclusions and future work are reported in Section VI.

II. MACROSCOPIC TRAFFIC FLOW MODEL

A validated second-order traffic flow model is used for the description of traffic flow on motorway networks to provide the modeling part of the optimal control problem formulation. Since traffic assignment (routing) aspects of the traffic process are not necessary when the only type of control measure applied is ramp metering, the traffic assignment modeling part will not be presented (see [15], for details).

The network is represented by a directed graph whereby the links of the graph represent motorway stretches. Each motorway stretch has uniform characteristics, i.e., no on-/off-ramps and no major changes in geometry. The nodes of the graph are placed at locations where a major change in road geometry occurs, as well as at junctions, on-ramps, and off-ramps.

The time and space arguments are discretised. The discrete-time step is denoted by T (typically $T = 10$–15 s). A
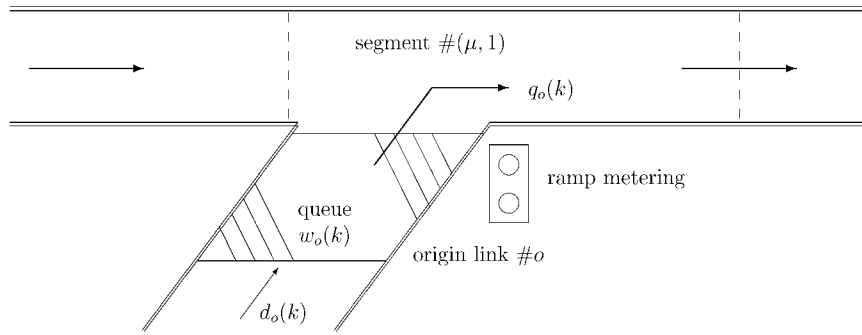
Fig. 1. Origin-link queue model.

motorway link m is divided into N_m segments of equal-length L_m (typically $L_m \approx 500$m). In principle, such a link may be divided into unequal segments as long as the stability condition $L_{m,i} \geq T \cdot v_{f,m}$ holds, where $v_{f,m}$ is the free-flow speed of link m corresponding to the speed at close to zero-density traffic conditions. This condition ensures that no vehicle travelling with its free speed will pass through a segment during one simulation time step. However, an uneven spatial discretization scheme does not provide any advantage over the equal length division. It is for our convenience that we use equal segment lengths for each motorway link. Each segment i of link m at time $t = kT, k = 0, \ldots, K$, is macroscopically characterized via the following variables: The *traffic density* $\rho_{m,i}(k)$ (veh/lane-km) is the number of vehicles in segment i of link m at time $t = kT$ divided by L_m and by the number of lanes Λ_m; the *mean speed* $v_{m,i}(k)$ (km/h) is the mean speed of the vehicles included in segment i of link m at time kT; and the *traffic volume* or *flow* $q_{m,i}(k)$ (veh/h) is the number of vehicles leaving segment i of link m during the time period $[kT, (k+1)T]$, divided by T. For each segment i of link m at each time step k we have the following:

$$\rho_{m,i}(k+1) = \rho_{m,i}(k) + \frac{T[q_{m,i-1}(k) - q_{m,i}(k)]}{L_m \Lambda_m} \quad (1)$$

$$q_{m,i}(k) = \rho_{m,i}(k) v_{m,i}(k) \Lambda_m \quad (2)$$

$$v_{m,i}(k+1) = v_{m,i}(k) + \frac{T\{V[\rho_{m,i}(k)] - v_{m,i}(k)\}}{\tau}$$
$$+ \frac{T[v_{m,i-1}(k) - v_{m,i}(k)]v_{m,i}(k)}{L_m}$$
$$- \frac{\nu T}{\tau L_m} \frac{\rho_{m,i+1}(k) - \rho_{m,i}(k)}{\rho_{m,i}(k) + \kappa} \quad (3)$$

$$V[\rho_{m,i}(k)] = v_{f,m} \exp\left[-\frac{1}{a_m}\left(\frac{\rho_{m,i}(k)}{\rho_{cr,m}}\right)^{a_m}\right] \quad (4)$$

where $\rho_{cr,m}$ denotes the critical density per lane of link m (the density where the maximum flow in the link occurs), and a_m is a parameter of the fundamental diagram (4) of link m; the fundamental diagram expresses a nonlinear relationship between the mean speed and the traffic density in link m which is needed in the second term of the right-hand side of (3). Furthermore, τ, a time constant, ν, an anticipation constant, and κ, are constant parameters, which are equal for all network links. This is due to the fact that the traffic model has relatively low sensitivity to these parameters, see [16] and [17]. Additionally, the mean speed resulting from (3) is limited from below by the minimum

speed in the network $v_{\min}$ in order to avoid unrealistically low flows during congestion.

Two additional terms are added to (3) in order to consider the speed decrease caused by merging phenomena at a junction and by lanes drops, respectively, see [16] for details.

For origin links, i.e., links that receive traffic demand and forward it into the motorway network, a simple queue model is used (Fig. 1). The outflow q_o of an origin link o depends on the traffic conditions of the corresponding mainstream segment $(\mu, 1)$ and the existence of ramp metering control measures. If ramp metering is applied, then the outflow $q_o(k)$ that is allowed to leave origin o during period k, is a portion $r_o(k)$ of the maximum outflow $\hat{q}_o(k)$ that would leave o and enter the mainstream in absence of ramp metering. Thus, $r_o(k) \in [r_{\min,o}, 1]$ is the metering rate for the origin link o, i.e., a control variable, where $r_{\min,o}$ is a minimum admissible value; typically, $r_{\min,o} > 0$ is chosen so as to avoid ramp closure. If $r_o(k) = 1$, no ramp metering is applied, else $r_o(k) < 1$. The queueing model is described by the following:

$$w_o(k+1) = w_o(k) + T[d_o(k) - q_o(k)] \quad (5)$$

where $w_o(k)$ is the queue length (veh) in origin o at time kT, and $d_o(k)$ is the demand (veh/h) at o at the same period. The outflow $q_o(k)$ is determined as follows:

$$q_o(k) = r_o(k)\hat{q}_o(k) \quad (6)$$

with

$$\hat{q}_o(k) = \min\{\hat{q}_{o,1}(k), \hat{q}_{o,2}(k)\} \quad (7)$$

and

$$\hat{q}_{o,1} = d_o(k) + w_o(k)/T \quad (8)$$

$$\hat{q}_{o,2} = Q_o \min\left\{1, \frac{\rho_{\max} - \rho_{\mu,1}(k)}{\rho_{\max} - \rho_{cr,\mu}}\right\} \quad (9)$$

where Q_o is the on-ramp's capacity (veh/h), i.e., the on-ramp's maximum possible outflow under free-flow traffic conditions in the mainstream; and $\rho_{\max}$ (veh/lane-km) is the maximum density in the network segments. Thus, the maximum outflow $\hat{q}_o(k)$ is determined by the current origin demand if $\hat{q}_{o,1} < \hat{q}_{o,2}$ [see (7), (8)], or by the geometrical ramp capacity Q_o if the mainstream density is undercritical, i.e., $\rho_{\mu,1}(k) < \rho_{cr,\mu}$ [see (9)], or by the reduced capacity due to congestion of the mainstream, if $\rho_{\mu,1}(k) > \rho_{cr,\mu}$ [see (9)]. Thus, (9) models the reduction of the origin link's capacity due to mainstream congestion (Fig. 2).
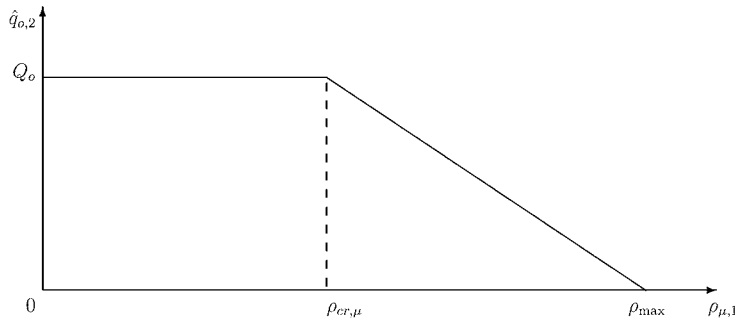
Fig. 2. Reduction of origin link capacity due to mainstream congestion.

A model similar to (5)–(9) applies to motorway-to-motorway interchanges.

Motorway bifurcations and junctions (including on-ramps and off-ramps) are represented by nodes. Traffic enters a node n through a number of input links and is distributed to the output links according to

$$Q_n(k) = \sum_{\mu \in I_n} q_{\mu,N_\mu}(k) \tag{10}$$

$$q_{m,0}(k) = \beta_n^m(k) Q_n(k) \quad \forall m \in O_n \tag{11}$$

where I_n is the set of links entering node n, O_n is the set of links leaving n, $Q_n(k)$ is the total traffic volume entering n at period k, $q_{m,0}(k)$ is the traffic volume that leaves n via outlink m, and $\beta_n^m(k)$ is the portion of $Q_n(k)$ that leaves the node through link m. $\beta_n^m(k)$ are the *turning rates* of node n and are assumed to be known for the entire time horizon. Equations (10) and (11) provide $q_{m,0}(k)$ required in (1) for $i = 1$.

If a node n has more than one leaving link, then the upstream influence of density (which is necessary for the modeling of congestion spillback) is taken into account in the last segment of the incoming link by an appropriate calculation of ρ_{m,N_m+1} which is required in (3) for $i = N_m$. When a node n has more than one entering links, then the downstream influence of speed is taken into account by appropriately calculating $v_{m,0}$ required in (3) for $i = 0$. For more details see [15].

It should be noted here that the investigations reported in this paper were based on turning rate information instead of OD information. Since it is very rare for neighboring on-ramps to have significantly different OD rates, something that would seriously affect the controller's performance, the use of turning rates instead does not pose any problem. AMOC can deal with OD information (see also [10] and [11]) but at the expense of the required computation effort due to the more detailed description of the traffic process, see also [15].

III. PROBLEM FORMULATION

The coordinated ramp metering control problem is formulated as a dynamic optimal control problem with constrained control variables which can be solved numerically over a given time horizon. The motorway traffic flow is considered as a process under control via the various ramp meters installed at the network entrances. The state of the process is described by the state vector $\mathbf{x} \in \Re^N$ and its evolution depends on the system dynamics and the input variables. The input variables are distinguished into control variables $\mathbf{u} \in \Re^M$ and the

uncontrollable external disturbances $\mathbf{d} \in \Re^D$. In the following, we will introduce a general problem formulation whereby the value of each control measure may change less frequently than at each model sample time T. Assume that the M different control measures have p distinct control sample times which are assumed to be multiples of the model sample time T, i.e., $T_\ell = z_\ell T, z_\ell \in \mathcal{N}, \ell = 1, \ldots, p, p \leq M = \dim(\mathbf{u})$. Let $k_\ell = \text{integer } [k/z_\ell]$ and let j_ℓ denote the number of control measures that have sample time T_ℓ and are organized in the vector $\mathbf{u}_\ell$. Then $\mathbf{u}(k) = [\mathbf{u}_1(k_1)^T \cdots \mathbf{u}_p(k_p)^T]^T$. The general discrete-time formulation of the optimal control problem reads

Minimize

$$J = \vartheta[K] + \sum_{k=0}^{K-1} \varphi[\mathbf{x}(k), \mathbf{u}_1(k_1), \ldots, \mathbf{u}_p(k_p), \mathbf{d}(k)] \tag{12}$$

subject to

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k), \mathbf{u}_1(k_1), \ldots, \mathbf{u}_p(k_p), \mathbf{d}(k)] \tag{13}$$

$$\mathbf{x}(0) = \mathbf{x}_0$$

$$\mathbf{u}_{\ell,\min} \leq \mathbf{u}_\ell(k_\ell) \leq \mathbf{u}_{\ell,\max} \quad \text{and } \ell \in \{1, \ldots, p\} \tag{14}$$

where K is the considered time horizon, and ϑ, φ are arbitrary, twice differentiable, nonlinear cost functions.

Based on the previous section, it may be seen that by substituting (2), (10), and (11) into (1); (4) into (3); and (6)–(9) into (15), the traffic flow model equations take the form of (13). In this case, the state vector $\mathbf{x}$ consists of the densities $\rho_{m,i}$, the mean speeds $v_{m,i}$ of every segment i of every link m, and the queues w_o of every origin o. The control vector $\mathbf{u}$ consists of the ramp metering rates r_o of every on-ramp o under control, with $r_{o,\min} \leq r_o(k) \leq 1$ according to (14). We assume here that all ramp meters have the same sample time $T_1 = z_1 T$. Finally, the disturbance vector consists of the demands d_o at every origin of the network, and all the turning rates β_n^m at the network's bifurcations. The disturbance trajectories $\mathbf{d}(k)$ are assumed known over the time horizon K. For practical applications, these values may be predicted with sufficient accuracy based on historical data and, if necessary, on real-time estimations, see, for example, [18].

The chosen cost criterion aims at minimizing the total time spent (TTS) of all vehicles in the network (including the waiting time experienced inside the network queues). It can be proved that the minimization of the TTS is equivalent to the maximization of the time weighted network outflow, see [3] and [14]. The minimization of the TTS is a natural objective for the traffic systems considered here, since it represents the total time that all users spent in the network. Furthermore, there is a widespread

use of the TTS as an evaluation criterion of control strategies that do not explicitly aim at its minimization, such as feedback strategies. More precisely, the cost criterion is as follows:

$$
J = T \sum_k \left\{ \sum_m \sum_i \rho_{m,i}(k) L_m \Lambda_m + \sum_o w_o(k) \right.
$$
$$
+ a_f \sum_o [r_o(k_1) - r_o(k_1 - 1)]^2
$$
$$
\left. + a_w \sum_o \psi[w_o(k)]^2 \right\} \tag{15}
$$

with

$$
\psi[w_o(k)] = \max\{0, w_o(k) - w_{o,\max}\} \tag{16}
$$

where a_f, a_w are weighting factors. The first two terms of (15) correspond exactly to the TTS. The term with weight a_f is included in the cost criterion to suppress high-frequency oscillations of the control trajectories. The last penalty term is included in the cost criterion in order to enable the control strategy to limit the queue lengths at the origins if and to the level desired. The parameters $w_{o,\max}$ are predetermined constants that express the maximum permissible number of vehicles at any time period in origin o's queue. The weights a_f and a_w were adjusted via trial-and-error striking a balance between acceptable time-variations in the optimal control trajectories and queue constraint violations on one hand and efficiency and fast convergence to the optimum on the other.

IV. NUMERICAL SOLUTION ALGORITHM

Define the quantities

$$
K_\ell = \begin{cases} \frac{K}{z_\ell} - 1, & \text{if } K \bmod z_\ell = 0 \\ \text{integer}\left(\frac{K}{z_\ell}\right), & \text{otherwise.} \end{cases} \tag{17}
$$

For a given admissible trajectory $\mathbf{u}_\ell(k_\ell), k_\ell = 0, \dots, K_\ell$, the state trajectory $\mathbf{x}(k + 1)$ can be found by solving (13) starting with the known initial state $\mathbf{x}(0)$ and, hence, the cost criterion can be regarded as depending on the control variables only, i.e., $J = \overline{J}(\mathbf{u})$. The gradient of $\overline{J}$ with respect to $\mathbf{u}_\ell(k_\ell)$ on the equality constraints surface for the time period k_ℓ is given by

$$
\mathbf{g}_\ell(k_\ell) = \sum_{k=k_\ell z_\ell}^{\min\{(k_\ell+1)z_\ell-1, K-1\}}
$$
$$
\left\{ \frac{\partial \varphi[\mathbf{x}(k), \mathbf{u}_1(k_1), \dots, \mathbf{u}_\ell(k_\ell), \dots, \mathbf{u}_p(k_p), \mathbf{d}(k)]}{\partial \mathbf{u}_\ell(k_\ell)} \right.
$$
$$
+ \frac{\partial \mathbf{f}[\mathbf{x}(k), \mathbf{u}_1(k_1), \dots, \mathbf{u}_\ell(k_\ell), \dots, \mathbf{u}_p(k_p), \mathbf{d}(k)]^T}{\partial \mathbf{u}_\ell(k_\ell)}
$$
$$
\left. \boldsymbol{\lambda}(k+1) \right\} \tag{18}
$$

where the costate vector $\boldsymbol{\lambda} \in \Re^N$ satisfies

$$
\boldsymbol{\lambda}(k) = \frac{\partial \mathbf{f}[\mathbf{x}(k), \mathbf{u}_1(k_1), \dots, \mathbf{u}_p(k_p), \mathbf{d}(k)]^T}{\partial \mathbf{x}(k)} \boldsymbol{\lambda}(k+1)
$$
$$
+ \frac{\partial \varphi[\mathbf{x}(k), \mathbf{u}_1(k_1), \dots, \mathbf{u}_p(k_p), \mathbf{d}(k)]}{\partial \mathbf{x}(k)}
$$
$$
k = 0, \dots, K-1 \tag{19}
$$

and

$$
\boldsymbol{\lambda}(K) = \frac{\partial \vartheta[\mathbf{x}(K)]}{\partial \mathbf{x}(K)}. \tag{20}
$$

A projected gradient $\boldsymbol{\xi}_\ell(k_\ell)$ is defined to have its components $\xi_{\ell,i}(k_\ell)$ equal to $g_{\ell,i}(k_\ell)$ if none of the corresponding bounds (14) is active, and $\xi_{\ell,i}(k_\ell) = 0$ else. Furthermore we define a saturation vector function $\mathbf{sat}(\boldsymbol{\eta})$ with components

$$
\text{sat}_i(\boldsymbol{\eta}) = \begin{cases} \eta_{i,\max}, & \text{if } \eta_i > \eta_{i,\max} \\ \eta_{i,\min}, & \text{if } \eta_i < \eta_{i,\min} \\ \eta_i, & \text{else} \end{cases}
$$

where $\eta_{i,\max}$ and $\eta_{i,\min}$ are the upper and lower bounds, respectively, of the variable η_i.

The necessary conditions of optimality are given by (13), (14), (19), (20), and $\xi_{\ell,i}(k_\ell) = 0 \, \forall i, k_\ell, \ell$. A well known solution algorithm based on feasible directions can be described as follows:

Step 1) select an admissible initial control trajectory $\mathbf{u}^{(0)}(k), k = 0, \dots, K - 1$; set the iteration index $\iota = 0$;

Step 2) using $\mathbf{u}^{(\iota)}(k), k = 0, \dots, K - 1$, solve (13) from known initial condition $\mathbf{x}(0)$ to obtain $\mathbf{x}^{(\iota)}(k + 1)$; using $\mathbf{x}^{(\iota)}(k + 1)$ and $\mathbf{u}^{(\iota)}(k)$ solve (19) from terminal condition (20) to obtain $\boldsymbol{\lambda}^{(\iota)}(k + 1)$;

Step 3) using $\mathbf{x}(0), \mathbf{x}^{(\iota)}(k+1), \mathbf{u}^{(\iota)}(k)$, and $\boldsymbol{\lambda}^{(\iota)}(k+1), k = 0, 1, \dots, K - 1$, calculate the gradients $\mathbf{g}^{(\iota)}(k)$ and $\boldsymbol{\xi}^{(\iota)}(k)$;

Step 4) specify a search direction $\mathbf{p}^{(\iota)}(k)$ (see below);

Step 5) apply an one-dimensional search routine along the $\mathbf{p}^{(\iota)}$-direction to obtain a new, improved admissible control trajectory $\mathbf{u}^{(\iota+1)}(k)$, i.e.,

$$
\alpha^{(\iota)} = \arg\min_\alpha \overline{J}\left[\mathbf{sat}\left(\mathbf{u}^{(\iota)} + \alpha \mathbf{p}^{(\iota)}\right)\right]
$$

where $\alpha^{(\iota)}$ is the scalar step length, and

$$
\mathbf{u}^{(\iota+1)}(k) = \mathbf{sat}\left(\mathbf{u}^{(\iota)}(k) + \alpha \mathbf{p}^{(\iota)}(k)\right);
$$

Step 6) if, for a given scalar σ, the condition $(J^{(\iota+1)} - J^{(\iota)})/J^{(\iota)} < \sigma$ is satisfied, stop; otherwise set $\iota := \iota + 1$ and go back to Step 2.

Several techniques for specifying a search direction $\mathbf{p}^{(\iota)}$ in Step 4 of the algorithm can be applied, including steepest descent, conjugate gradient, and variable metric, see [19] and [20]. Whichever of these seach methods is used, a periodic restart of the search algorithm takes place, i.e., after a certain number of iterations, the search direction is set equal to the steepest-descent direction so as to accelerate the algorithm's speed of convergence to a minimum. For more details on the overall numerical solution algorithm, see [21] and [20].

As an alternative solution algorithm, Step 4 may be omitted while Step 5 may be replaced by the derivative backpropagation method RPROP (resilient backPROPagation) proposed in [22] for neural network training. RPROP does not require the line-search routine used in the above algorithm, since it calculates the necessary changes of the control variables at each iteration based only on the signs of the gradient components $g_i(k)$. To the best of our knowledge, this is the first time that the RPROP method is used for large-scale nonlinear optimization.

When the RPROP method is used (slightly modified as compared to its original form in [22]), Steps 4 and 5 of the above algorithm are replaced by the following calculations:

$$\mathbf{u}^{(\iota+1)}(k) = \mathbf{sat}\left(\mathbf{u}^{(\iota)}(k) + \Delta\mathbf{u}^{(\iota)}(k)\right)$$

where the control variable increments $\Delta_i^{(\iota)}(k)$ are calculated based on the sign of the gradient $g_i^{(\iota)}(k)$ and the increment $\Delta_i^{(\iota-1)}(k)$ of the previous iteration, as shown in the equation at the bottom of the page, where $0 < \eta^- < 1 < \eta^+$. Thus, if no change of sign of $g_i(k)$ occured between iterations $(\iota - 1)$ and (ι), the corresponding increment $\Delta u_i^{(\iota)}(k)$ is increased as compared to $\Delta u_i^{(\iota-1)}(k)$ by a factor η^+ (typically $\eta^+ = 1.2$). If a sign change of $g_i(k)$ occured, then the algorithm has stepped over a minimum in the corresponding direction, hence, the new increment $\Delta u_i^{(\iota)}(k)$ is opposite in sign and reduced in size (typically $\eta^- = 0.5$) as compared to $\Delta u_i^{(\iota-1)}(k)$. The algorithm starts with $\Delta u_i^{(0)}(k) = \Delta_i$; the calculated $\Delta u_i^{(\iota)}(k)$ at each iteration may be restricted to lie in a prespecified interval $[\Delta_{\min}, \Delta_{\max}]$.

The RPROP method preserves feasibility of the overall algorithm but cannot guarantee a decrease of the objective function value at each iteration.

The previously described numerical optimization algorithm does not guarantee convergence to a global minimum. Previous experiences, indicate that a physically satisfactory minimum is always achieved. In some cases, when the algorithm was started with different initial control trajectories, convergence to different local minima occurred. The corresponding difference in the cost criterion's value, however, was insignificant.

V. APPLICATION EXAMPLE

A. Site Description

The previously described approach to network-wide optimal ramp metering has been applied to the Amsterdam ring road with the use of AMOC.

The Amsterdam orbital motorway (A10) is shown in Fig. 3. The A10 simultaneously serves local, regional, and inter-regional traffic and acts as a hub for traffic entering and exiting North Holland. There are four main connections with other motorways, the A8 at the north, the A4 at the southwest, the A2 at the south, and the A1 at the southeast. The A10 contains two tunnels, the Coen Tunnel at the northwest and the Zeeburg Tunnel at the east.

For the purposes of our study, only the counterclockwise direction of the A10, which is about 32 km long, is considered. There are 21 on-ramps on this motorway, including the connections with the A8, A4, A2, and A1 motorways, and 20 off-ramps, including the junctions with A4, A2, A1, and A8. The topological network model may be seen in Fig. 4. It is assumed that ramp metering may be performed at each of the on-ramps.

The model parameters for this network were determined from validation of the network traffic flow model against real
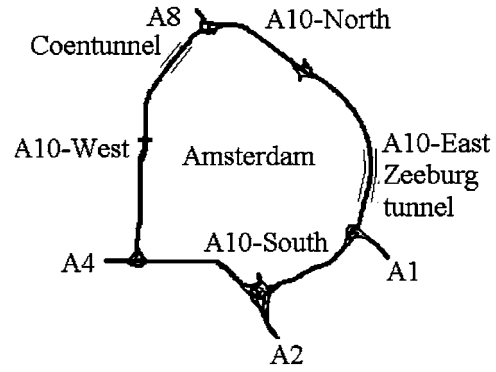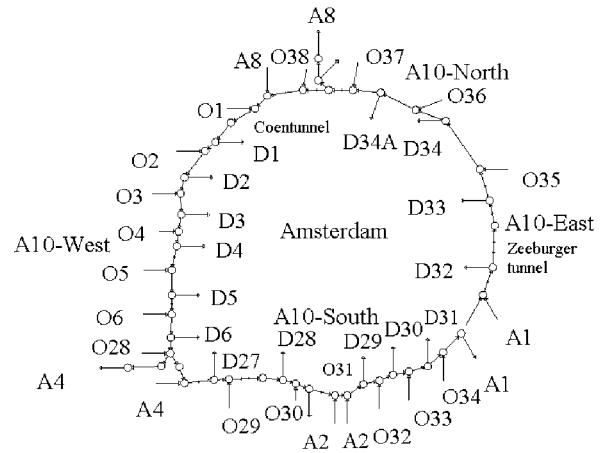

Fig. 3. Amsterdam ring road.

Fig. 4. Amsterdam ring road model.

data taken from the motorways. The validation was conducted in two phases. The first phase was the quantitative validation where several motorway stretches from the ring road were selected and a rigorous optimization-based validation procedure was carried out for each of them estimating the model parameters $v_{\min}, \tau, \nu, \rho_{\max}, \kappa, \phi, \delta$, and $a_m, v_{f,m}$, and $\rho_{\mathrm{cr},m}$ for every link m. The second phase of the validation involved the manual tuning of other parameters, such as certain turning rates, that enabled the model to capture the network-wide dynamics. In fact, the network considered here is only a part of the whole network considered in [23] and [24] where the detailed results of the modeling and validation of the Amsterdam network are reported.

The ring road was divided in 76 segments with average length 421 m. This means that the state vector is 173-dimensional (including the 21 on-ramp queues). If ramp metering is applied to all on-ramps, the control vector is 21-dimensional, while the disturbance vector is 41-dimensional. With a time step $T = 10$ s we have, for a horizon of 4 h, $K = 1440$. This means that for a control sample time of 1 min and all on-ramps metered, there are 254, 160 variables included in the resulting nonlinear optimization problem.

$$\Delta u_i^{(\iota)}(k) = \begin{cases} -\mathrm{sign}\left(g_i^{(\iota)}(k)\right)\eta^+\Delta u_i^{(\iota-1)}(k), & \text{if } g_i^{(\iota-1)}(k)g_i^{(\iota)}(k) > 0 \\ -\eta^-\Delta u_i^{(\iota)}(k), & \text{else} \end{cases}$$
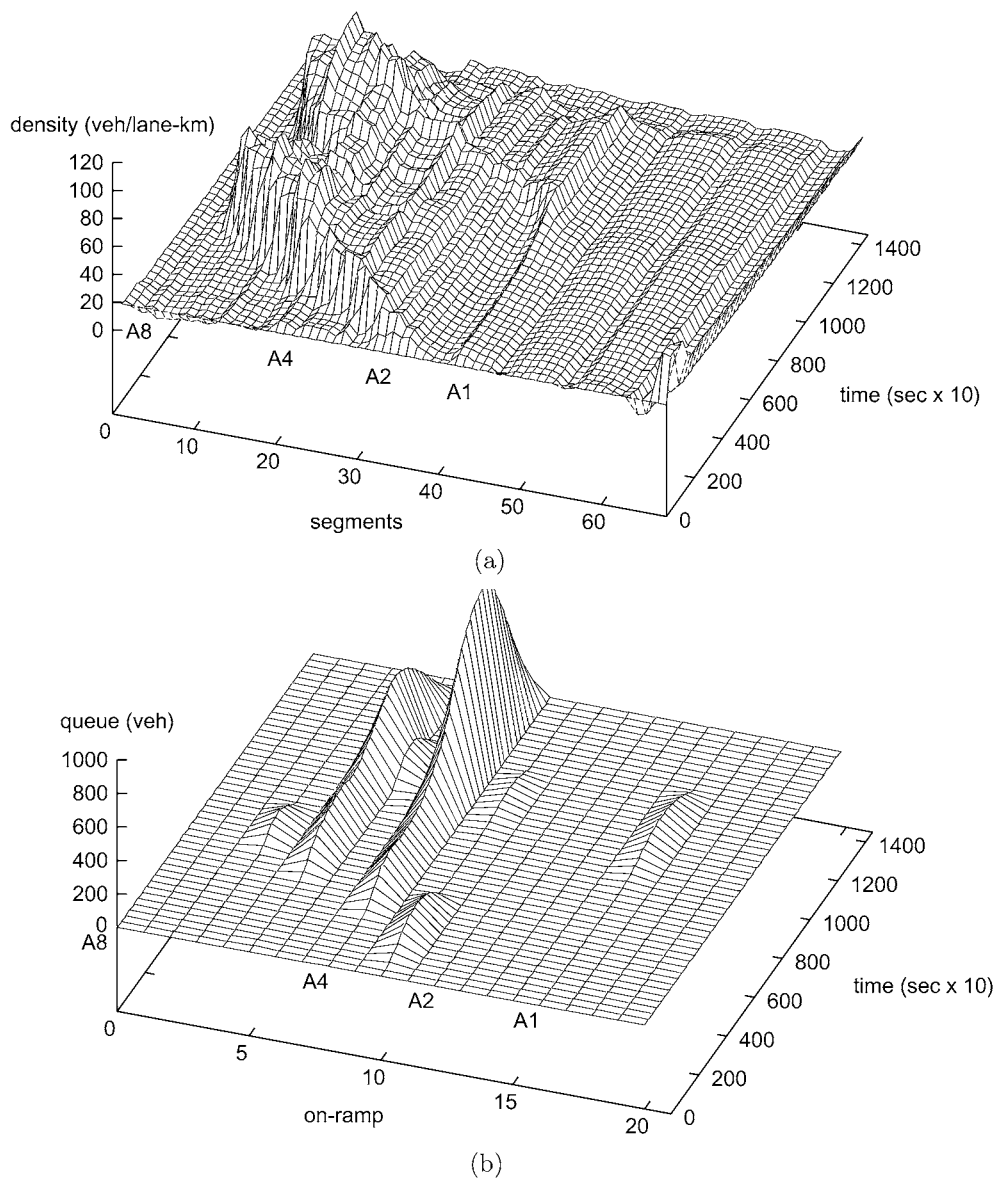
Fig. 5. No control. (a) Density. (b) On-ramp queues.

Four different scenarios are considered for this network. For all scenarios it is assumed that, even in case of long on-ramp queues, no rerouting takes place. Scenario 0 corresponds to the no-control case, i.e., when no ramp metering control measures are applied. During the four hour period in this scenario, which corresponds to the evening rush hours, the whole spectrum of traffic conditions (free, critical, and congested) appears on the ring road. Starting from initially uncongested conditions, the build up of congestion inside the motorway and the build up of queues at the network's on-ramps may be observed, and we can follow their evolution until they are resolved. This way, the no-control case provides a sound base-case, based on which the control strategy performance is evaluated under different scenarios. In Scenario 1, the maximum permissible storage $w_{o,\max}$ in (16) is set to 40 vehicles for the urban on-ramps and 100 vehicles for the motorway-to-motorway ramps; in Scenario 2 these storage capacities are equal to 80 and 120 vehicles, respectively, while in Scenario 3 no maximum queue constraints are considered, i.e., $a_w = 0$ in (15).

B. Optimal Results

When no control measures are applied, the excessive demand coupled with the uncontrolled entrance of drivers in the mainstream causes congestion from the beginning of the time horizon [Fig. 5(a)]. This congestion originates at the junction of A2 with A10 and propagates upstream blocking the A4 and a large part of the A10-West. By the time this congestion begins to dissolve, a new one appears at the junction of A10 with A1 which begins to propagate upstream until it reaches the first congestion whose trend of resolving is reversed and both are united into a single more severe congestion. This strong congestion keeps the A4 entrance to the A10 blocked, something which results in the accumulation of many vehicles in the motorway-to-motorway (mtm) on-ramp of A4 (i.e., a spillback of the congestion from A10 onto the A4 motorway) and in the surrounding on-ramps [Fig. 5(b)]. As a result the TTS becomes 13 226 veh·h.
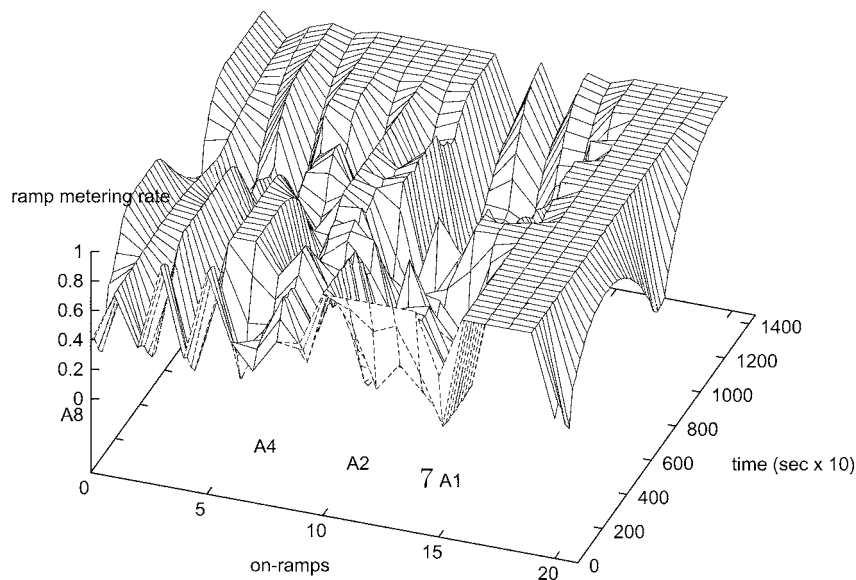
When optimal control is applied to the network under Scenario 1, the TTS becomes 9 032 veh·h, a 31.7% improvement.
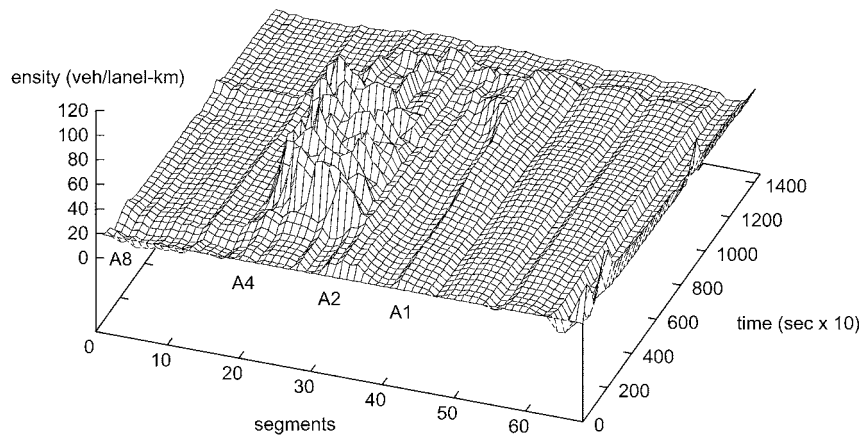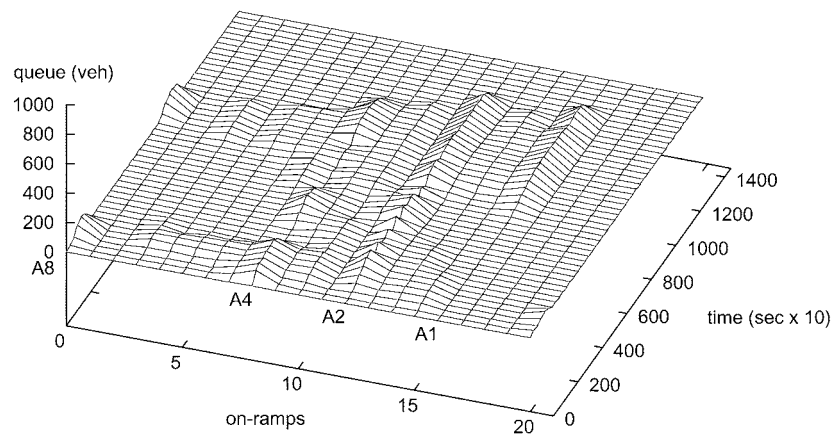
Fig. 6. Scenario 1: (a) density, (b) on-ramp queues, and (c) optimal ramp metering rates.

This improvement is evident in Fig. 6 which depicts the density and queue evolution profiles as well as the optimal ramp metering rates for Scenario 1. The large ramp queues that occur in the no-control case are not present any more, but queues are spread to many on-ramps so as to counteract the formation of congestion. It can be seen that the control strategy manages to comply with the queue constraints imposed and at the same time reduce the cost criterion, by distributing the queues spatially and temporally almost in the same pattern as the density's propagation in Scenario 0 [Fig. 5(a)].
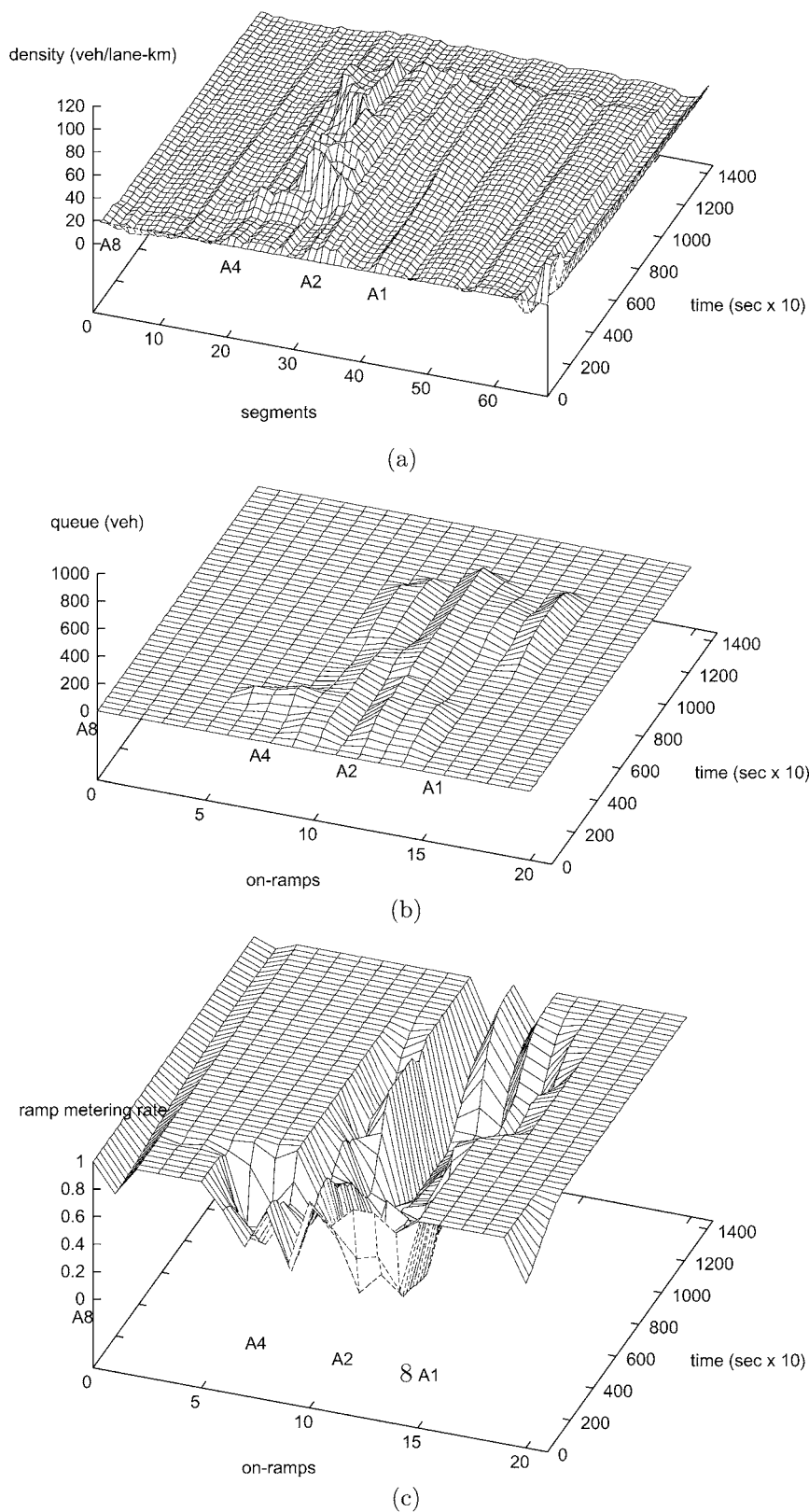
Fig. 7. Scenario 2: (a) density, (b) on-ramp queues, and (c) optimal ramp metering rates.

Scenario 2 assumes that there is more ramp storage capacity at the strategy's disposal (80 vehicles for urban on-ramps and 120 vehicles for mtm on-ramps). When optimal control is applied under these conditions, the TTS becomes 8 230 veh · h, a 37.8% improvement over the no-control case. This larger im-

provement is evident in Fig. 7(a) where the density profile is seen to be much flatter than that of Scenario 1 [Fig. 6(a)]. Because larger storage space is available to the strategy, larger queues are formed, but less on-ramps are used for storage purposes, see Fig. 7(b). Once the strategy has dealt with the primary

cause of congestion effectively, due to the increased storage capacity available, keeping the demand of the A10-West metered would only increase the delays without any benefit to the overall cost criterion, hence, ramp metering is released. The control strategy is intelligent enough to create queues only when and only to the extent they may contribute to the reduction of the TTS under the imposed queue constraints, as can also be seen from the optimal ramp metering rates in Fig. 7(c).

In the case of Scenario 3, the absence of any queue constraints leads to a TTS equal to 7 466 veh · h, an improvement of 43.5% over the no-control case. As can be seen from Fig. 8(a), the density evolution profile is completely flat in this case, indicating the fact that there is no congestion present in the ring road. The control strategy achieves this impressive amelioration of traffic conditions by creating large queues at the on-ramps that are located in the critical bottleneck area where congestion originates, see Fig. 8(b). The queues created, when the optimal ramp metering rates depicted in Fig. 8(c) are applied, prevent the excessive demand from entering into the motorway and degrading the network's capacity. Since there is no limit to the queues that are allowed to occur, the queues created in the bottleneck area of A10-South are quite large. No queues are required for the on-ramps further upstream since the primary cause of congestion is dealt with locally. The optimal results for each scenario are summarized in Table I.

The amount of TTS reduction depends on the on-ramp storage capacity available to the control strategy. The behavior exhibited by optimal control application in the three scenarios indicates that the most efficient way to deal with bottlenecks and potential congestion is to perform ramp metering at the on-ramps immediately upstream of the primary bottleneck location, as was the case in Scenario 3. The unlimited storage capacity assumed in this scenario made the optimal control indifferent to the creation of large queues in the vicinity of the bottleneck. However, once the storage capacity becomes limited, as in the cases of Scenarios 1 and 2, queues are created further upstream from the bottleneck location, and at the same time some congestion appears in the mainstream whose extent depends on the storage capacity assumed available by each scenario. Despite the fact that further storage is available at the on-ramps far upstream from the bottleneck, e.g., in A10-North and East (and in the case of Scenario 2 in A10-West as well), congestion is tolerated because large queues in these on-ramps would impose delays to drivers exiting the network before reaching the bottleneck. In essence, the control strategy aims at establishing an optimal tradeoff between the delay reduction due to the decrease of the congestion extent, and the delay increase due to the metering of vehicles that exit before reaching the bottleneck.

C. Equity

Let us now examine the queues formed when the control strategy is applied for the three scenarios considered. Figs. 6(b), 7(b), and 8(b) differ in the size and location of the queues formed which is natural because different storage capacities are at the strategy's disposal in each case. In Scenario 1, the short queues are spread over the network, and they endure for a large part of the time horizon. In Scenario 2, the queues are longer and

TABLE I
TTS FOR EACH SCENARIO

Scenario	TTS (veh·hours)	Improvement
No control	13,226	–
1	9,032	31.7%
2	8,230	37.8%
3	7,466	43.5%

less spread; they are concentrated around the critical area, and they endure for less time than in Scenario 1. In Scenario 3, long queues occur at selected on-ramps where excessive demand ultimately creates the congestion problem, and their duration is shorter. The reason for these different behaviors is the queue constraints.

The control strategy distributes the metering burden for reducing the TTS among the on-ramps subject to the maximum-queue constraints. When no such constraints are imposed, the burden of improving the traffic conditions is assigned to the on-ramps that ultimately create the problem because all corresponding users will cross the bottleneck location, hence, nobody is delayed unnecessarily. Thus, imposing maximum queue constraints can be seen as a way of distributing the delays experienced by drivers while waiting in the controlled on-ramps. This way the cost of ameliorating the traffic conditions in the overall network is shared more or less fairly among the drivers that enter the motorway from various on-ramps, depending on how strict the constraints are. This equity aspect, however, is achieved at the expense of a lower improvement of the traffic conditions, hence, equity and efficiency are partially competing properties of the control strategy. For a discussion on the equity properties of ramp metering strategy, see also [25] and [26].

Fig. 9 depicts the average (over the time horizon) travel times at each on-ramp, needed for ramp-queueing and travelling a 6.5 km mainstream distance. For each on-ramp o the average travel time $\bar{t}_o$ is calculated according to

$$\bar{t}_o = \frac{1}{K} \sum_{k=0}^{K-1} t_o(k)$$

$$= \frac{1}{K} \sum_{k=0}^{K-1} \left[\frac{w_o(k)}{q_o(k)} + \sum_{i=\zeta_1}^{\zeta_2} \sum_{j=1}^{\xi} \frac{L_i}{v_{i,j}(k)} \right] \qquad (21)$$

where ζ_1 is the index number of the link downstream of o, and ζ_2 is the link index number where the considered mainstream distance of 6.5 km ends, and $\xi \leq N_i$.

It can be seen that, in absence of control measures, the on-ramps of A10-West are the most disadvantaged ones because the congestion occuring at the A10-South creates queues and severe density waves traveling upstream [Fig. 5(a)]. Since the ring road from the A1 up to the A8 (counterclockwise) is not congested, the average travel time for the corresponding on-ramps is small.

The smallest average travel times are achieved in Scenario 3, where no maximum queue constraints exist. However, in this scenario, larger travel times occur at the specific on-ramps between A1 and A4 where ramp metering is applied [Fig. 8(b)]. By focusing in this area, the control strategy clearly induces a
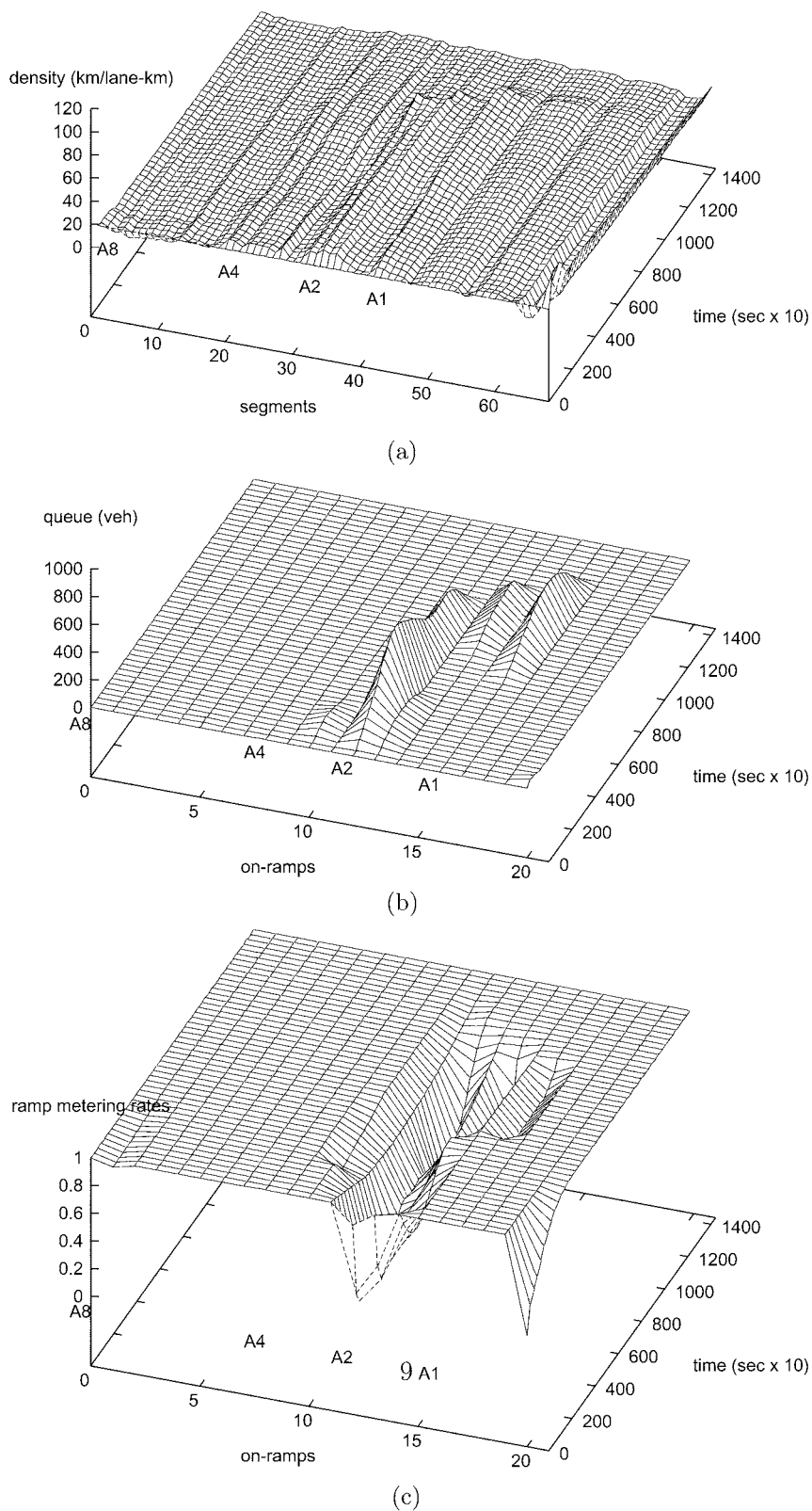
Fig. 8. Scenario 3: (a) density, (b) on-ramp queues, and (c) optimal ramp metering rates.

disadvantage to the drivers that enter the motorway from the specific on-ramps, over the drivers that enter the ring road from the rest on-ramps. By doing so, however, it achieves the smallest TTS compared with Scenarios 1 and 2.

The effects of queue constraints in Scenarios 1 and 2 may be seen in Fig. 9. The average travel times become larger as the maximum queue constraints become smaller, but this increase is distributed more evenly to other on-ramps also, not only to those
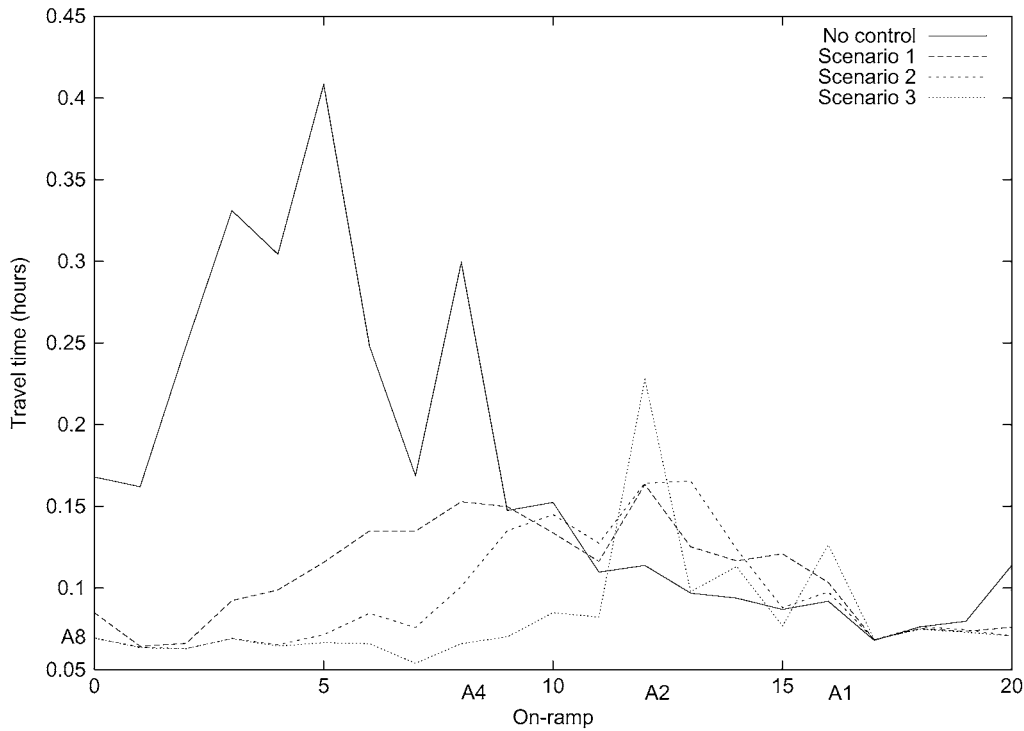
Fig. 9. Average travel times $\bar{t}_o$ at each on-ramp o.

in the bottleneck area that are extensively metered in Scenario 3, thus, making the control strategy to behave more fairly.

As a further equity index, the spatial variance of the travel times at every discrete time step k is calculated according to

$$\text{Var}_t(k) = \frac{\sum_{o=1}^n [\tilde{t}(k) - t_o(k)]^2}{n} \tag{22}$$

where $\tilde{t}(k) = \sum_{o=1}^n t_o(k)/n$ and n is the number of origins. The average variance is given by $\text{Var}_t = \sum_{k=0}^{K-1} \text{Var}_t(k)/K$. In the no-control case, the variance of the travel times follows the pattern of the congestion formation (Fig. 10) illustrating the large travel times experienced by the drivers that enter the mainstream from on-ramps directly affected by the congestion, and the average variance is $0.0183\,\text{h}^2$. In the cases of Scenarios 1, 2, and 3, the variance of the travel times becomes much smaller as a result of the applied ramp metering, while the total variance becomes $0.0022\,\text{h}^2$, $0.0020\,\text{h}^2$, and $0.0030\,\text{h}^2$, respectively. Comparing the three control scenarios, Scenario 3 has the greatest average variance, and, in the beginning of the time horizon, it has even greater variance than the no-control case. This result is due to the increased delays incurred by the large queues formed at the beginning of the time horizon. The variance of travel times for the other two scenarios is more or less at the same level, with Scenario 1 having slightly larger average variance than Scenario 2. This result is due to the fact that a larger congestion is allowed to occur in the mainstream under Scenario 1 than under Scenario 2. The delays caused by this congestion result in higher variance of the travel times due to longer mainstream travel times for ramps affected by the congestion.

Under these terms, Figs. 9 and 10 illustrate also the partially competitive nature of equity and efficiency. Compared to no control, all control scenarios are substantially more efficient and

fair. Among the control scenarios, however, Scenario 3 is the most efficient achieving a 43.5% improvement of the TTS, but also the most unfair; scenario 1 is the most fair from the considered scenarios but at the cost of achieving a "mere" 31.7% reduction of the TTS; finally, Scenario 2 is in the middle of Scenarios 1 and 3 concerning efficiency and equity.

It should be noted that these results were not obtained from the explicit optimization of the travel time variance (22). From (21) and (22) it can be seen that the variance is a function of the problem's state variables, hence, this equity index could be optimized explicitly by including it as an additional weighted factor in the cost criterion (15).

D. Computational Effort

The computation time required to obtain the optimal solutions is moderate and depends upon the search method used and the specific parameters of each algorithm. It was found that for the search methods that use line optimization, a restart had to take place every four iterations for best convergence, while for RPROP the parameters used where $\eta^+ = 1.2, \eta^- = 0.5$, $\Delta_{\max} = 0.1$, and $\Delta_{\min} = 10^{-7}$. Fig. 11 depicts the cost criterion against the CPU time for various solution methods for Scenario 3. It can be seen that RPROP is much faster than the search direction methods which use line optimization. It may be seen that with RPROP, particularly near the optimum, the cost criterion value does not necessarily decrease monotonically, but the major part of the cost criterion improvement is typically achieved very fast. The computation time for the 4-h horizon is 20 min for the bulk of the 43.5% improvement (more than 40.5%) on a Sun Ultra5 with a Sparc IIi-360 MHz processor workstation. Furthermore, the inclusion of queue penalty terms does not have any significant effect on the RPROP algorithm's
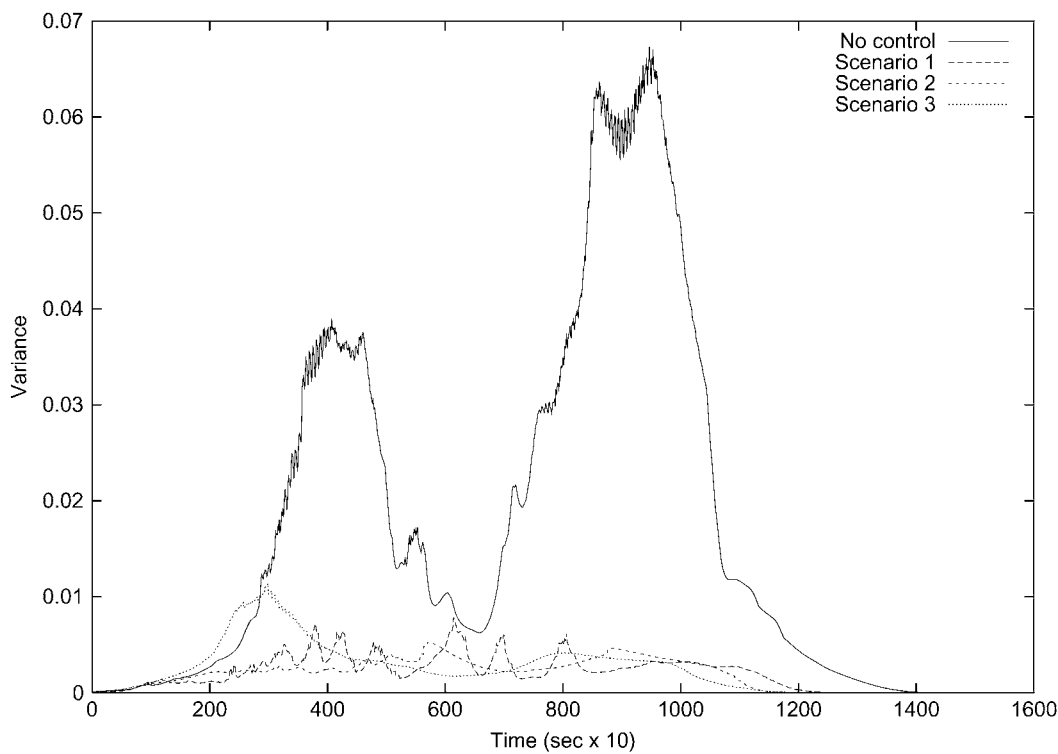
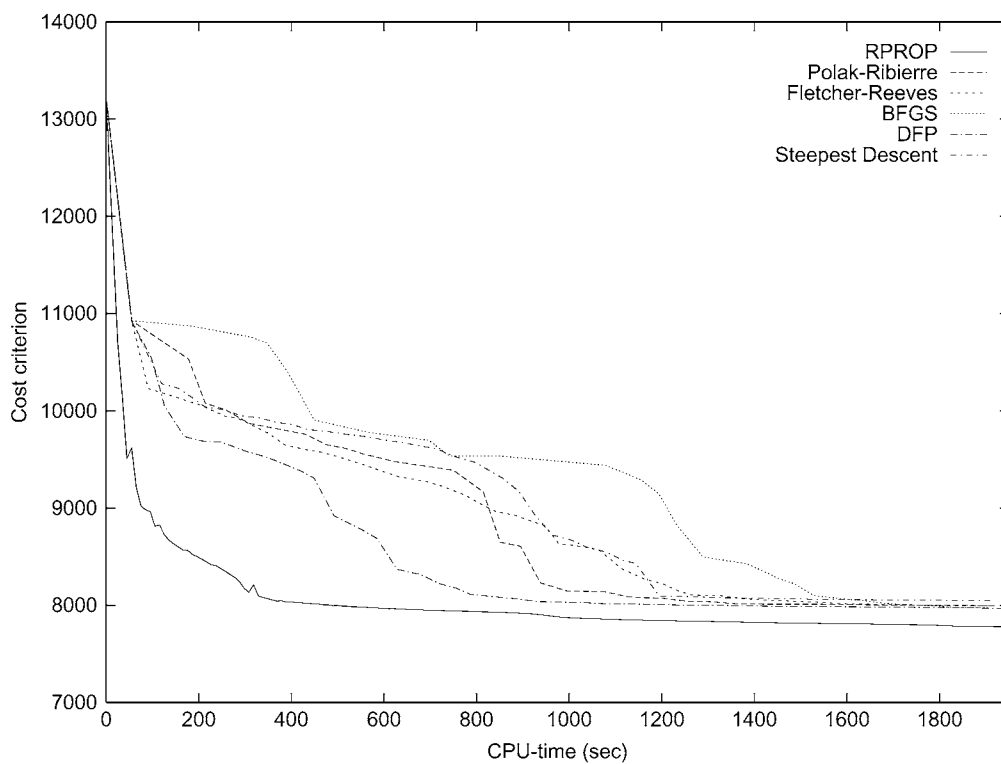Fig. 10. Variance of travel times $\mathrm{Var}_t(k)$.

Fig. 11. Cost criterion versus CPU time.

convergence, although the convergence speed becomes some-what slower. The queue penalty terms have a more important effect on the search methods that use line optimization by re-ducing their speed of convergence.

Finally, for different initial feasible trajectories, relatively dif-ferent optimal control trajectories have been observed in some cases, but the differences of the corresponding optimal traffic states was negligible, which means that there are local minima,

but they do not constitute a serius handicap for the practical efficiency of the proposed algorithm.

VI. Conclusion and Further Developments

The reported results demonstrate that the uncontrolled utilization of the motorway infrastructure strongly degrades the available infrastructure. An impressive amelioration of traffic conditions in motorway networks (including the ramps and motorway intersections) is possible with the use of optimal ramp metering by increasing the network throughput. The highest efficiency is achieved if only the on-ramps closest to the bottlenecks are strongly metered, which however creates long queues and disbenefits the corresponding users for the sake of the general efficiency. There is a tradeoff of efficiency versus equity which the optimal control strategy addresses implicitly via consideration of the available ramp storage space, something which may be used as a tool to establish a desired policy.

The formulation of the problem of coordinated ramp metering control as a discrete-time optimal control problem, allows the application of well known concepts from automatic control theory, and allows the consideration of other control measures as well. Further control measures such as speed control and route guidance may be readily integrated cooperatively due to the flexible nature of the problem formulation, see [13], [27] and [11], [28], respectively. The control trajectories obtained may be used as strategic decision in the sense of providing optimal and fair set values over a long time horizon (e.g., 4 h) for subordinate reactive ramp metering, using, e.g., local ramp metering control strategies such as ALINEA (see [29]). This strategic role can be further enhanced by use of a rolling horizon framework whereby the optimal control problem is solved repeatedly in real time, with updated initial state, demand predictions, and turning rates, as well as with inclusion of possible incidents.

Finally, it has been demonstrated that the RPROP search direction method is superior to the search methods that use line optimization for this particular problem.

Acknowledgment

References

[1] M. Blinkin, "Problem of optimal control of traffic flow on highways," *Automat. Remote Control*, vol. 37, pp. 662–667, 1976.

[2] M. Papageorgiou and R. Mayr, "Optimal decomposition methods applied to motorway traffic control," *Int. J. Control*, vol. 35, pp. 269–280, 1982.

[3] M. Papageorgiou, *Application of Automatic Control Concepts in Traffic Flow Modeling and Control*. New York: Springer-Verlag, 1983.

[4] N. Bhouri, M. Papageorgiou, and J. M. Blosseville, "Optimal control of traffic flow on periurban ringways with application to the Boulevard Périphérique in Paris," in *Proc. 11th IFAC World Congress*, vol. 10, Tallinn, Estonia, 1990, pp. 236–243.

[5] N. Bhouri, "Commande d'un système de traffic autoroutier: application au Boulevard Périphérique de Paris," Ph.D. dissertation, Univ. Paris-Sud, Centre d'Orsay, Paris, France, 1991.

[6] Y. Stephanedes and K. K. Chang, "Optimal control of freeway corridors," *ASCE J. Transportat. Eng.*, vol. 119, pp. 504–514, 1993.

[7] H. Zhang, S. Ritchie, and W. Recker, "Some general results on the optimal ramp metering control problem," *Transportat. Res. C*, vol. 4, pp. 51–69, 1996.

[8] O. Chen, A. Hotz, and M. Ben-Akiva, "Development and evaluation of a dynamic metering control model," in *Proc. 8th IFAC/IFIP/IFORS Symp. Transportation Systems*, M. Papageorgiou and A. Pouliezos, Eds., Chania, Greece, 1997, pp. 1162–1168.

[9] H. M. Zhang and W. W. Recker, "On optimal freeway ramp control policies for congested traffic corridors," *Transportat. Res. B*, vol. 33, no. 6, pp. 417–436, 1999.

[10] A. Kotsialos, M. Papageorgiou, and A. Messmer, "Optimal coordinated and integrated motorway network traffic control," in *Proc. 14th Int. Symp. Transportation Trafic Theory (ISTTT)*, A. Ceder, Ed., Jerusalem, 1999, pp. 621–644.

[11] A. Kotsialos, M. Papageorgiou, M. Mangeas, and H. Haj-Salem, "Coordinated and integrated control of motorway networks via nonlinear optimal control," *Transportat. Res. C*, vol. 10, no. 1, pp. 65–84, 2002.

[12] A. Kotsialos, M. Papageorgiou, and F. Middelham, "Optimal coordinated ramp metering with AMOC," in *Proc. 80th Transportation Research Board Annu. Meeting*, Washington, DC, 2001, Paper No. 01-3125.

[13] A. Alessandri, A. Di Febbaro, A. Ferrara, and E. Punta, "Optimal control of freeways via speed signalling and ramp metering," *Control Eng. Practice*, vol. 6, pp. 771–780, 1998.

[14] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: An overview," *IEEE Trans. Intell. Transport. Syst.*, vol. 3, pp. 271–281, Dec. 2002.

[15] A. Messmer and M. Papageorgiou, "METANET: A macroscopic simulation program for motorway networks," *Traffic Eng. Control*, vol. 31, no. 8/9, pp. 466–470, 549, 1990.

[16] M. Papageorgiou, J. M. Blosseville, and H. Hadj-Salem, "Modeling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris. Part I: Modeling," *Transportat. Res. A*, vol. 24, pp. 345–359, 1990.

[17] M. Papageorgiou, J. M. Blosseville, and H. Hadj-Salem, "Modeling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris. Part II: Coordinated on-ramp metering," *Transportat. Res. A*, vol. 24, pp. 361–370, 1990.

[18] Y. Wang, M. Papageorgiou, and A. Messmer, "RENAISSANCE: A real-time motorway network traffic surveillance tool," in *Proc. 10th IFAC Symp. Control Transportation Systems*, Tokyo, Japan, 2003, pp. 235–240.

[19] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. New York: Wiley, 2000.

[20] M. Papageorgiou, *Optimierung*, 2nd ed. Munich, Germany: Oldenbourg Verlag, 1996.

[21] M. Papageorgiou and M. Marinaki, "A feasible direction algorithm for the numerical solution of optimal control problems," Dynamic Syst. Simulation Lab., Tech. Univ. Crete, Chania, Greece, 1995.

[22] M. Riedmiller and H. Braun, "A direct adaptive method for faster back-propagation learning: The RPROP algorithm," in *Proc. IEEE Int. Conf. Neural Networks (ICNN)*, H. Ruspini, Ed., San Francisco, CA, 1993, pp. 586–591.

[23] A. Kotsialos, Y. Pavlis, F. Middelham, C. Diakaki, G. Vardaka, and M. Papageorgiou, "Modeling of the large scale motorway network around Amsterdam," in *Proc. 8th IFAC Symp. Large Scale Systems Theory Applications*, Patras, Greece, 1998, pp. 354–360.

[24] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, "Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool METANET," *IEEE Trans. Intell. Transport. Syst.*, vol. 3, pp. 282–292, Dec. 2002.

[25] L. Benmohamed and S. M. Meerkov, "Feedback control of highway congestion be a fair on-ramp metering," in *Proc. 33rd IEEE Conf. Decision Control*, vol. 3, Lake Buena Vista, FL, 1994, pp. 2437–2442.

[26] A. Kotsialos and M. Papageorgiou, "Efficiency versus fairness in network-wide ramp metering," in *Proc. IEEE 4th Int. Conf. Intelligent Transportation Systems*, Oakland, CA, 2001, pp. 1190–1195.

[27] A. Hegyi, B. De Schutter, and J. Hellendoorn, "MPC-based optimal coordination of variable speed limits to suppress shock waves in freeway traffic," *Proc. 2003 Amer. Control Conf.*, 2003.

[28] T. Bellemans, "Traffic control on motorways," Ph.D. dissertation, Katholieke Univ. Leuven, Leuven, Belgium, May 2003.

[29] M. Papageorgiou, H. Haj-Salem, and J. M. Blosseville, "ALINEA: A local feedback control law for on-ramp metering," *Transportat. Res. Record*, vol. 1320, pp. 58–64, 1991.

Apostolos Kotsialos was born in Larissa, Greece, in 1972. He received the Dipl.-Eng. and M.Sc. degrees in production and management engineering from the Technical University of Crete, Greece, in 1995 and 1998, respectively. He is currently working toward the Ph.D. degree at the Dynamic Systems and Simulation Laboratory (DSSL), Technical University of Crete.

Since 1995, he has worked as a Research Associate at DSSL, where he has been involved in numerous research projects. His main research interests include traffic control of large-scale freeway networks, numerical optimization and its applications.

Markos Papageorgiou (M'82–SM'90–F'99) received the Dipl. Ing. and Dr. Ing. (honors) degrees in electrical engineering from the Technical University of Munich, Germany, in 1976 and 1981, respectively.

He is currently a Professor and Director of the Dynamic Systems and Simulation Laboratory, Technical University of Crete, Chania, Greece. From 1988 to 1994, he was a Professor of Automation at the Technical University of Munich. He is the author of the books *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control* (Berlin, Germany: Springer, 1983) and *Optimierung* (Munich, Germany: R. Oldenbourg, 1991; 1996), and the editor of the *Concise Encyclopedia of Traffic and Transportation Systems* (Oxford, U.K.: Pergamon Press, 1991). His research interests include automatic control, optimization, and their application to traffic and transportation systems, water networks and further areas.

Dr. Papageorgiou is an Associate Editor of *Transportation Research-Part C* and of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and Chairman of the IFAC Technical Committee on Transportation Systems.