Nonminimal state space approach to multivariable ramp metering control of motorway bottlenecks

C.J.Taylor
P.C.Young
A.Chotai
J.Whittaker

Indexing terms: Motorway traffic flow, Automatic control, Ramp metering, Nonminimal state space, Proportional integral plus, Linear quadratic

Abstract: The paper discusses the automatic control of motorway traffic flows utilising ramp metering, i.e. traffic lights on the on-ramp entrances. A multivariable ramp metering system is developed, based on the nonminimal state space (NMSS) approach to control system design using adaptive proportional-integral-plus, linear quadratic (PIP–LQ) optimal controllers. The controller is evaluated on a nonlinear statistical traffic model (STM) simulation of the Amsterdam motorway ring road near the Coen Tunnel.

1 Introduction

The electronics revolution has led to enormous advances in transportation technology. In particular, communication advancements, coupled with induction loops implanted in road surfaces, allow data on traffic behaviour at distant points in a large network to be utilised in real time. Automatic data processing techniques together with intelligent systems for effectively utilising the information are now required. In this regard, a dynamic state space modelling technique, which provides transport management with a 'data assimilation' tool for monitoring and prediction of interurban motorway networks, has recently been developed [1]. Although developed primarily to provide short term forecasts, the resulting statistical traffic model (STM) also provides a good simulation model of the traffic system suitable for investigating new control measures.

During congestion, mean velocities fall dramatically and traffic flow becomes lower than the capacity of the road. In this case, the motorway is not utilised in the most efficient manner and overall travel times and queue lengths are higher than their theoretically optimum values. Ramp metering utilises traffic lights at the on-ramp entrances to limit access to the main carriageway of the motorway in an attempt to maintain flow close to the capacity. Most commonly, an occupancy set point is employed. Occupancy, which is usually measured by induction loops embedded in the road surface, is defined as the proportion of time during which vehicles are above the loop and is linearly related to the density of traffic on the carriageway [[2], p. 141].

The proposed approach to ramp metering control is based on earlier research into nonminimal state space (NMSS) control system designs [3, 4]. This approach yields proportional-integral-plus (PIP) controllers which can be considered as logical successors to conventional two- and three-term controllers, but exploit model-based predictive control action [5]. There are numerous advantages to this more advanced approach: the vagaries of manual tuning are replaced by model-based computation of the control gains; its structure always exploits fully the power of state variable feedback (SVF) methods, thus allowing for the implementation of optimal control strategies, such as the PIP linear quadratic (PIP–LQ) controller discussed in this paper; and it does not rely on digitisation of a continuous-time design but works directly with discrete-time transfer function (TF) models identified from normal operational data.

The present paper, which follows up earlier research for the European Community DACCORD project [6], utilises STM simulations of the motorway network on the Amsterdam motorway ring road to evaluate the control designs. This ring road is a highly monitored network and is the focus of a broad traffic management programme in The Netherlands. At the Coen Tunnel, on the western part of the ring, there is often a traffic jam in the southern direction during the morning peak hour, and in the northern direction during the afternoon peak hour. Ramp metering has been applied at this location since 1989, when the first metering system in The Netherlands was installed. One particularly successful and well known approach, for example, is the ALINEA strategy [7], based on a simple *integral* feedback controller.

When ramp metering was first introduced near the Coen Tunnel, it improved the flow of traffic at this location. As might be expected, however, there was a corresponding increase in flow from upstream, unmetered (at that time) on-ramps. Clearly, by metering several on-ramps, the traffic can be more equally distributed over the entrances to the motorway, in principle resulting in fewer disturbances and smaller queues. This simple example illustrates the importance of coordination between different on-ramps in a region.

A degree of coordination may be obtained with essentially local controllers, such as ALINEA, by intro-

IEE Proceedings online no. 19982383
Paper first received 4th February 1998 and in revised form 6th May 1998
The authors are with the Centre for Research on Environmental Systems and Statistics, Lancaster University, UK

ducing simple rule-based decisions. However, a full multivariable approach, such as the linear quadratic integral (LQI) control law of Papageorgiou *et al.* [8], or the related PIP–LQ method discussed in the present paper, provides a mechanism for inherent coordination. The LQI controller, which has been available for many years [9], is a standard LQ control law, with the addition of an integral control component to ensure zero steady state offset, even in the presence of output disturbances to the system. As Papageorgiou *et al.* emphasise, such an optimal approach provides a powerful design tool which can be applied directly to nonlinear systems, such as traffic systems, if they can be adequately linearised around a desired steady state. In the present research, such linearisation is achieved through the use of data-based identification and estimation techniques, where simple linear TF models are obtained by recursive estimation based on the telemetered traffic flow measurements.

2 The statistical traffic model (STM) and linearisation

The STM [1] includes the following components: (i) a state vector that holds two variables per link, namely traffic density (r, vehicles/km) and flow (q, vehicles/minute), ordered by the corresponding measurement station; (ii) an observation vector that holds information from the monitoring stations: (iii) a vector of exogenous inputs consisting of the flows from the on-ramp entrances; and (iv) variables that specify the behaviour of the system, such as measures of link length and capacity. The state transition matrix depends on traffic dynamics, which classically follow differential equations analogous to those of fluid dynamics.

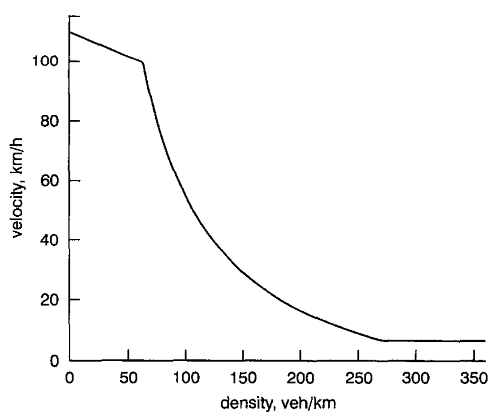

Fig. 1 *Typical fundamental traffic diagram utilised in the solution*

In particular, the three types of equations which comprise the model are as follows: a balance equation, which ensures that vehicles are conserved; a flow equation based on the identity $q = v \cdot r$ where v is the velocity (km/hour); and a highly nonlinear, fundamental traffic diagram (FTD) relating the velocity to the density. The FTD takes the shape illustrated in Fig. 1, where it is clear that during 'free flow' conditions at low densities, the corresponding velocities are high, as would be expected. However, once the density exceeds a critical level, linear flow conditions end and the velocity falls dramatically due to developing congestion. It is this critical density, defining the capacity of the road, that is of most importance when defining the set point for control design. The STM equations hold for on-ramps (or merging motorways) and off-ramps (or diverging motorways), with the latter requiring the estimation of a *turning fraction* between the two downstream choices. The network is built up from a number of such links, e.g. for sections of motorway between two adjacent measurement stations (typically 0.5km).

Although in previous publications the STM has been used as a basis for data assimilation using the Kalman Filter [1], the present paper utilises the STM purely as a simulation tool. For this purpose, the STM parameters are selected to represent particular flow conditions of interest, such as free flow, congestion or lane closure, with realistic upstream flows, on-ramp demands and turning fractions introduced to ensure authentic traffic behaviour. For example, in order to bring the results as close to real traffic conditions as possible, measured data are employed as inputs to the system (i.e. defining the boundary conditions). When used for simulation in this manner, the STM yields traffic behavioural patterns similar to those encountered on several road networks in The Netherlands [1].

In its basic form, the high order STM is much too complex for use in analytical control system design. In order to develop a linear optimal control algorithm, a linearised, *small perturbation*, representation of the traffic system is required and this is normally obtained by conventional linearisation of the nonlinear model equations. We have found, however, that the behaviour of the nonlinear traffic system can be approximated well by simple, reduced order linearised time variable parameter (TVP) transfer function models (see eqn. 1 in Section 3) that are identified and recursively estimated *directly* from simulated or real data using the simplified refined instrumental variable (SRIV) algorithm [10]. Such linearised models may be obtained for any of the macroscopic traffic variables measured on motorways, including flow, velocity and occupancy. This novel approach to combined linearisation and model reduction is particularly valuable because it not only identifies directly the range of applicability of the linearised model but it also facilitates the subsequent design of an adaptive controller using online recursive estimation of the model parameters [4].

Taylor *et al.* [6] successfully apply these modelling techniques to actual telemetered traffic data from motorways in The Netherlands and discuss how the parameters change over time according to current levels of congestion. In fact, it was found that the model fit is the strongest during the most congested part of the data set. This result is worth stressing: it suggests that the linear model most adequately describes the data when traffic control is most needed.

3 Nonminimal state space (NMSS) control design: the PIP–LQ controller

As an illustration of the proposed control methodology, we consider the coordinated ramp metering of northbound traffic on the Amsterdam ring road, south of the Coen tunnel. The algorithm is implemented on an STM simulation of the corridor, assuming that measurements are available from each of the eight STM links, as in Fig. 2. Note that the STM is based on density variables rather than occupancy, but since the two are linearly related the algorithm may be developed in terms of occupancy without any additional complications. As the system illustrated in Fig. 2 consists of four on-ramps ($U1$–$U4$), an associated four locations may be controlled using SVF methods. These are selected as the 'bottlenecks' immediately following the sections where the on-ramps merge with the motorway (i.e. $L2$, $L4$, $L6$ and $L8$).

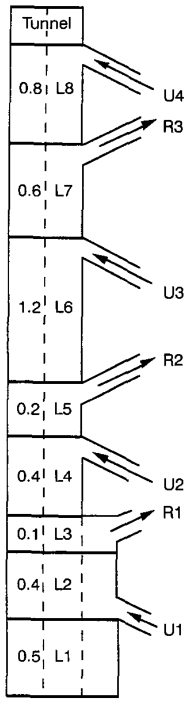

Fig. 2 *Schematic diagram of the STM simulation of the Amsterdam A10 West showing the relative locations of the links (L1–L8), on-ramps (U1–U4) and off-ramps (R1–R3)*
the numbers on the left indicate the link lengths in kilometres

Although most of the corridor consists of two lanes of traffic, at the upstream end of Fig. 2 the links $L1$ to $L3$ have three lanes; while the on-ramps have one lane of traffic, except for $U3$ which has two. The ramp metering system is intended to allow one car to pass the traffic lights for each green period, usually with a cycle time in the range 4.5 to 12 seconds, yielding on-ramp flows of 5 to 13.33 veh/min.

3.1 State space representation of the system

The local linear model for occupancy at location s takes the following general form:

$$o_{s,t} = a_s o_{s,t-1} + b_s o_{s-1,t-1} + c_s o_{s+1,t-1} + \{d_s u_{s,t-1}\} \tag{1}$$

where $o_{s-1,t}$ and $o_{s+1,t}$ are the occupancies at locations immediately upstream and downstream of $o_{s,t}$; $u_{s,t}$ is the *flow* from the associated on-ramp (if one exists); and a_s, b_s, c_s, d_s are the time variable parameters. In the notation of eqn. 1, $o_{2,t}$ is the occupancy at link $L2$ after the on-ramp, while $u_{2,t}$ is the on-ramp flow $U1$ at that location. Of course, if the data-based identification analysis suggests that a higher order model than eqn. 1 is appropriate, this introduces no problems since the PIP control system design methods can be applied to TF models of any order, with any number of sampled time delays [3–5]. However, analysis of telemetered data from The Netherlands suggest that higher order models are not necessary in this case [6].

The associated state space system is described by the following nonminimal state space (NMSS) equations that are derived directly from the linear model (eqn. 1) and its associated parameter estimates:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_{t-1} + \mathbf{D}\mathbf{y}_{d,t} \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t \end{aligned} \tag{2}$$

Here $\mathbf{x}_t$ is the nonminimal state vector composed only of the measured variables which, in the case of the A10 West, takes the following form,

$$\mathbf{x}_t = \begin{bmatrix} o_{1,t} & o_{2,t} & o_{3,t} & o_{4,t} & o_{5,t} & o_{6,t} \\ & o_{7,t} & o_{8,t} & z_{2,t} & z_{4,t} & z_{6,t} & z_{8,t} \end{bmatrix}^T \tag{3}$$

in which $z_{s,t}$, the *integral of error* between the occupancy at location s $\{s = 2, 4, 6, 8\}$ and the set point or command input for that location $d_{s,t}$, is defined as follows:

$$z_{s,t} = z_{s,t-1} + \{d_{s,t} - o_{s,t}\} \tag{4}$$

The inclusion of $z_{i,t}$ terms in this manner ensures inherent type-one servomechanism performance; i.e. the state variable in question (selected by means of the matrix $\mathbf{D}$) will always achieve its desired set point in the steady equilibrium state, provided the closed–loop system is stable and there are no input constraints. In the multivariable or 'coordinated' situation, this also ensures steady state decoupling.

In the state equations (eqn. 2), the vectors of controlled outputs and on-ramp inputs are, respectively

$$\begin{aligned} \mathbf{y}_t &= [o_{2,t} \quad o_{4,t} \quad o_{6,t} \quad o_{8,t}]^T \\ \mathbf{u}_t &= [u_{2,t} \quad u_{4,t} \quad u_{6,t} \quad u_{8,t}]^T \end{aligned} \tag{5}$$

while the vector of set points takes the following form:

$$\mathbf{y}_{d,t} = [d_{2,t} \quad d_{4,t} \quad d_{6,t} \quad d_{8,t}]^T \tag{6}$$

Note that the formulation does not include the off-ramps in the state space model since the controller is designed to be robust to such disturbances. If necessary, however, these can be included as measured feed-forward variables in the NMSS form. Also, it is clear that, although the state matrices (eqn. 7) are specific to the network structure illustrated in Fig. 2, any modifications needed to handle 'missing' section measurements or other motorway structures are straightforward and require only the ability to identify appropriate linear models.

Finally, the NMSS matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ are defined below.

$$\mathbf{A} = \begin{bmatrix} a_1 & c_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_2 & a_2 & c_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_3 & a_3 & c_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_4 & a_4 & c_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_5 & a_5 & c_5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_6 & a_6 & c_6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & b_7 & a_7 & c_7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_8 & a_8 & 0 & 0 & 0 & 0 \\ -b_2 & -a_2 & -c_2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -b_4 & -a_4 & -c_4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -b_6 & -a_6 & -c_6 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -b_8 & -a_8 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$
\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ d_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & d_4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & d_6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_8 \\ -d_2 & 0 & 0 & 0 \\ 0 & -d_4 & 0 & 0 \\ 0 & 0 & -d_6 & 0 \\ 0 & 0 & 0 & -d_8 \end{bmatrix}
$$

$$
\mathbf{C}^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}
$$

3.2 The PIP–LQ control algorithm and practical implementation

Having established a suitable NMSS model of the system, the appropriate gains defining the control law can be determined by optimisation of a standard linear quadratic (LQ) cost function in the form,

$$
J = \sum_{t=1}^{t=\infty} \{\mathbf{x}_t^T \mathbf{Q} \mathbf{x}_t + \mathbf{u}_t^T \mathbf{R} \mathbf{u}_t\} \tag{8}
$$

where $\mathbf{Q}$ and $\mathbf{R}$ are symmetric positive semidefinite and symmetric positive definite weighting matrices, respectively. The SVF control gains are obtained from the steady state solution of the well known discrete time Riccati equation, with the control law then defined as follows in conventional SVF form:

$$
\mathbf{u}_t = -\mathbf{K}^T \mathbf{x}_t \tag{9}
$$

where $\mathbf{K}$ is the SVF control gain matrix (of dimension 4 by 12 in this case). However, the PIP control law is always implemented in its equivalent incremental form where, in this case,

$$
\mathbf{u}_t = \mathbf{u}_{t-1} - \mathbf{F}\{\mathbf{O}_t - \mathbf{O}_{t-1}\} + \mathbf{K}_I \{\mathbf{y}_{d,t} - \mathbf{y}_t\} \tag{10}
$$

Here, $\mathbf{O}_t$ is a vector of occupancy measurements (the first eight states), $\mathbf{F}$ is a 4 by 8 feedback matrix and $\mathbf{K}_I$ is a 4 by 4 matrix of integral gains. eqn. 10 is not only the most obvious and convenient form for digital implementation, but it also provides an inherent means of avoiding 'integral windup' when the controller is subject to constraints on the actuator signal. In this regard, the following correction is utilised for each input variable:

$$
\begin{cases} \text{if } u_{i,t} \geq \max(u_{i,t}), & u_{i,t} = \max(u_{i,t}) \\ \text{if } u_{i,t} \leq \min(u_{i,t}), & u_{i,t} = \min(u_{i,t}) \end{cases} \quad i = 2, 4, 6, 8
$$

where $\max(u_{i,t})$ and $\min(u_{i,t})$ are the maximum and minimum flow rates allowed at each location. This is a standard approach to the problem of integral wind-up and has been evaluated in practice on many real systems [7, 11].

Note that the results obtained in Section 4 (below) have been obtained with a fixed parameter PIP control law of the above kind (eqn. 9). For all of the simulated traffic conditions investigated so far, this has yielded satisfactory regulation of the traffic flows. However, the recursive SRIV estimation results suggest that the parameters of the linear model may change significantly when extreme motorway conditions are encountered (see [6]). In such situations, it is advantageous to implement an adaptive PIP controller in which the Riccati equation and the resulting control gain matrix are both updated at each sample on the basis of the latest recursive SRIV estimates of the linearised model parameters (see [4]).

In practice, the final stage in the algorithmic design is the translation of the calculated $\mathbf{u}_t$ into a practical control input to the system: namely a traffic light sequence. Most ramp metering systems, such as those utilised in The Netherlands, are designed to allow one vehicle onto the motorway at a time, so that the control input signal is simply inverted to obtain the required 'seconds per vehicle' between each green phase (the cycling time). However, since the inputs to the STM are flow signals, this step is not required for the simulation study.

It is worth noting that, due to the special structure of the non-minimal state vector, the LQ weighting matrices in the cost function (eqn. 8) have particularly simple interpretation, since the diagonal elements directly define weights assigned to the measured variables and integral of error states. In the general multivariable case, however, these matrices are not readily translated into exact closed-loop response characteristics, such as decoupling, overshoot and rise times. In this regard, one recently developed technique that automatically maps the various technical characteristics into the elements of $\mathbf{Q}$ and $\mathbf{R}$, is multi-objective optimisation in its goal attainment form [12]. Rather than directly optimising the control *gains* themselves, this method optimises the parameters of the Cholesky factors of $\mathbf{Q}$, which has the advantage of generating only guaranteed stable optimal solutions. When applied to the problem of traffic control, this multi-objective approach may be utilised to obtain PIP–LQ controllers that are robust to a specified range of traffic conditions on the basis of repeated simulations with the STM.

4 Simulation experiments

In the simulation experiments used to evaluate the optimal PIP–LQ ramp metering control system designs, the STM is solved in MATLAB/SIMULINK™ at a 10 second sampling interval with a realistic set of traffic inputs. Here, measured flow from the A10 motorway at the appropriate location south of the Coen Tunnel (15 May 1995) is employed as the upstream input; while the on-ramp inputs are fully constrained within the saturations currently employed on the real system. In addition, a realistic set of on-ramp demands are used in the simulation, so that the final on-ramp flow is the minimum of the calculated control input and the flow actually approaching the on-ramp.

4.1 No on-ramp control

A typical simulation of the STM with the above inputs and without any on-ramp control is shown in Figs. 3

and 4. Since this simulation illustrates the 'uncontrolled' response of the traffic system from 2pm until 8:45pm, it includes the afternoon 'rush hour', which occurs during the middle part of the simulation, as can be seen quite clearly in the on-ramp flows of Fig. 4. It is also clear that, while no congestion occurs in link $L2$, the remaining links become severely congested, as represented by the dramatic increase in density illustrated in Fig. 3. In fact, the congestion is so severe that the associated velocities for links $L6$ and $L8$ remain below 20km/h for most of the time and, in practice, the vehicles may become stationary.

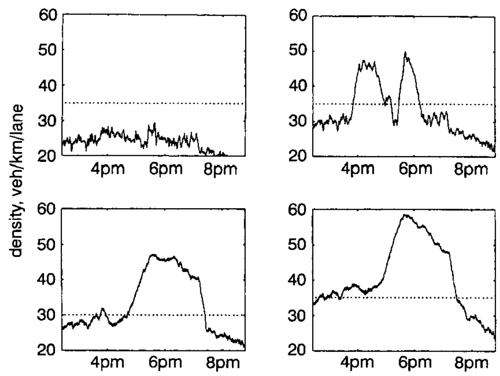

Fig. 3 *Uncontrolled density at locations L2, L4, L6 and L8*
(top left, top right, bottom left, bottom right, respectively). Set point used in Fig. 5 (dotted)

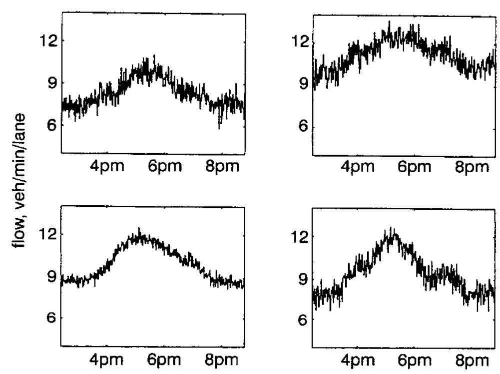

Fig. 4 *Uncontrolled on-ramp flow at locations U1, U2, U3 and U4*
(top left, top right, bottom left, bottom right, respectively)

4.2 Unrealistic unconstrained on-ramp control

As an initial evaluation step, the PIP–LQ controller is first simulated without any constraints applied to the input signals. Furthermore, it is assumed that the specified control input signals can be exactly achieved, i.e. there are always vehicles waiting at the on-ramp, ready to be released when the controller demands this. In this highly simplified and unrealistic situation, the multivariable PIP–LQ control algorithm successfully maintains the specified densities and, in fact, is able to completely decouple the densities at each of the four bottlenecks, showing the efficacy of the controller when used in an unconstrained situation. Such perfect performance is not, of course, possible in the real world: in particular, on-ramp $U2$ requires negative flows, while the flow at $U1$ greatly exceeds the demand actually waiting to join the motorway. The next stage in the design exercise, therefore, is to evolve a modified and realistic PIP–LQ strategy which maintains smooth, uncongested traffic flows on the motorway, while avoiding congestion on the approach roads.

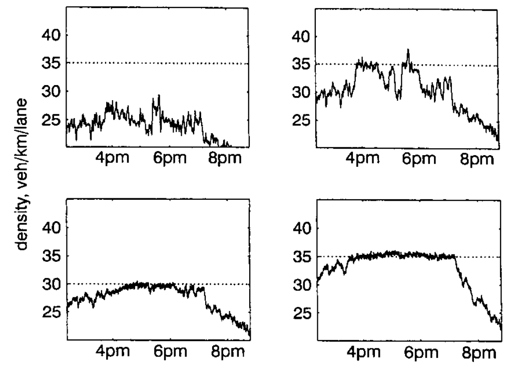

Fig. 5 *Density for a simulation with input constraints at locations L2, L4, L6 and L8*
(top left, top right, bottom left, bottom right, respectively). Set point (dotted)

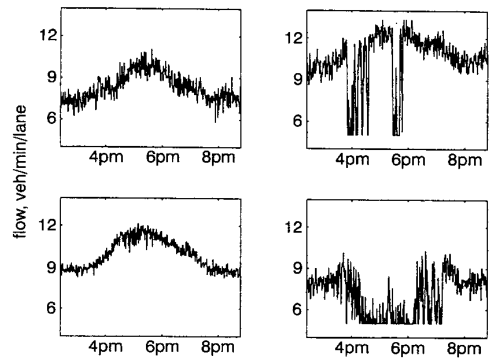

Fig. 6 *On-ramp flow for a simulation with input constraints at locations U1, U2, U3 and U4*
(top left, top right, bottom left, bottom right, respectively)

4.3 Realistic constrained PIP–LQ control

Figs. 5 and 6 illustrate the performance of the PIP–LQ controller when input saturations and the realistic on-ramp demands are taken into account. In this case the **Q** and **R** weighting matrices are simply left as identity matrices and the set points at the four locations are chosen according to the capacity of the road under these conditions (dotted traces in Fig. 5). In this constrained implementation of the controller, the final on-ramp flow is the minimum of the calculated control input and the flow actually approaching the on-ramp. For example, the controller selects the maximum on-ramp flow (13.33 veh/min/lane) for $U1$, since the density at $L2$ is below the set point for the entire simulation. In this case, however, the actual demand at this on-ramp is somewhat lower (initially about 8 veh/min/lane, rising to 10–11 veh/min/lane during the middle of the simulation) and it is this lower figure that is actually utilised as the on-ramp flow in the simulation. By contrast, at on-ramps $U2$ and $U4$, the input is often constrained at its minimum value (5 veh/min/lane).

Despite these input constraints, the multivariable control algorithm is able to maintain good regulation of the density at the four bottlenecks and it clearly prevents the congestion seen in Figs. 3 and 4. Of course, when the demands on the network are low during the early afternoon and evening periods, there is no need to restrict entry onto the main carriageway. Here, the densities fall well below the set points so the control algorithm specifies the maximum allowed input signals.

In this case, therefore, the on-ramp inputs depend on the actual demand approaching the motorway from the urban network.

It is worth noting that when the on-ramp inputs are operating within their allowable limits, the control is very good indeed. For example, the PIP–LQ controller maintains tight control of the density close to the Coen tunnel at $L6$ and $L8$. Similarly, the algorithm reacts quickly to deal with the sudden large peaks in the density at $L4$, automatically engaging and disengaging the ramp metering as required (see Fig. 6).

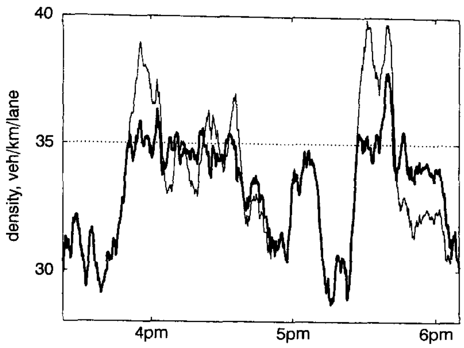

Fig. 7 *PIP–LQ control (——) compared to local integral control (—) at location L4*

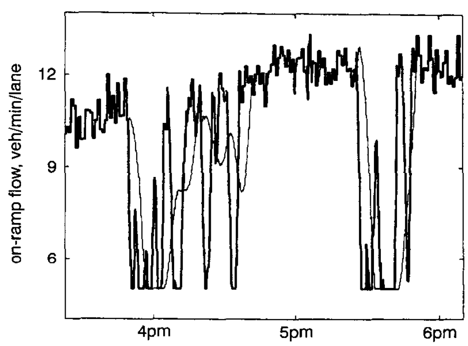

Fig. 8 *PIP–LQ control (——) compared to local integral control (——) at location U4*

4.4 Comparison with local integral control

In Figs. 7 and 8, the performance of the multivariable PIP–LQ design is compared with one of the approaches currently being employed on the real network; namely local integral control at each location in the corridor, with the integral control gain tuned manually to obtain the best performance. Note that the downstream on-ramp ($U4$) is saturated for much of the time during the simulation, as illustrated in Fig. 8. For example, during the period 5:35pm to 5:45pm, the flow at $U4$ is held at the minimum allowed value of 5 veh/min/lane, explaining the overshoot of the set point during this part of the day for both the PIP–LQ and especially for the integral controller. However, the superiority of the multivariable design is clear: as would be expected, if the gain on the local integral controller is too low then the response to disturbances is poor, as in Fig. 7; while if the gain is too high, then the input signal tends to be very volatile and there is a danger of instability. The advantage of the PIP–LQ approach is that it is able to achieve rapid response, without these problems.

4.5 Other possibilities

Further evaluation of the constrained PIP–LQ control system reveals a potential problem with the controller in its present form; namely that certain traffic conditions can arise such that 'poorly chosen' set points are not achievable in practice. For example, if the set point at $L6$ in Fig. 5 is increased from 30 to 35 veh/km/lane, then the flow to the downstream links is increased to a point where the density at $L8$ exceeds the set point during the most congested part of the day. The input $U4$ is unable to prevent this offset since it becomes saturated at its minimum allowed value (5 veh/min/lane); while the coordination between locations does not yield the necessary reduction in the input from the next upstream on-ramp $U3$. This limitation arises because, although it is only the positive error that may lead to congestion on the main carriageway, LQ optimisation penalises a negative error (when the density is below the set point, as for $L6$ during the first 2 hours of the simulation) to the same degree as a positive error. This is simply because the PIP–LQ controller is basically a linear design, albeit modified to handle the practical constraints.

One solution to this kind of problem is to introduce a simple, rule-based, upper level controller into the system. Here, when an input is saturated at its minimum value for a specified period of time, the set point at the next upstream location is gradually ramped down. This reduces the density downstream because of the natural linkages in the motorway system. We have found that such a hierarchical controller, programmed to coordinate the set points at this second level, works very well in this example and is based on reasonable heuristics. Another approach which has proven effective is to simply modify the weighting matrices employed in the LQ optimisation. For example, the effect of greatly reducing the diagonal weighting terms associated with the *integral-of-error* states for $L2$, $L4$ and $L6$, is to force the control algorithm to concentrate on preventing congestion at $L8$. This is a reasonable aim in practice, since the greatest problems of congestion occur close to the Coen Tunnel, downstream of the four on-ramps. In effect, the controller is able to successfully maintain tight control at this location by automatically restricting entry to the upstream on-ramps $U1$, $U2$ and $U3$ when necessary.

This latter kind of redesign requires manual adjustment of the (usually diagonal) weighting matrices. A superior alternative, presently under investigation, is to apply multiple objective optimisation (see Section 3.2) and so determine whether improved solutions are possible by exploiting fully the off-diagonal elements of the weighting matrices. The authors are presently investigating ways of effectively utilising such multi-objective optimisation in the context of dynamic traffic management. In this regard, the next stage in the research will be to extend the STM model to provide global evaluation measures such as total travel times and fuel consumption.

Of course, all of the possibilities mentioned above entail modification of the *linear* PIP–LQ control system design to allow it to handle the constraints and nonlinearities of the real world. This is entirely consistent with previous on-ramp control system designs and should lead to further improvement in the multivariable performance. However, such developments should be considered as a prelude to the investigation of truly nonlinear PIP, or related model-based predictive controllers, which take explicit account of the nonlinearities and constraints at the design stage. Such approaches are currently receiving attention [13].

5 Conclusions

This paper has shown how it is possible to linearise the essential dynamics of the nonlinear traffic system and so represent a given region of the traffic network in a nonminimal state space (NMSS) form. This means that state variable feedback techniques based on this NMSS description, such as the proportional-integral-plus (PIP) controller, can be applied to the traffic control problem and so yield self-adaptive, on-ramp metering controllers which provide a considerable improvement over more conventional designs. Here, traffic lights on motorway on-ramps, operated by the PIP controller, limit access to the main carriageway, to prevent the density of traffic at key bottlenecks from exceeding critical values and so causing congestion.

Ramp metering represents a particularly difficult control problem because of the highly constrained, nonlinear traffic system. Furthermore, the success of any ramp metering methodology is ultimately measured in terms of the overall travel times, which depend on many factors in addition to the on-ramp flows. Consequently, the choice of the occupancy (or other) set point for the controller is of crucial importance. Offline calculation of a constant set point based on the fundamental traffic diagram, contrasts with more sophisticated approaches utilising online data, where the set points are obtained within a hierarchical system based on network wide optimisation, as discussed, for example, in [14]. The importance of the multivariable PIP controllers, within this wider context, is that the higher level objectives will only be effective if there is the possibility of very tight coordinated control at the local level.

Computer-based simulation models are a well established means of testing new approaches to traffic control before they are implemented on real networks. In this tradition, the optimised PIP–LQ designs discussed in this paper have been evaluated comprehensively on a detailed MATLAB/SIMULINK™ simulation of the Amsterdam ring road, south of the Coen Tunnel. The simulation results illustrate the effectiveness of the PIP–LQ ramp metering algorithms, both under conditions of recurrent congestion (caused by high demand exceeding the capacity of the motorway), and in reacting to traffic incidents such as lane closure. In particular, it is in the response to disturbances away from the immediate area of the on-ramp itself that the PIP algorithm is particularly useful, since it includes a coordinated feedback of all the available measurements on the corridor and so can anticipate the onset of congestion. In this sense, it provides a 'second generation' approach to ramp metering which provides one particular logical improvement on the mainly local controllers used heretofore.

6 Acknowledgments

The research described in this paper is partly funded by the European Community DACCORD project, TR1017 in the Fourth Framework Telematics Applications Programme, sector Transport. The authors are also grateful for the support of the UK Engineering and Physical Sciences Research Council (EPSRC), grant number GR/J10136.

7 References

1 WHITTAKER, J., GARSIDE, S., and LINDVELD, K.: 'Tracking and predicting a network traffic process', *Int. J. Forecasting*, 1997, **13**, pp. 51–61

2 PAPAGEORGIOU, M. (Ed.): 'Concise encyclopedia of traffic and transportation systems' (Pergamon Press, Oxford, 1991)

3 YOUNG, P.C., BEHZADI, M.A., WANG, C.-L., and CHOTAI, A.: 'Direct digital and adaptive control by input-output, state variable feedback pole assignment', *Int. J. Control*, 1987, **46**, pp. 1867–1881

4 CHOTAI, A., YOUNG, P.C., and BEHZADI, M.A.: 'Self-adaptive design of a nonlinear temperature control system', special issue on self-tuning control, *IEE Proc. D*, 1991, **38**, pp. 41–49

5 TAYLOR, C.J., YOUNG, P.C., and CHOTAI, A.: 'On the relationship between GPC and PIP control', *in* CLARKE, D.W. (Ed.): 'Advances in Model-Based Predictive Control' (Oxford University Press, Oxford, 1994), pp. 53–68

6 TAYLOR, J., WHITTAKER, J., YOUNG, P., and CHOTAI, A.: 'Forecasting and control of interurban traffic networks using a state space formulated traffic model', *in* PAPAGEORGIOU, M., and POULIEZOS, A. (Eds.): '8th IFAC symposium on Transportation systems, 16–18 June 1997, Chania, IFAC/IFIP/IFORS preprint', 1997 vol. 3, pp. 1054–1059

7 PAPAGEORGIOU, M., HAJ-SALEM, H., and BLOSSEVILLE, J.-M.: 'ALINEA: A local feedback control law for on-ramp metering', *Transp. Res. Rec.*, 1991, **1320**, pp. 58–64

8 PAPAGEORGIOU, M., BLOSSEVILLE, J.-M., and HADJ-SALEM, H.: 'Modelling and real time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part I: Modelling, Modelling and real time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part II: Coordinated on-ramp metering', *Transp. Res. A*, 1990, **24A**, (5), pp. 345–359, 361–370

9 YOUNG, P.C., and WILLEMS, J.C.: 'An approach to the linear multivariate servomechanism problem', *Int. J. Control*, 1972, **15**, pp. 961–976

10 YOUNG, P.C.: 'The instrumental variable method: a practical approach to identification and system parameter estimation', *in* BARKER, H.A., and YOUNG, P.C. (Eds.): 'Identification and System Parameter estimation' (Pergamon, Oxford, 1985), vol. 1, pp. 1–16

11 YOUNG, P.C., LEES, M., CHOTAI, A., TYCH, W., and CHALABI, Z.S.: 'Modelling and PIP control of a glasshouse micro-climate', *Control Eng. Pract.*, 1994, **2**, pp. 591–604

12 CHOTAI, A., YOUNG, P.C., MCKENNA, P.G., and TYCH, W.: 'Proportional integral-plus (PIP) design for delta (δ) operator systems: Part 2, MIMO systems', *Int. J. Control*, 1998, **70**, pp. 149–168

13 YOUNG, P.C.: 'A general approach to identification, estimation and control for a class of nonlinear dynamic systems', *in* FRISWELL, M.I., and MOTTERSHEAD, J.E. (Eds.): 'Identification in Engineering Systems' (University of Wales, Swansea, 1996), pp. 436–445

14 CHEN, O.J., HOTZ, A.F., BEN-AKIVA, M.: 'Development and evaluation of a dynamic ramp metering control model', *in* PAPAGEORGIOU, M., and POULIEZOS, A. (Eds.): '8th IFAC symposium on Transportation systems, 16–18 June 1997, Chania, IFAC/IFIP/IFORS preprint', 1997, vol. 3, pp. 1162–1167