ASME 2012 5th Annual Dynamic Systems and Control Conference joint with the JSME 2012 11th Motion and Vibration Conference
DSCC2012-MOVIC2012
October 17-19, 2012, Fort Lauderdale, Florida, USA

DSCC2012-MOVIC2012-8851

MODEL PREDICTIVE CONTROL OF A FREEWAY NETWORK WITH CAPACITY DROPS

Ajith Muralidharan *
Department of Mechanical Engineering
University of California, Berkeley
Berkeley, CA 94720
ajith@berkeley.edu

Roberto Horowitz
Department of Mechanical Engineering
University of California, Berkeley
Berkeley, CA 94720
horowitz@berkeley.edu

Pravin Varaiya
Department of Electrical Engineering
University of California, Berkeley
Berkeley, CA 94720
varaiya@eecs.berkeley.edu

ABSTRACT

In this paper, we present a model predictive controller to reduce road traffic congestion in freeway networks. The model predictive controller regulates traffic in the freeway through the use of ramp metering and variable speed limits. The controller uses a Link-Node Cell transmission model (LN-CTM) to represent freeway dynamics. We modify the standard LN-CTM to account for the capacity drop phenomenon, which is observed as a discontinuous decrease in flow throughput when traffic density exceeds a critical value. The resulting optimal control problem with a modified model, which accounts for the capacity flow phenomenon, is non-convex. We present heuristic restrictions on the solution trajectories, which allow us to solve the problem efficiently. This enables us to obtain the solution of the actual optimal control problem by solving a sequence of relaxed linear programs. We describe the procedure which can be used to map the optimal solution of this relaxed problem to the solution of the actual optimal control problem. Finally, we demonstrate the application of the model predictive controller on a simulated example, and discuss the characteristics of the controller.

1 Introduction

Traffic congestion in metropolitan areas has been increasing over the last decade, leading to large losses in productivity due to increased commute time. Congestion, both recurrent and non-recurrent, has been triggered by ever increasing demand. Due to significant investments involved, infrastructure expansions are not always feasible even though they provide the best means to tackle traffic congestion. As a result, transportation engineers rely on intelligent operational management of the existing infrastructure to increase system efficiency. The most commonly used operational management strategy is traffic control. For freeways, traffic control involves the use of ramp metering and variable speed control. In ramp metering, traffic entering the on-ramp is regulated, with the objective of allowing the freeway to operate at maximum efficiency. Speed control (variable speed limits) is primarily used for improving safety [1], but it has also been found to be useful for congestion alleviation when ramp metering is insufficient [2].

Traditionally, traffic control using ramp metering was carried out using simple feed-forward or feedback strategies (Eg. [3]). These commonly used ramp metering controllers independently operated on single on-ramps to perform local congestion control. Lack of co-ordination between local controllers limit the performance benefits that can be obtained from ramp metering. Recently, heuristic controllers [4], have been successfully deployed to show the benefits of co-ordinated ramp metering in networks. With improving detection coverage and the advent of accurate traffic modeling, model based traffic control has the potential to further improve ramp metering benefits. Particularly, optimal controllers which simultaneously specify ramp metering rates and variable speed limits can allow for efficient management of freeways.

Optimal control methods based on macroscopic models have been extensively reported in literature [2, 5–7]. For freeway networks, first order models (Cell Transmission model, CTM [8]) and second order models (METANET [5]) are commonly used to describe the traffic dynamics within these controllers. Second order models have an advantage over the first order models in incorporating the capacity drop. However, the optimization

*Address all correspondence to this author.

problems based on the second order models are non-linear, computationally intensive and the solutions obtained are usually only locally optimal. The former proves to be a drawback when the controller is embedded as a part of a model predictive framework, since this requires fast optimizations to be executed repeatedly [5].

Optimization approaches based on CTM hold more promise in terms of computational efficiency and global optimality. In [6], the authors present an optimal controller for coordinated on-ramp metering of a freeway network, which minimizes total travel time (TTT) and delay in a freeway network including the time spent on the on-ramp queues, based on the Asymmetric Cell Transmission model (ACTM), a simplified version of the CTM. The authors solve the optimal ramp metering problem using a linear program. Recently, we presented an optimal controller for coordinated on-ramp metering and variable speed limits in a freeway network, which also minimizes total travel time (TTT) and delay in a freeway network including the time spent on the on-ramp queues, based on the Link-Node Cell Transmission model (LN-CTM) [9]. The LN-CTM uses a more accurate model of link merges as compared to the ACTM. This makes the LN-CTM suitable for simulating on-ramp merges even when on-ramp inflows are appreciable (eg. freeway-freeway interconnections). The optimal control formulation for minimizing total travel time/ delay in a freeway network based on the LN-CTM model is non-linear and non-convex, and we presented a relaxation procedure which enabled us to solve the problem using an equivalent linear program. The solution of this linear program is used to derive the optimal control trajectories of the original problem. However, these models do not include the notion of capacity drop, which is one of the important causes of recurrent congestion in freeway networks.

In this paper, we extend the results presented in [9] to cover situations where we observe the capacity drop phenomenon. Capacity drop denotes the reduction in the (maximum) flow throughput of a freeway section when traffic density at the section increases beyond a known threshold. Capacity drop is usually observed in locations of geometric discontinuities like lane drops along the freeways. In order to model capacity drop within the LN-CTM, we propose the use of a discontinuous demand function. The optimal control problem based on this modified LN-CTM cannot be solved using the techniques presented in [9]. We identify restrictions under which the optimal control problem can be efficiently solved. First, we assume that the freeway can be divided into regions (consisting of a sequence of freeway sections). Only the last section (i.e. the most downstream link) of each region experiences capacity drop. We assume that each region will be controlled using a independent controller, which will coordinate the controllers operating on various sections belonging to the region. Secondly, we propose a simple technique to solve this optimal control problem for each region. We apply heuristic restrictions which poses additional constraints on the solutions/control trajectory. With this assumption, we show that the solution to the optimal control problem can be obtained using a sequence of linear programs.

The paper is organized as follows. We present the LN-CTM model in Section 2 and extend it to model capacity drop. We present the optimal control formulation based on this model in Section 3. We also present our solution technique to solve this optimal control problem. In Section 4, we present an MPC controller based on the developed framework, and demonstrate its application on a simulated scenario.

2 Link Node Cell transmission model

In this paper, we adopt a modified version of the Link-Node Cell Transmission model to describe the traffic dynamics. The Link-Node Cell Transmission Model (LN-CTM) is an extension of the CTM [8], which can be used to simulate traffic in any road network. This model is implemented as a part of the suite of Tools for Operations Planning (TOPL) [10] that is currently under development by our research team at the Partners for Advanced Transportation Technology (PATH program at U.C. Berkeley) in Aurora, a simulation software ([11]). The traffic network is represented as a directed graph of links connecting nodes in the LN-CTM. Figure 1 shows the directed graph representation of a freeway network. In this network, links represent road segments and belong to three types : (a) normal (b) source and (c) sink. Nodes are located at link junctions and they are specified with a (possibly) time-varying split-ratio matrix (routing parameters), which represents the portion of traffic moving from a particular input link to an output link. A fundamental diagram (which specifies the flow-speed-density relationship) is specified for each link, while the source links are also specified with an input demand profile. We assume that the simulation steps are chosen such that $0 < V_i, W_i < 1$. For typical freeways with minimum link length of $1500ft$, a time step of $10s$ is appropriate.

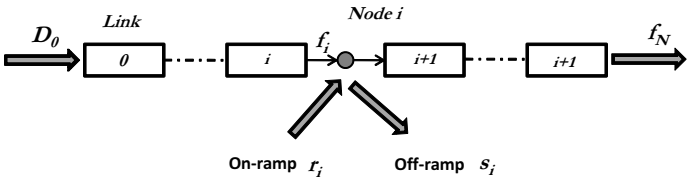

Figure 1. Freeway with N links. Each Node contains a maximum of one on- and one off-ramp. Note that Node i is upstream of Link i

The LN-CTM algorithm is explained in [11]. The model update is executed in two steps : (1) density update on each link (2) flow calculation in each node. At each node, the input (i.e. upstream) links specify demands (the vehicular flow that wants to leave the input link and reach the intended output link), while the output (i.e. downstream) links specify the supplies (the maximum flow that the output link can accept). Additionally, split

Symbol	Name	Unit
F_i	Flow capacity of Link i	veh/period
V_i	maximum free flow speed of Link i	section/period
W_i	congestion wave speed of Link i	section/period
n_i^c	critical density of Link i	section/period
n_i^J	jam density of Link i	veh/section
k	period number	dimensionless
$\beta_i(k)$	split ratio at node i	dimensionless
$f_i(k)$	flow out of Link i	veh/period
$n_i(k)$	number of vehicles in Link i	veh/period
$s_i(k), r_i(k)$	off-ramp, on-ramp flow in node i	veh/period
$d_i(k)$	demand for on-ramp i	veh/period
$l_i(k)$	queue for on-ramp i	veh/period
$c_i(k)$	Ramp metering rate for on-ramp i	veh/period
C_i	Flow capacity for on-ramp i	veh/period
L_i	Queue capacity for on-ramp i	veh/period
$Q_i(k)$	input flow for on-ramp i	veh/period
K	optimization time horizon	period

Table 1. Model variables and parameters.

ratios, which denote the portion of flows that travel from a given input link to another specified output link, is also given for each node. The flow update model uses the demands and split ratios to ascertain the total demand destined for each output link. If the total demand for any output link exceeds the supply, then all input demands are reduced, in order for the total outflow to match the supply. For each freeway node, the downstream link supply is compared with the total demands from the on-ramp joining the node and the flow from the previous link, and the final output flow is calculated. Once the input and output flows of each link are obtained from the node calculations, the density update is carried out using a conservation equation.

2.1 Nominal Model

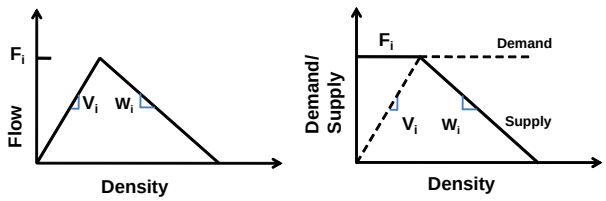

Figure 2. Fundamental Diagram, Demand function (dashed line) and the Supply function (solid line).

In the nominal model, the demand functions and the supply functions of the freeway link are derived from the fundamental diagram. Figure 2 shows the triangular fundamental diagram and the demand and supply functions. These functions are completely characterized by the maximum free-flow speed (V_i), congestion wave speed (W_i) and Capacity flow F_i. Under nominal conditions (without Variable Speed Limits (VSL)), the demand function $\bar{D}_i(n_i(k)) = \min(n_i(k)V_i, F_i)$ defines the number of vehicles available to move out of the link, while the supply function $S_i(n_i(k)) = \min(W_i(n_i^J - n_i(k)), F_i)$ defines the maximum number of vehicles that can flow into the link i. Variable speed limits only affect the demand function while the supply function is unchanged. The actual demand function under speed control is given by $D_i(n_i(k)) = \min(n_i(k)v_i(k), F_i) = \min(n_i(k)v_i(k), \bar{D}_i(n_i(k)))$, ie. speed control (or reduction) can only lead to a decrease in link throughput. For on-ramps, we implement a point queue model, and hence its demand function is $d_i(k) = \min(l_i(k), c_i(k))$ and it accepts any input flow (i.e. it has infinite supply).

2.2 Capacity Drop Model

In any freeway section, congestion originates at bottlenecks and propagate upstream. A (latent) bottleneck exists when the capacity of the link upstream exceeds the capacity of a link downstream. Natural bottlenecks can occur due to lane drops, ramp merges and also less typically in graded locations, turns etc. Bottlenecks are said to be activated when the demand feeding into the bottleneck section (link) exceeds the capacity downstream. As a result, vehicle buildup occurs in the link and congestion propagates upstream.

Bottlenecks are sometimes characterized by a capacity drop, wherein, the section experiences a reduced throughput once congestion forms. The nominal model presented above does not simulate capacity drop. We extend the model to simulate the capacity drop by incorporating a discontinuous link demand function, defined as

$$\bar{D}_i(k) = \begin{cases} \min(n_i(k)V_i, F_i) & \text{if } n_i(k) \le n_i^c, \\ \bar{F}_i & \text{if } n_i(k) > n_i^c. \end{cases}$$

where n_i^c is a critical density above which capacity drop occurs and $\bar{F}_i < F_i$. The flow out of any link is the minimum of its demand, and the supply imposed by the downstream link. Hence, to derive the effective capacity drop, one needs to consider the capacity imposed by the downstream supply, since the effective capacity of any junction is given by $\min(F_i, F_{i+1})$. In case the current link (Link i) and the next downstream link (Link $i+1$) do not have ramps in between, the effective capacity drop (for Link i) is defined as $\min(F_i, F_{i+1}) - \min(\bar{F}_i, F_{i+1})$, which is different from $F_i - \bar{F}_i$. Clearly, even with a discontinuous demand function, unless $F_{i+1} > \bar{F}_i$, the link will not experience capacity drop. Figure 3 shows an example of a section with a discontinuous demand function. In this figure, we illustrate that capacity

drop occurs at the density corresponding to the apex of the fundamental diagram. In general, the critical density used for the capacity drop can be located beyond this value, and the definition naturally extends in the case of a trapezoidal fundamental diagram. Finally, in case speed control is applied, the demand function is given by $D_i(n_i(k)) = \min(n_i(k)v_i(k), \bar{D}_i(n_i(k)))$.

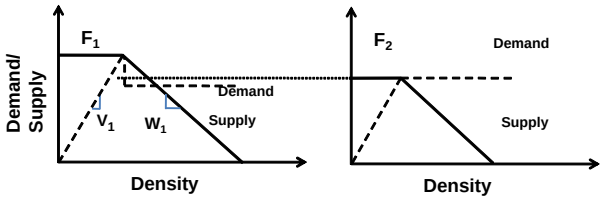

Figure 3. Demand (dashed line) and Supply (solid line) functions of two consecutive sections. The first section (left) experiences a capacity drop

2.3 Complete model

Let I denote the set of all sections considered, while I_d denote the freeway sections where discontinuous capacity model is used. The density and flow update equations are given by

$$
\begin{aligned}
&\underline{\text{Mainline/Queue Conservation Equation}} \\
&n_0(k+1) = n_0(k) + Q_0(k) - f_0(k) \\
&n_i(k+1) = n_i(k) + f_{i-1}(k)(1-\beta_{i-1}(k)) + r_{i-1}(k) - f_i(k) \\
&l_i(k+1) = l_i(k) + Q_i(k) - r_i(k) \quad i = 1, \cdots, N \qquad (1) \\
&\underline{\text{Flow Equations}} \\
&f_N(k) = D_n(k) \\
&f_i(k) = D_i(k) \times \frac{\min(R_i(k), S_{i+1}(k))}{R_i(k)} \\
&r_i(k) = d_i(k) \times \frac{\min(R_i(k), S_{i+1}(k))}{R_i(k)} \quad i = 1, \cdots, N \\
&\textit{where} \\
&D_i(k) = \min(n_i(k)v_i(k), F_i), \forall i \in I \setminus I_d \\
&D_i(k) = \begin{cases} \min(n_i(k)v_i(k), F_i) & \text{if } n_i(k) \leq \max(n_i^c, \bar{F}_i/v_i(k)) \\ \bar{F}_i & \text{if } o.w. \end{cases}, \\
&\qquad \forall i \in I_d \\
&R_i(k) = D_i(k)(1-\beta_i(k)) + d_i(k), \\
&S_{i+1}(k) = \min(W_{i+1}(n_{i+1}^J - n_{i+1}(k)), F_{i+1}) \quad i = 0, \cdots, N-1 \\
&d_i(k) = \min(c_i(k), l_i(k)) \qquad (2)
\end{aligned}
$$

Here $\beta_i(k) = s_i(k)/f_i(k)$ is a (possibly) time-varying split ratio matrix that describes the portion of vehicles exiting through a particular off-ramp. $n_i(k)$ represents the density (in units of number of vehicles/section) of vehicular traffic at any link i. The flow out of link i is represented by the variable $f_i(k)$. All the variables and parameters used above are explained in Table 1. Figure 1 shows the indexing scheme used for refereing nodes, freeway links, on-ramps and off-ramps.

In the model equations, we have assumed that the last cell is in free-flow. This can usually be done by choosing boundaries appropriately. The input to the freeway is supplied through a source link (Link 0) which implements a point queue model. The supply for Link $i+1$ is given by $S_{i+1}(k)$, while the total demand into Link $i+1$ ($R_i(k)$) is given by the sum of demand from the previous link i ($\min(D_i(k)(1-\beta_i(k)))$, not exiting through the off-ramp and the demand from the on-ramp for node i ($d_i(k)$). In free-flow conditions, the total demand is less than the supply and the flow equals the demand. In congested conditions, the supply exceeds the demand, and the available supply is shared by the on-ramp and the previous link. In the LN-CTM model, the available supply is shared proportionally to the demands (i.e $f_i(k)/r_i(k) = D_i(k)/d_i(k)$. This results in the above closed form expressions for the flow computation at each node. In this paper we assume that the on-ramp demands $d_i(k)'s$ can be controlled by manipulating the ramp metering rates $c_i(k)'s$ and link free flow speeds $v_i(k)'s$ can be controlled through the use of Variable Speed Limits (VSL). VSL control allows us to change the input demand which affect the flows at the nodes. Therefore, the control variables are the ramp metering rates, $c_i(k)'s$ and the mainline link speeds, $v_i(k)'s$.

3 Optimal ramp metering and speed control

In this section we will present the optimal control problem for freeway congestion control based on the LN-CTM model presented above. The controller will provide ramp metering rates for all on-ramps and speed limit profiles for all links. For freeway traffic control, we choose the total delay experienced by all the vehicles as our optimization objective. Total delay can be expressed as a linear combination of the macroscopic flow and density variables as follows.

$$J = \sum_{k,i} \left(n_i(k) + l_i(k) - \frac{1}{V_i} f_i(k) \right) \quad \text{Total Delay} \qquad (3)$$

where $k = 1 \cdots K$ denotes the time period and $i = 0 \cdots N$ denotes the link or ramp index. In [9], the authors proposed a optimal control problem based on the objective given above, based on the LN-CTM without the modified demand model. In that paper, the authors presented an equivalent relaxed linear program, whose solution can be used to obtain the optimal solution of the original optimal control problem. However, the procedure does not work with the LN-CTM model with the discontinuous demand function.

As noted in [12], given a set of (stationary) ramp demands, the freeway can be divided into regions, with each region consisting of multiple sections. Only the most downstream section

of a region acts as a bottleneck. This is also the case with usually observed time varying demands, and these bottlenecks are classified as recurrent bottlenecks. In most cases, locations with lane drops with sufficient demand acts as a recurrent bottleneck.

To develop a computationally efficient controller, we divide the freeway into regions, with each region consisting of a only one bottleneck (with a modified demand function) as its most downstream segment (link). A controller based on a optimal control framework will be described for each region, where the controller will prescribe ramp metering rates and speed limits for all links belonging to the section. Thus without loss of generality, we state the following assumption.

Assumption 3.1. *The freeway section considered has only one bottleneck described using the modified demand function. This bottleneck with the capacity drop will be located in the most downstream section, ie. Link N.*

Initially, we will assume that the location downstream of the bottleneck is in free-flow. Later, the controller will then be modified to account for congestion downstream. In the problem formulations listed below, the following parameters and initial conditions must be specified for each link and on-ramp:

Link i Fundamental Diagram Parameters : Capacity F_i, Free-flow speed V_i, Congestion wave speed W_i, Reduced capacity $\bar{F}_i$ and critical density n_i^c
On-ramp i parameters : Ramp flow capacity C_i, and maximum queue length L_i
Off-ramp i split ratios : $\beta_i(k)$
Initial Conditions : $n_0(0), n_i(0), l_i(0) \quad i = 1, \cdots, N$
Flow Demands : $Q_i(k) \quad i = 0, \cdots, N\,,\, k = 0, \cdots, K$

We now define three optimal control problems. **Problem A** states the actual(original) non-linear optimal control problem, and **Problem B** adds additional restrictions on the optimal trajectory and reduces the problem to a mixed integer program. Finally, **Problem C** is presented as a simple and effective method to solve **Problem B** through a sequence of relaxed linear programs.

Problem A Actual delay minimization Problem

$$
\begin{aligned}
\min: \quad & J, \text{ given by Eq. (3)} \\
S.t. \quad : \quad & For \quad k = 1, \cdots, K \\
& \underline{\text{Conservation Equations}} \\
& \text{Equations (1)} \\
& \underline{\text{Flow equations}} \\
& \text{Equations (2)} \\
& \underline{\text{Constraint equations}} \\
& 0 \le v_i(k) \le V_i \\
& 0 \le c_i(k) \le \min(C_i, l_i(k)) \\
& l_i(k) \le L_i \\
& n_i(k), l_i(k), f_i(k), r_i(k) \ge 0 \\
& \text{with given initial conditions/parameters.} \qquad (4)
\end{aligned}
$$

From the formulation above, we can see that the control inputs for the optimal controller include the ramp metering rate $c_i(k)$ and the speed profile $v_i(k)$ for each link. From the flow equations, we also see that $d_i(k) = \min(c_i(k), l_i(k))$. Given the optimal ramp demand profile($d_i(k)$), we can choose $c_i(k) = d_i(k)$ as our ramp metering rate without changing the optimality of the solution. Hence, in the following problems, we will not explicitly consider ramp metering rate as a optimization variable, and replace $c_i(k)$ with $d_i(k)$ in any corresponding constraints. Once the ramp demand profile is known, we get the ramp metering rate as $c_i(k) = d_i(k)$. We highlight that the solution of **Problem A** involves non-linear optimization, primarily due to non-linear equalities in the flow equations and the presence of capacity drop. In the above formulation, in case on-ramps are absent the corresponding variables are removed from the formulation.

Problem A can be converted to a Mixed Integer Program. Consider an integer variable $\mu(k) \in \{0,1\}$, such that $\mu(k) = 0 \Leftrightarrow n_i(k) \le n_N^c$. Then, the constraint corresponding to the modified demand function can be replaced by

$$
\begin{aligned}
& D_N(k) = \min(n_i(k)v_i(k), F_i + (\bar{F}_i - F_i)\mu(k)) \\
& n_i(k) \le n_c^i(1 - \mu(k)) + n_i^J \mu(k) \\
& n_i(k) \ge n_c^i \mu(k) \\
& \mu(k) \in \{0,1\} \qquad (5)
\end{aligned}
$$

We can consider $\mu(k)$ to be the variable representing the "mode" of the last (i.e. most downstream) link link, corresponding to either free-flow or capacity drop. In order to define **Problem B**, we replace the above mixed integer constraint for the modified demand function. In addition, we also make the following assumption.

Assumption 3.2. *For the freeway section considered, we restrict the system evolution such that once the downstream bottleneck link switches back to the "free-flow" mode after having been congested, it remains in the free-flow mode.*

This heuristic restriction is expected to produce an optimal cost , which is almost the same as the optimal cost of **Problem A**, since the free-flow mode is more efficient as it allows vehicles to exit the section at a much faster rate. Hence, once the system switches into the free-flow mode, it is usually optimal for the controller to maintain this mode for maximum throughput. Using this assumption we get,

Problem B Modified mixed integer Problem

$$
\begin{aligned}
\min: \quad & J, \text{ given by Eq. (3)} \\
S.t. \quad : \quad & For \quad k = 1, \cdots, K \\
& \underline{\text{Conservation Equations}} \\
& \text{Equations (1)} \\
& \underline{\text{Flow equations}} \\
& \text{Equations (2) with} \\
& \qquad \text{(5) replacing the modified demand function} \\
& \underline{\text{Constraint equations}} \\
& 0 \leq v_i(k) \leq V_i \\
& 0 \leq d_i(k) \leq \min(C_i, l_i(k)) \\
& l_i(k) \leq L_i \\
& n_i(k), l_i(k), f_i(k), r_i(k) \geq 0 \\
& \mu(k) \geq \mu(k+1)\, k = 1, \cdots, K-1 \\
& \mu(k) \in \{0,1\}\, k = 1, \cdots, K \\
& \text{with given initial conditions/parameters.} \qquad (6)
\end{aligned}
$$

In the above formulation, the constraints $\mu(k) \geq \mu(k+1)\ k = 1, \cdots, K-1$ is equivalent to $\exists j \in \{1 \cdots K\}$ s.t. $\mu(k) = 1,\ k = 1, \cdots, j$ and $\mu(k) = 0,\ k = j+1, \cdots, K$. This interpretation is used to formulate **Problem C**. For a given j, we can formulate an equivalent linear program along the lines of the methodology adopted in [9]. To accomplish this, we convert the non-linear equality constraints in the flow equations to a set of linear inequality constraints. This is done by removing the variables $v_i(k)$ and $d_i(k)$ from the formulation. The final optimal control problem solves a linear program for each j and computes the minimum cost and corresponding control action.

Problem C Modified linear Problem

$$
\begin{aligned}
\min_{j=0 \cdots K}: \quad & J_j^*, \\
where & \\
J_j^* = \min \quad & J, \text{ given by Eq. (3)} \\
S.t. \quad & For \quad k = 1, \cdots, K \\
& \underline{\text{Conservation Equations}} \\
& \text{Equations (1)} \\
& \underline{\text{Relaxed Flow equations}} \\
& \bar{f}_i(k) \leq \bar{n}_i(k) V_i \quad i = 1, \cdots, N \\
& \bar{f}_i(k) \leq F_i \quad i = 1, \cdots, N \\
& \bar{f}_i(k)(1-\beta_i(k)) + \bar{r}_i(k) \leq F_{i+1} \quad i = 1, \cdots, N-1 \\
& \bar{f}_i(k)(1-\beta_i(k)) + \bar{r}_i(k) \\
& \qquad \leq W_{i+1}(n_{i+1}^J - \bar{n}_{i+1}(k)) \quad i = 1, \cdots, N-1 \\
& \underline{\text{Constraint equations}} \\
& 0 \leq \bar{r}_i(k) \leq \min(C_i, \bar{l}_i(k)) \quad i = 1, \cdots, N \\
& \bar{l}_i(k) \leq L_i \\
& For \quad k = 1, \cdots, j \\
& \bar{n}_N(k) \geq n_N^c \\
& \bar{f}_N(k) \leq \bar{F}_i \\
& For \quad k = j+1, \cdots, K \\
& \bar{n}_N(k) \leq n_N^c \\
& \text{with the same initial conditions/parameters.} \qquad (7)
\end{aligned}
$$

Notice that we have chosen to use an upper bar to denote the optimization variables in **Problem C** (e.g. $\bar{n}_i(k)$, $\bar{f}_i(k)$, $\bar{r}_i(k)$) in order to distinguish them from their counterparts in **Problem B**. Each subproblem of **Problem C** is a linear program as we do not explicitly consider the link velocity variables (e.g $\bar{v}_i(k)$) and the on-ramp demands (e.g. $\bar{d}_i(k)$). The j^{th} subproblem captures the situation when the system is in the capacity drop mode for the first j time instants and thereafter switches to the free-flow mode. In essence, Problem C is an exhaustive search of the solution to the Mixed Integer Problem B using linear programs.

Let $j^* = \underset{j=0..K}{\text{argmin}} \quad J_j^*$, denote the subproblem that produces the optimal cost. We denote the corresponding optimal trajectory as $(\bar{n}_i^*(k), \bar{f}_i^*(k), \bar{l}_i^*(k), \bar{r}_i^*(k))$. Using the same technique developed in [9], we outline the methodology to extract ramp metering rates and speed limit profiles, along with the equivalent system trajectory corresponding to **Problem B** using **Algorithm A** given below.. We define $(n_i^*(k), f_i^*(k), l_i^*(k), r_i^*(k), v_i^*(k), d_i^*(k))$ as the trajectory corresponding to **Problem B** as generated by **Algorithm A** from the optimal solution trajectories $(\bar{n}^*(k) etc.)$ to **Problem C**.

Algorithm A

For each time period k and link $0 \leq i \leq N$,
$n_i^*(k) = \bar{n}_i^*(k)$
$f_i^*(k) = \bar{f}_i^*(k)$
$l_i^*(k) = \bar{l}_i^*(k)$
$r_i^*(k) = \bar{r}_i^*(k)$
For each time period k and link $0 \leq i < N-1$,
$if \quad f_i^*(k) = \min(n_i^*(k) V_i, F_i))$
$\quad v_i^*(k) = V_i$
$\quad d_i^*(k) = r_i^*(k)$
$else\, if \quad f_i^*(k)(1-\beta_i(k)) + r_i^*(k) < S_{i+1}^*(k)$
$\quad d_i^*(k) = r_i^*(k)$
$\quad v_i^*(k) = f_i^*(k)/n_i^*(k)$

$$
\begin{aligned}
&\textit{else if} \quad \frac{r_i^*(k)}{S_{i+1}(k)} < \\
&\qquad \frac{\min\left(C_i, l_i(k)\right)}{\min(n_i^*(k)V_i, F_i)(1-\beta_i(k)) + \min\left(C_i, l_i(k)\right)} \\
&\qquad v_i^*(k) = V_i \\
&\qquad d_i^*(k) = r_i^*(k) \times \frac{\min(n_i^*(k)V_i, F_i)(1-\beta_i(k))}{S_{i+1}^*(k) - r_i^*(k)} \\
&\textit{else} \\
&\qquad v_i^*(k) = \frac{\min\left(C_i, l_i(k)\right)}{n_i^*(k)(1-\beta_i(k))} \times \left(\frac{S_{i+1}^*(k)}{r_i^*(k)} - 1\right) \\
&\qquad d_i^*(k) = \min\left(C_i, l_i(k)\right) \\
&\textit{where} \quad S_i^*(k) = \min\left(W_i(n_i^J - \bar{n}_i^*(k)), F_i(k)\right) \\
&\text{and for each time period } k \\
&\tilde{F}_N(k) = \begin{cases} F_N & \text{if } n_N(k) \le n_N^c \\ \bar{F}_N & \text{o.w} \end{cases}, \\
&\textit{if } f_N^*(k) = \min(n_N^*(k)V_N, \tilde{F}_N) \\
&\qquad v_N^*(k) = V_N \\
&\textit{else} \\
&\qquad v_N^*(k) = f_N^*(k)/n_N^*(k)
\end{aligned}
\tag{8}
$$

We state, for completeness, the following theoretical results that prove the equivalence of solutions to **Problem C** and **Problem B**. The proof of this theorem is similar to the one presented in [9] and is omitted here.

Theorem 3.1. *Let* $C = \bar{n}_i^*(k), \bar{f}_i^*(k), \bar{l}_i^*(k), \bar{r}_i^*(k)$ *be an optimal solution of* ***Problem C*** *and* $B = n_i^*(k), f_i^*(k), l_i^*(k), r_i^*(k), v_i^*(k), d_i^*(k)$ *be the solution derived using* ***Algorithm A****. Then B is an optimal solution for* ***Problem B****.*

Now we consider the effect of boundary conditions downstream of the bottleneck. We make the following assumption.

Assumption 3.3. *For the freeway section considered, the downstream boundary condition can be represented using a constant boundary flow restriction* F^d.

The boundary flow restriction means that flows from the most downstream location cannot exceed F^d. When the boundary is in free-flow, we have $F^d = F_N$. However, as the link downstream of our downstream boundary begins to get congested, we have $F^d < F_N$. We assume that the downstream flow restriction is constant, even though the downstream flow restriction will be usually time varying. In fact, the downstream boundary condition is a function of the upstream flows, and is usually indeterminate. However, when the optimal controller is used as a part of a model predictive control strategy, we can use the current downstream flow measurement to provide an estimate of F^d, which we can assume to be constant. When the control update step in the model predictive control strategy is sufficiently small, this assumption is usually valid.

When $F_N \ge F^d > \bar{F}$, we can replace F_N by F^d, and solve the optimal control problem as before. In this case, even when $F_N > F^d$, we recognize that maintaining the first link in the free-flow mode would increase the current section throughput. However, when $\bar{F}_n > F^d$ this is no longer true, since both modes are equally efficient with respect to maximizing discharge from the particular section. In fact, in this case, it is efficient to switch to the congested mode, since this allows the freeway section to store more vehicles in the freeway (due to the increased density prevalent in the capacity drop mode). We replace F_N by F^d and solve the following linear program to obtain the optimal solution in this case.

$$
\begin{aligned}
&\min \quad J, \text{ given by Eq. (3)} \\
&S.t. \quad \textit{For} \quad k = 1, \cdots, K \\
&\qquad \underline{\text{Conservation Equations}} \\
&\qquad \text{Equations (1)} \\
&\qquad \underline{\text{Relaxed Flow equations}} \\
&\qquad \bar{f}_i(k) \le \bar{n}_i(k)V_i \quad i = 1, \cdots, N \\
&\qquad \bar{f}_i(k) \le F_i \quad i = 1, \cdots, N \\
&\qquad \bar{f}_i(k)(1-\beta_i(k)) + \bar{r}_i(k) \le F_{i+1} \quad i = 1, \cdots, N-1 \\
&\qquad \bar{f}_i(k)(1-\beta_i(k)) + \bar{r}_i(k) \\
&\qquad\qquad \le W_{i+1}(n_{i+1}^J - \bar{n}_{i+1}(k)) \quad i = 1, \cdots, N-1 \\
&\qquad \underline{\text{Constraint equations}} \\
&\qquad 0 \le \bar{r}_i(k) \le \min(C_i, \bar{l}_i(k)) \quad i = 1, \cdots, N \\
&\qquad \bar{l}_i(k) \le L_i \\
&\qquad \text{with the same initial conditions/parameters.}
\end{aligned}
\tag{9}
$$

The optimal solution of the above linear program can be used to obtain speed limit profiles and ramp metering rates using **Algorithm A**, with a slight modification. We need to replace the calculations for determining the speed limit profile in the last section in **Algorithm A** by

$$
\begin{aligned}
&\textit{if } f_N^*(k) = \min(n_N^*(k)V_N, F_N) \\
&\qquad v_N^*(k) = V_N \\
&\textit{else} \\
&\qquad v_N^*(k) = f_N^*(k)/n_N^*(k)
\end{aligned}
\tag{10}
$$

We have observed that the optimal controller does not specify a restrictive speed limit profile (ie. $v_N(k) = V_N$) for the last link in any of the cases mentioned above, as these would increase the delay of vehicles in the freeway section considered.

3.1 Remarks

In our final optimal control problem, we do not explicitly consider traffic speed variables. At any section, we can calculate the speed of the traffic as $v_i^s(k) = f_i(k)/n_i(k)$. One concern is to ensure that the variable speed limits do not lead to sudden changes in speed at an particular section. We adopt an indirect method to limit speed variations at any particular section. Let ΔV_i denote the nominally allowed speed variation within which we would like to operate. Then we add the following constraint

$$-\bar{\zeta}_i - \frac{\Delta V_i}{2}(n_i(k)+n_i(k+1)) \leq f_i(k+1) - f_i(k) \leq \bar{\zeta}_i + \frac{\Delta V_i}{2}(n_i(k)+n_i(k+1))$$

We will also add a penalty term $\bar{C}\sum_{i,k}\bar{\zeta}_i(k)$ to our cost function. The above constraint indirectly limits speed variations at a particular section by limiting flow variations across different time steps.

When we use this optimal control formulation within the model predictive framework, we assume that the split ratios are constant (and equal to the split ratios, or its estimate at the current time instant). In [9], we discuss the caveats in using a time-varying split ratios in the above formulation in detail. The constant split ratio assumption works well in practice when the control update step used in the model predictive controller using the formulation above is small.

Finally, in deriving **Problem C**, we have used a simple but effective enumeration method to list a sequence of linear programs that need to be solved. Alternatively, a Mixed Integer Linear Program(MILP) can also be formulated with the heuristic constraints on mode changes, which was described in **Problem B**. In theory, this could be solved with various MILP solution methods, like the branch and cut methods. Our optimal control problems are large-scale in nature, and we noticed that a few available MILP solvers did not converge to a solution in some cases. Also, the computational efficiency obtained through the sequence of linear programs was more than sufficient for the use of these optimal controllers for use in realistic scenarios.

4 Model Predictive ramp metering and speed control

In this section we demonstrate the application of a model predictive controller based on the optimal control formulation presented in the previous section. The Model Predictive Controller solves an open loop optimal control problem online based on a plant model at each sampling time, using the state information measured at the current sampling time. The controller implements the control steps of the obtained optimal control profile till the next sampling time, and then the process is repeated [13].

Let T and N_p denote the model time step and prediction horizon used in the optimization problem respectively. We execute the MPC every $T_c = N_c \times T$ time instants (here we assume that N_p, N_c are natural numbers). Figure 4 represents the geometry of the freeway section (approx 3.1 mile length) which is considered for our simulation studies. In this portion of the freeway, Link 9 is the only link which experiences a capacity drop. The fundamental diagram parameters are listed in Table 2. We can see that the maximum throughput of Link 9 is 7600vph in free-flow conditions. This decreases to 7300vph once the density in Link 9 exceeds the critical density of 121vpm. This represents a capacity drop of 3.9% which is modest compared to nominal capacity drop reported in literature, which is typically 5-10 %. We assign our optimal controller to operate on Links 1 to Link 9. A constant split ratio $\beta = 0.15$ is chosen for both the off-ramps during the entire time period considered. Figure 5 shows the on-ramp demands we use for the simulation. The on-ramp demands are chosen such that the freeway is congested between T=1hr to T=2hr. We use a constant flow of 8000vph as the input flow into the first link on the freeway. We use the LN-CTM model explained in the paper to perform our simulations. In the first simulation, we assume that the boundary downstream of Link 11 is in free-flow. Figure 6 (top) shows the velocity contours that result when no control action is applied. In this simulation, we can see that the freeway starts to get congested around T=1hr. We also simulate the freeway, with the model predictive controller is

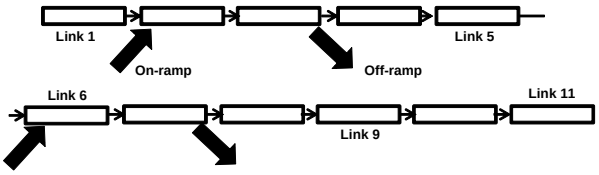

Figure 4. Freeway geometry with location of on-ramps and off-ramps. Link 9 experiences capacity drop.

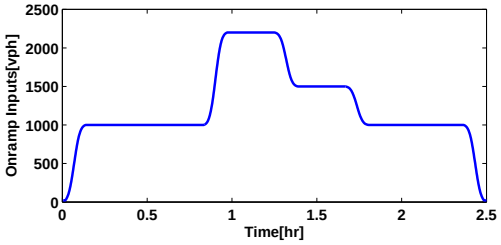

Figure 5. On-ramp demands for both ramps.

Table 2. Link parameters

Links	V	W	F	$\bar{F}$
1-8	65mph	20mph	10500vph	n.a
9	65mph	20mph	7900vph	7300vph
10-12	65mph	20mph	7600vph	n.a

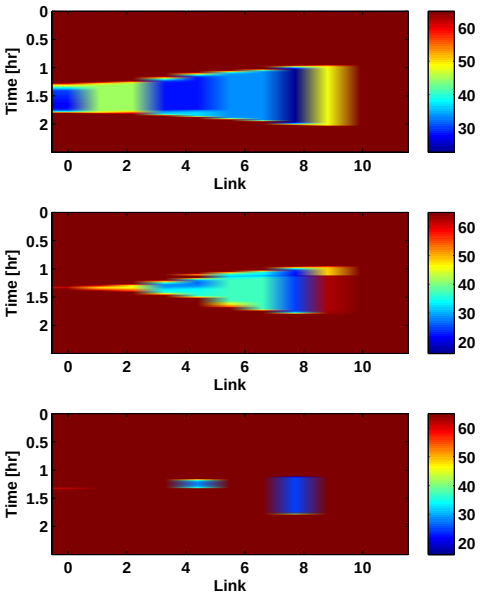

Figure 6. Top : Simulated Velocity contours [mph] - no control scenario. Middle : Simulated Velocity contours [mph] - optimal controller. Bottom : Optimal speed limit profile [mph].

active. The controller is initially inactive, and we start applying the controller at T=1.11hr, by which time Link 9 already experiences capacity drop. We choose $N_p = 30$,$T = 10s$,$N_c = 6$ for our controller. In this case, the controller has to solve at-most 30 (and in most cases, less) linear programs. Figure 6 (middle), shows results of the simulation in which our model predictive controller is used. We can clearly see the benefits of applying our optimal controller. Figure 6 (bottom) shows the variable speed limit profiles specified by our optimal controller. We can see that optimal controller specifies a speed limit profile for Link 8, which helps decrease the density of Link 9 to below critical density. Thereafter, it still maintains the speed limit which enables Link 9 to stay in free-flow. Finally, we see from 7 that the optimal controller maintains the ramp queue limits. In the no-control scenario(not shown here), we observed that the queues extend to 100 on one of the on-ramp. We see that the optimal controller mostly uses the ramp space in the most downstream ramp, which is closest to the location of the bottleneck till it reaches its maximum queue limit. The total travel time and the delay experienced by all users in the no-control scenario are 1253vh (vehicle-hours) and 257vh respectively. In contrast, these reduce to 1133vh and 137vh respectively when the controller is used. This leads to a substantial delay reduction of 46.7%. In the second simulation, we assume that the boundary downstream of Link 11 is initially in free-flow. At 1.39hr, the boundary (Link 11) begins to get congested and this congestion propagates back onto Link 9 soon after. At 1.6hr, the boundary becomes free-flowing. All

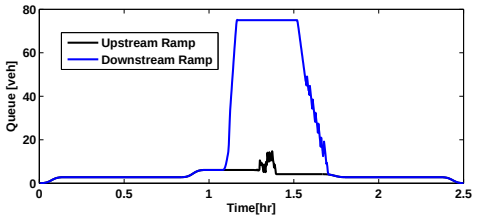

Figure 7. On-ramp queues for simulation 1 with optimal controller. Each ramp has a maximum queue size of 75veh.

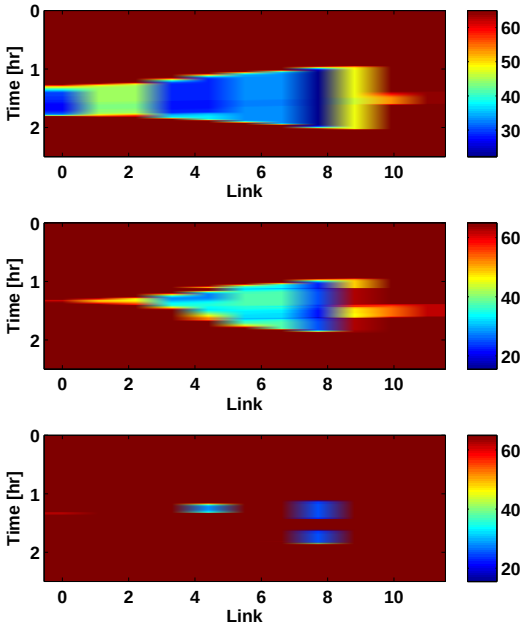

Figure 8. Top : Simulated Velocity contours [mph] - no control scenario. Middle : Simulated Velocity contours [mph] - optimal controller. Bottom : Optimal speed limit profile [mph].

other parameters and demands are same as the first simulation. Figure 8 (top) shows the velocity contours under the no-control scenario. The congestion is more widespread as compared to the simulation with the optimal controller (Figure 8 (middle)). From Figure 8 (bottom), we note the optimal speed limit profiles specified by the controller. In this case, we see that the optimal controller specifies a speed limit profile until the boundary congestion reaches the location downstream of Link 9. Thereafter, the optimal controller only resumes the speed limit control when the congestion due to the downstream section has dissipated. The total travel time and the delay experienced by all users in the no-control scenario are 1259vh (vehicle-hours) and 264vh respectively. In contrast, these reduce to 1164vh and 169vh respectively when the controller is applied. This leads to a delay reduction of 35.97%.

5 Conclusion and comments

In this paper, we presented a framework for optimal congestion control for freeway networks, using ramp metering and variable speed limits. This paper extends the results of [9] to include situations when there is a significant capacity drop at any particular segment in the freeway considered. We introduce a discontinuous capacity drop function to model capacity reductions due to the capacity drop phenomena. We present various assumptions that are needed to allow us to solve the optimal control problem efficiently. As our first assumption, we divide the freeway into regions and assume that each region is controlled using an independent controller which controls all sections within that region. We recognize that one of its drawbacks is that we cannot fully co-ordinate all controllers present in the freeway, which might limit performance improvements. We plan to explore supervisory schemes which improve controller co-ordination in this scenario. According to our second assumption, the optimal trajectory does not switch back from the free-flow mode to the capacity drop mode. This assumption is usually found to be a valid, and it does not lead to any appreciable drop in performance of the controller. Our final assumption provides us a tractable method to calculate the optimal control strategy under congested boundary conditions. The issues related to this assumption are similar to the ones discussed for our first assumption. We note that these assumptions allow us to efficiently solve the optimal control problem for its use in a model predictive controller. We can solve our optimal control problem for the scenario shown in this paper within 2 sec for each model predictive control step, using the MOSEK linear programming solver. This is a fraction of the controller time horizon. We also see that our sequence of linear programs can be solved completely in parallel. One single iteration of the controller can be executed within 10-20s for a realistic sized freeway, like the one shown in [9] when we exploit the inherent parallelism.

We presented simulation results where we compared the freeway characteristics with and without the application of our model predictive controller. In both the scenarios studies, our controller leads to a substantial reduction of delay experienced by all travelers in the freeway. From our experience of simulating freeway sections with and without capacity drop, we determine that capacity drop is usually the single most important factor that contributes to delay in the freeway, if present. Accounting for capacity drop in our controllers, whenever they are present can help us significantly improve traveler experience on the controlled freeways.

6 ACKNOWLEDGMENTS

This work is partially supported by the California Department of Transportation through the California PATH and by the National Science Foundation through grant CDI-0941326. The first author gratefully acknowledges support from the John and Janet McMurtry Fellowship. We would also like to acknowledge Gabriel Gomes for his insightful comments during discussions.

REFERENCES

[1] van den Hoogen, E., and Smulders, S., 1994. "Control by variable speed signs: results of the dutch experiment". In Road Traffic Monitoring and Control, 1994., Seventh International Conference on, pp. 145 –149.

[2] Carlson, R. C., Papamichail, I., Papageorgiou, M., and Messmer, A., 2010. "Optimal mainstream traffic flow control of large-scale motorway networks". *Transportation Research Part C: Emerging Technologies,* **18**(2), pp. 193 – 212.

[3] Papageorgiou, M., Hadj-Salem, H., and Middleham, F., 1997. "ALINEA local ramp metering: Summary of field results". *Transportation Research Record,* **1603**, pp. 90–98.

[4] Papamichail, I., Papageorgiou, M., Vong, V., and Gaffney, J., 2010. "Heuristic ramp-metering coordination strategy implemented at monash freeway, australia". *Transportation Research Record,* **2178**, pp. 10–20.

[5] Kotsialos, A., and Papageorgiou, M., 2004. "Nonlinear optimal control applied to coordinated ramp metering". *Control Systems Technology, IEEE Transactions on,* **12**(6), nov., pp. 920 – 933.

[6] Gomes, G., and Horowitz, R., 2006. "Optimal freeway ramp metering using the asymmetric cell transmission model". *Transportation Research, Part C,* **14**(4), pp. 244–262.

[7] Hegyi, A., Schutter, B. D., and Hellendoorn, H., 2005. "Model predictive control for optimal coordination of ramp metering and variable speed limits". *Transportation Research Part C: Emerging Technologies,* **13**(3), pp. 185 – 209.

[8] Daganzo, C., 1994. "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory". *Transportation Research, Part B,* **28**(4), pp. 269–287.

[9] Muralidharan, A., and Horowitz, R., 2012. "Optimal control of freeway networks based on the link node cell transmission model". To appear in the Proceedings of the American Control Conference, Montreal, Canada.

[10] TOPL. Tools for operations planning website. http://path.berkeley.edu/topl/, accessed 9/15/2011.

[11] Kurzhanskiy, A., 2007. "Modeling and software tools for freeway operational planning". PhD thesis, University of California, Berkeley.

[12] Gomes, G., Horowitz, R., Kurzhanskiy, A. A., Varaiya, P., and Kwon, J., 2008. "Behavior of the cell transmission model and effectiveness of ramp metering". *Transportation Research Part C: Emerging Technologies,* **16**(4), pp. 485 – 513.

[13] Borrelli, F., Bemporad, A., and Morari, M. Predictive control. http://www.mpc.berkeley.edu/mpc-course-material, accessed 6/1/2012.