Globally Optimal Solutions to the On-ramp Metering Problem - Part II

Gabriel Gomes and Roberto Horowitz
Department of Mechanical Engineering
University of California at Berkeley
Email: gomes@me.berkeley.edu , horowitz@me.berkeley.edu

***Abstract*— The main results of Part I are summarized and extended. The problems of rapidly decaying mainline cost weights, non-zero on-ramp cost weights, and the lack of on-ramp queue length constraints in the original formulation are resolved by introducing two new assumptions: 1) congestion on the mainline does not spill onto the on-ramps, and 2) $\underline{r}_i^c = 0$. A numerical example based on a 14-mile stretch of a congested freeway is used to demonstrate the technique. The example predicts a 8.4% savings in Total Travel Time, with queue constraints, over the 5-hour peak period.**

I. Introduction

On-ramp metering is a method of traffic control that seeks to improve the performance of freeway systems by regulating the flow of vehicles from the on-ramps. The scope of the *system* may be the freeway alone, or it may include alternative routes. The improvement in *performance* is usually measured in terms of total travel time savings. The task is an extremely delicate one for several reasons. First, Zhang concluded in [1] that freeways are best left uncontrolled (i.e. no improving controllers exist) whenever they are either "uniformly congested" or "uniformly uncongested", meaning that the state of congestion cannot be affected by on-ramp control. This makes intuitive sense: metering an uncongested freeway introduces unnecessary delays. However, even when the freeway is in a state of "mixed congestion", and can therefore benefit from on-ramp control, there exist only a few mechanisms for reducing travel time. Banks identified four in [2]: 1) increasing bottleneck flow, 2) diverting traffic to alternative routes, 3) preventing accidents, and 4) preventing the obstruction of off-ramps by congestion on the mainline. The second and third mechanisms are difficult phenomena to model and verify, and are not considered in this and many other optimal control designs. Increasing bottleneck and off-ramp flow, both related to the avoidance of mainline congestion, are left as the two principle mechanisms for optimizing travel time. However, congestion can only be reduced by storing the surplus vehicles in the on-ramp queues, and this often conflicts with the limited storage available in the on-ramps. These can typically hold up to 30 vehicles each, which is a small number compared to the number of vehicles on a congested freeway. The on-ramp metering problem is thus recognized as one of careful allocation of on-ramp storage space and timely release of vehicles. The example in this paper shows that optimal travel time savings with limited on-ramp storage is on the order of only a few percentage points. Given the small margins, the quality of the numerical solution becomes a very important factor, in addition to the validity of the model and its calibration. Using a local solution to the problem may in fact be detrimental to system performance. In this paper we seek a global or near-global solution by solving a relaxed linear problem. The model and problem statements of Part I are reviewed in Section II. Section III describes modifications and enhancements to the original problem formulation and solution. The technique is demonstrated in Section IV with an example.

II. Overview of Part I

Part I [3] described the Asymmetric Cell Transmission Model (ACTM), the formulation of the nonlinear ($\mathcal{P}_A$) and linear ($\mathcal{P}_B$) optimization problems, and sufficient conditions on the cost weights of the objective function for the two problems to have identical solution sets. The sufficient conditions, given by Eqs. (23) and (24) in Part I, required that $(\mathcal{I} + |\mathcal{E}n|) \times \mathcal{K}$ linear combinations of the cost weights with $(\mathcal{I} + |\mathcal{E}n|) \times \mathcal{K}$ Maximal Worst-Case Causal (MWCC) perturbations evaluate to negative numbers. The causal nature of the MWCC perturbation allowed the use of a simple backstepping method for finding cost weights that met the conditions. A small example tested the algorithm and exposed a couple deficiencies in its solution. Below are reproduced some of the important equations of Part I for reference in this paper.

A. Notation

$\mathcal{I}$: set of all freeway sections. $\mathcal{I} = \{0 \ldots \mathrm{I}-1\}$
$\mathcal{K}$: set of time intervals. $\mathcal{K} = \{0 \ldots \mathrm{K}-1\}$

$\mathcal{E}n$: set of sections with on-ramps. $\mathcal{E}n \subseteq \mathcal{I}$
$\mathcal{E}n^+$: set of sections with metered on-ramps. $\mathcal{E}n^+ \subseteq \mathcal{E}n$
$\rho_i[k]$: vehicles in section i at time $k\Delta t$.
$l_i[k]$: vehicles queueing in on-ramp i at time $k\Delta t$.
$f_i[k]$: vehicles going from i to $i+1$ during interval k.
$r_i[k]$: vehicles entering i from an on-ramp during k.
$r_i^c[k]$: metering rate for on-ramp i.
$d_i[k]$: demand for on-ramp i.
$\beta_i[k]$: dimensionless split ratio for off-ramp i.
v_i : normalized freeflow speed $\in [0,1]$
w_i : congestion wave speed $\in [0,1]$
$\bar{\rho}_i$: jam density [veh]
$\bar{f}_i$: mainline capacity [veh]
$\bar{s}_i$: off-ramp capacity [veh]
α_i : influence parameter $\in [0,1]$
γ_i : influence parameter $\in [0,1]$
ξ_i : influence parameter $\in [0,1]$

$$\delta_i \triangleq \begin{cases} 1 & \text{if} \quad i \in \mathcal{E}n \\ 0 & \text{else} \end{cases}$$

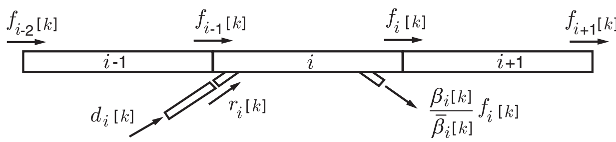

Fig. 1. Interpretation of model variables.

B. Traffic Model

The four components of the ACTM are: (1) mainline conservation, (2) on-ramp conservation, (3) mainline flow, and (4) on-ramp flow. Off-ramp flows are modelled as a known proportion $\beta_i[k]$ of the total flow leaving the section. $\bar{\beta}_i[k]$ is defined for convenience as $1-\beta_i[k]$.

$$\rho_i[k+1] = \rho_i[k] + f_{i-1}[k] + \delta_i r_i[k] - f_i[k]/\bar{\beta}_i[k] \tag{1}$$

$$l_i[k+1] = l_i[k] + d_i[k] - r_i[k] \tag{2}$$

$$f_i[k] = \min\Big\{ \bar{\beta}_i[k] v_i(\rho_i[k] + \delta_i\gamma_i r_i[k]) \; ; \; \bar{f}_i \; ; \; w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1}[k]) - \delta_{i+1}\alpha_{i+1} r_{i+1}[k] \; ; \; \frac{\bar{\beta}_i[k]}{\beta_i[k]}\bar{s}_i \Big\} \tag{3}$$

$$r_i[k] = \begin{cases} \min\{ l_i[k]+d_i[k] \; ; \; \xi_i(\bar{\rho}_i - \rho_i[k]) \; ; \; r_i^c[k] \} & i \in \mathcal{E}n^+ \\ \min\{ l_i[k]+d_i[k] \; ; \; \xi_i(\bar{\rho}_i - \rho_i[k]) \} & i \in \mathcal{E}n \backslash \mathcal{E}n^+ \end{cases} \tag{4}$$

Parameters γ_i and α_i determine, respectively, the influence of the on-ramp flow on mainline flows either downstream or upstream of the on-ramp. ξ_i determines the portion of available space on the freeway $(\bar{\rho}_i - \rho_i[k])$ that is allocated to vehicles entering from an on-ramp. A portion w_i of the total available space can be filled by vehicles coming from an upstream freeway section.

C. Problem Formulation

The objective is to minimize a linear combination of on-ramp and mainline flows called the *Generalized Total Travel Time* (gTTT):

$$\texttt{gTTT} \triangleq -\sum_{k\in\mathcal{K}} \left[\sum_{i\in\mathcal{I}} a_i[k] f_i[k] + \sum_{i\in\mathcal{E}n} b_i[k] r_i[k] \right] \tag{5}$$

Minimizing the actual Total Travel Time (TTT) is equivalent to minimizing gTTT with:

$$a_i[k] = \begin{cases} (\mathrm{K}-k)\, \beta_i[k]/\bar{\beta}_i[k] & i < I-1 \\ (\mathrm{K}-k)\left(\beta_i[k]/\bar{\beta}_i[k] + 1\right) & i = I-1 \end{cases} \tag{6}$$
$$b_i[k] = 0$$

The nonlinear and relaxed linear optimization problems are stated below.

<u>**Problem $\mathcal{P}_A$**</u>: $\quad \psi_A^* = \arg\min_{\psi\in\Omega_A} \texttt{gTTT}(\psi)$ (7)

$$\Omega_A = \Big\{ \psi = \{\rho_i[k], l_i[k], f_i[k], r_i[k], r_i^c[k]\} :$$
Dynamic equations : $(1), (2)$
Concave fundamental diagram : $(3)-(4)$,
Control bounds : $\underline{r}_i^c \le r_i^c[k] \le \bar{r}_i^c \Big\}$

<u>**Problem $\mathcal{P}_B$**</u>: $\quad \psi_B^* = \arg\min_{\psi\in\Omega_B} \texttt{gTTT}(\psi)$ (8)

$$\Omega_B = \Big\{ \psi = \{\rho_i[k], l_i[k], f_i[k], r_i[k], r_i^c[k]\} :$$
Dynamic equations : $(1), (2)$,
Relaxed constitutive relations : $(9)-(14)$,
Control bounds : $\underline{r}_i^c \le r_i^c[k] \le \bar{r}_i^c \Big\}$

$\forall\, k\in\mathcal{K}, i\in\mathcal{I}$:

$$f_i[k] \le \bar{\beta}_i[k] v_i(\rho_i[k] + \delta_i\gamma_i r_i[k]) \tag{9}$$
$$f_i[k] \le w_{i+1}(\bar{\rho}_{i+1} - \rho_{i+1}[k]) - \delta_{i+1}\alpha_{i+1} r_{i+1}[k] \tag{10}$$
$$f_i[k] \le \min\left\{ \bar{f}_i \; ; \; \frac{\bar{\beta}_i[k]}{\beta_i[k]}\bar{s}_i \right\} \tag{11}$$

$\forall\, k\in\mathcal{K}\,,\; i\in\mathcal{E}n:$

$$r_i[k] \le l_i[k] + d_i[k] \tag{12}$$
$$r_i[k] \le \xi_i(\bar{\rho}_i - \rho_i[k]) \tag{13}$$

$\forall\, k\in\mathcal{K}\,,\; i\in\mathcal{E}n^+:$

$$r_i[k] \le r_i^c[k] \tag{14}$$

D. The cost weights synthesis problem

The non-linearities in Eqs. (3) and (4) make $\mathcal{P}_A$ difficult to solve globally. $\mathcal{P}_B$ on the other hand, is easy to solve, but not necessarily useful to the solution of $\mathcal{P}_A$. The goal of the cost weights synthesis (CWS) problem is to find weights $a_i[k]$ and $b_i[k]$ that render problems $\mathcal{P}_A$ and $\mathcal{P}_B$ *equivalent* $(\mathcal{P}_A \equiv \mathcal{P}_B)$, in the sense that their solution sets are identical. Thus, when $\mathcal{P}_B$ is posed with weights provided by the CWS, its global solution is also the desired global solution to $\mathcal{P}_A$.

The CWS problem can be stated as having to find cost weights such that any feasible point ψ in Ω_B but not in Ω_A is *not* a minimizer of gTTT, and can therefore be improved upon by some "upward" feasible perturbation. The CWS problem was solved in Part I by dividing the set $\Omega_B \setminus \Omega_A$ into $(\mathcal{I} + |\mathcal{E}n|) \times \mathcal{K}$ subsets and defining a *Maximal Worst-Case Causal* (MWCC) perturbation for each subset, with the characteristic that the MWCC was feasible for all points in its subset. The CWS problem was thus reduced to finding cost weights such that each of the $(\mathcal{I} + |\mathcal{E}n|) \times \mathcal{K}$ MWCC perturbations was also an improvement on gTTT (Eqs. (23) and (24) in Part I). A simple backstepping algorithm was designed for computing the cost weights, based on the *causality* of the MWCC perturbation.

III. Changes to the original formulation

A. Additional Assumptions

The example presented in Part I revealed two basic deficiencies in the proposed solution to the CWS problem. Ideally we would like the computed cost weights to equal the TTT costs of Eq. (6). Short of that, we would at least want the cost weights to be similar to Eq. (6) in some quantifiable way. The weights of the example in Part I differed from Eq. (6) in two ways: 1) the $a_{i[k]}$'s decayed faster than a straight line, and 2) the $b_{i[k]}$'s were not identically zero (see Figure 3 in Part I).

A third deficiency, this one in the formulation of the optimization problem, was the omission of upper bounds on the on-ramp queue lengths:

$$l_{i[k]} \leq \bar{l}_i \quad , \quad \forall\, i \in \mathcal{I}\,,\; k \in \mathcal{K} \tag{15}$$

The reason for excluding the queue length constraints was that they made impossible the definition of a generalized feasible perturbation, such as the MWCC perturbation. To illustrate, consider ψ a feasible candidate solution to problem $\mathcal{P}_B$ for which none of Eqs. (9) through (11) are active for a particular $i = \iota$ and $k = \kappa$. Such a candidate solution would be classified in Part I as a member of $\mathrm{I}_{\iota\kappa}$. Assume also that ψ has:

$$\begin{aligned} f_{\iota+1[\kappa]} &= \bar{f}_{\iota+1} \\ r_{\iota+1[\kappa+1]} &= \xi_{i+1}(\bar{\rho}_{\iota+1} - \rho_{\iota+1[\kappa+1]}) \\ l_{\iota+1[\kappa+2]} &= \bar{l}_{\iota+1} \end{aligned}$$

In this situation, any positive perturbation to $f_{\iota[\kappa]}$ produces an increase in $\rho_{\iota+1[\kappa+1]}$, which in turn forces $r_{\iota+1[\kappa+1]}$ to decrease, and $l_{\iota+1[\kappa+2]}$ to grow beyond $\bar{l}_{\iota+1}$. We have demonstrated that no positive causal perturbation exists that is feasible for *all* members of $\mathrm{I}_{\iota\kappa}$, whenever queue length constraints are included.

Two additional assumptions are sufficient to remedy these three shortcomings:

<u>Assumption #1</u>: Eq. (13) is not active in the optimal solution for any i or k.

In other words, we assume that congestion on the mainline does not obstruct vehicles entering from the on-ramps, when optimal metering is used. Obviously, this assumption does not hold for every freeway. It therefore limits the apllicability of the proposed controller to freeways where it does – presumably most well-designed freeways. The assumption permits the removal of Eq. (13) from the problem statement, and the "$-\xi_i \Delta\rho_{i[k]}$" terms from the definitions of $\bar{\boldsymbol{\Delta}}^{\mathrm{I}}_{\iota\kappa}$ and $\bar{\boldsymbol{\Delta}}^{\mathrm{II}}_{\iota\kappa}$ (Eqs. (21) and (22) in Part I). Notice that then $\Delta r_{i[k]} = \Delta l_{i[k]} = 0$ in $\bar{\boldsymbol{\Delta}}^{\mathrm{I}}_{\iota\kappa}$, and $\Delta l_{i[k]} \leq 0$ in $\bar{\boldsymbol{\Delta}}^{\mathrm{II}}_{\iota\kappa}$. There is no longer a conflict between the MWCC perturbation and the queue length constraint, which requires only $\Delta l_{i[k]} \leq 0$. As we shall see in Section IV, this assumption also eliminates the problem of rapidly decaying $a_{i[k]}$'s. Section IV also shows that the assumption is not overly restrictive, and in fact holds for the I-210 model.

<u>Assumption #2</u>: $\underline{r}^c_i = 0$.

We assume that the on-ramp meters are able to completely shut off on-ramp flow. This is never true in practice: the actual minimum metering rate on I-210 is 180 vph, which corresponds to 1 vehicle every 20 seconds. The optimal control plan generated with $\underline{r}^c_i = 0$ must therefore be modified before implementation by increasing all optimal metering rates to at least 180 vph. This is presumably not too large a sacrifice on global optimality. The assumption is useful because it enables the formulation of the following simplified but equivalent form of Problem $\mathcal{P}_B$:

<u>Problem $\mathcal{P}_C$</u>: $\psi^*_C = \arg\min_{\psi \in \Omega_C} \mathrm{gTTT}(\psi)$ (16)

$$\Omega_C = \Big\{\ \psi = \{\rho_{i[k]}, l_{i[k]}, f_{i[k]}, r_{i[k]}\}\ :$$
Dynamic equations : (1), (2),
Relaxed fundamental diagram : (9) – (11)
On-ramp flow constraints : (17) – (19) $\Big\}$

$$\forall\, k \in \mathcal{K}\,,\; i \in \mathcal{E}n \backslash \mathcal{E}n^+ : \qquad r_{i[k]} = d_{i[k]} \tag{17}$$
$$\forall\, k \in \mathcal{K}\,,\; i \in \mathcal{E}n^+ : \qquad r_{i[k]} \leq l_{i[k]} + d_{i[k]} \tag{18}$$
$$\forall\, k \in \mathcal{K}\,,\; i \in \mathcal{E}n^+ : \qquad 0 \leq r_{i[k]} \leq \bar{r}^c_i \tag{19}$$

Having used Assumption #1 to eliminate Eq. (13) from $\mathcal{P}_B$, the flow from uncontrolled on-ramps is now known to equal the demand (Eq. (17)). Due to both assumptions, the model for metered on-ramps becomes:

$$r_i[k] \;=\; \min\{l_i[k] + d_i[k]\;,\; r_i^c[k]\} \tag{20}$$
$$0 \le r_i^c[k] \;\le\; \bar{r}_i^c \tag{21}$$

$r_i^c[k]$ can be seen to be an unnecessary variable. It appears nowhere else in the problem, and its only role in Eqs. (20) and (21) is to limit the on-ramp flow to no greater than $\bar{r}_i^c$. Furthermore, its value is somewhat arbitrary: when $r_i[k] = l_i[k] + d_i[k]$ in Eq. (20), $r_i^c[k]$ can be given any value between $l_i[k] + d_i[k]$ and $\bar{r}_i^c$ without affecting the cost. A valid metering plan can be derived once the solution to $\mathcal{P}_C$ is found, with:

$$r_i^c[k] = \max\{\; r_i[k]\;;\; 180 \text{ vph }\} \tag{22}$$

The main advantage of $\mathcal{P}_C$ over $\mathcal{P}_B$ is that neither metered nor uncontrolled $r_i[k]$'s in $\mathcal{P}_C$ are required to fall on their "upper boundary". Perturbations to $r_i[k]$'s and the resulting non-zero $b_i[k]$'s are therefore not needed.

The CWS problem is solved for Problem $\mathcal{P}_C$ by setting all $b_i[k]$'s to zero, and computing $a_i[k]$'s with the procedure developed in Part I, with $\bar{\boldsymbol{\Delta}}^{\mathrm{I}}_{\iota\kappa}$'s adjusted as per Assumption #1.

B. Dual Time Scales

A modification to the original model that was found useful for reducing the size of the LP problem, was to consider different time intervals for the mainline and on-ramp variables. Consistency of the mainline conservation equation (Eq. (1)) requires:

$$v_i \Delta t_m \le L_i \quad \forall\, i \in \mathcal{I}$$

where Δt_m is the duration of the discrete time interval, and L_i is the length of the ith section. Consistency of the on-ramp model (Eq. (2)) on the other hand, does not depend on the duration of the time interval. The example of Section IV uses a freeway partition with a shortest section of 1000 ft. At a freeflow speed of 65 mph, this corresponds to a maximum Δt_m of 10.5 seconds. Most freeway control systems only update the metering rates at intervals of 1 to 5 minutes, making it unnecessary to model the on-ramps at the faster rate.

The modified dual-scale ACTM assumes that the on-ramp interval, Δt_o, is an integer multiple of Δt_m, with $p \triangleq \Delta t_o / \Delta t_m$. The modified model equations are obtained by replacing the $r_i[k]$ terms in Eqs. (1) and (3) with $r_i[k]/p$. To retain the important properties given by the theorem of Part I, Eq. (4) must be modified by including the p separate $\xi_i(\bar{\rho}_i - \rho_i[k])$-like terms of the *upcoming* Δt_o interval. That is, $r_i[k]$ at time $t = k{\cdot}\Delta t_o = k\cdot p\Delta t_m$ is computed as the minimum of $l_i[k] + d_i[k]$ and $\min_x\{\xi_i p(\bar{\rho}_i - \rho_i[x])\}$, with $x = k{\cdot}p \ldots (k{+}1)p{-}1$. This, unfortunately, destroys the causality of the model, as $r_i[k]$ now depends on future values of $\rho_i[x]$. A specialized iterative algorithm was created to integrate the non-causal model equations.

IV. Experiments and Results

In this section we report on simulation tests using the suggested procedure for solving the CWS problem and resulting LP problem, with the modifications outlined in this paper. Model parameters and traffic data were taken from a 14-mile stretch of Interstate 210 WB in Pasadena, California. This site contains 20 metered on-ramps and a single uncontrolled freeway connector from I-605 NB. A speed contour plot constructed from loop detector data is shown in Figure 2. The darker shaded areas indicate average speeds below 40 mph. A study of the traffic characteristics of this site [4] identified three recurring bottlenecks. The severest of the three can be seen in the contour plot to affect the first third of the test section over a time period of about 4 hours (6:30 to 10:30 am).

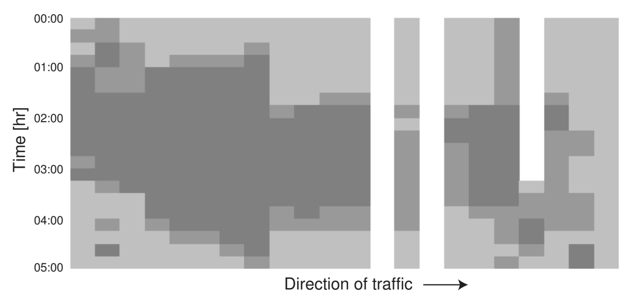

Fig. 2. Measured speed contour plot [mph]

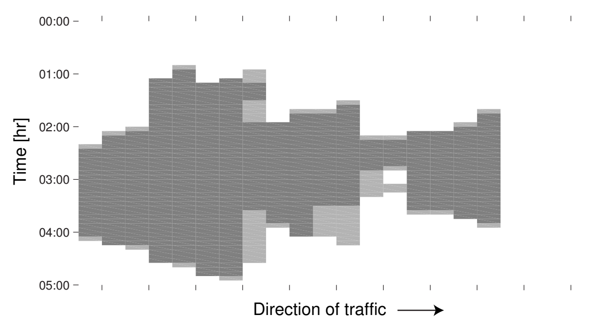

Fig. 3. Simulated uncontrolled contour

A manual calibration of the model parameters (v_i, w_i, $\bar{f}_i$, $\bar{s}_i$, $\bar{\rho}_i$, α_i, γ_i, ξ_i) was performed, with resulting speed contour plot shown in Figure 3. The speed variable used in Figure 3 was calculated with:

$$\text{speed}_i[k] \triangleq \frac{f_i[k]/\bar{\beta}_i[k]}{\rho_i[k] + \gamma_i\; r_i[k]/p}\left(\frac{L_i}{\Delta t_m}\right) \tag{23}$$

This equation ensures $\text{speed}_i[k] = v_i$ when the freeway section is free flowing.

The CWS and LP problems were solved for 1, 2 and 5-hour time horizons. In all cases, an additional half-hour “cool-down” period was appended to the end of the simulation period. The optimizations were performed over the entire 1.5, 2.5, and 5.5 hour time windows. During the cool-down period, all traffic demands were set to zero, and the freeway was allowed to empty completely.

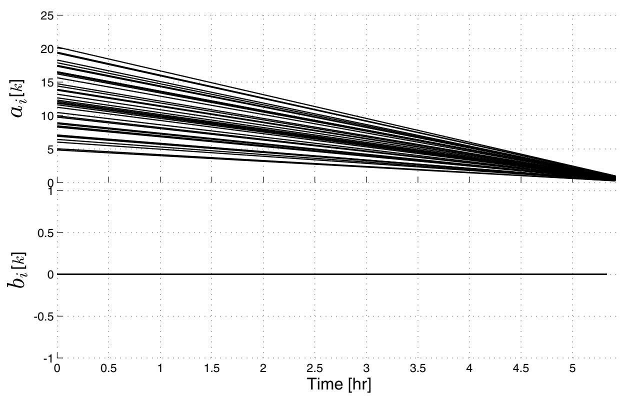

Fig. 4. Computed cost weights

The cost weights generated by the CWS problem, with Assumptions #1 and #2, are shown in Figure 4. As opposed to the result of Part I, these cost weights have the desired TTT-like properties of linear decay and $b_i[k] \equiv 0$. The optimal solution to Problem $\mathcal{P}_C$ posed with these weights is also a global solution to $\mathcal{P}_A$ (with the same objective), and probably a near-optimal controller in terms of TTT.

Problem $\mathcal{P}_C$ was solved using the commercial LP solver MOSEK 3.0. Each of the three time horizons was solved with and without on-ramp queue length restrictions (Eq. (15)). The size of the problem ranged from 92,310 constraints and 41,480 variables for the 1-hour problem without queue constraints, to 352,950 constraints and 158,600 variables for the 5-hour problem with queue constraints. Percent improvements in TTT are reported in Table I. TTT was calculated from the optimal solution with the following formula:

$$\text{TTT} = \sum_k \sum_{i \in \mathcal{I}} \rho_i[k] \Delta t_m + \sum_k \sum_{i \in \mathcal{E}n} l_i[k] \Delta t_o \qquad (24)$$

This computation included the cool-down period. It was confirmed in every case that the solution to the LP problem satisfied the equations of the model to a high degree of precision - i.e. $\psi^* \in \Omega_A$.

The validity of the two assumptions was also confirmed. For Assumption #1, it was verified that the optimal $r_i[k]$'s never exceeded $\xi_i p(\bar{\rho}_i - \rho_i[x])$. Assumption #2 was found to have little effect on the solution. This was confirmed by generating an *implementable* metering plan from the optimal solution using Eq. (22). The TTT for the implementable plan was found by running it through the model. TTT values for the optimal and implementable controllers are shown in Table I. These results show that increasing the minimum metering rate from 0 (optimal) to 180 (implementable) induces only a small loss in travel time savings (0.04% without queue constraints and 1.12% with queue constraints). It should be noted that applying Eq. (22) can never cause the queue constraint to be violated, since increasing the metering rate will only make the queues shorter.

TABLE I

TRAVEL TIME SAVINGS AND RUN TIMES

Period [hr:min]	TTT no control	TTT optimal	% saved	TTT implem.	% saved
Without queue constraints					
1:00	1,716	1,715	0.06%	1,716	0.00%
2:00	4,080	4,035	1.10%	4,036	1.08%
5:00	13,075	11,535	11.78%	11,540	11.74%
With queue constraints					
1:00	1,716	1,715	0.06%	1,716	0.00%
2:00	4,080	4,053	0.66%	4,060	0.49%
5:00	13,075	11,824	9.56%	11,971	8.44%

It is also interesting to note that the 1-hour and 2-hour time horizons yielded almost no improvement over no control. This is because, as can be seen in Figure 3, congestion only begins after the first hour, and starts to dissipate in the fourth hour. These two experiments tend to corroborate Zhang's observations in [5]. The 1-hour experiment demonstrates that an uncongested freeway should not be metered. In the 2-hour case there is no post-peak period. Hence, vehicles retained in the on-ramps cannot be released without increasing congestion. Not much can be gained by metering in these two situations. In the 5-hour experiment, however, the optimizer is able to shift the surplus demand to the post-peak period by holding vehicles on the on-ramps. Only the 5-hour time horizon produced a substantial improvement over no control: 11.78% travel time savings without queue constraints, and 9.56% with queue constraints. This result also emphasizes the importance of using a numerical technique that is efficient enough to produce optimal plans for sufficiently long time horizons, within a relatively short computation time.

Optimized speed contour plots and queue lengths for the 5-hour experiments are shown in Figures 5 and 6. Figure 5 shows that the optimal strategy, when on-ramp queue lengths are left unrestricted, is to keep the freeway almost completely uncongested by storing large numbers of vehicles in the on-ramps. In this situation, one of the on-ramp queues accumulates over 500 vehicles. Figure 6 shows that congestion cannot be avoided when the on-

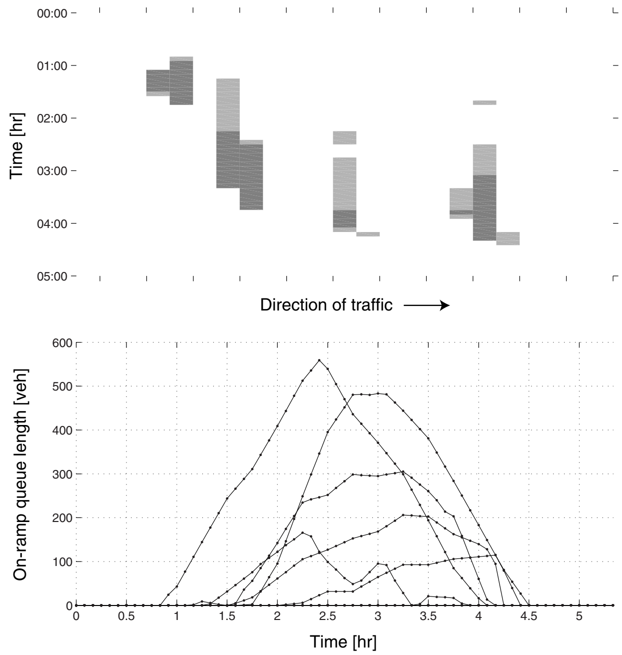

Fig. 5. Congestion and queue lengths without queue constraints.

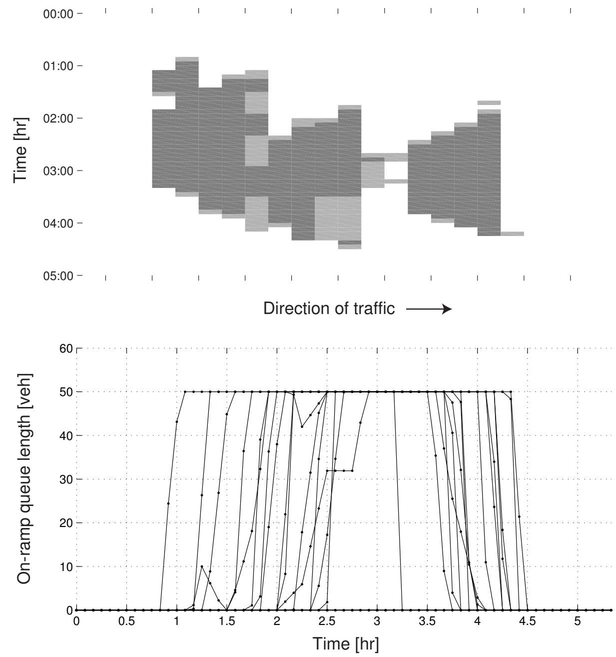

Fig. 6. Congestion and queue lengths with queue constraints.

ramp queues are limited to at most 50 vehicles. The implementable metering plan nevertheless achieves a reduction of 8.44% in TTT in this case.

V. Conclusions

This two part paper has outlined a complete methodology for solving the feedforward optimal metering problem efficiently and with near-global optimality. The unadulterated result was described in the first part, and was found to have a few drawbacks. This second part has focussed on solving those problems, and on testing the approach on realistic freeway setup. The repairs required the use of two additional assumptions. The first, that congestion does not propagate from the mainline to the on-ramps, can be easily verified in the optimal solution. Future work will consider the question of what to do if this assumption is temporarily violated. The second assumption was that the on-ramp flows could be reduced to zero by the on-ramp meters. This is never true, but it was found to require only a small sacrifice of global optimality. It was found that, under these two assumptions, the cost weights generated by the CWS problem were qualitatively similar to the values that minimize total travel time, in that 1) the weights on the mainline flows decayed linearly in time, and 2) the on-ramp flow weights were all zero. Also, the assumptions enabled the enforcement of queue length constraints, and allowed the formulation of a simplified but equivalent problem ($\mathcal{P}_C$).

This technique has several advantages over many other predictive on-ramp metering designs. First, it requires only to solve a single linear program, which can be done with extreme efficiency using any modern LP solver. Second, it takes on-ramp storage constraints explicitly into account. Finally, the optimal solution is a guaranteed global optimum with respect to a cost function that is qualitatively similar to total travel time. We envision this methodology as part of a larger and more robust traffic-responsive control structure. The complete freeway control system will include the optimizer within a “rolling-horizon” framework, and will update the model parameter values using the on-line parameter estimation of [6], [7].

References

[1] M. Zhang, S. Ritchie, and W. Recker. Some general results on the optimal ramp control problem. *Trans. Res. Part C.*, 4(2), 1996.

[2] J. Banks. Are minimization of delay and minimization of freeway congestion compatible ramp metering objectives? *Transportation Research Record*, 1727, 2000.

[3] G. Gomes and R. Horowitz. Globally optimal solutions to the onramp metering problem - Part I. *IEEE Conference on ITS*, 2004.

[4] G. Gomes, A.D. May, and R. Horowitz. Calibration of Vissim for a congested highway. In *83rd Annual Meeting of the TRB*, 2004.

[5] M. Papageorgiou. An integrated control approach for traffic corridors. *Transportation Research C*, 3(1), 1995.

[6] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez. Traffic density estimation with the cell transmission model. In *American Control Conference Proceedings*, 2003.

[7] X. Sun, L. Muñoz, and R. Horowitz. Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control. In *Proc. 42nd IEEE Conference on Decision and Control*, 2003.