Freeway Traffic Management Using Linear Programming

Denis Jacquet
Protoptim
1 rue lieutenant Chanaron
38000 Grenoble, France
(e-mail: jacquet@protoptim.fr).

Abstract: We present in this paper a linear programming framework to address freeway control applications such as ramp metering. After showing the equivalence between the LWR model and a linear optimization problem, several extensions are introduced to model the ramp queues and the capacity drop phenomenon. A wide range of objective functions which are relevant in traffic engineering are then introduced and several optimization-based control strategies are discussed. The effectiveness and versatility of this method is illustrated on a class of objective functions for a ramp metering problem.

Keywords: Transportation control, road traffic, control oriented models, linear programming, multiple-criterion optimization.

1. INTRODUCTION

Throughout the world, roadways are notorious for their congestions, from dense urban networks to large freeway systems. In addition, this situation seems to get worse over time, a tendency mainly due to the continuous increase of transportation demand. The most obvious impacts of traffic congestions are the increases of travel times, accidents and fuel consumption. An other critical and less obvious effect is that the infrastructures are not operated at their capacity when a congestion occurs. As a consequence, they serve less vehicles that they were designed for and the related investments are partially wasted. The main alternatives to reduce the global congestion bill are threefold: build more roads which is often not possible due to the lack of space and public acceptability, use less vehicles by promoting public transportation or improve the traffic operation, which is the topic of this paper. Freeways being less subject to the driver freewill than other roads, they constitute a natural setting to design Intelligent Transportation Systems (ITS) able to improve the infrastructure efficiency. Papageorgiou and Schmidt [1983] proposed an early contribution in this direction in the 80’s and several control strategies have been proposed since then. Ramp metering, which consists in controlling the amount of vehicles released on the freeway at on-ramps, is a control measure that have proven to be efficient in practice and is used in many states. Nevertheless, the field study of Levinson and Zhang [2006] claims that some improvements are still necessary so that real time algorithms really fulfill the traffic engineering requirements, in particular by balancing the mainlane and ramp delays. Relying on a suitable reinterpretation of the LWR macroscopic model introduced by Lighthill and Whitham [1955], we propose in this paper a generic linear programming framework able to tackle a whole family of freeway management problems. As an application, we present a ramp metering algorithm that precisely addresses the shortcomings mentioned by Levinson and Zhang [2006]. An other direct application is the production of guidelines for freeway dimensioning using the sensitivity information contained in the Lagrange multipliers. Working with linear programs is very comfortable here as it ensures that the optimal solution is global and it can be computed efficiently on a computer using largely available softwares. One of the main ingredients used to obtain a linear program is the fundamental diagram relaxation introduced by Gomes and Horowitz [2006]. This relaxation procedure is applied at the on-ramps as well in order to model correctly the ramp flows and the associated queue length. Moreover, given the importance of the capacity drop phenomenon at bottlenecks as illustrated by Cassidy and Bertini [1999], we introduce a linear capacity drop model that improve the relevance of the proposed ramp metering algorithm. One of the main property of our model is it ability to address weighted multi-objective criteria in the problem formulation, such as the Vehicle Miles Traveled (VMT), the Total Travel Time (TTT) on the mainlane, the Total Served Vehicle (TSV) at on-ramps and the Total Waiting Time (TWT) at ramps.

2. TRAFFIC FLOW AS A LINEAR PROGRAM

2.1 The LWR model and its discretization

The LWR model as introduced by Lighthill and Whitham [1955] is the simplest macroscopic freeway model. It takes the form of a nonlinear partial differential equation involving only the traffic density $\rho$ [veh/km] and the traffic flow $\phi$ [veh/h] distributions along the freeway. This models is parameterized by the fundamental diagram $\phi = \Phi(\rho)$, which is a phenomenological relationship between the density and the flow that may depend on space and time. The fundamental diagram is taken to be concave in general
and its maximum value $\Phi_m$ [veh/h] is called the capacity. Along with its initial and boundary conditions, the LWR model on an interval $x \in (x_L, x_R)$ writes

$$\begin{align*}
\partial_t \rho + \partial_x \Phi(\rho) &= 0 \\
\rho(0, x) &= \rho_0(x) \\
\rho(t, x_L) &= \rho_L(t) \quad \text{and} \quad \rho(t, x_R) = \rho_R(t)
\end{align*}$$

(1)

Though Equation (1) may look simple, it exhibits complex behaviors such as the interaction of forward and backward waves, the propagation of discontinuous shock waves and the possible non-applicability of boundary conditions. These features require to use specific numerical schemes as described by LeVeque [1992]. Introducing the grids $x = \{x_i\}_{i=1,...,N}$ and $t = \{t_k\}_{k=1,...,M}$ with cells of size $dx_i \times dt$, the discrete form of the LWR model can always be written

$$\begin{align*}
\rho^{k+1}_i &= \rho^k_i + \frac{dt}{dx_i} [\phi^{k+1}_{i-1} - \phi^k_i] \\
\rho^0_i &= \rho^k_i \\
\rho^k_i &= \rho^k_{i-1}, \quad \rho^k_N = \rho^k_R
\end{align*}$$

(2)

The discrete conservation law (2), with $\phi^k_i$ the flow at the interface between cells $i$ and $i + 1$ in the time interval $(t_k, t_{k+1})$, rules the evolution of a piecewise constant approximation of the solution of (1). Solving the LWR model thus consists in providing a suitable expression for $\phi^k_i$. The Godunov scheme (see LeVeque [1992]) is a classical choice for this class of equations, its convergence being guaranteed if $\min_i \{dx_i/dt\} > \max_i \{\Phi(\xi)\}$. The demand and supply paradigm introduced by Daganzo [1994] and Lebacque [1996] for concave flow diagrams is a interesting interpretation of the Godunov scheme often used in practice. Defining the demand as the increasing part of $\Phi(\rho)$ by

$$D(\rho) = \int_0^\rho \max(\Phi'(\xi), 0)d\xi$$

(3)

and the supply as its decreasing part by

$$S(\rho) = \Phi_{\text{max}} + \int_0^\rho \min(\Phi'(\xi), 0)d\xi$$

(4)

the Godunov flow can be shown to be equivalent to

$$\phi^k_i = \min \{D(\rho^k_{i-1}), S(\rho^k_{i+1})\}$$

(5)

In particular, the triangular fundamental diagram introduced by Daganzo [1994] gives the CTM model

$$\phi^k_i = \min \{v^k_i \rho^k_i, \rho^k_i c^k_i, \rho^k_{i+1} - w^k_i + \rho^k_{i+1} c^k_{i+1}\}$$

with $v^k_i$ the free flow speed, $w^k_i$ the congestion wave speed, $\rho^k_i$ the product of the maximal density with $c^k_i$ the capacity. The presence of the indices $i$ and $k$ on these parameters allow to model spatial topology changes along the freeway or to take into account an accident that alters the fundamental diagram in time and space.

More generally, the same procedure can be applied to any piecewise affine (PWA) fundamental diagram as the one illustrated on Figure 1. Using the notations of this figure and removing the possible dependency of the parameters on time and space for readability, the Godunov flow writes

$$\phi^k_i = \min \{a_1^k \rho^k_i, a_2^k \rho^k_i + b_2^k, ..., a_N^k \rho^k_i + b_N^k, b_0^k, a_1^k \rho^k_i + b_1^k, ..., a_N^k \rho^k_i + b_N^k\}$$

Fig. 1. Piecewise affine fundamental diagram.

2.2 Concave relaxation of the LWR model

A third interpretation of the Godunov scheme can be obtained by noticing that Equation (5) is equivalent to the maximization problem

$$\phi_i^k = \max \{\phi_i^k\}$$

with

$$\begin{align*}
\xi^k_i &\leq D(\rho^k_i) \\
\xi^k_i &\leq S(\rho^k_{i+1})
\end{align*}$$

Replacing $\xi^k_i$ by $\phi^k_i$ and disregarding the maximization problem, we obtain a concave relaxation similar to the one introduced by Gomes and Horowitz [2006] where the equality $\phi = \Phi(\rho)$ is replaced by the inequality $\phi \leq \Phi(\rho)$. This relaxation writes

$$\begin{align*}
\phi^k_i &\leq D(\rho^k_i) \\
\phi^k_i &\leq S(\rho^k_{i+1})
\end{align*}$$

and the correct fundamental diagram is recovered when the flow is maximized. Combining this concave relaxation with the time stepping given is Equation (2), the solution of the LWR model can be computed by solving the following optimization problem

$$\max \sum_{i,k} \phi^k_i$$

subject to

$$\begin{align*}
\phi^k_i &\leq D(\rho^k_i) \\
\phi^k_i &\leq S(\rho^k_{i+1}) \\
\rho^{k+1}_i &= \rho^k_i + \frac{dt}{dx_i} [\phi^{k+1}_{i-1} - \phi^k_i] \\
\rho^0_i &= \rho^k_i \\
\rho^k_i &= \rho^k_{i-1}, \quad \rho^k_N = \rho^k_R
\end{align*}$$

(6)

The traffic engineering interpretation of this reformulation is that traffic evolves such that the aggregated flow is maximized. Equation (6) is a variational formulation of the traffic dynamics and is a linear program for piecewise affine demand and supply functions. It will prove to be useful to treat freeway management problems by introducing additional control variables.

2.3 Extensions in the presence of on and off ramps

On and off ramps can be modeled by adding a set of exogenous flow contributions in the optimization framework described previously. Only the on-ramp case is presented in this section, the off-ramp case being treated without difficulty using exit flows or split ratios. To stay consistent with the original Godunov scheme, the ramp flow $r^k_i$ released on the freeway should enter at the cell boundaries as illustrated in Figure 2. In this setting, the demand at the on-ramp location becomes $D_r(\rho^k_i, r^k_i) = D(\rho^k_i) + r^k_i$ with $D(\rho^k_i)$ the classical demand defined in Equation (3). The
supply function (4) remains unchanged and the equations for the discrete model becomes

\[
\rho_i^{k+1} = \rho_i^k + \frac{dt}{dx_i} \left[ \phi_i^{k+1} - \phi_i^k \right]
\]

\[
\phi_i^k = \min \{ D(\rho_i^k) + r_i^k, S(\rho_i^{k+1}) \}
\]

\[
\phi_i^{k+1} = \phi_i^k - r_i^k
\]

with \(\phi_i^{k+1}\) and \(\phi_i^k\) respectively the flow entering and leaving cell \(i\) as illustrated on Figure 2. For more general freeway management applications, the traffic demands and the ramp queues should be modeled for each on-ramp. Given the substantial imprecisions of the field data and the freeway geometry, the model complexity should be kept to its strict minimum in order to facilitate the parameter tuning. To be relevant, the on-ramp model should:

- be equivalent to its LWR counterpart with no ramp,
- accounts for the on-ramp flow contribution,
- ensures that the maximal density is not exceeded,
- ensures that the queue length is always positive.

Several merge models have been proposed in the literature by Gomes and Horowitz [2006] and Daganzo [1995] to name a few but they usually require additional parameters that may be cumbersome to set in practice. The model presented in this section only relies on the flow conservation principle and a set of inequations that bound the flows involved in the system. The variables used in this model with a CTM fundamental diagram can be classified according to the following tables.

### Unknowns variables

| \(\rho_i^k\) | vehicle density on mainlane | veh/km |
| \(\phi_i^k\) | vehicle flow on mainlane | veh/h |
| \(r_i^k\) | vehicles released from the on-ramp | veh/h |
| \(q_i^k\) | vehicles queueing at the on-ramp | veh |

### Geometric and discretization parameters

| \(dx_i\) | cell length | km |
| \(I(j)\) | cell upstream of ramp \(j\) | |
| \(J(i)\) | ramp downstream of cell \(i\) | |
| \(dt\) | time step | hour |

### Constitutive parameters

| \(v_i^k\) | free wave speed | km/h |
| \(w_i^k\) | congestion wave speed | km/h |
| \(p_i^k\) | \(w_i^k \times (\text{maximal density})\) | veh/h |
| \(c_i^k\) | mainlane capacity | veh/h |
| \(f_i^k\) | onramp capacity | veh/h |

### Scenario parameters

| \(\rho_{i_0}^k\) | initial density condition | veh/km |
| \(\rho_{i_0}^k\) | upstream density condition | veh/km |
| \(\rho_{i_0}^k\) | downstream density condition | veh/km |
| \(q_i^k\) | initial queue condition | veh |
| \(d_i^k\) | traffic demand | veh/h |

In the above tables, the index \(i\) is for the mainlane cells, the index \(j\) for the ramps and the index \(k\) for the time sample. The scenario parameters, which are needed in the computations, are measured or estimated in practice. With the notations introduced above, the relaxed form of the extended model writes

\[
\max \sum \phi_i^k \quad \text{subject to}
\]

\[
\forall i: \begin{cases}
\phi_i^k & \leq \frac{q_i^k}{v_i^k} + r_i^k \\
\phi_i^k & \leq \rho_i^{k+1} + r_i^k \\
\phi_i^k & \leq \frac{q_i^k}{v_i^k} + w_i^k \\
\phi_i^k & \leq \rho_i^{k+1} + w_i^k
\end{cases}
\]

\[
\forall j: \begin{cases}
\phi_j^k & \leq \frac{q_j^k}{v_j^k} + d_j^k \\
\phi_j^k & \leq \frac{q_j^k}{v_j^k} + d_j^k \\
\phi_j^k & \leq \frac{q_j^k}{v_j^k} + d_j^k
\end{cases}
\]

The constraints on \(r_j^k\) in Equation (9) describe the admissible domain for the flow of vehicles released on the mainlane from ramp \(j\). First, it should be positive. Second, it should be applicable in the sense that it is smaller than the available supply. Third, each ramp flow is bounded by the ramp capacity noted \(F_j^k\) [veh/h]. Finally, the flow leaving a ramp at most empties its queue, which is the last inequality in (9). The inequalities in (8) are the ones in Equation (6) for the CTM fundamental diagram. Equations (10) and (11) implement the conservation of vehicles on the mainlane and on the ramps as time evolves.

### 2.4 Capacity drop model

When a congestion occurs at a bottleneck such as an on-ramp, it has been noticed by Cassidy and Bertini [1999] that the flow downstream of the congestion is often lower (up to 10%) than the freeway capacity. Taking this capacity drop phenomenon into account is particularly important when designing control strategies as this is precisely the bottleneck outflow that regulates its discharge. Considering the mutual embrace between the drivers coming from the mainlane and the on-ramp, Haut et al. [2005] proposed a capacity drop model able to predict flows lower than the capacity during congestions. We propose here a similar capacity drop model that can be used in the linear programming framework developed so far. First, the mutual embrace is taken into account by lowering the capacity when the on-ramp flow is larger than zero. Second, the capacity drop is assumed to increase with the mainlane demand, which seems reasonable from the traffic engineering perspective. With these assumptions, the capacity drop model for ramp \(j\) only consists in adding the following flow constraint

\[
\phi_{I(j)}^k \leq c_{I(j)}^{k+1} + \left( \phi_{I(j)}^k - \frac{\phi_{I(j)}^k - \phi_{I(j)}^k}{v_{I(j)}^k} \right)
\]

with \(z_j^k\) a parameter measuring the severity of the capacity drop phenomenon which should be set to reproduce field...
data near the on-ramps. Equation (12) means that the freeway capacity can be achieved as soon as the total demand is lower than capacity. Otherwise the capacity is decreased proportionally to the ramp flow and the mainlane flow.

3. TRAFFIC MANAGEMENT APPLICATIONS

3.1 Generic objective functions

Many traffic management problems can be casted as an optimization problem using the objective function

\[
\max_{(\rho, \phi, r, q)} \sum_{i,k} \alpha_i^k \rho_i^k + \sum_{i,k} \beta_i^k \phi_i^k + \sum_{j,k} \gamma_j^k r_j^k + \sum_{j,k} \delta_j^k q_j^k
\]

(13)

which is completely described by the value of the parameters \(\alpha_i^k, \beta_i^k > 0, \gamma_j^k\) and \(\delta_j^k\). The constraint on \(\beta_i^k\) is needed so that the solution fulfills the fundamental diagram relationship as discussed previously. Classical performance measures in traffic engineering are summarized in Table 1. The Vehicle Miles Traveled (VMT) and the Total Served Vehicles (TSV) criteria are related to the infrastructure efficiency whereas the Total Travel Time (TTT), the Total Waiting Time (TWT) and Total Time Spent (TSS) are indicators of the quality of service for the freeway users. Weighing these criteria using scalarization in a composite optimization formulation can anticipate the propagation of a congestion and lead to a local structure for freeways. We call this approach Local Instantaneous Control (LIC).

3.2 Control strategies

There are basically 2 practical ways of using an optimization algorithm to control a process:

(1) Either perform the optimization repetitively on a predefined finite time horizon, possibly with a forgetting factor. It is called Model Predictive Control (MPC).

(2) Or perform the optimization for the current time only, which leads to a local structure for freeways. We call this approach Local Instantaneous Control (LIC).

By considering a sufficiently long time horizon, the MPC method can anticipate the propagation of a congestion and take appropriate measures to reduce its effect. However, it requires the knowledge of the upstream and downstream conditions as well as the traffic demands at all ramps for the full time horizon. As these data should be predictable in practice, the validity of the MPC approach can be questionable for large horizons. On the other hand, the LIC method only needs to know in real time the traffic density upstream and downstream of the on-ramp, noted \(\rho_L\) and \(\rho_R\) respectively. The optimization problem writes

\[
\max \phi + \sigma r
\]

subject to

\[
\begin{align*}
0 \leq r & \leq \min \{ \rho_R - w_R (\rho_R + \phi) + f, q/dt + d \} \\
0 \leq \phi & \leq \min \{ v_{L}L, c_L \} + r \\
0 \leq \phi & \leq \min \{ \rho_L - w_L (\rho_L + \phi) + c_R \}
\end{align*}
\]

(15)

with \(\sigma\) a parameter weighting the ramp flow with respect to the mainlane flow. The solution of (14)-(15) can be computed explicitly which provides a closed-form controller. For instance, if we want to maximize the mainlane flow, the objective function should be \((\phi - r) + \phi\), which gives \(\sigma = -1/2\). In that case, the ramp flow is the maximum admissible flow that do not generate a congestion.

3.3 Sensitivity analysis

Even when no composite objective function is used, the optimization formulation (7)-(11) provides additional valu-
able information with respect to the more classical Godunov method. As for all convex optimization problems (Boyd and Vandenberghe [2004]), the constraints of this problem give rise to Lagrange multipliers, also called shadow prices or marginal prices. When dimensioning a freeway system, these Lagrange multipliers can be very useful in guiding traffic engineers as they tell how the relaxation of a given constraint would affect the objective function value. For instance, looking at the Lagrange multipliers associated to the supply inequality in (8) informs directly on the gains we can expect by adding a new lane.

3.4 Simulations for the ramp metering problem

We have developed a Java software called Karrus that implements the optimization problems described in this paper. As an illustration, we consider a virtual freeway section constituted of 3 links of 500 meters, each of them being discretized in 3 cells. The initial and boundary conditions are taken such that a congestion enters from the downstream boundary and two metered on-ramps are considered, one between each link. The time horizon is taken to be 5 minutes, which leads to a problem of 1495 variables that is solved in less than a second. Without ramp metering, the simulated density given in Figure 3 clearly shows the congestion propagation. When considering the VMT objective only, the optimizer is able to stop the congestion at the first on-ramp as shown in Figure 4. But Figure 5 shows that a large queue is building up at the second on-ramp. To alleviate the drawbacks of the VMT criterion, an objective function composed of the sum of the VMT and the TSV is considered. As shown in Figure 7, the queues are now emptied at the end of the optimization horizon but at the expense of a congestion on the mainlane, as illustrated on Figure 6. It is remarkable in this case that the optimizer behaves similarly to a bang-bang controller. Using a series of weights for the TSV objective, we obtain the discrete front of optimal solutions depicted on Figure 8. This figure provides a compact representation of the degree of freedom available to the designer to balance the different competing objectives.

4. CONCLUSION

We presented in this paper a linear programming framework which is relevant to treat freeway management prob-

Fig. 3. Density on the freeway without ramp metering.

Fig. 4. Density on the freeway with the VMT objective.

REFERENCES

Fig. 5. Ramp flows with the VMT objective.

Fig. 7. Ramp flows with the VMT and TSV objectives.

Fig. 6. Density with the VMT and TSV objectives.

Fig. 8. Evolution of the VMT and TSV costs with the weighting parameter (0 on the right and 1 on the left).