Delft University of Technology

Fac. of Information Technology and Systems **Control Systems Engineering**

Technical report bds:99-02

Optimal control of a class of linear hybrid systems with saturation*

B. De Schutter

If you want to cite this report, please use the following reference instead:
B. De Schutter, "Optimal control of a class of linear hybrid systems with saturation," *Proceedings of the 38th IEEE Conference on Decision and Control*, Phoenix, Arizona, pp. 3978–3983, Dec. 1999.

Control Systems Engineering
Faculty of Information Technology and Systems
Delft University of Technology
Delft, The Netherlands
Current URL: https://www.dcsc.tudelft.nl

Optimal control of a class of linear hybrid systems with saturation

Bart De Schutter
Control Laboratory, Fac. of Information Technology and Systems
Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands
email: `b.deschutter@its.tudelft.nl`

Abstract

We consider a class of queueing systems that can operate in several modes; in each mode the queue lengths exhibit a linear growth until a specified upper or lower level is reached, after which the queue length stays at that level until the end of the mode. We present some methods to determine the optimal switching time instants that minimize a criterion such as average queue length, worst case queue length, average waiting time, and so on. We show that if there is no upper saturation then for some objective functions the optimal switching scheme can be computed very efficiently.

1 Introduction

Hybrid systems arise from the interaction between continuous variable systems and discrete event systems. In general we could say that a hybrid system can be in one of several modes whereby in each mode the behavior of the system can be described by a system of difference or differential equations, and that the system switches from one mode to another due to the occurrence of events. There are many frameworks to model, analyze and control hybrid systems (see, e.g. [1, 7, 2]). We shall consider a special class of hybrid systems that can be analyzed using a special mathematical programming problem that is called the extended linear complementarity problem (ELCP).

This paper is organized as follows. In Section 2 we introduce the class of first order linear hybrid system with saturation. We derive a model that describes the queue lengths at the switching time instants. In Section 3 we show that computing the optimal switching time instants in general leads to an non-convex optimization problem or to an ELCP. Next we show that if there is no upper saturation then for some objective functions the optimal switching scheme can be computed very efficiently. Furthermore, by making some approximations the problem becomes a linear programming problem. The resulting approximate solution can be used as the initial point for solving the original optimization problem. We conclude this paper with an illustrative example.

2 A class of switched linear systems with saturation

Consider a system consisting of several queues. The evolution of the system is characterized by consecutive phases. In each phase or mode the length of each queue exhibits a linear growth or decrease until a certain upper or lower saturation level is reached; then the queue length stays constant until the end of the phase. A system the behavior of which satisfies this description will be called a *first order linear hybrid system with saturation*. A typical example of such a system is a traffic signal controlled intersection provided that we use a continuous approximation for the queue lengths (see Section 5 and [4]). For a traffic signal controlled intersection the lower bound for the queue lengths is 0. The upper bound could correspond to the maximal available storage space due to the distance to the preceding junction or to the layout of the intersection. We could assume that if this upper bound is reached then newly arriving cars take another route to get to their destination. Another example of a first order linear hybrid system with saturation is a system consisting of several fluid containers that are connected by tubes with valves and that have two outlets — one at the bottom (with a tube that leads to another fluid container), and one at the top (so that the fluid level in the containers can never exceed a given level), — provided that we assume that the increase or decrease of the fluid levels is linear if the system is not saturated.

In analogy with a traffic signal controlled intersection, we will use the word "queue lengths" to refer to the state variables of a first order linear hybrid system. Note however that our definition of a first order linear hybrid system with saturation is not limited to queuing systems only. Let M be the number of "queues". The length of queue i at time t is denoted by $q_i(t)$. Let $\alpha^{\mathrm{a}}_{i,k}$, $\alpha^{\mathrm{d}}_{i,k}$, $l_{i,k}$ and $u_{i,k}$ be respectively the arrival and tdeparture rate for queue i in phase k, and the lower and upper bound for the queue length q_i in phase k. The net queue length growth rate $\alpha_{i,k}$ for queue i in phase k is given by $\alpha_{i,k} = \alpha^{\mathrm{a}}_{i,k} - \alpha^{\mathrm{d}}_{i,k}$. The evolution of the system begins at time t_0. Let $t_1, t_2, \ldots$ be the time instants at which the system switches from one phase to another. The length of the kth phase is equal to $\delta_k = t_{k+1} - t_k$. We assume that $0 \leqslant l_{i,k+1} \leqslant q_i(t_{k+1}) \leqslant u_{i,k+1}$ for all i,k such that the queue lengths are always nonnegative and such that there are no sudden jumps in the queue lengths due a change in the saturation level at one of the switching time instants.

For queue i we have

$$\frac{dq_i(t)}{dt} = \begin{cases} \alpha_{i,k} & \text{if } l_{i,k} < q_i(t) < u_{i,k} \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

for $t \in (t_k, t_{k+1})$. This implies that the evolution of the queue lengths at the switching time instants is given by

$$q_i(t_{k+1}) = \max\big(\min(q_i(t_k) + \alpha_{i,k}\delta_k, u_{i,k}), l_{i,k}\big)$$

for $k = 0, 1, \ldots$ So if we define $q_{i,k} = q_i(t_k)$ and if we introduce the dummy variables $z_{i,k}$, we obtain

$$\begin{aligned} z_{i,k+1} &= \min(q_{i,k} + \alpha_{i,k}\delta_k, u_{i,k}) \\ q_{i,k+1} &= \max(z_{i,k+1}, l_{i,k}) \ . \end{aligned}$$

If we define column vectors $z_k, q_k, \alpha_k, l_k, u_k \in \mathbb{R}^{M\times 1}$ such that $(z_k)_i = z_{i,k}$, $(q_k)_i = q_{i,k}$, and so on, then this results in

$$z_{k+1} = \min(q_k + \alpha_k\delta_k, u_k) \tag{2}$$

$$q_{k+1} = \max(z_{k+1}, l_k) \tag{3}$$

for $k = 0, 1, \ldots$

3 Optimal switching schemes for linear hybrid systems with saturation

Let $N \in \mathbb{N}$. Now we want to compute an optimal switching sequence $t_0, \ldots, t_N$ that minimizes a criterion J such as:

- (weighted) average queue length over all queues:

$$J_1 = \sum_{i=1}^{M} w_i \frac{1}{t_N - t_0} \int_{t_0}^{t_N} q_i(t)\,dt \ , \tag{4}$$

- (weighted) average queue length over the worst queue:

$$J_2 = \max_i \left(w_i \frac{1}{t_N - t_0} \int_{t_0}^{t_N} q_i(t)\,dt \right) \ , \tag{5}$$

- (weighted) worst case queue length:

$$J_3 = \max_{i,t} \big(w_i q_i(t)\big) \ , \tag{6}$$

- (weighted) average "waiting time" over all queues:

$$J_4 = \sum_{i=1}^{M} w_i \frac{\displaystyle\int_{t_0}^{t_N} q_i(t)\,dt}{\displaystyle\sum_{k=0}^{N-1} \alpha^{\mathrm{a}}_{i,k}\delta_k} \ , \tag{7}$$

- (weighted) average "waiting time" over the worst queue:

$$J_5 = \max_i \left(w_i \frac{\displaystyle\int_{t_0}^{t_N} q_i(t)\,dt}{\displaystyle\sum_{k=0}^{N-1} \alpha^{\mathrm{a}}_{i,k}\delta_k} \right) \ , \tag{8}$$

where $w_i > 0$ for all i. We can impose extra conditions such as minimum and maximum durations for the switching time intervals, minimum or maximum queue lengths, and so on. This leads to the following problem:

$$\underset{\delta_0,\ldots,\delta_{N-1}}{\text{minimize}}\, J \tag{9}$$

subject to

$$\delta_{\min,k} \leqslant \delta_k \leqslant \delta_{\max,k} \quad \text{for } k = 0, \ldots, N-1, \tag{10}$$

$$q_{\min,k} \leqslant q_{k+1} \leqslant q_{\max,k} \quad \text{for } k = 0, \ldots, N-1 \tag{11}$$

$$z_{k+1} = \min(q_k + \alpha_k\delta_k, u_k) \quad \text{for } k = 0, \ldots, N-1, \tag{12}$$

$$q_{k+1} = \max(z_{k+1}, l_k) \quad \text{for } k = 0, \ldots, N-1. \tag{13}$$

where $\delta_{\min,k}$ ($\delta_{\max,k}$) is the minimum (maximum) length of the kth switching time interval (t_k, t_{k+1}), and $(q_{\min,k})_i$ $\big((q_{\max,k})_i\big)$ is the minimum (maximum) queue length for queue i at time instant t_{k+1}.

Now we discuss some methods to solve problem (9) – (13). First we consider (12) for an arbitrary index k. This equation can be rewritten as follows:

$$z_{k+1} \leqslant q_k + \alpha_k\delta_k$$

$$z_{k+1} \leqslant u_k$$

$$z_{i,k+1} = q_{i,k} + \alpha_{i,k}\delta_k \ \text{ or } \ z_{i,k+1} = u_{i,k} \quad \text{for all } i,$$

or equivalently

$$q_k + \alpha_k\delta_k - z_{k+1} \geqslant 0 \tag{14}$$

$$u_k - z_{k+1} \geqslant 0 \tag{15}$$

$$(q_k + \alpha_k\delta_k - z_{k+1})_i\,(u_k - z_{k+1})_i = 0 \qquad \text{for all } i. \tag{16}$$

Since a sum of nonnegative numbers is equal to 0 if and only if all the numbers are equal to 0, (16) is equivalent to:

$$(q_k + \alpha_k\delta_k - z_{k+1})^T\,(u_k - z_{k+1}) = 0 \ . \tag{17}$$

We can repeat this reasoning for (13) and for each index k. So if we define

$$x_q = \begin{bmatrix} q_1 \\ \vdots \\ q_N \end{bmatrix}, \ x_z = \begin{bmatrix} -z_1 \\ \vdots \\ -z_N \end{bmatrix}, \ x_\delta = \begin{bmatrix} \delta_0 \\ \vdots \\ \delta_{N-1} \end{bmatrix},$$

it is easy to verify that we finally get a problem of the form

$$\underset{x_\delta}{\text{minimize}}\, J \tag{18}$$

subject to

$$Ax_q + Bx_z + Cx_\delta + d \geqslant 0 \tag{19}$$

$$Ex_q + Fx_z + g \geqslant 0 \tag{20}$$

$$Hx_q + Kx_\delta + l \geqslant 0 \tag{21}$$

$$(Ax_q + Bx_z + Cx_\delta + d)^T(Ex_q + Fx_z + g) = 0 \ . \tag{22}$$

Equations (19), (20) and (22) correspond to (14), (15) and (17) respectively and to similar equations derived from (13), and the system of linear inequalities (21) contains the conditions (10) and (11).

The system (19) – (22) is a special case of an Extended Linear Complementarity Problem (ELCP) [3]. In [3] we have developed an algorithm to compute a parametric description of the complete solution set of an ELCP. Once this parametric description has been obtained, we can compute for which combination of the parameters the objective function J reaches a global minimum. Our computational experiments have shown that in most cases the determination of the optimal values of the parameters for the objective functions J_1, J_2, J_3, J_4 and J_5 is a well-behaved problem in the sense that using a local minimization routine starting from different initial points almost always yields the same numerical result (within a certain tolerance). However, the general ELCP is an NP-hard problem [3] and the algorithm of [3] to compute the solution set of a general ELCP requires exponential execution times. This implies that the full-ELCP approach sketched above is not feasible if the number of phases N is large. The following approaches can be used to compute suboptimal switching schemes for cases where the full-ELCP approach is not practicable:

- **multi-start local optimization**:
  The objective functions J_1 up to J_5 do not explicitly depend on x_q and x_z since for given q_0, $\alpha_{i,k}$'s, $l_{i,k}$'s and $u_{i,k}$'s, the components of x_q and x_z are uniquely determined by x_δ. Therefore we can consider (9) – (13) as a constrained optimization problem in x_δ where the constraints (11) – (13) are nonlinear constraints. Alternatively, these constraints can be taken into account by adding an extra penalty term to the objective function J if $q_{i,k} < (q_{\min,k})_i$ or $q_{i,k} > (q_{\max,k})_i$. If we use the penalty functions approach, the only remaining constraints on x_δ are the simple upper and lower bound constraints (10). The major disadvantage of these two approaches is that in general the minimization routine will only return a local minimum. Our computational experiments have shown that it is necessary to run the constrained local minimization algorithm several times — each time with a different initial starting point —in order to obtain the global minimum.
- **multi-ELCP approach**:
  If N is large, we could consider a smaller number N_s of phases, compute the optimal switching strategy for the first N_s phases using the full-ELCP method, implement the first step(s) of this strategy, afterwards compute the optimal switching strategy for the next N_s phases, implement the first step(s) of this strategy, and so on. We call this approach the *multi-ELCP approach*. Since the ELCPs for a horizon of N_s phases will be much smaller than the ELCP for N phases, the multi-ELCP approach will be tractable in practice even if N is large. Note that in general this approach will only give a suboptimal solution.

Note that we can also use a linear hybrid system with saturation as an approximate model if we have a hybrid system with saturation in which the queue length growth or decrease rates are slowly time-varying: we can approximate time-varying rate functions by piecewise-constant functions. Although in general we do not know the exact behavior of these functions in advance the behavior can often be predicted on the basis of historical data and measurements. Also note that we do not know the lengths of the phases in advance. In order to determine the average rates for each phase, we could therefore first assume that all phases have equal length. Then we compute an optimal or suboptimal switching scheme and use the result to get better estimates of the lengths of the phases and thus also of the average queue length growth rates in each phase, which can then be used as the input for another optimization run. If necessary we could repeat this process in an iterative way.

4 Optimal and suboptimal switching schemes for systems with saturation at a lower level only

In this section we consider linear hybrid systems with saturation at the lower level only (So $u_{i,k} = \infty$ for all i,k.) Furthermore, we assume that there are only upper bound constraints for the queue lengths; so we do not impose extra lower bound conditions on the queue lengths (or equivalently we assume that $q_{\min,k} \leqslant l_k$ for all k). In that case the optimal switching problem (9) – (13) reduces to

$$\underset{x_\delta}{\text{minimize}}\ J \tag{23}$$

subject to

$$\delta_{\min,k} \leqslant \delta_k \leqslant \delta_{\max,k} \quad \text{for } k = 0,\ldots,N-1, \tag{24}$$

$$q_{k+1} \leqslant q_{\max,k} \quad \text{for } k = 0,\ldots,N-1 \tag{25}$$

$$q_{k+1} = \max(q_k + \alpha_k \delta_k,\, l_k) \quad \text{for } k = 0,\ldots,N-1. \tag{26}$$

We call this problem $\mathscr{P}$. We define the "relaxed" problem $\tilde{\mathscr{P}}$ corresponding to the problem $\mathscr{P}$ as:

$$\underset{x_q, x_\delta}{\text{minimize}}\ J \tag{27}$$

subject to

$$\delta_{\min,k} \leqslant \delta_k \leqslant \delta_{\max,k} \quad \text{for } k = 0,\ldots,N-1, \tag{28}$$

$$q_{k+1} \leqslant q_{\max,k} \quad \text{for } k = 0,\ldots,N-1 \tag{29}$$

$$q_{k+1} \geqslant q_k + \alpha_k \delta_k \quad \text{for } k = 0,\ldots,N-1, \tag{30}$$

$$q_{k+1} \geqslant l_k \quad \text{for } k = 0,\ldots,N-1. \tag{31}$$

So compared to the original problem we have replaced (26) by relaxed equations of the form (14) – (15) without taking (16) or (17) into account. As a consequence, x_q and x_δ are not directly coupled any more. Note that in general it is easier to solve the relaxed problem $\tilde{\mathscr{P}}$ than the problem $\mathscr{P}$ since the set of feasible solutions of $\tilde{\mathscr{P}}$ is a convex set, whereas the set of feasible solutions of $\mathscr{P}$ is in general not convex since (26) is a non-convex constraint.

We say that the function J is a monotonic function of x_q if for every x_δ with positive components and for every $\tilde{x}_q, \hat{x}_q$ with $\tilde{x}_q \leqslant \hat{x}_q$, we have $J(\tilde{x}_q, x_\delta) \leqslant J(\hat{x}_q, x_\delta)$. The following proposition shows that for monotonic objective functions any optimal solution of the relaxed problem $\tilde{\mathscr{P}}$ can be transformed into an optimal solution of the problem $\mathscr{P}$.

Proposition 4.1 *Let the objective function J be a monotonic function of x_q and let (x_q^*, x_δ^*) be an optimal solution of $\tilde{\mathscr{P}}$. If we define $x_q^\sharp$ such that*

$$q_1^\sharp = \max\left(q_0 + \alpha_k \delta_0^*, l_0\right) \tag{32}$$

$$q_{k+1}^\sharp = \max\left(q_k^\sharp + \alpha_k \delta_k^*, l_k\right) \quad \textit{for } k = 1, \ldots, N-1. \tag{33}$$

then $(x_q^\sharp, x_\delta^*)$ is an optimal solution of the problem $\mathscr{P}$.

Proof: Let (x_q^*, x_δ^*) be an optimal solution of $\tilde{\mathscr{P}}$ and let $x_q^\sharp$ be defined by (32) – (33). Clearly, $(x_q^\sharp, x_\delta^*)$ is a feasible solution of $\tilde{\mathscr{P}}$. Since x_q^* satisfies (30) – (31), we have $q_{k+1}^* \geqslant \max(q_k^* + \alpha_k \delta_k^*, l_k)$ for all k. Hence, $q_1^\sharp \leqslant q_1^*$ and thus also $q_k^\sharp \leqslant q_k^*$ for all k. As a consequence, we have $x_q^\sharp \leqslant x_q^*$ and thus also $J(x_q^\sharp, x_\delta^*) \leqslant J(x_q^\sharp, x_\delta^*)$ since J is a monotonic function of x_q. Since $(x_q^\sharp, x_\delta^*)$ is a feasible solution of $\tilde{\mathscr{P}}$ and since (x_q^*, x_δ^*) is an optimal solution of $\tilde{\mathscr{P}}$, this implies that $(x_q^\sharp, x_\delta^*)$ is also an optimal solution of $\tilde{\mathscr{P}}$.
The set of feasible solutions of $\mathscr{P}$ is a subset of the set of feasible solutions of $\tilde{\mathscr{P}}$. Hence, the minimal value of J over the set of feasible solutions of $\tilde{\mathscr{P}}$ will be less than or equal to the minimal value of J over the set of feasible solutions of $\mathscr{P}$. Since $(x_q^\sharp, x_\delta^*)$ is a feasible solution of $\mathscr{P}$ and an optimal solution of $\tilde{\mathscr{P}}$, this implies that $(x_q^\sharp, x_\delta^*)$ is an optimal solution of $\mathscr{P}$. $\square$

Since the objective functions J_1, J_2, J_3, J_4 and J_5 do not explicitly depend on x_q, they are by definition monotonic functions of x_q. This implies that we can use Proposition 4.1 to transform the optimal switching problem for the objective functions J_1 up to J_5 into an optimization problem with a convex feasible set. Although the objective functions J_1 up to J_5 are in general not convex functions of x_δ, our computational experiments have shown that they are smooth enough, so that selecting different starting points for the local minimization routine almost always leads to more or less the same numerical result. Furthermore, the minimization routine converges quickly and the resulting solution is almost always optimal (see also Section 5).

In [4] we have considered a special class of first order linear hybrid systems with saturation at the lower level only. Although we did not yet use Proposition 4.1 there, we made some approximations that also lead to suboptimal switching schemes that can be computed very efficiently. The approximation techniques used in [4] can easily be extended to the class of first order linear hybrid systems with saturation at the lower level only that is considered in this paper. We shall present the main results of this technique here.

For a given q_0 and t_0, we define the function $\tilde{q}_i(\cdot, x_q, x_\delta)$ as the piecewise-affine function with breakpoints $(t_k, q_{i,k})$ for $k = 0, \ldots, N$. The approximate objective functions $\tilde{J}_1$, $\tilde{J}_2$, $\tilde{J}_3$, $\tilde{J}_4$ and $\tilde{J}_5$ are also defined by (4) – (8) but with q_i replaced by $\tilde{q}_i$. Now it can be shown that $\tilde{J}_1$ and $\tilde{J}_4$ are *strictly* monotonic functions of x_q and in that case any optimal solution of $\tilde{\mathscr{P}}$ is also an optimal solution of $\mathscr{P}$. We can even make a further approximation of $\tilde{J}_1$ and $\tilde{J}_4$ that will lead to a problem that can be solved very efficiently. Since $\tilde{q}_i$ is a piecewise-affine with breakpoints $(t_k, q_{i,k})$ for $k = 0, \ldots, N$, we have

$$\int_{t_k}^{t_{k+1}} \tilde{q}_i(t, x_q, x_\delta)\,dt = \frac{\delta_k}{2}\left(q_{i,k} + q_{i,k+1}\right) .$$

Hence,

$$\tilde{J}_1(x_q, x_\delta) = \sum_{i=1}^{M} w_i \frac{\sum_{k=0}^{N-1} \delta_k (q_{i,k} + q_{i,k+1})}{2(\delta_0 + \ldots + \delta_{N-1})} . \tag{34}$$

Sometimes we already have a good idea about the relative lengths of the different phases (in a traffic signal situation we know, e.g. that the green phases will be much longer than the amber phases). If we assume that $\delta_k = \rho_k \bar{\delta}$ for all k and for some yet unknown $\bar{\delta}$, then (34) leads to:

$$\tilde{J}_1(x_q, x_\delta) \approx \sum_{i=1}^{M} w_i \Big(\frac{\rho_0}{2R} q_{i,0} + \sum_{k=1}^{N-1} \frac{\rho_{k-1} + \rho_k}{2R} q_{i,k} + \frac{\rho_{N-1}}{2R} q_{i,N} \Big) \stackrel{\text{def}}{=} \hat{J}_1(x_q) .$$

with $R = \sum_{k=0}^{N-1} \rho_k$. Note that $\hat{J}_1$ is an affine function of x_q. We can use a similar reasoning to obtain an affine approximation of the objective function $\tilde{J}_4$. Since $w_i > 0$ for all i and $\rho_k > 0$ for all k, $\hat{J}_1$ and $\hat{J}_4$ are strictly monotonic functions of x_q. As a consequence, any optimal solution of $\tilde{\mathscr{P}}$ with objective function $\hat{J}_1$ and $\hat{J}_4$ will also be an optimal solution of $\mathscr{P}$. This implies that the optimal switching problem then reduces to a linear programming problem, which can be solved efficiently using a simplex method or an interior point method. Note that the assumption on the relative lengths is only used to simplify the objective function; it will not be included explicitly in the linear programming problem. As a consequence, the optimal δ_k's do not necessarily have to satisfy the assumption on the relative lengths.

Note that the approximate solutions obtained using the objective functions $\tilde{J}_l$ or $\hat{J}_l$ with $l \in \{1, 4\}$ can be used as starting points for a local minimization routine applied to the problem $\tilde{\mathscr{P}}$ with the objective function J_l.

5 Example

In order to illustrate the effectiveness of Proposition 4.1 we shall use the different approaches presented in this paper to design an optimal switching scheme for a traffic signal controlled intersection and compare the results.

We consider an intersection of two two-way streets (see Figure 1). There are four lanes L_1, L_2, L_3 and L_4, and on each corner of the intersection there is a traffic signal (T_1, T_2, T_3 and T_4). For each traffic signal there are three subsequent

Period	T_1	T_2	T_3	T_4
t_0–t_1	red	green	red	green
t_1–t_2	red	amber	red	amber
t_2–t_3	green	red	green	red
t_3–t_4	amber	red	amber	red
t_4–t_5	red	green	red	green
t_5–t_6	red	amber	red	amber
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 1: The traffic signal switching scheme.

phases: green, amber, and red. The switching scheme for the intersection is given in Table 1. Since all the cars will leave the queue in lane L_i provided that we make the length of the green phase in lane L_i large enough, we have $l_k = 0$ for all k. We assume that there is no saturation at the upper level, either due to the fact that there is enough buffer space before the traffic signal in each lane or due to the fact that we impose additional maximal queue length conditions such that $q_{\max,k} \leqslant u_k$.

In order to obtain a model that is amenable to mathematical analysis, we shall make two extra assumptions that will result in a simple model that can be analyzed very easily and for which we can efficiently compute (sub)optimal traffic signal switching schemes using the methods presented in Section 4. From now on we make the following assumptions (see also [4]):

- the queue lengths are continuous variables,
- the average arrival and departure rates of the cars are constant or slowly time-varying.

Let λ_i be the average arrival rate of cars in lane L_i, and let

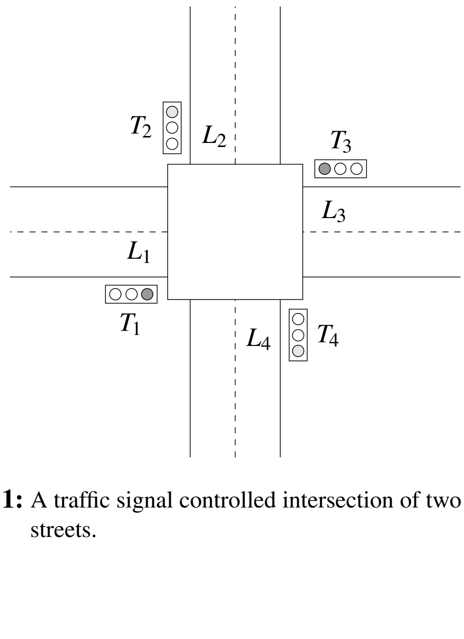

Figure 1: A traffic signal controlled intersection of two two-way streets.

μ_i^{green} (μ_i^{amber}) be the departure rate of cars in lane L_i when the traffic signal T_i is green (amber). If we define

$$\alpha_{i,k}^{\text{a}} = \lambda_i$$

$$\alpha_{i,k}^{\text{d}} = \begin{cases} 0 & \text{if } T_i \text{ is red in } (t_k, t_{k+1}) \\ \mu_i^{\text{green}} & \text{if } T_i \text{ is green in } (t_k, t_{k+1}) \\ \mu_i^{\text{amber}} & \text{if } T_i \text{ is amber in } (t_k, t_{k+1}) \end{cases}$$

for all i,k, then the relation between the switching time instants and the queue lengths is described by a system of equations of the form (26) and then we can use the techniques presented in Sections 3 and 4 to compute optimal and suboptimal traffic signal switching schemes.

Now consider the intersection of Figure 1 with the switching scheme of Table 1 and with the following data[1] $\lambda_1 = 0.24$, $\lambda_2 = 0.12$, $\lambda_3 = 0.17$, $\lambda_4 = 0.13$, $\mu_1^{\text{green}} = 0.5$, $\mu_2^{\text{green}} = \mu_4^{\text{green}} = 0.4$, $\mu_3^{\text{green}} = 0.45$, $\mu_1^{\text{amber}} = \mu_3^{\text{amber}} = 0.05$, $\mu_2^{\text{amber}} = \mu_4^{\text{amber}} = 0.03$, $q_0 = [\,21\ \ 17\ \ 14\ \ 9\,]^T$ and $q_{\max,k} = [\,25\ \ 20\ \ 25\ \ 20\,]^T$ for all k. The minimum and maximum length of the green phases are respectively 6 and 60, and the length of the amber phase is fixed at 3. Let $w = [\,2\ 1\ 2\ 1\,]^T$. We want to compute a traffic signal switching sequence $t_0, \ldots, t_7$ that minimizes J_4, the weighted average waiting time over all queues.

We have computed an optimal solution $x^*_{\delta,\text{ELCP}}$ obtained using the ELCP method, a solution $x^*_{\delta,\text{constr}}$ using constrained optimization with nonlinear constraints, a solution $x^*_{\delta,\text{penalty}}$ using constrained optimization with a penalty function, a multi-ELCP solution $x^*_{\delta,\text{multi}}$ with $N_\text{s} = 3$, a solution $x^*_{\delta,\text{relaxed}}$ for the relaxed problem $\tilde{\mathscr{P}}$ with objective function J_4, a solution $x^*_{\delta,\text{approx}}$ for the relaxed problem $\tilde{\mathscr{P}}$ with approximate objective function $\tilde{J}_4$, and a linear programming solution $x^*_{\delta^*,\text{linear}}$ that minimizes $\hat{J}_4$ with the linear objective function obtained by assuming that the length of the green phases is 10 times the length of the amber phases. In Table 2 we have listed the value of the objective function J_4 for the various switching interval vectors x^*_δ and the CPU time needed to compute the switching interval vectors on a Sun Ultra 10 300 MHz workstation with the optimization routines called from MATLAB and implemented in C or Fortran. The CPU time values listed in the table are average values over 10 experiments. For $x^*_{\delta,\text{constr}}$ and $x^*_{\delta,\text{penalty}}$ we have listed the best solution over respectively 5 and 20 runs[2] with random initial points; the indicated CPU time is the time needed for the total number of runs. For $x^*_{\delta,\text{relaxed}}$ and $x^*_{\delta,\text{approx}}$ different starting points always lead to more or less the same numerical value of the final objective function. Therefore, we have only performed one run with an arbitrary random initial point here.

In this example the ELCP solution is only given as a ref-

[1] All times will be expressed in seconds and all rates in vehicles per second.

[2] This choice for the number of runs is based on the typical variation in the value of the final values of objective values for the different runs.

x^*_δ	$J_4(x^*_\delta)$	CPU time
$x^*_{\delta,\mathrm{ELCP}}$	343.54	1824.83
$x^*_{\delta,\mathrm{constr}}$	343.54	144.35
$x^*_{\delta,\mathrm{penalty}}$	345.70	147.58
$x^*_{\delta,\mathrm{multi}}$	346.29	19.44
$x^*_{\delta,\mathrm{relaxed}}$	343.54	1.12
$x^*_{\delta,\mathrm{approx}}$	344.58	1.12
$x^*_{\delta,\mathrm{linear}}$	352.77	0.59

Table 2: The values of the objective functions J_4 and the CPU time needed to compute the (sub)optimal switching interval vectors of the example of Section 5.

erence since the CPU time needed to compute the optimal switching interval vector using the ELCP algorithm of [3] increases exponentially as the number of phases N increases. This implies that the full-ELCP approach should never be used in practice, but one of the other methods should be used instead. If we look at Table 2 then we see that if we take the trade-off between optimality and efficiency into account, then the $x^*_{\delta,\mathrm{relaxed}}$ solution — which is based on Proposition 4.1 — is clearly the most interesting.

In the simple traffic signal set-up discussed above we did not make a distinction between cars that turn left, right or that go straight ahead. However, the approach presented in this paper can also be applied to more complex traffic signal switching schemes such as, e.g. one consisting of four main phases with amber phases in between where in the first main phase cars on the north-south axis can go straight ahead or turn right, in the next main phase they can turn left, and in the next two main phases the same process is repeated for the traffic on the east-west axis. As a consequence, the method of this paper can also be used to efficiently compute suboptimal traffic signal switching schemes for a more complex intersection and/or switching scheme than the one presented above.

For more information on other traffic models and on traffic signal control the interested reader is referred to [5, 6, 8]. and to the references given therein.

6 Conclusions and future research

We have introduced a class of hybrid systems with first order linear dynamics subject to saturation and derived a model that describes the evolution of the queue lengths at the switching time instants. We have shown how the Extended Linear Complementarity Problem (ELCP) can be used to determine optimal switching schemes. We have also discussed several other techniques to compute (sub)optimal switching schemes for systems with saturation at the lower level only. If the objective function is a monotonic function of the queue lengths, then the optimal switching problem can be transformed into an optimization problem with a convex feasible set and then the optimal switching scheme can be computed very efficiently. We have illustrated these approach by computing (sub)optimal switching schemes for a traffic signal controlled intersection.

An important topic for future research is the extension of the results obtained in this paper to networks of dependent queues, i.e. to situations in which the outputs of some queues are connected to the inputs of some other queues. If we use a moving horizon strategy in combination with a *decentralized control* solution, we can still apply the approach given in this paper and use measurements from one queue to predict the arrival rates at the other queues provided that we know the routing rates and the traveling times from one queue to another. Other topics for further research include: development of other efficient algorithms and/or approximations to compute optimal switching strategies for other objective functions than the ones considered in this paper, development of efficient algorithms for the special cases of the ELCP that appear in the analysis of specific classes of first order linear hybrid systems with saturation, and investigation of the use of the ELCP and approximations to control other classes of hybrid systems.

References

[1] P. Antsaklis, W. Kohn, M.D. Lemmon, A. Nerode, and S. Sastry, eds., *Hybrid Systems V*, vol. 1567 of *Lecture Notes in Computer Science*, Berlin, Germany: Springer-Verlag, 1999. (Proceedings of the 5th International Hybrid Systems Workshop, Notre Dame, Indiana, Sept. 1997).

[2] P.J. Antsaklis and A. Nerode, eds., "Special issue on hybrid systems," *IEEE Transactions on Automatic Control*, vol. 43, no. 4, Apr. 1998.

[3] B. De Schutter and B. De Moor, "The extended linear complementarity problem," *Mathematical Programming*, vol. 71, no. 3, pp. 289–325, Dec. 1995.

[4] B. De Schutter and B. De Moor, "Optimal traffic light control for a single intersection," *European Journal of Control*, vol. 4, no. 3, pp. 260–276, 1998.

[5] N.H. Gartner, J.D.C. Little, and H. Gabbay, "Simultaneous optimization of offsets, splits, and cycle time," *Transportation Research Record*, no. 596, pp. 6–15, 1976.

[6] A.D. May, *Traffic Flow Fundamentals*. Englewood Cliffs, New Jersey: Prentice-Hall, 1990.

[7] A.S. Morse, C.C. Pantelides, S. Sastry, and J.M. Schumacher, eds., "Special issue on hybrid systems," *Automatica*, vol. 35, no. 3, Mar. 1999.

[8] E.S. Park, J.H. Lim, I.H. Suh, and Z. Bien, "Hierarchical optimal control of urban traffic networks," *International Journal of Control*, vol. 40, no. 4, pp. 813–829, Oct. 1984.