Delft University of Technology

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

Technical report bds:99-01

Optimizing acyclic traffic signal switching sequences through an extended linear complementarity problem formulation*

B. De Schutter

If you want to cite this report, please use the following reference instead:

B. De Schutter, "Optimizing acyclic traffic signal switching sequences through an extended linear complementarity problem formulation," *European Journal of Operational Research*, vol. 139, no. 2, pp. 400–415, June 2002.

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

*This report can also be downloaded via `https://pub.deschutter.info/abs/99_01.html`

Optimizing Acyclic Traffic Signal Switching Sequences through an Extended Linear Complementarity Problem Formulation

Bart De Schutter

Systems and Control Engineering
Faculty of Information Technology and Systems
Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands
tel: +31-15-278.51.13, fax: +31-15-278.66.79
email: b.deschutter@its.tudelft.nl

Keywords: Traffic, Control, Mathematical Programming, Optimization

Abstract

In this paper we first show how the Extended Linear Complementarity Problem, which is a mathematical programming problem, can be used to design optimal switching schemes for a class of switched systems with linear dynamics subject to saturation. More specifically, we consider the determination of the optimal switching time instants (the switching sequences are acyclic, but the phase sequence is pre-fixed). Although this method yields globally optimal switching time sequences, it is not feasible in practice due to its computational complexity. Therefore, we also discuss some approximations that lead to suboptimal switching time sequences that can be computed very efficiently and for which the value of the objective function is close to the global optimum. Finally we use these results to design optimal switching time sequences for a traffic signal controlled intersection so as to minimize criteria such as average queue length, worst case queue length, average waiting time, and so on.

1 Introduction

1.1 Overview

As the number of vehicles grows and the need for mobility increases, the frequency and duration of traffic jams in and around major cities increase. In the short term the most effective measures in the battle against traffic congestion seem to be a selective construction of new roads — an option which is often not viable due to lack of space and/or budgetary means — and a more efficient use of the existing infrastructure through traffic management. One of the strategies that is used in traffic management is traffic signal control to regulate access to highways or main roads (ramp metering) or to obtain a smoother or better circulation of traffic in urban areas (using, e.g., "green waves").

In this paper we study the design of optimal acyclic traffic signal switching time sequences for a traffic signal controlled intersection of several streets. We start from a model that describes the evolution of the queue lengths in the different lanes as a function of the switching time instants and the average arrival and departure rates of the cars, which can be measured, approximated or predicted using induction loop detectors, ultrasonic sensors or cameras, possibly in combination with historical data. In general, when we make a model of a system, an important issue to consider is the trade-off between accuracy of the model and the (computational) complexity of the analysis of the given model. In this paper we opt for an approximate model that can be analyzed easily and that allows us to efficiently design (sub)optimal traffic signal switching time sequences. The method proposed here is to be used for on-line optimization of switching time sequences. Therefore, we only consider a finite number of switchings in the optimization procedure and we combine this with a moving horizon approach. The latter also allows us to take changes in the average arrival and departure rates into account.

This paper is organized as follows. After introducing some notation in Section 1.2, we present the Extended Linear Complementarity Problem (ELCP) — which is a mathematical programming problem — in Section 2. In Section 3 we consider a general class of switched systems with linear dynamics subject to saturation and we show that the ELCP can be used to design optimal switching time sequences for this class of systems. Since in general the ELCP is an NP-hard problem, we consider a special subclass of switched systems with linear dynamics subject to lower saturation only and we show that for this class of systems we can derive very efficient methods to compute (sub)optimal switching time sequences. These results will then be used in Section 4 to efficiently design (sub)optimal traffic signal switching time sequences for traffic signal controlled intersections. We also illustrate our approach with a worked example. Finally we present some conclusions and topics for future research.

1.2 Notation

In this paper we use vector as a synonym for "column vector" or "matrix with one column".

Let a and b vectors with n components. We use a_i or $(a)_i$ to denote the ith component of a. We use $a \geqslant b$ to indicate that $a_i \geqslant b_i$ for all i. The maximum operator on vectors is defined as follows: $\big(\max(a,b)\big)_i = \max(a_i, b_i)$ for all i. The minimum operator on vectors is defined analogously. The zero vector with n components is denoted by 0_n, or by 0 if the dimension is clear from the context.

The set of the real numbers is denoted by $\mathbb{R}$, the set of the integers by $\mathbb{N}$, and the set of the positive integers by $\mathbb{N}_0$ (so $\mathbb{N}_0 = \{1, 2, 3, \dots\}$).

2 The Extended Linear Complementarity Problem

The Linear Complementarity Problem (LCP) is defined as follows [1]:

> Given $M \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$, find w, $z \in \mathbb{R}^n$ such that $w \geqslant 0$, $z \geqslant 0$, $w = q + Mz$ and $z^T w = 0$.

The LCP has numerous applications such as quadratic programming problems, the determination of the Nash equilibrium of a bimatrix game problem, the market equilibrium problem, the optimal invariant capital stock problem, the optimal stopping problem, etc. [1]. This makes the LCP one of the fundamental problems of mathematical programming.

The Extended Linear Complementarity Problem (ELCP) is an extension of the LCP and is defined as follows:

> Given $A \in \mathbb{R}^{p\times n}$, $B \in \mathbb{R}^{q\times n}$, $c \in \mathbb{R}^p$, $d \in \mathbb{R}^q$ and m subsets ϕ_1, ϕ_2, ..., ϕ_m of $\{1, 2, \ldots, p\}$, find $x \in \mathbb{R}^n$ such that
>
> $$\sum_{j=1}^{m} \prod_{i\in\phi_j} (Ax - c)_i = 0$$
>
> subject to $Ax \geqslant c$ and $Bx = d$, or show that no such x exists.

The formulation of the ELCP arose from our work in the study of discrete event systems, typical examples of which are flexible manufacturing systems, subway traffic networks, parallel processing systems, telecommunication networks and logistic systems. Many fundamental problems in the system theory of a subclass of discrete event systems — the max-plus-linear discrete event systems — can be recast as an ELCP [3, 4].

The ELCP can be considered as a system of linear equations and inequalities ($Ax \geqslant c$, $Bx = d$), where there are m groups of linear inequalities (one group for each index set ϕ_j) such that in each group at least one inequality should hold with equality. In [2] we have developed an algorithm to compute the complete solution set of an ELCP. Our algorithm yields a compact representation of the solution set of an ELCP by vertices, extreme rays and a basis of the linear subspace corresponding to the largest affine subspace of the solution set. In that way it provides a geometrical insight in the entire solution set of the ELCP and related problems. In [2] we have also shown that the general ELCP is NP-hard.

In the next section we shall show that the ELCP can be used to determine optimal switching time instants for a class of switched linear systems with saturation.

3 Optimal switching time sequences for a class of switched linear systems with saturation

3.1 Switched systems with linear dynamics subject to saturation

Consider a system consisting of several queues. The evolution of the system is characterized by consecutive phases. In each phase each queue length exhibits a linear growth or decrease until a certain upper or lower 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 switched system with linear dynamics subject to saturation — or *switched linear system with saturation* for short.

A typical example of a switched linear system with saturation is a traffic signal controlled intersection provided that we use a continuous approximation for the queue lengths (see Section 4) and that we assume that the arrival and departure rates of cars at the intersection is constant during the time period under consideration (see Remark 3.2 and Section 4). For a traffic signal controlled intersection the lower bound for the queue length is equal to 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 switched linear system with saturation is a system consisting of several

fluid containers that are connected by tubes with valves and that have an opening 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.

Now we derive the equations that describe the evolution of the queue lengths in a switched linear system with saturation. 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 rate for queue i in phase k, the departure rate for queue i in phase k, the lower bound for the queue length q_i in phase k and the upper bound for the queue length q_i in phase k. The 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, ... be the switching time instants, i.e., the time instants at which the system switches from one phase to another. The length of the kth phase is equal to $\delta_k \stackrel{\text{def}}{=} t_{k+1} - t_k$. Note that $\delta_k > 0$ for all 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 $t \in (t_k, t_{k+1})$ with $k \in \mathbb{N}$. This implies that the evolution of the queue lengths at the switching time instants is given by

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

for $k = 0, 1, 2, \ldots$ 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

$$z_k = \begin{bmatrix} z_{1,k} \\ z_{2,k} \\ \vdots \\ z_{M,k} \end{bmatrix}, \ q_k = \begin{bmatrix} q_{1,k} \\ q_{2,k} \\ \vdots \\ q_{M,k} \end{bmatrix}, \ \alpha_k = \begin{bmatrix} \alpha_{1,k} \\ \alpha_{2,k} \\ \vdots \\ \alpha_{M,k} \end{bmatrix}, \ l_k = \begin{bmatrix} l_{1,k} \\ l_{2,k} \\ \vdots \\ l_{M,k} \end{bmatrix}, \ u_k = \begin{bmatrix} u_{1,k} \\ u_{2,k} \\ \vdots \\ u_{M,k} \end{bmatrix},$$

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 $k = 0, 1, 2, \ldots$

3.2 Optimal switching time sequences

Now we consider the problem of determining optimal switching time instants. Since our approach is to be used for on-line optimization of switching time sequences, we only consider

a finite number of switchings in the optimization procedure, say N, in combination with a moving horizon approach (i.e., for a giving starting time t_0 we compute an optimal switching time sequence $t_1, t_2, \ldots, t_N$; then we fix the time instant at the first phase switching will occur at t_1; next we compute an optimal switching time sequence $t_2, t_3, \ldots, t_{N+1}$; we fix the time instant at the second phase switching will occur at t_2; and so on). Note that by using a moving horizon approach we can also take changes in the average arrival and departure rates into account. Note that in general the optimal switching time sequence will be acyclic, i.e., the optimal phase lengths will not be equal.

So now we consider the following problem: for a given integer N and a given starting time t_0 we want to compute an optimal sequence t_1, t_2, $\ldots$, t_N of switching time instants that minimizes a criterion 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:

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

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

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

where $w_i > 0$ for all i.

Remark 3.1 The reason for introducing the factor $t_N - t_0$ in J_1 and J_2 is that in our approach the time horizon is not fixed in advance. Using criteria based on time averaged values has the advantage that we maintain finite values for the objective functions even if N or t_N go to ∞ (provided that the queue lengths remain finite).

We can impose extra conditions such as minimum and maximum durations for the switching time intervals, minimum or maximum queue lengths (which could be useful in order to prevent saturation at the lower or upper level for some queues), and so on. This leads to the following problem:

$$\underset{\delta_0,\delta_1,\ldots,\delta_{N-1}}{\text{minimize}} \; J \tag{9}$$

subject to

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

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

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

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

where $\delta_{\min,k}$ and $\delta_{\max,k}$ are respectively the minimum and the maximum length of the kth switching time interval (t_k, t_{k+1}) for queue i, and $(q_{\min,k})_i$ and $(q_{\max,k})_i$ are respectively the minimum and the maximum queue length for queue i at time instant t_{k+1}. Since the entries of q_k correspond to queue lengths, we may assume without loss of generality that $q_{\min,k} \geqslant 0$ for all k.

If the lower and upper bounds for the queue lengths are finite, then the feasible set (10)–(13) of the optimization problem is not convex since the evolution equations (12)–(13) contain both minimization and maximization operations. Furthermore, this also implies that the queue lengths are not a convex function of the phase lengths. Hence, the objective functions J_1 up to J_5 are not convex. As a consequence, the optimization problem (9)–(13) is a nonlinear non-convex optimization problem. In general such a problem can have many local minima and is considered to be hard to solve. In the next section we discuss some methods to solve the optimization problem (9)–(13).

3.3 The Extended Linear Complementarity Problem and optimal switching time sequences

Now we show that the system (10) – (13) can be reformulated as an ELCP. First consider (12) for an arbitrary index k. This equation can be rewritten as follows:

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

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

$$z_{i,k+1} = q_{i,k} + \alpha_{i,k}\delta_k \quad \text{or} \quad z_{i,k+1} = u_{i,k} \qquad \text{for } i = 1, 2, \ldots, M\,,$$

or equivalently

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

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

$$(q_k + \alpha_k \delta_k - z_{k+1})_i \, (u_k - z_{k+1})_i = 0 \qquad \text{for } i = 1, 2, \ldots, M \tag{16}$$

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:

$$\sum_{i=1}^{M} (q_k + \alpha_k \delta_k - z_{k+1})_i \, (u_k - z_{k+1})_i = 0$$

or

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

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

$$x_q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{bmatrix}, \quad x_z = \begin{bmatrix} -z_1 \\ -z_2 \\ \vdots \\ -z_N \end{bmatrix}, \quad x_\delta = \begin{bmatrix} \delta_0 \\ \delta_1 \\ \vdots \\ \delta_{N-1} \end{bmatrix},$$

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

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

subject to

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

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

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

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

with appropriately defined matrices A, B, C, E, F, H, K, and vectors d, g, l. Equations (19), (20) and (22) correspond to (14), (15) and (17) respectively, and the system of linear inequalities (21) contains the conditions (10) and (11). It is easy to verify that the system (19) – (22) is (a special case of) an ELCP. It is easy to verify that additional constraints such as, e.g., a (maximum) total duration for the N phases ($\delta_0 + \delta_1 + \cdots + \delta_N = T_{\text{tot}}$ or ($\delta_0 + \delta_1 + \cdots + \delta_N \leqslant T_{\max}$) or (maximum) total durations for two consecutive phases ($\delta_{2k} + \delta_{2k+1} = T_k$ or $\delta_{2k} + \delta_{2k+1} = T_{\max,k}$), will still lead to an ELCP.

In order to determine the optimal switching time sequence we could minimize the objective function J over the solution set of the ELCP (19) – (22) as follows. If we assume that x_q and x_δ are bounded — note that a sufficient condition for this is that $\delta_{\max,k}$ is defined and finite for all k, — then the solution set of the system (19) – (22) consists of a union of faces of the polyhedron defined by (19) – (21). Each face of the polyhedron can be represented by its vertices, and the points of the face can be written as convex combinations of these vertices. We could for each face determine for which convex combination of the vertices the objective function J reaches a global minimum over the face and afterwards select the overall minimum. Our computational experiments have shown that in most cases the determination of the minimum value of the objective functions given above is a well-behaved problem in the sense that using a local minimization routine (that uses, e.g., sequential quadratic programming) starting from different initial points almost always yields the same numerical result (within a certain tolerance).

The algorithm of [2] 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. Therefore, we shall now discuss two approaches to compute suboptimal solutions in a reasonable amount of time. In the next section we shall discuss some other approaches to compute approximate solutions if there is no saturation at the upper level.

The following approaches can be used to compute suboptimal switching time sequences for cases where the full-ELCP approach is not tractable:

- **multi-start local optimization**:
  The objective functions defined in Section 3.2 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.
  So if there are no extra constraining bounds on the q_k's (i.e., if $(q_{\min,k})_i \leqslant l_{i,k}$ and $(q_{\max,k})_i \geqslant u_{i,k}$ for all i,k), then problem (9) – (13) reduces to a constrained optimization problem in x_δ, which could be solved using a constrained local minimization algorithm. Note that the constraints on x_δ are simple upper and lower bound constraints on the components of x_δ.
  If $(q_{\min,k})_i > l_{i,k}$ or $(q_{\max,k})_i < u_{i,k}$ for some pairs of indices (i,k), we can still use this constrained local minimization approach by adding an extra penalty term to J if $q_{i,k} < (q_{\min,k})_i$ or $q_{i,k} > (q_{\max,k})_i$.
  The major disadvantage of this approach 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. This suboptimal solution can be used as the starting point for a local minimization routine applied to the original full problem.

In practice there is always some uncertainty and variation in time of the queue length growth rates, which makes that in general computing the exact optimal switching time sequence is utopian. Moreover, in practice we are more interested in quickly obtaining a good approximation of the optimal switching time sequence than in spending a large amount of time to obtain the exact optimal switching time sequence. In the next section we shall consider a subclass of switched linear systems that only have saturation at a lower bound. In that case it is possible to make some extra approximations that lead to very efficient algorithms to compute suboptimal switching time sequences.

Remark 3.2 Note that we can also use a switched linear system with saturation as an approximate model if we have a switched system subject to 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 time sequence 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. Note that this iterative procedure is an heuristic approach, which is not guaranteed to converge.

3.4 Suboptimal switching time sequences for systems with saturation at a lower level only

In this section we consider switched linear systems with saturation at the lower level only. So $u_{i,k}$ is equal to ∞ for all i, k, or equivalently $(q_{\max,k})_i \leqslant u_{i,k}$ for all i, k, i.e., the switching times will be selected such that the upper saturation level will never be reached. Furthermore, we assume that $q_{\min,k} \leqslant l_k$ for all k, i.e., we do not impose extra lower bound conditions on the queue lengths.

If there is no upper saturation and if there are no extra lower bound conditions on the queue lengths, the optimal switching problem (9) – (13) reduces to

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

subject to

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

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

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

We call this problem $\mathcal{P}$.

Recall that the objective functions J_1, J_2, J_3, J_4 and J_5 defined in Section 3.2 do not explicitly depend on x_q since for given q_0, $\alpha_{i,k}$'s and $l_{i,k}$'s the components of x_q are uniquely determined by x_δ. The approximate objective functions that we will introduce next will depend explicitly on x_q and x_δ.

For a given q_0 and t_0, we define the function $\tilde{q}_i(\cdot, x_q, x_\delta)$ — or $\tilde{q}_i(\cdot)$ for short — as the piecewise-linear function that interpolates in the points $(t_k, q_{i,k})$ for $k = 0, 1, \ldots, N$. The approximate objective functions $\tilde{J}_1$, $\tilde{J}_2$, $\tilde{J}_3$, $\tilde{J}_4$ and $\tilde{J}_5$ are defined as in Section 3.2 but with q_i replaced by $\tilde{q}_i$. It is easy to verify that the values of J_3 and $\tilde{J}_3$ coincide. Now let $l \in \{1, 2, 4, 5\}$. Note that the value of the objective functions J_l and $\tilde{J}_l$ depends on the surface under the functions q_i and $\tilde{q}_i$ respectively. In a situation such as the one represented in Figure 1 where the queue lengths alternatively decrease and increase the surface under the function $\tilde{q}_i$ will be a reasonable approximation of the surface under the function q_i. Note that this will certainly be the case for a traffic signal controlled intersection where the traffic signals alternate between green and red (with a short amber phase in between) since an optimal traffic signal switching time sequence implies the absence of long periods in which no cars wait in one lane while in the other lanes the queue lengths increase. So if we have an optimal traffic signal switching time sequence, then the periods during which the queue length in some lane is equal to 0 are in general short and then the optimal value of $\tilde{J}_l$ will be a good approximation of the optimal value of J_l.

Now we show that the use of the approximate objective functions $\tilde{J}_1$ or $\tilde{J}_4$ leads to an optimization problem that can be solved more efficiently than the original problem in which J_1 or J_4 is used. We define the "relaxed" problem $\tilde{\mathcal{P}}$ corresponding to the problem $\mathcal{P}$ as:

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

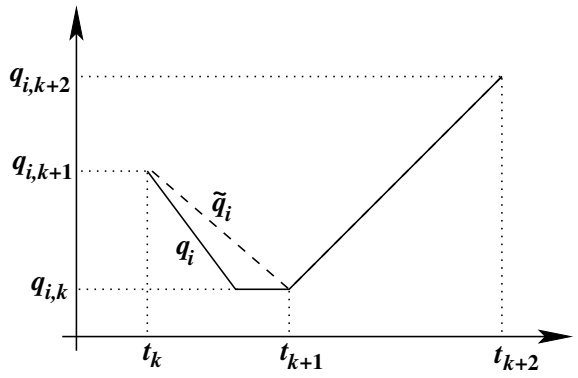

Figure 1: The functions q_i (full line) and $\tilde{q}_i$ (dashed line) for a queue with a decrease phase and a subsequent positive growth phase and without saturation at an upper level. Note that in the growth phase the functions q_i and $\tilde{q}_i$ coincide.

subject to

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

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

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

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

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 $\tilde{\mathcal{P}}$ than the problem $\mathcal{P}$ since the set of feasible solutions of $\tilde{\mathcal{P}}$ is a convex set, whereas the set of feasible solutions of $\mathcal{P}$ is in general not convex since it consists of a union of faces of the polyhedron defined by the system of inequalities (28) – (31).
The following propositions show that for the objective functions $\tilde{J}_1$ and $\tilde{J}_4$ any optimal solution of the relaxed problem $\tilde{\mathcal{P}}$ is also an optimal solution of the problem $\mathcal{P}$. Another advantage of $\tilde{J}_1$ and $\tilde{J}_4$ is that using formula (32) given below we can compute the gradient of these objective functions analytically which in general speeds up the minimization algorithm significantly.

Proposition 3.3 *If J is a strictly monotonous function of x_q — i.e., if for any x_δ with positive components and for all $\tilde{x}_q, \hat{x}_q$ with $\tilde{x}_q \leqslant \hat{x}_q$ and with $(\tilde{x}_q)_j < (\hat{x}_q)_j$ for at least one index j, we have $J(\tilde{x}_q, x_\delta) < J(\hat{x}_q, x_\delta)$ — then any optimal solution of the relaxed problem $\tilde{\mathcal{P}}$ is also an optimal solution of the problem $\mathcal{P}$.*

Proof: Let (x_q^*, x_δ^*) be an optimal solution of $\tilde{\mathcal{P}}$. Now we show by contradiction that if J depends strictly monotonously on x_q, then (x_q^*, x_δ^*) also satisfies (26), i.e., (x_q^*, x_δ^*) is a feasible solution of $\mathcal{P}$.
Assume that (x_q^*, x_δ^*) does not satisfy (26). Let p be the smallest index such that

$$q_{p+1}^* \geqslant \max\left(q_p^* + \alpha_p \delta_p^*,\, l_p\right) \quad \text{and} \quad (q_{p+1}^*)_i = q_{i,p+1}^* \neq \max\left(q_{i,p}^* + \alpha_{i,p}\delta_p^*,\, l_{i,p}\right)$$

for some index i.
Now define $x_\delta^\sharp = x_\delta^*$ and $x_q^\sharp$ such that

$$\begin{aligned} q_k^\sharp &= q_k^* && \text{for } k = 0, 1, \ldots, p, \\ q_{k+1}^\sharp &= \max\left(q_k^\sharp + \alpha_k \delta_k^\sharp, l_k\right) && \text{for } k = p, p+1, \ldots, N-1. \end{aligned}$$

Note that $q_{p+1}^\sharp \leqslant q_{p+1}^*$. Since $\delta_k^\sharp = \delta_k^*$ for all k, this implies that $q_k^\sharp \leqslant q_k^*$ for $k = p+2, p+3, \ldots, N$. Hence, $x_q^\sharp \leqslant x_q^*$. Since $q_k^\sharp \leqslant q_k^* \leqslant q_{\max,k}$ for all k, this implies that $(x_q^\sharp, x_\delta^\sharp)$ is also a feasible solution of $\tilde{\mathcal{P}}$.
Since $x_q^\sharp \leqslant x_q^*$ and $(x_q^\sharp)_j < (x_q^*)_j$ for at least one index $j = Mp + i$ and since $x_\delta^\sharp = x_\delta^*$, we have $J(x_q^\sharp, x_\delta^\sharp) < J(x_q^*, x_\delta^*)$ which implies that (x_q^*, x_δ^*) is not an optimal solution of $\tilde{\mathcal{P}}$. Since this is a contradiction, our initial assumption that (x_q^*, x_δ^*) does not satisfy (26) is wrong. Hence, $(x_q^*,\ x_\delta^*)$ also is a feasible solution of the problem $\mathcal{P}$ and since the set of feasible solutions of $\mathcal{P}$ is a subset of the set of feasible solutions of $\tilde{\mathcal{P}}$ this implies that $(x_q^*,\ x_\delta^*)$ is also an optimal solution of $\mathcal{P}$. □

Proposition 3.4 *For given q_0, $\alpha_{i,k}$'s, $l_{i,k}$'s and a given x_δ the functions $\tilde{J}_1$ and $\tilde{J}_4$ are strictly monotonous functions of x_q.*

Proof: Let $x_\delta \in \mathbb{R}^N$ with $x_\delta > 0$ and let $x_q^*, x_q^\sharp \in \mathbb{R}^{MN}$ with $x_q^*, x_q^\sharp \geqslant 0$. If $x_q^\sharp \leqslant x_q^*$ and $(x_q^\sharp)_j < (x_q^*)_j$ for at least one index $j = M(k-1) + i$ with $i, k \in \mathbb{N}_0$ then we have $\tilde{q}_i(t, x_q^\sharp, x_\delta^\sharp) \leqslant \tilde{q}_i(t, x_q^*, x_\delta^*)$ for all $t \in [t_0, t_N]$, and $\tilde{q}_i(t, x_q^\sharp, x_\delta^\sharp) < \tilde{q}_i(t, x_q^*, x_\delta^*)$ in some non-empty interval $(t_k - \eta, t_k + \eta) \cap [t_0, t_N]$ with $\eta > 0$. Hence,

$$\int_{t_0}^{t_N} \tilde{q}_i(t, x_q^\sharp, x_\delta^\sharp)\, dt \ < \ \int_{t_0}^{t_N} \tilde{q}_i(t, x_q^*, x_\delta^*)\, dt \ ,$$

which implies that $\tilde{J}_1(x_q^\sharp, x_\delta^\sharp) < \tilde{J}_1(x_q^*, x_\delta^*)$ and $\tilde{J}_4(x_q^\sharp, x_\delta^\sharp) < \tilde{J}_4(x_q^*, x_\delta^*)$. □

It is easy to verify that $\tilde{J}_2$, $\tilde{J}_3$ and $\tilde{J}_5$ are in general not *strictly* monotonous functions of x_q for a given x_δ. Although in general $\tilde{J}_1$ and $\tilde{J}_4$ are not convex functions of x_q and x_δ, our computational experiments have shown that the objective functions $\tilde{J}_1$ and $\tilde{J}_4$ are smooth enough, so that selecting different starting points for the optimization routine almost always leads to more or less the same numerical result (see also [5]).

We shall now discuss a further approximation of $\tilde{J}_1$ and $\tilde{J}_4$ that will lead to a problem that can be solved very efficiently. Since

$$\int_{t_k}^{t_{k+1}} \tilde{q}_i(t, x_q, x_\delta) dt \ = \ \frac{\delta_k}{2}(q_{i,k} + q_{i,k+1}) \ , \tag{32}$$

we have

$$\tilde{J}_1(x_q, x_\delta) \ = \ \sum_{i=1}^{M} w_i \, \frac{\displaystyle\sum_{k=0}^{N-1} \delta_k (q_{i,k} + q_{i,k+1})}{2(\delta_0 + \delta_1 + \ldots + \delta_{N-1})} \ . \tag{33}$$

If we assume that $\delta_k \approx \frac{t_N - t_0}{N}$ for all k then (33) leads to:

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

Note that $\hat{J}_1$ is an affine function of x_q. Since scaling of an affine objective function and subtracting a constant term $\left(\sum_{i=1}^{M} \frac{w_i}{2N} q_{i,0} \right)$ does not change the location of the minimum, this implies that we can minimize $J_{\text{lin}} = \omega^T x_q$ with

$$\omega = \left[w_1 \ w_2 \ \ldots w_M \ w_1 \ w_2 \ \ldots \ w_M \ w_1 \ w_2 \ \ldots \ w_M \ \frac{w_1}{2} \ \frac{w_2}{2} \ \ldots \ \frac{w_M}{2} \right]^T$$

in order to find a minimum of $\hat{J}_1$. We can use a similar reasoning to obtain a linear approximation of the objective function $\tilde{J}_4$. Since $w_i > 0$ for all i, J_{lin} is a strictly monotonous function of x_q. As a consequence, Proposition 3.3 implies that for $J = J_{\text{lin}}$ problem $\mathcal{P}$ reduces to a linear programming problem, which can be solved efficiently using (variants of) the simplex method or using an interior point method (see, e.g., [10]).

It is easy to verify that the reduction to a linear objective function is still possible if instead of assuming all δ_k's to be equal, we impose relative durations for the lengths of the phases, i.e., we select coefficients ρ_k such that $\delta_k = \rho_k \bar{\delta}$ for some, yet unknown, $\bar{\delta}$. The assumption considered above that all δ_k's are equal would correspond to $\rho_k = 1$ for all k. 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 (see, e.g., Example 4.2).

Remark 3.5 Note that the approximate solutions obtained using the methods presented in this section can be used as starting points for a local minimization routine applied to the problem $\tilde{\mathcal{P}}$ (with the objective functions $\tilde{J}_1$ or $\tilde{J}_4$) or to the original problem $\mathcal{P}$ (with the objective functions J_1 or J_4).

4 Optimal traffic signal control

In this section we show that the models and methods presented in this paper can be used to compute optimal traffic signal switching time sequences. We shall first illustrate this for a simple set-up consisting of an intersection of two two-way streets (see Figure 2). 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 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 $l_k = q_{\text{min,k}} = 0$ 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 $q_{\text{max},k} \leqslant u_k$.

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 time sequences using the methods presented in Section 3.4. From now on we make the following assumptions:

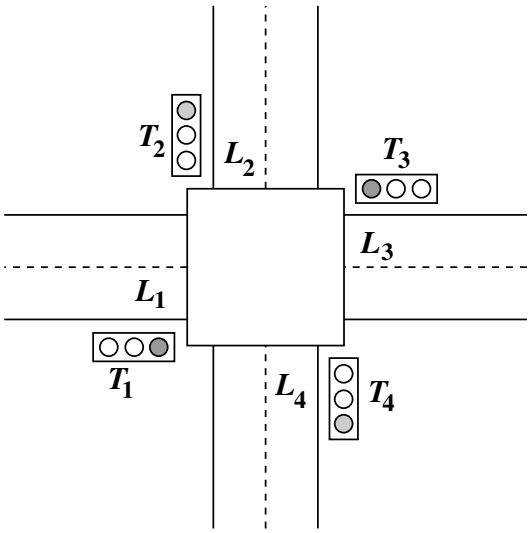

Figure 2: A traffic signal controlled intersection of two two-way streets.

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.

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

These assumptions deserve a few remarks:

- Recall that the main purpose is to compute optimal traffic signal control schemes. Designing optimal traffic signal switching schemes is only useful if the arrival and departure rates of vehicles at the intersection are high. In that case, approximating the queue lengths by continuous variables only introduces small errors. Furthermore, there is also some uncertainty and variation in time of the arrival and departure rates, which makes that in general computing the exact optimal traffic signal switching time sequence is utopian. Moreover, in practice we are more interested in quickly obtaining a good approximation of the optimal traffic signal switching time sequence than in spending a large amount of time to obtain the exact optimal traffic signal switching time sequence.

- If we keep in mind that one of the main purposes of the model that we shall derive, is the design of optimal traffic signal switching time sequences, then assuming that the average arrival and departure rates are constant is not a serious restriction, provided that we use a *moving horizon strategy*: we compute the optimal traffic signal switching time sequence for, say, the next $N = 10$ phases, based on a prediction of the average arrival and departure rates (using data measured during the previous phases) and we apply this scheme during the first of the 10 phases, meanwhile we update our estimates of the arrival and departure rates and compute a new optimal switching time sequence for the next 10 phases, and so on.
  We can also combine this approach with the iterative procedure discussed in Remark 3.2.

Let λ_i be the average arrival rate of cars in lane L_i, and let μ_i^{green} and μ_i^{amber} be the departure rates of cars in lane L_i when the traffic signal T_i is green respectively amber. If we define

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

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 3.2 and 3.4 to compute optimal and suboptimal traffic signal switching time sequences.

Remark 4.1 Note that the model derived above can accommodate varying amber durations. However, in many countries the amber time is fixed by regulation (e.g., to 3 s in France). If we assume that the duration of the amber phase is fixed, then we can adapt our model and reduce the number of variables (see also [5]).

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 set-ups or more complex traffic signal switching schemes such as, e.g., the one depicted in Figure 3 which consists 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.

The following traffic signal control example illustrates that using the approximations for the objective function J_1 that have been introduced in Section 3.4 leads to good suboptimal solutions that can be computed very efficiently. All times will be expressed in seconds and all rates in vehicles per second. The numerical results will be given up to 3 decimal places unless explicitly stated otherwise.

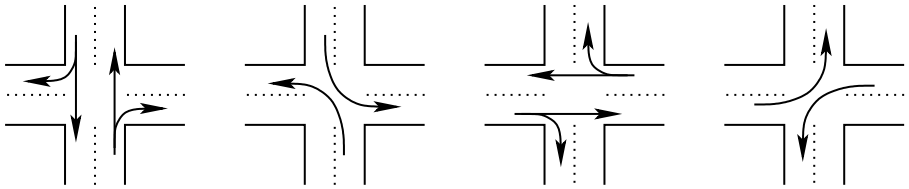

Figure 3: The four main phases of a more complex traffic signal switching scheme. The arrows indicate possible directions for the cars that receive a green signal.

Example 4.2 Consider the intersection of Figure 2 with the switching scheme of Table 1 and with the following data: $\lambda_1 = 0.22$, $\lambda_2 = 0.11$, $\lambda_3 = 0.19$, $\lambda_4 = 0.13$, $\mu_1^{\text{green}} = \mu_3^{\text{green}} = 0.51$, $\mu_2^{\text{green}} = \mu_4^{\text{green}} = 0.42$, $\mu_1^{\text{amber}} = \mu_3^{\text{amber}} = 0.05$, $\mu_2^{\text{amber}} = \mu_4^{\text{amber}} = 0.03$, $t_0 = 0$, $q_0 = [\,21\ \ 16\ \ 9\ \ 7\,]^T$ 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 minimum and maximum length of the amber phases are respectively 2 and 5. Let $w = [\,2\ \ 1\ \ 2\ \ 1\,]^T$.

Suppose that we want to compute a traffic signal switching sequence $t_1, t_2, \ldots, t_{10}$ that minimizes J_1 — the weighted average queue length over all queues.
We have computed an optimal solution $x^*_{\delta,\text{ELCP}}$ using the full-ELCP method, a suboptimal solution $x^*_{\delta,\text{penalty}}$ using constrained optimization with a penalty function for queue lengths that exceed $q_{\max,k}$, a multi-ELCP suboptimal solution $x^*_{\delta,\text{multi}}$ with $N_\text{s} = 5$, a solution $x^*_{\delta,\text{approx}}$ that minimizes the approximate objective function $\tilde{J}_1$ (using the relaxed problem $\tilde{\mathcal{P}}$), and a linear programming solution $x^*_{\delta,\text{linear}}$ (also using the relaxed problem $\tilde{\mathcal{P}}$ and with the linear objective function obtained by assuming that the length of the green phases is 10 times the length of the amber phases). This results in

$$
\begin{aligned}
x^*_{\delta,\text{elcp}} &= [\,10.226\ \ 3.000\ \ 60.000\ \ 3.000\ \ 43.188\ \ 3.000\ \ 60.000\ \ 3.000\ \ 52.496\ \ 3.000\,]^T \\
x^*_{\delta,\text{penalty}} &= [\,10.354\ \ 3.000\ \ 60.000\ \ 3.000\ \ 43.063\ \ 3.000\ \ 60.000\ \ 3.000\ \ 51.846\ \ 3.000\,]^T \\
x^*_{\delta,\text{multi}} &= [\,10.226\ \ 3.000\ \ 60.000\ \ 3.000\ \ 43.188\ \ 3.000\ \ 60.000\ \ 3.000\ \ 31.818\ \ 3.000\,]^T \\
x^*_{\delta,\text{approx}} &= [\,10.226\ \ 3.000\ \ 60.000\ \ 3.000\ \ 43.188\ \ 3.000\ \ 59.245\ \ 3.000\ \ 44.189\ \ 5.000\,]^T \\
x^*_{\delta,\text{linear}} &= [\,15.182\ \ 3.000\ \ 60.000\ \ 3.000\ \ 38.232\ \ 3.000\ \ 59.245\ \ 3.000\ \ 6.000\ \ 3.000\,]^T
\end{aligned}
$$

Note that for the optimal linear solution $x^*_{\delta,\text{linear}}$ the 10 to 1 ratio for the lengths of the green versus the amber phases does not hold. The reason for this is that the assumption on the relative lengths of the green and amber phases was only used to simplify the objective function and was not included explicitly in the linear programming problem.

In Table 2 we have listed the values of the objective functions J_1, $\tilde{J}_1$ and $\hat{J}_1$ 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 640 MB RAM and 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 $x^*_{\delta,\text{penalty}}$ we have listed the best solution over 20 runs with random initial points (The mean of the objective values of the local minima returned by the minimization routine was 50.123 with a standard deviation of 3.275); the indicated CPU time is the time needed for the 20 runs. For $x^*_{\delta,\text{approx}}$ different starting points always lead to more or less the same numerical value of the optimal objective function (In an experiment with 20 random starting points the first 12 decimal places of the

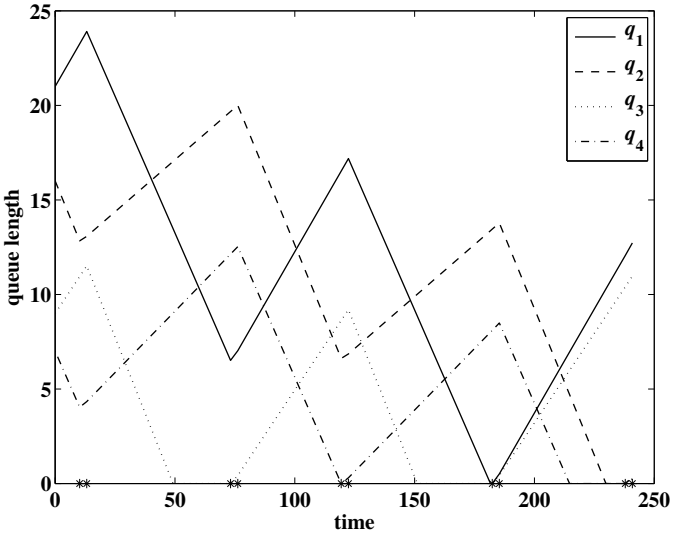

Figure 4: The queue lengths in the various lanes as a function of time for the traffic signal switching sequence that corresponds to the switching interval vector $x^*_{\delta,\mathrm{ELCP}}$ of Example 4.2. The * signs on the time axis correspond to the switching time instants.

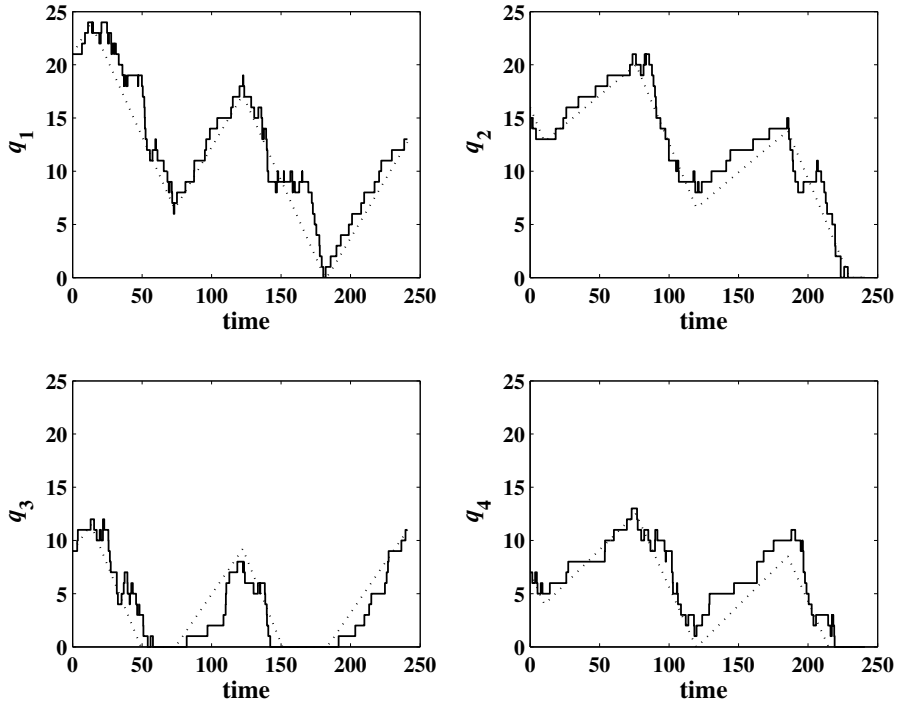

Figure 5: The queue lengths in the various lanes as a function of time for an integer queue length simulation for the traffic signal switching sequence that corresponds to the switching interval vector $x^*_{\delta,\mathrm{ELCP}}$. The integer queue length functions are plotted in full lines and their continuous approximations in dotted lines.

final objective function always had the same value). 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 reference since the CPU time needed to compute the optimal switching interval vector using the ELCP algorithm of [2] 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 suboptimal methods given in Sections 3.3 and 3.4 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 $x^*_{\delta,\text{approx}}$ solution is clearly the most interesting.

The evolution of the queue lengths for the optimal control strategy that corresponds to $x^*_{\delta,\text{ELCP}}$ is represented in Figure 4. In Figure 5 we have plotted the results of an integer queue length simulation for the traffic signal switching strategy that corresponds to optimal switching interval vector $x^*_{\delta,\text{ELCP}}$. The effective average queue length over all lanes for this simulation is 50.642.

The main purpose of the computational simulations of this example is to illustrate the effectiveness of using Propositions 3.3 and 3.4 to transform the original, hard optimization problem into a relaxed, easy optimization problem. Simulation results for the objective function J_4 lead to similar results as for J_1. Note that we have not considered the objective functions J_2, J_3 and J_5 since for these objective functions we cannot apply Proposition 3.3. □

For more information on other models for the evolution of the queue lengths at a traffic signal controlled intersection and on optimal traffic signal control the interested reader is referred to [6, 7, 8, 9, 11, 12] and to the references given therein. The main difference between the traffic model presented in this paper and the models used by most other researchers is that in our approach the green-amber-red cycles lengths may vary from cycle to cycle, which — in contrast to a fixed cycle length — adds an extra degree of freedom to obtain an optimal switching time sequence.

x^*_δ	$J_1(x^*_\delta)$	$\tilde{J}_1(x^*_\delta)$	$\hat{J}_1(x^*_\delta)$	CPU time
$x^*_{\delta,\text{ELCP}}$	47.367	50.402	55.294	58601.61
$x^*_{\delta,\text{penalty}}$	47.376	50.385	55.229	57.76
$x^*_{\delta,\text{multi}}$	48.105	50.774	53.871	240.19
$x^*_{\delta,\text{approx}}$	47.497	50.153	54.533	2.87
$x^*_{\delta,\text{linear}}$	51.160	53.941	52.798	1.84

Table 2: The values of the objective functions J_1, $\tilde{J}_1$ and $\hat{J}_1$ (up to 3 decimal places) and the CPU time (up to 2 decimal places) needed to compute the (sub)optimal switching interval vectors of Example 4.2.

5 Conclusions and future research

We have introduced the Extended Linear Complementarity Problem (ELCP) and indicated how it can be used to determine optimal switching time sequences for a class of switched systems with linear dynamics subject to saturation. Since the ELCP is NP-hard, we have also discussed several techniques to efficiently compute suboptimal and approximate switching time sequences. Next we have applied these results to design (sub)optimal switching time sequences for traffic signal controlled intersections.

In this paper we have derived methods to optimize performance measures such as average or worst case waiting times and queue lengths for a switched system with linear dynamics subject to saturation. So we have considered quantitative properties of the system. If we are more interested in qualitative properties such as, e.g., safety, we could use the techniques presented in [13].

An important topic for future research is the extension of the results obtained in this paper to networks of dependent queues, i.e., a situation where the outputs of some queues will be connected the inputs of some other queues. If we use a moving horizon strategy in combination with a *decentralized control* solution, we can apply still 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 r_{ij} (i.e., the amount of cars, fluid, ... that will be routed from the output of queue i to the input of queue j) and the traveling times t_{ij} from one queue to another.

Also note that we have assumed that the phase sequence order was pre-fixed. We could use the switching time optimization presented in this paper as an inner loop in a discrete optimization outer loop that also optimizes the phase order.

Other topics for further research include: development of other efficient algorithms and/or approximations to compute optimal switching strategies for the class of systems discussed in this paper, development of efficient algorithms for the special cases of the ELCP that appear in the analysis of specific classes of switched linear systems with saturation, and investigation of the use of the ELCP to model and to control other classes of switched systems.

References

[1] R.W. Cottle, J.S. Pang, and R.E. Stone, *The Linear Complementarity Problem*. Boston: Academic Press, 1992.

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

[3] B. De Schutter and B. De Moor, "Minimal realization in the max algebra is an extended linear complementarity problem," *Systems & Control Letters*, vol. 25, no. 2, pp. 103–111, May 1995.

[4] B. De Schutter and B. De Moor, "A method to find all solutions of a system of multivariate polynomial equalities and inequalities in the max algebra," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 6, no. 2, pp. 115–138, Mar. 1996.

[5] 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.

[6] 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.

[7] H.R. Kashani and G.N. Saridis, "Intelligent control for urban traffic systems," *Automatica*, vol. 19, no. 2, pp. 191–197, Mar. 1983.

[8] J.H. Lim, S.H. Hwang, I.H. Suh, and Z. Bien, "Hierarchical optimal control of oversaturated urban traffic networks," *International Journal of Control*, vol. 33, no. 4, pp. 727–737, Apr. 1981.

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

[10] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, Pennsylvania: SIAM, 1994.

[11] 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.

[12] M.G. Singh and H. Tamura, "Modelling and hierarchical optimization for oversaturated urban road traffic networks," *International Journal of Control*, vol. 20, no. 6, pp. 913–934, 1974.

[13] H. Wong-Toi, "The synthesis of controllers for linear hybrid automata," in *Proceedings of the 36th IEEE Conference on Decision and Control*, San Diego, California, pp. 4607–4612, Dec. 1997.