Optimization of signal timing transition period

R. Mussa & M. F. Selekwa
FAMU-FSU College of Engineering, USA

Abstract

This paper reports on the development of a transition procedure based on the quadratic optimization method. The procedure is aimed at reducing disutility measures to motorists during the transition period. The transition is modeled as a linear dynamic process, and the disutility measures are modeled as the sum of squares of the deviations of the coordination parameters—that is, cycle length, phase split, and offset—from the optimal values during the transition. Optimal control techniques are used to determine the step size and optimal number of steps necessary to complete the transition with minimum disruption to traffic flow. The proposed transition procedure does not go beyond the current and the target cycle lengths thus eliminating the need for the user to specify minimum and maximum values of splits and cycle lengths to accommodate pedestrians and other local intersection needs.

A simulation study showed that the proposed transition method reduced the queue delay for minor streets in one scenario by an average of 5.88 percent, compared to the immediate transition method embedded in CORSIM. However, the method slightly increased delay on the major street by approximately 1.43% compared to CORSIM’s immediate method. The paper further discusses techniques that could improve the performance of the proposed algorithm under various geometric and traffic conditions.

Keywords: signal timing, quadratic optimization, signalized intersections, simulation.

1 Introduction

Proper handling of the transition between signal timing plans is important because improper coordination parameters during the transition can cause traffic to operate suboptimally during the transition period, resulting in increased delays and stops for vehicles. A number of studies have documented the deleterious
effects of the transition [1,2,3,4,5]. In addition, Basu [1] reported that under changing conditions of demand, a timing plan change is warranted only if the benefits provided by a new, more appropriate signal timing plan outweigh the deleterious effects of transitioning between timing plans. Also, suboptimal coordination parameters during the transition can cause traffic perturbations that can outlast the transition period itself.

2 Defining coordination parameters

Before developing the proposed transition optimization procedure, in this section the coordination parameters are first defined. The parameters that generally define coordination are cycle length \( (c) \), phase split \( (s) \), and offset \( (o) \). The cycle length is the time needed to complete a rotation through all phases. The choice of the cycle length is an optimization problem that seeks to maximize the bandwidth or seeks to minimize a combination of disutility measures. The cycle length is typically in the range of \( 0.75C_o \leq c \leq 1.5C_o \) where \( C_o \) is the optimum cycle length for the intersection determined through an optimization procedure. The phase split is the percentage of cycle length allocated to a particular movement or phase while offset is defined as the start of green time of the coordinated phase relative to the start of green time at a reference intersection.

The majority of intersections have left, through, and right turn movements on the major street and sometimes a similar number of movements on the minor street. Most modern traffic signal controllers are capable of pairing nonconflicting phases and flexibly transferring the green from one approach to another using dual ring phasing. Depending on the volume of traffic, the controller may terminate the green time early for one conflicting movement and overlap the second high-volume conflicting phase with the next nonconflicting movement. Thus, a controller can theoretically implement from one to over eight phases per cycle depending on a number of factors including presence of pedestrians, phase sequencing, and other local geometric and traffic factors. Normally, nonconflicting movements are made to run simultaneously; therefore, the relationship between the cycle length \( (c) \) and the phase splits \( (s_i) \) for \( n \) coordinated movements with \( m \) nonconflicting movements is

\[
c = \left( \sum_{i=1}^{n} s_i - \sum_{j=1}^{m} s_j \right) + (n-m) \gamma ,
\]

(1)

where \( s_j, j = 1,2,\ldots,m \) are the short phase splits for the nonconflicting movements that run simultaneously. The relationship in eqn (1) can be modified in different ways depending on the signal timing engineer’s preference. Changes in one phase split \( s_i \) can also affect other phase splits \( s_j, j \neq i \) or the node
cycle $c$ in conformity with eqn (1). The problem addressed in this paper is that of finding an optimal method of changing from one set of coordination parameters to the next set.

3 Modeling and optimization of the transitioning

The main results of this paper are presented in this section. The section begins by discussing the modeling of the transition as a dynamic system. Then it applies the principles of optimal control to determine the optimal transitioning procedure.

3.1 Modeling

The problem of optimizing the transitioning is viewed here as that of determining the optimal path for moving the coordination parameters from one timing plan to another in such a way that disruption of traffic flow is minimized and all operational constraints defined in the preceding subsection are satisfied. This is a dynamic optimization problem, see [6, 7], which can be formulated to have an explicit solution by using linear quadratic methods [7, 8]. To formulate the problem in linear quadratic terms, define the signal coordination parameter vector (SCPV) $q_i$ for each node $i$ with $n_i$ controlled movement as

$$q_i = \begin{bmatrix} o_i & s_{i_1} & s_{i_2} & \cdots & s_{i_{n_i}} & c_i \end{bmatrix},$$

where $o_i$ is the offset, $s_{i_j}$ are the phase splits and $c_i$ is the cycle length. For generality, it is assumed that these parameters are different for each signal and $c_i$'s are integer common factors. If the network has $n$ controlled signals, then the network coordination parameter vector (NCPV) $x$ is formed by concatenating all the SCPVs in the network, i.e.,

$$x = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix},$$

Let the NCPV for the current timing plan be $x_j$ and that of the next plan be $x_{j+1}$. The transition from $x_j$ to $x_{j+1}$ follows the evolution described by

$$x(k+1) = x(k) + \Delta x(k),$$

where $x(0) = x_j$ and $\Delta x(k)$ is the gradual change in the NCPV. It is desired to find an optimal sequence of $\Delta x(k)$ and the number of steps $m+1$ which minimize the traffic disruption such that $x(m) = x_{j+1}$.

Since both $x(0) = x_j$ and $x(m) = x_{j+1}$ are known optimal plans, we express the optimality of the transitioning in terms of the squares of deviations from these optimal values. This is an abstract quantity chosen to reflect the penalty imposed to the traffic because of the signal parameters not being at the planned values. At any instant $k$, the deviation from the terminal values is given by...
\[ x(k) - x(0) = \sum_{i=0}^{k-1} \Delta x(i), \]  
(5) 

and

\[ x(m) - x(k) = \sum_{i=k}^{m-1} \Delta x(i). \]  
(6) 

The cost function for optimality \( J \)—which is the sum of the squares of the deviations—becomes

\[ J = \sum_{i=0}^{k-1} \Delta x(i)^T \Delta x(i) + \sum_{i=k}^{m-1} \Delta x(i)^T \Delta x(i), \]

\[ = \sum_{i=0}^{m-1} \Delta x(i)^T \Delta x(i). \]  
(7) 

The optimization problem now is that of finding \( \Delta x(k) \), and the number of steps \( m \) which minimize the squares of the deviations in eqn (7). The next section describes how this problem is solved.

### 3.2 Optimization

Before expressing this problem as a linear quadratic optimization problem, it is worth mentioning that since the cycle length for each node is an affine combination of the phase splits and some phase splits run simultaneously, the SCPV and NCPV defined in the previous section contain redundant elements that need not be optimized separately. In particular, the optimal cycle update \( \Delta c_i \) can be computed directly from the optimal split updates \( \Delta s_i \) by using eqn (2) as

\[ \Delta c_i = \left[ \sum_{j=1}^{n_i} \Delta s_{ij} - \sum_{j=1}^{m_i} \Delta s_{ij} \right] \]  
(8) 

where \( n_i \) is the total number of phase splits at node \( i \), and \( m_i \) is the number of phase splits that run simultaneously in both the previous and the next coordination plans. Phase splits that run simultaneously in both the previous and the next TOD plan do not have to be optimized separately; they can be optimized simultaneously. In view of that, the optimizable SCPV (OSCPV) for each node can now be defined as

\[ \vec{Q}_i = \begin{bmatrix} \theta_i & s_{i1} & s_{i2} & \cdots & s_{in_i-m_i} \end{bmatrix} \]  
(9) 

If the network is comprised of \( n^* \) intersections, then OSCPV becomes

\[ \vec{x} = \begin{bmatrix} \vec{Q}_1 & \vec{Q}_2 & \cdots & \vec{Q}_n \end{bmatrix} \]  
(10) 

where \( n = n - n^* \). Therefore, eqn (2) can be rewritten in terms of \( \vec{x} \) as

\[ \vec{x}(k+1) = A\vec{x}(k) + Bu(k), \]  
(11)
which is a dynamic expression of a linear time invariant sampled data system where \( A \) is an identity matrix of dimension
\[
\dim(A) = \sum_{i} (n_i - m_i) + \bar{n}.
\] (12)

The matrix \( B \) is diagonal with \( \dim(B) = \dim(A) \), and the quantity \( Bu(k) \) represents the amount of update in the ONCPV, i.e.,
\[
\Delta \bar{x}(k) = Bu(k).
\] (13)
The diagonal elements of \( B \) can be assigned any values; for simplicity, \( B \) is chosen to be an identity matrix. If the transition is done in \( m+1 \) steps, the cost function \( J \) becomes,
\[
J = \sum_{k=0}^{m} \Delta \bar{x}(k)^T \Delta \bar{x}(k) = \sum_{k=0}^{m} u(k)^T B^T Bu(k),
\] (14)

Now, the optimization problem is that of determining the number of steps and the step length, i.e., \( m \) and \( u(k) \), respectively that minimize \( J \) in eqn (14) such that for \( \bar{x}(0) = \bar{x}_j, \ \bar{x}(m+1) = \bar{x}_{j+1} \). Note that by eliminating the redundant elements in the coordination vector, the dimension of the problem has been highly reduced. In fact, this linear quadratic optimization problem is well established [7,8]. Its optimal solution is given by
\[
u_{opt}(k) = (B^T B)^{-1} B^T (A^T)^{m-1} G^{-1}_{(0,m)} \left[ \bar{x}_{j+1} - A^m \bar{x}_j \right],
\] (15)
where
\[
G_{(0,m)} = \sum_{i=0}^{m-1} A^{m-i-1} B (B^T B)^{-1} B^T (A^T)^{m-i-1}.
\] (16)

The optimal solution in eqn (15) exists only if matrix \( G_{(0,m)} \) is invertible, i.e., is square and nonsingular. Since \( (B^T B)^{-1} \) is a symmetric matrix, and \( G_{(0,m)} \) in eqn (16) is a sum of the \( (B^T B)^{-1} \) quadratics, then \( G_{(0,m)} \) is a square matrix. For it to be a nonsingular matrix it requires matrix \( R \), defined by
\[
R = \begin{bmatrix} B, AB, \cdots, A^{m-2} B, A^{m-1} B \end{bmatrix},
\] (17)
to be of rank \( \dim(A) \) [6,7,8,9,10]. Alternatively, system \( (A,B) \) must be reachable. This requires \( m \) to be at least equal to the number of rows of \( A \). For speedy transition, \( m \) is chosen to be the minimum of its admissible values, i.e.,
\[
m = \dim(A).
\] (18)
Since \( A \) is an identity matrix and the product \( B(B^T B)^{-1} B^T \) for any \( B \) is always an identity matrix, then
\[
G_{(0,m)} = \dim(A) x A.
\] (19)
Therefore
\[ u_{opt}(k) = (B^T B)^{-1} B^T \left( \frac{\bar{x}_{j+1} - \bar{x}_j}{\dim(A)} \right), \tag{20} \]

which yields
\[ \Delta \bar{x}_{opt}(k) = \left[ \frac{\bar{x}_{j+1} - \bar{x}_j}{\dim(A)} \right]. \tag{21} \]

This result states that the optimal transition step length \( \Delta \bar{x}(k) \) is obtained by dividing the range between the target ONCPV \( \bar{x}_{j+1} \) and the current ONCPV \( \bar{x}_j \) into \( \dim(A) \) equal parts. Frequently, the split \( (s_i) \) and the cycle length \( (c_i) \) are limited to minimum and maximum values such that \( s_{min} \leq s_i \leq s_{max} \) and \( c_{min} \leq c \leq c_{max} \). Note that if the operational conditions for minimum and maximum values of \( s_i \) and \( c_i \) are satisfied by the optimal timing plans \( \bar{x}_j \) and \( \bar{x}_{j+1} \) then they will also be satisfied by \( \bar{x}(k) \) throughout the transition.

4 Simulation

To test the efficacy of the proposed transition optimization procedure a simulation analysis was conducted. To increase the realism of the analysis, a real corridor in the City of Tallahassee was simulated using actual field data. The coordinated corridor actually had seven intersections, but in order to simplify the analysis only three intersections were simulated. Typical weekday data were collected in the corridor and used to develop time of the day (TOD) timing plans. Three timing plans—i.e., pre-pm peak, pm peak, and evening peak—were used in the analysis to represent conditions of increasing traffic flow and decreasing traffic flow. These conditions also represent increasing cycle length and decreasing cycle length. The timing plans developed were for pre-pm peak, pm peak, and evening peak. SYNCHRO PRO signal optimization software was used to determine optimum coordination parameters for each TOD timing plan. SYNCHRO PRO begins by optimizing the cycle length \( (c_i) \), which is achieved by minimizing the performance index, \( PI \), defined by:
\[ PI = \frac{d + 10s_i + 100q_p}{3600} \tag{22} \]

where \( d \) = percentile signal delay (in seconds), \( s_i \) = number of vehicle stops, and \( q_p \) = queue penalty for vehicles affected. The optimization process continues by optimizing intersection phase splits \( (s_i) \) followed by optimizing intersection offsets \( (o_i) \). The green time is divided based on each lane group’s traffic volume divided by the adjusted saturation flow rate. The phase splits are optimized subject to various constraints such as defining the minimum pedestrian walk time. The offset optimization is basically achieved by evaluating delays.
resulting from varying the offset in increments of 8 seconds above or below the cycle length. Then the optimizer varies the offset by step size increments (1 or 4 seconds) around the 8-second choices with the lowest delays or close to the lowest delays. This process finds the offset with the lowest delays, even if there are multiple "valleys" with low delay [11]. Eqn (21) derived above was used to determine the stepwise increment or decrement of coordination parameters during the transition period. According to the derived optimization procedure, there are 17 time steps (since \( \dim(A) = 16 \)) required each with a unique set of cycle length, splits, and offsets for a 3-node corridor. The signalization data together with the geometric and traffic data for each intersection were input into CORSIM (CORridor SIMulation) software to determine the requisite performance measures during the transition period. These performance measures were compared to the performance measures resulting from one of the transition procedures embedded in CORSIM. The CORSIM simulation model gives the user three options of transitioning—i.e., immediate, two-cycle, and three-cycle options [12]. The proposed method was compared to the immediate option, which is akin to the dwell method discussed earlier. Table 1 shows the results of the comparison. The results are for transitioning from “pre-pm peak” period to “pm peak” period—that is, a scenario of increasing traffic volumes. The “pre-pm peak” period was simulated for 900 seconds and thus the transitioning to “pm peak” period starts at 900 seconds. The measure of performance used for comparison is the queue delay in seconds per vehicle. The queue delay is defined in CORSIM as “delay calculated by taking vehicles having acceleration rates less than 2 feet per second" and speed less than 9 feet per second.”

The results in Table 1 shows that the proposed transition optimization procedure reduces queue delay for minor streets traffic when compared to the CORSIM immediate method of transitioning. The improvement is by an average of 5.88 percent. This is not surprising because the immediate method in CORSIM involves transitioning to a new setting with the controller dwelling in the major street green until the new offset is attained. This strategy inevitably increases delay to the minor street movements as they face a longer red interval. Table 1 further shows that the proposed transition optimization procedure increases delay slightly (-1.43%) on the major streets approaches. This can be attributed to the fact that the “dwell” method employed by CORSIM favors the major street more than the proposed method.

As indicated earlier, the use of eqn (21) resulted in 17 time steps of adjusted offset, split, and cycle length given that the corridor had three simulated intersections. The results shown in Table 1 are for the first time step simulated for 15 minutes. Graphical analysis (not shown) of the delays showed increasing delay with increase in the transition time for both current and proposed method. The increase might be attributed to the transition settings and/or to the increasing traffic flow in the pm peak period compared to mid-day peak period. Additional research is needed to ascertain this phenomenon. It is noteworthy that a trend similar to the one shown in Table 1 was found when simulating transition from “pm peak” period to “post-pm peak” period—i.e., transition from high volume conditions to low volume conditions.
Table 1: Simulation comparison of the transitioning methods.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Proposed Procedure</th>
<th>CORSIM’s Immediate Procedure</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Major Street</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approach 1</td>
<td>228.5</td>
<td>239.3</td>
<td>4.51%</td>
</tr>
<tr>
<td>Approach 2</td>
<td>296.4</td>
<td>297.7</td>
<td>0.44%</td>
</tr>
<tr>
<td>Approach 3</td>
<td>400.5</td>
<td>388.6</td>
<td>-3.06%</td>
</tr>
<tr>
<td>Approach 4</td>
<td>196.2</td>
<td>184.5</td>
<td>-6.34%</td>
</tr>
<tr>
<td>Approach 5</td>
<td>304.4</td>
<td>296.2</td>
<td>-2.77%</td>
</tr>
<tr>
<td>Approach 6</td>
<td>146.6</td>
<td>144.1</td>
<td>-1.73%</td>
</tr>
<tr>
<td>Average</td>
<td>262.1</td>
<td>258.4</td>
<td>-1.43%</td>
</tr>
<tr>
<td><strong>Minor Streets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approach 7</td>
<td>310.7</td>
<td>288.9</td>
<td>-7.55%</td>
</tr>
<tr>
<td>Approach 8</td>
<td>317.8</td>
<td>300.8</td>
<td>-5.65%</td>
</tr>
<tr>
<td>Approach 9</td>
<td>444.5</td>
<td>485.4</td>
<td>8.43%</td>
</tr>
<tr>
<td>Approach 10</td>
<td>338.5</td>
<td>367.2</td>
<td>7.82%</td>
</tr>
<tr>
<td>Approach 11</td>
<td>353.1</td>
<td>383.6</td>
<td>7.95%</td>
</tr>
<tr>
<td>Approach 12</td>
<td>346.6</td>
<td>417.1</td>
<td>16.90%</td>
</tr>
<tr>
<td>Average</td>
<td>351.9</td>
<td>373.9</td>
<td>5.88%</td>
</tr>
</tbody>
</table>

5 Discussion

Previous studies of the transition process developed procedures that generally were aimed at minimizing the disruption of traffic flow by minimizing the transition period itself—i.e., the quicker the transition is achieved, the better. The transition procedure developed herein takes a different approach. Instead of concentrating on minimizing the transition period, the procedure finds incremental values of coordination parameters that are to be implemented in an optimal number of steps that define the transition period. The procedure takes advantage of the fact that the starting and ending points (i.e., terminal constraints) are the known optimal coordination parameters that have already been optimized using offline or online optimization techniques.

The simulation results discussed above show some promising results especially for transition from low to high traffic conditions. However, these results are preliminary and further simulation and field studies are needed. Different geometric and traffic scenarios need to be simulated in both saturated and undersaturated conditions. A major concern of the proposed method is that, as revealed by eqn (21), the number of steps required to achieve synchronization...
is directly proportional to the number of intersections in the corridor. A corridor with a large number of intersections can result in a long transition period though this is counterbalanced by the fact that synchronous phases in a corridor need to be optimized only once thus reducing the number of time steps. Yet still, there is a need for a careful investigation of this phenomenon using longer and shorter cycle lengths. Additionally, further studies are needed to determine if there is a benefit in starting the transition earlier in TOD timing plans. Such an approach is likely to reduce overall delay when the transition is from low to high traffic flow.

References


