Timing Design of Traffic Signals

MASAMI SAKITA

ABSTRACT

A method of optimizing signal timing by the use of a linear programming technique is described. The linear program is based on the premise that the green time given to each traffic movement in the signal cycle must be long enough to handle the traffic coming into the intersection in one cycle. The linear programming method gives minimum signal phase times, minimum cycle time, and critical movements. The optimum signal phase times and cycle time can also be obtained in a nondirect manner. The linear programming method may be used as an off-line tool to analyze signal timing or it may be adopted as the basis for developing a new local controller for critical intersections of centrally controlled systems. When used as an off-line tool, the linear programming method is very effective for "what if" analysis of traffic signal timing problems.

In recent years, renewed attention has been paid to analyzing capacity and signal timing at signalized intersections. Two representative works are those of Akcelik (1) and JHK and Associates (2). The former emphasizes the analysis of signal timing and the latter, capacity applications in the real world. In both analyses, however, critical movements play a key role. The importance of critical movements in the analysis of intersection capacity and signal timing may be due to the introduction of movement-oriented signal controllers as alternatives to phase-oriented signal controllers.

The old-fashioned phase-oriented controllers are very inflexible relative to signal phasing: a fixed sequence of phases is repeated until the traffic engineer changes the signal phase sequence. Usually not many overlapping movements of traffic can be accommodated between phases. In contrast, the movement-oriented traffic controllers are much more flexible in the selection of phase sequence, and they can handle overlapping traffic movements from one signal phase to another.

To illustrate the point, Figure 1 is a diagram of an eight-phase signal controller. Each box indicates a phase that consists of a combination of nonconflicting traffic movements. As Figure 1 shows, many phase sequences are possible with this type of local controller. In addition, the overlapping of movements between phases permits reduction of cycle time by allowing additional phases without increasing the total lost time in a cycle. (Lost time is time not efficiently used to handle traffic.)

The cycle-by-cycle local-actuation method is most commonly used as for signal timing and phase sequencing for isolated signals. Both pretimed and actuated types are commonly used for centrally controlled signals. A linear programming method for optimizing signal timing and phase sequencing for pretimed local controllers is discussed here. This method may be used as an off-line tool for signal timing and capacity analysis, similar to other methods developed for the analysis of signalized intersections (3), or it may be used as a basis for on-line phase sequence triggering and signal timing plan generation incorporated within a local controller. The utility of the linear programming method is demonstrated by applying it to a real-world problem.

Masami Sakita and Associates, 1259 El Camino Real No. 121, Menlo Park, Calif. 94025.

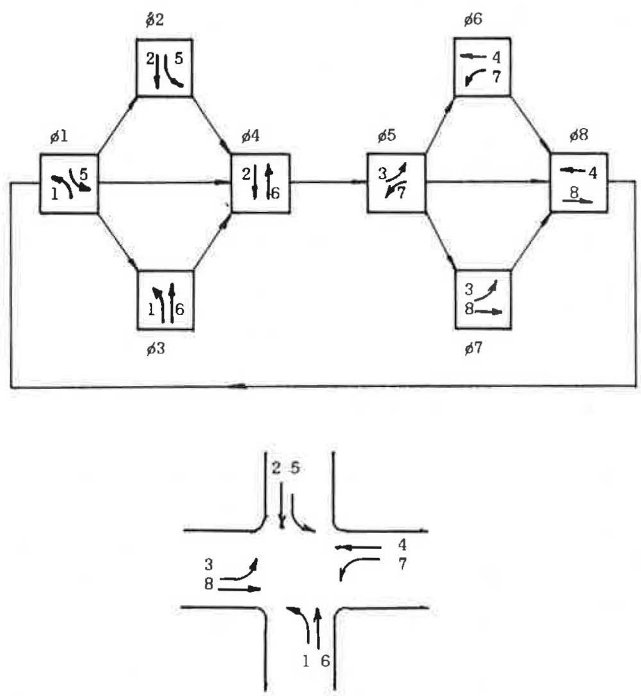


FIGURE 1 Typical phase sequence diagram for an eight-phase controller.

LINEAR PROGRAMMING METHOD

Consider a single lane in the approach to a signalized intersection. The traffic in that lane gets a green signal for a certain duration of time in one cycle of signal. The incoming flow during the entire cycle length must be discharged during the green time interval. It is assumed that m traffic movements are controlled at this intersection and that the signal controller handles pretimed n-phases in one cycle. Further, the traffic movement is the ith movement, and this movement receives the green signal indication in a set of signal phases identified as i. The minimum green time duration assigned to traffic movement i is expressed as (4)

$$G_i \geq C\, q_i/s_i \qquad (1)$$

where

G_i = green time assigned to traffic movement i,
C = cycle time,
q_i = incoming traffic flow rate, and
s_i = practical saturation flow rate for traffic movement i.

The practical saturation flow rate is the rate that can be maintained throughout the signal cycle under the prevailing conditions.

Inequality 1 merely means that the total signal phase times assigned to handle traffic movement i are long enough to handle all the vehicles (of that movement) coming into the intersection in one cycle time. Traffic flow in a movement can be handled either by one signal phase or by multiple signal phases. For example, movement i may be handled by signal phases j and j + 1, whereas movement i + 1 may be handled by phases j + 1, j + 2, and j + 3. In this case both movements i and i + 1 are handled by multiple phases, and phase j + 1 handles both i and i + 1 movements.

If x_j denotes the phase time for signal phase j and X_i denotes the sum of signal phase times that handle movement i,

$$G_i = X_i - L_i = \sum_{\forall j \epsilon \phi_m i} x_j - L_i \quad (2)$$

where $\phi_m i$ is a set of signal phases that handles traffic movement i and L_i is lost time for traffic movement i.

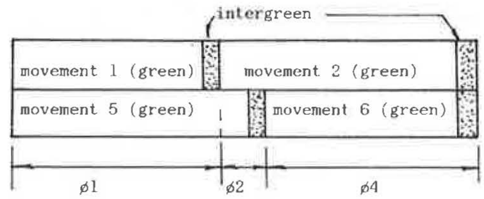
The lost time corresponding to a phase change consists of the starting delay ending lag and all-red duration. It is closely related to the intergreen time following the green time that carries the movement--the amber duration (when no all-red duration is used) or the amber and all-red duration time. The lost time is associated with the traffic movement, and the intergreen time is associated with the signal phase. In reality, however, the intergreen time could be considered as being defined by the movement itself, because of the relationship between the movement and the signal phases. In this paper the terms "signal phase time" and "movement phase time" are used. The signal phase time denotes the time duration consisting of the green time and the intergreen time following the green time, or, in some cases, the green time itself. The movement phase time is the effective green time plus the lost time for that movement.

If the signal phase time is longer than the intergreen time, the phase time for signal phase j, x_j, is expressed as

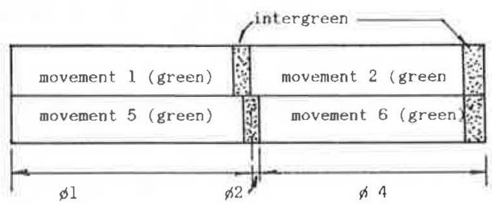
$$x_j = g_j + l_j \quad (3)$$

where g_j is green-time duration of signal phase j and l_j is intergreen-time duration between signal phases j and j + 1.

Sometimes a phase time is so short that the ending of the intergreen time for one traffic movement overlaps with the starting time of the intergreen time that handles another traffic movement. This situation occurs rather frequently with movement-oriented local controllers. The phase sequence $\phi 1$, $\phi 2$, and $\phi 4$ in Figure 1 is used to illustrate this very short phase time. Figure 2 shows two phase-time cases for illustrative purposes. In Case a, $\phi 1$, $\phi 2$, and $\phi 4$ have phase times longer than the intergreen times. In Case b, the phase time of $\phi 2$ is shorter than the intergreen time between $\phi 2$ and $\phi 4$. When the phase time is shorter than the intergreen time, the traffic movements in that short



(a) All Signal Phase Times Are Longer than the Intergreen Times



(b) One Signal Phase Time (Phase 2) Is Shorter than the Intergreen Time

FIGURE 2 Example of phase sequence combinations.

phase do not get green time simultaneously. Under such circumstances, Equation 3 is rewritten as

$$x_j = g_j \quad (3a)$$

From Equations 1 and 2, signal timing constraints can be expressed as

$$\sum_{\forall j \epsilon \phi_m i} x_j - L_i \geq C\, q_i/s_i \quad \text{for } i = 1, 2, \ldots, m \quad (4)$$

The cycle time (C) is expressed as

$$C = x_1 + x_2 + \ldots + x_n \quad (5)$$

The minimum cycle time that satisfies Constraints 4 can be obtained by solving a linear program the objective function of which is to minimize the cycle time as expressed in Equation 5 subject to Constraints 4. (The minimum cycle time is the cycle time that is theoretically just sufficient for the traffic to pass through the intersection.)

Constraints 4 can be simplified by introducing an m x n matrix (P) expressing the relationship between signal phases and traffic movements, as follows:

$$P = (a_{ij}) \quad (6)$$

The P-matrix indicates the movements handled by each of the n phases of the signal cycle. The matrix consists of elements that are either 0 or 1. In the matrix $a_{ij} = 1$ indicates that the ith movement gets a green indication in the jth phase, and $a_{ij} = 0$ indicates otherwise.

Then the linear program shown in Constraints 4 and Equation 5 can be rewritten as follows:

To find nonnegative $x_1, x_2, \ldots, x_n$ that minimize

$$C = \sum_{j=1}^{m} x_j \quad (7)$$

subject to

$$\sum_{j=1}^{n} (a_{ij} - q_i/s_i)\, x_j \geq L_i \quad \text{for } i = 1, \ldots, m \quad (8)$$

The dual problem is expressed as follows:

To find nonnegative $y_1, y_2, \ldots, y_m$ that maximize

$$C = \sum_{i=1}^{m} L_i \, y_i \tag{9}$$

subject to

$$\sum_{i=1}^{m} (a_{ij} - q_i/s_i) \, y_i \leq 1 \quad \text{for } j = 1, 2, \ldots, n \tag{10}$$

The unknown variables $y_1, y_2, \ldots, y_m$ of the dual problem are dimensionless. The nonzero values of y_1's obtained as the solution indicate that the corresponding movements are the critical ones that define the signal phase times and the cycle time. Those nonzero y_1's, however, do not reflect the movement phase times of the critical movements. In fact, all nonzero y_i's will have the same value as long as no other types of constraints are introduced into the linear program.

OPTIMUM SIGNAL TIMING

So far, the discussion has dealt with the minimum required green times and cycle time. The direct mathematical programming approach to solve the optimum timing of signals (least delay for all vehicles) is not simple because the objective function for the problem cannot be expressed as a linear combination of unknown variables (4). The approach used here is an indirect one. The linear programming method has been combined with the findings from earlier signal timing studies.

The minimum cycle time in seconds, when all the critical movements are known, is expressed as (5)

$$C_m = L/(1 - Y) \tag{11}$$

where L is the total lost time in a cycle (in seconds) and Y is the sum of degrees of saturation for all critical movements. The total lost time is not the sum of all the intergreen times but is the sum of lost times for all the critical movements. The minimum cycle time (Equation 11) can easily be verified by the use of the linear programming method.

Webster and Cobbe (5) showed, by the use of computer simulation, that the optimum cycle time (C_o) that gives the least average delay to all vehicles using the intersection is

$$C_o = (1.5L + 5)/(1 - Y) \tag{12}$$

where C_o and L are in seconds.

Thus, from Equations 11 and 12 the optimum cycle time can be expressed as a function of the minimum cycle time and the lost time:

$$C_o = [(1.5\,L + 5)/L] \, C_m = r \, C_m \tag{13}$$

The C_o/C_m ratio (r) is not sensitive to variation of the L value: At L = 16 sec (which is considered to be a long total lost time): At L = 16 sec (which is considered to be a short total lost time), r = 2.13. In most cases, the r-value is in the range of 1.8 to 2.2. The findings by Webster and Cobbe (5) show that, in a practical sense, the optimum cycle time is reasonably well approximated by twice the minimum cycle time.

The optimum signal phase times and cycle time can be obtained by solving the primal problem of the linear program identical to the one described in expressions 7 and 8, except that the right-hand side of Inequality 8 will be rL_i instead of L_i. In the corresponding dual problem, expressions 9 and 10, rL_i will also replace L_i. Thus, the optimum timings can be obtained either in two stages or in one. In the first stage of the two-stage approach, the minimum signal phase times and cycle times are solved to obtain the r-value; and in the second stage, the optimum signal times are obtained with the r-value. The one-stage approach is to solve the linear program by using an approximate value for r, for example, 2.0.

SIGNAL PHASE SELECTION AND OTHER CONSIDERATIONS

The construction of a phase sequence chart is essential in the signal timing design process. The chart is used identify signal phases and traffic movements and their relationships before construction of linear programs. The chart is also used to identify the phase sequence after the linear program has been run because the linear program will not directly identify the phase sequence: It will select the phases and critical movements that satisfy the requirements. The chart should include all the possible signal phase sequences. For example, Figure 3 is a phase sequence chart of a typical two-way/two-way intersection, where right turns are allowed all the time and the left turns are handled by separate green arrow signals. In the figure, phase sequences can be identified by tracing a sequence of phases, following the arrows, and closing a loop that covers both north-south and east-west legs of the intersection. For example, $\phi 1$, $\phi 4$, $\phi 5$, and $\phi 8$ are a possible phase sequence. Similarly, a possible phase sequence chart can be generated for such configurations as T-intersections or five-legged intersections.

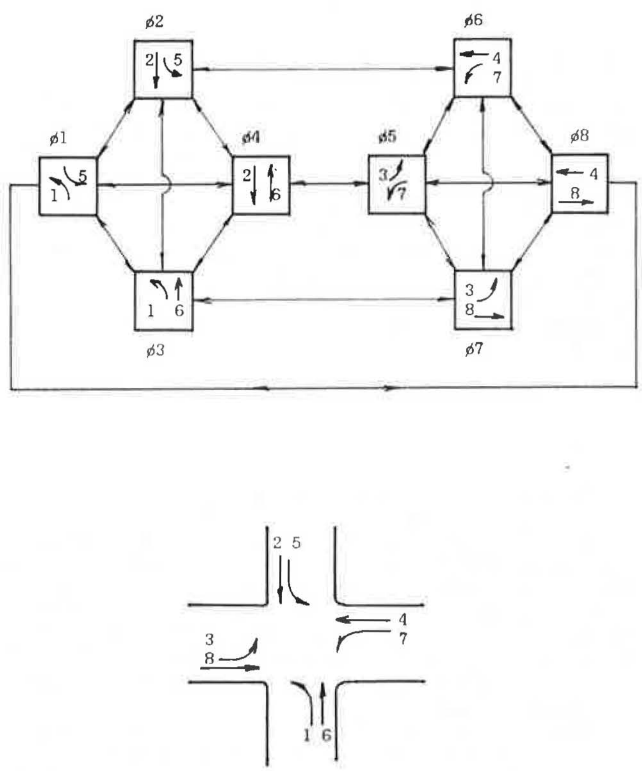


FIGURE 3 Possible phase sequences for a two-way/two-way intersection.

An important factor to consider in determining signal timing is the requirement for noncritical movements. Signal phase times must be determined to satisfy the timing requirements not only for critical movements but also for noncritical movements. To illustrate the problem, refer to the north-south half of the phase sequence diagram presented in Figure 3. If the selected phase sequence is $\phi 1$, $\phi 2$, and $\phi 4$, the critical movements that determine the phase times are either 1 and 2 or 5 and 6. If 1 and 2 are the critical movements, the linear program with constraints consisting of only those requirements for the critical movements will select signal phase times based on the timing requirements for traffic movements 1, 2, and 5 or for movements 1, 2, and 6; but the selected signal phase times will not be based on the requirements for all four traffic movements simultaneously. Thus, the selected signal phase times may yield a situation where $\phi 1$ and $\phi 2$ are both fully saturated in both directions but $\phi 4$ is saturated only in one direction (movement 2) or $\phi 4$ is saturated in both directions but both $\phi 1$ and $\phi 2$ are saturated only in one direction (movements 1 and 2). The requirement for signal-phase-time and movement-time selection is that it be in such a manner that the degrees of saturation for noncritical movements (movements 5 and 6 in the example) are equal when signal phase times are adjustable. This can be achieved by adding these requirements to the constraints of the linear program. The new constraints are expressed in any equality form. For example, in the north-south direction of the two-way/two-way intersection, this requirement for the minimum cycle time and phase times is expressed as

$$(x_1 + x_2 - L_3)/(q_3/s_3) = (x_3 + x_4 - L_4)/(q_4/s_4) \tag{14}$$

$$(x_1 + x_3 - L_1)/(q_1/s_1) = (x_2 + x_4 - L_4)/(q_2/s_2) \tag{15}$$

The same requirement for the optimum cycle time and phase times is expressed in the same manner as in Equations 14 and 15, except that L_1, L_2, L_3, and L_4 are replaced by rL_1, rL_2, rL_3, and rL_4, respectively.

Other signal timing considerations, such as pedestrian crossing time or minimum green time requirements, can also be put in an inequality form as a constraint for the linear program. These requirements, however, must be formulated as part of the constraints of the linear program that seeks for the optimum timing of signals and should not be incorporated in the linear program seeking for the minimum timing.

THE EXAMPLE

To show how the linear programming technique is applied, it was used to solve the following example of a real-world problem. The example is a typical four-leg intersection that has a total of six movements. The major street is controlled with full separate turning signals, and the minor street is controlled by one phase for all movements. Data required to compute signal timing are given in Table 1 and Figure 4 shows possible phasing schemes and the layout of the intersection.

The traffic data (Table 1) indicate that at least two phases are required to handle traffic for the north-south street, but only one phase is required to operate the east-west traffic. For this example, the phase-movement relationship matrix P will be

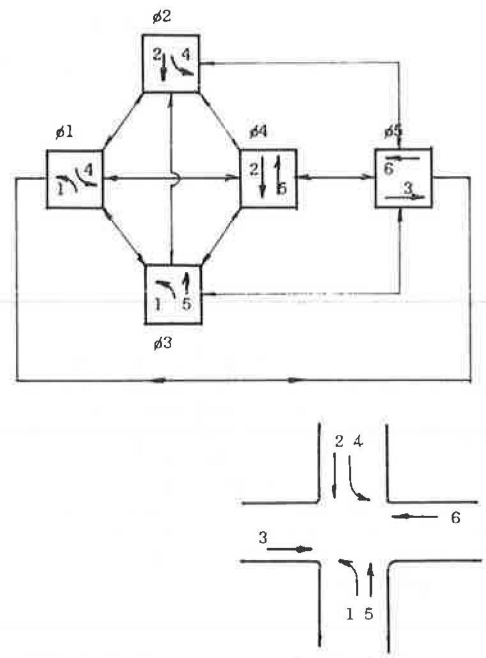


FIGURE 4 Phase sequence chart and intersection layout for text example.

TABLE 1 Traffic Characteristics Data for Example in Text

Movement	Volume (veh/hr)	Practical Saturation Flow Rate (veh/hr)	Lost Time[a] (sec)	Intergreen Time[b] (sec)
1	180	1,440	4	6 (2+4)
2	840	2,520	4	6 (2+4)
3	620	3,060	4	6 (2+4)
4	400	1,440	4	6 (2+4)
5	600	2,700	4	6 (2+4)
6	400	3,060	4	6 (2+4)

[a] Lost time corresponding to a phase change.
[b] Red + amber.

$$P = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{16}$$

The resultant linear program is to find nonnegative $x_1, \ldots, x_5$ that minimize

$$C = x_1 + x_2 + \ldots + x_5 \tag{17}$$

subject to the constraints

$$\begin{aligned}
&0.875x_1 - 0.125x_2 + 0.875x_3 - 0.125x_4 - 0.125x_5 \geq 4.0 \\
&-0.333x_1 + 0.667x_2 - 0.333x_3 + 0.667x_4 \\
&\quad - 0.333x_5 \geq 4.0 \\
&-0.202x_1 - 0.202x_2 - 0.202x_3 - 0.202x_4 \\
&\quad + 0.798x_5 \geq 4.0 \\
&0.722x_1 + 0.722x_2 - 0.278x_3 - 0.278x_4 - 0.278x_5 \geq 4.0 \\
&-0.222x_1 - 0.222x_2 + 0.778x_3 + 0.778x_4 \\
&\quad - 0.222x_5 \geq 4.0 \\
&-0.131x_1 - 0.131x_2 - 0.131x_3 - 0.131x_4 \\
&\quad + 0.869x_5 \geq 4.0
\end{aligned} \tag{18}$$

The linear program has two sets of solutions. The first set is x_1 = 9 sec, x_2 = 6 sec, x_4 = 13 sec, x_5 = 12 sec, and C = 40 sec. In the second set, the solutions are x_2 = 15 sec, x_3 = 9 sec, x_4 = 4 sec, x_5 = 12 sec, and C = 40 sec. In the first set Phases 1, 2, 4, and 5 are selected; and in the second set Phases 2, 3, 4, and 5 are selected by the linear program. In both cases the binding dual variables are y_3, y_4, and y_5, implying that movements 3, 4, and 5 are the critical movements. If any special phase sequence is desired, that requirement must be added to the constraints. For example, adding

$$x_3 = 0 \tag{19}$$

to the constraints will force the linear program to take the first set in the foregoing example.

In addition to the basic signal timing requirements, the timing requirement that the degrees of saturation of noncritical movement phase times be equal whenever possible is added. In this example, this is expressed as

$$x_1 - 0.375x_2 + x_3 - 0.375x_4 = 2.50 \tag{20}$$

The solution of the foregoing linear program with this additional requirement is x_1 = 9 sec, x_2 = 6 sec, x_4 = 13 sec, x_5 = 12 sec, and C = 40.0 sec. (The binding dual variables are naturally the same as those without the new constraints.)

The total lost time for a cycle is 12 sec. Thus, the r-value is 1.92. The optimum signal timings are x_1 = 17 sec, x_2 = 12 sec, x_4 = 25 sec, x_5 = 23 sec, and C = 77 sec.

In this specific case, the total lost time in a cycle is 12 sec for any phase combination. Therefore, it could have been possible to directly solve the linear program for the optimum timings as described in the previous section. Other requirements, such as minimum green time and pedestrian green times, are also possible constraints. If so required, it is also possible to predefine the phase sequence by putting in the constraints that phase times for nondesired phases are zero. In this manner, it is possible to conduct "what if" types of analyses by varying constraints in the linear program.

CONCLUSIONS

The linear programming method can be used not only to optimize signal timings but also to find critical movements. The linear programming method may be used as an off-line tool to analyze intersection signal timing and capacity, or it may be used as a basis for an on-line signal-timing-optimization and phase-sequence-triggering mechanism that can be installed within the local controllers.

REFERENCES

1. R. Akcelik. Traffic Signals: Capacity and Timing Analysis. Research Report 123. Australian Road Research Board, Nunawading, Victoria, Australia, March 1981.
2. Signalized Intersection Capacity Method. JHK and Associates, San Francisco, Calif.; Traffic Institute, Northwestern University, Evanston, Ill., May 1982.
3. A.D. May. Description of Some Fixed-Time Signal Control Computer Programs. Institute of Transportation and Traffic Engineering, University of California, Berkeley, Sept. 1973.
4. G.F. Newell. Lecture Notes for CE 263 Highway Traffic Control. University of California, Berkeley, Spring 1960.
5. F.V. Webster and B.M. Cobbe. Traffic Signals. Technical Paper 56. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1966.

Publication of this paper sponsored by Committee on Traffic Control Devices.