Advanced Public Transportation Systems Based On Pre-signal Management

Miran Khwais[1] and Jack Haddad[1]

[1] *Technion Sustainable Mobility and Robust Transportation (T-SMART) Laboratory, Faculty of Civil and Environmental Engineering, Technion–Israel Institute of Technology, Haifa, Israel*

Introduction

Urban area networks contain arterials which consist of several consecutive signalized intersections. While many factors affect the performance of public transportation systems in urban areas, the delays introduced by the operation of traffic signals account for around 10% to 25% of the total travel time of public transportation vehicles [10, 12, 2, 3].
The most known and used strategies, in attempt to reduce transit vehicle delay at signalized intersections, are transit signal priority (TSP) and dedicated bus lanes (DBLs). Generally, TSP alters the timing plans by either advancing or extending the green duration, in favour of the approaching bus. This is done, so that the bus can move out of the intersection without being hindered. While the DBL strategy dedicates a special lane for buses only in an attempt to give this mode priority, which eliminates the use of one lane.
However, within urban areas the car demand is usually much higher than the bus demand, which means that the capacity of DBLs might be over-reserved for buses and wasted for cars. Additionally, sharing the capacity of the arterial between both car and bus modes in such a fixed manner does not allow adjustable response to the varying traffic demand during the day [7].
Therefore, DBL strategy could be joined with TSP strategy to enhance the performance of both strategies [1]. The idea is to allow cars to use the downstream portion of the DBL and merge with buses at bottlenecks. This is done by installing additional traffic signals prior to the main traffic signal. These signals divide the lane dedicated to buses into two portions - the upstream portion being bus-use only and downstream portion being mix-use.
The additional traffic signals allow vehicles to use the full discharge capacity of the road, by allowing cars to use all lanes of the road to discharge, when no buses are present. When a bus is approaching, the pre-signal turns red for cars, to grant priority for buses to be the first in queue at the main signal. This reduces the interaction of buses with cars [5, 4]. Afterwards the pre-signal turns green for cars, letting them queue after the buses, but gives them the advantage of discharging from all lanes of the road.
It was found that the bus delays could be reduced by 25-70% with the use of the pre-signal strategy, opposing TSP which could reduce the total bus delay only by 20-25% [11]. These delay savings are obtained at each isolated signalized intersection. Therefore, the pre-signal strategy seems promising for bus savings along the full urban routes.
The pre-signal acts as a control tool for both buses and cars, through determining the priority given to each mode. An optimal pre-signal control system that optimizes signal settings based on minimization of person delay was presented in [8]. The developed model compromises of a *simplified* signalized intersection with *one approach*, where two pre-signal lights are located, one for bus lane and one for car lane. A typical layout of a pre-signal system can be seen in Fig. 1, which is the model that was solved and analyzed. In the developed model, the priority was translated into the green-split durations given to each mode. However, the model presented in [8] only considered the boundary limitation on the lanes of mixed queues of buses and cars. In this short paper the model considered boundary limitations on bus and car lanes as well.

Continuous-time model for controlled intersections with pre-signal lights

In this paper, a typical simplified controlled intersection with one pre-signal approach is dealt with. As shown in Fig. 1, there are two separate lanes, one for buses and second for cars, and each lane has a separate pre-signal green/red light. The exit flows from these two lanes move to the main signal section which includes mixed lanes. Hence, the evolution of queue lengths are considered for these three lanes, i.e. bus, car, and mixed lanes.

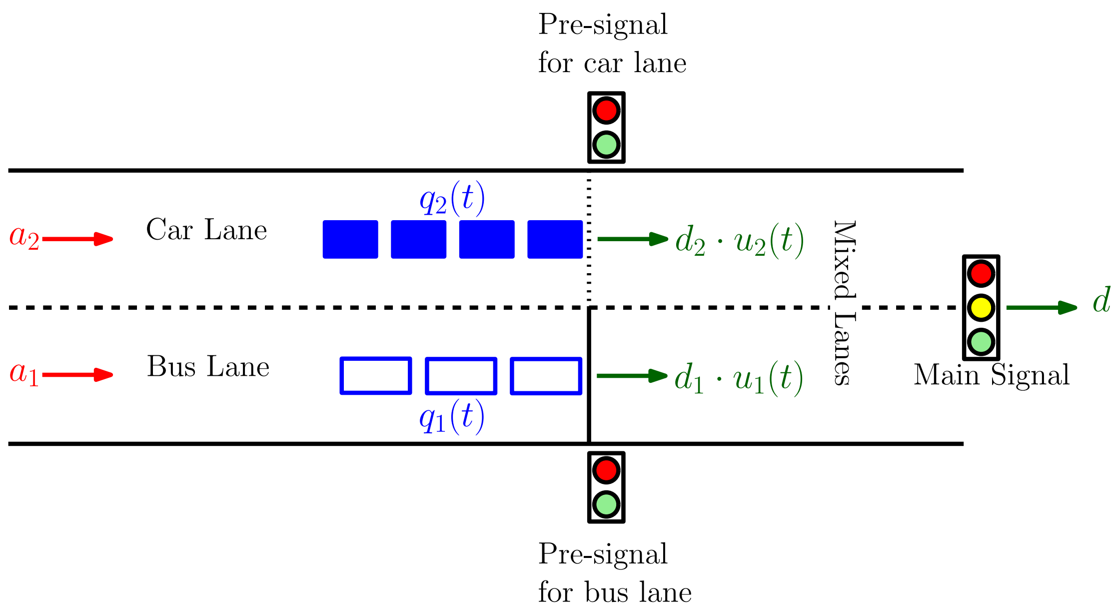

Figure 1: Pre-signal lights located at the upstream of a controlled intersection.

The bus queue length at time t, which is the number of buses stopping behind the stop line of the bus pre-signal, is denoted by $q_1(t)$ (buses), while $q_2(t)$ (cars) denotes the number of cars stopping behind the stop line of the car pre-signal. Let $a_1(t)$, $a_2(t)$ and $d_1(t)$, $d_2(t)$ be, respectively, the arrival and departure rates of buses (bus/s), cars (car/s). The green duration splits at the pre-signals for bus and car movements are respectively denoted by $u_1(t)$ $(-)$ and $u_2(t)$. The green split durations is defined as the ratio of green duration over the cycle length, hence, one can impose the following lower and upper constraints $0 \leq u_1(t),\, u_2(t) \leq 1$.

Then, the dynamic equations for a continuous-time model are as follows

$$\begin{aligned} \frac{\mathrm{d}q_1(t)}{\mathrm{d}t} &= a_1(t) - d_1(t) \cdot u_1(t)\,, \\ \frac{\mathrm{d}q_2(t)}{\mathrm{d}t} &= a_2(t) - d_2(t) \cdot u_2(t)\,, \end{aligned} \tag{1}$$

where

$$\underline{u}_1 \leq u_1(t) \leq \overline{u}_1, \qquad \underline{u}_2 \leq u_2(t) \leq \overline{u}_2 \tag{2}$$

in which the lower bounds $\underline{u}_1$, $\underline{u}_2$ correspond to the minimum green durations given at the pre-signals, and the upper bounds $\overline{u}_1$, $\overline{u}_2$ for the pre-signals are

$$\overline{u}_1 = 1,\, \overline{u}_2 = 1\,. \tag{3}$$

After the pre-signal lights the buses and cars are moving together. We assume equivalence of buses and cars in terms of the traffic space as one bus is equivalent to $k > 1$ cars, i.e. k (equivalent car/bus). The amount of equivalent cars $q(t)$ (car) after the pre-signal lights is determined by the equation

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = u_1(t) \cdot d_1(t) \cdot k + u_2(t) \cdot d_2(t) - d(t)\,, \tag{4}$$

where $d(t)$ (car/s) is an output capacity of a joint lane in terms of equivalent cars. The initial equivalent cars $q_0 = q(0)$ is determined by

$$q_0 = q_{1\mathrm{a},0} \cdot k + q_{2\mathrm{a},0}\,, \tag{5}$$

where $q_{1\mathrm{a},0}$, $q_{2\mathrm{a},0}$ are respectively the initial values of buses and cars after the pre-signal in the mixed lanes.

Because of physical limitation, upper and lower bound constraints on q_2 and q_1 and upper constraint on q are respectively imposed

$$q(t) \leq \bar{q},\ q_2(t) \leq \bar{q}_2. \tag{6}$$

$$q_1(t) \geq 0,\ q_2(t) \geq 0. \tag{7}$$

Let us denote the mean number of persons in bus by B (pers/equivalent car) and the mean number of persons in car by M (pers/car). Then, minimizing the total delay for all travelers is defined as

$$\int_0^T \big[B \cdot k \cdot q_1(t) + M \cdot q_2(t)\big]\mathrm{d}t \rightarrow \min\,, \tag{8}$$

where T (s) is the final time. From the traffic point of view this thesis considers that one bus is equivalent to k cars. However, real cars differ from the considered equivalent car, therefore it is assumed that there are M passengers in one car and B passengers in the equivalent car, meaning that the queue of q_2 for cars contains $M \cdot q_2$ passengers (pers) and the queue of q_1 for buses contains $B \cdot k \cdot q_1$ passengers (pers). The objective function therefore results with the units of (pers $\cdot$ sec) which represents the total person delay.

The following assumptions are made: (A1) the arrival and departure rates are known, (A2) the departure rates are constant, i.e. $d_1(t) = d_1$, $d_2(t) = d_2$, $d(t) = d$, (A3) the queue lengths (number of vehicles) are approximated by real numbers.

Optimal control solution

There are many feasible cases studied and solved for the proposed model, however in this short paper one case will be presented. We will call this case "Case A". First the general optimisation analysis for all the cases is shown then the summary of the solution of one case will be presented and verified.

Optimization analysis

The PMP principle is applied in order to solve the problem. The augmented Hamiltonian, [9] and [6], according to equations (1)-(8) have the following form

$$\begin{aligned} H(t) &= p_1(t) \cdot \big[a_1(t) - d_1 \cdot u_1(t)\big] + p_2(t) \cdot \big[a_2(t) - d_2 \cdot u_2(t)\big] \\ &+ p_3(t) \cdot \big[u_1(t) \cdot d_1 \cdot k + u_2(t) \cdot d_2 - d\big] - \lambda(t) \cdot \big[q(t) - \bar{q}\big] \\ &+ \lambda_{1L}(t) \cdot \big[q_1(t)\big] - \lambda_{1U}(t) \cdot \big[q_1(t) - \bar{q}_1\big] \\ &+ \lambda_{2L}(t) \cdot \big[q_2(t)\big] - \lambda_{2U}(t) \cdot \big[q_2(t) - \bar{q}_2\big] \\ &- q_1(t) \cdot k \cdot B - q_2(t) \cdot M\,. \end{aligned} \tag{9}$$

Here λ_{iL} are the Lagrange multipliers related to lower state constraints $q_i(t) \geq 0$, see (7). Note that Lagrange multiplier is non-negative and equal to zero if the corresponding constraint is not active. This form of the augmented Hamiltonian in (9) is suitable to the direct (with use of λ_i) method according to [6]. The costate equations accordingly are

$$
\begin{aligned}
\frac{\mathrm{d}p_1(t)}{\mathrm{d}t} &= k \cdot B + \lambda_{1U}(t) - \lambda_{1L}(t), \\
\frac{\mathrm{d}p_2(t)}{\mathrm{d}t} &= M + \lambda_{2U}(t) - \lambda_{2L}(t), \\
\frac{\mathrm{d}p_3(t)}{\mathrm{d}t} &= \lambda(t)\,.
\end{aligned}
\tag{10}
$$

Introducing the modified costate $P_1(t) = p_1(t)/k$, one gets

$$
\frac{\mathrm{d}P_1(t)}{\mathrm{d}t} = B + \tilde{\lambda}_{1U}(t) - \tilde{\lambda}_{1L}(t)\,. \tag{11}
$$

where $\tilde{\lambda}_{1U}(t) = \lambda_{1U}(t)/k$ and $\tilde{\lambda}_{1L}(t) = \lambda_{1L}(t)/k$.

The Hamiltonian must be maximized over the control variables $u_1(t)$ and $u_2(t)$ subject to the control constraints (2). The optimal control solution obtained by $\max_{u_1,u_2} H$ in (9) is

$$
u_1(t) = \begin{cases} \overline{u}_1 & \text{if } S_1(t) > 0 \\ \underline{u}_1 & \text{if } S_1(t) < 0, \end{cases} \quad \text{and } u_2(t) = \begin{cases} \overline{u}_2 & \text{if } S_2(t) > 0 \\ \underline{u}_2 & \text{if } S_2(t) < 0, \end{cases} \tag{12}
$$

where the coefficients of $u_1(t)$ and $u_2(t)$ in (9) are defined as the switching functions

$$
S_1(t) \triangleq d_1 \cdot k \cdot [p_3(t) - P_1(t)]\,, \tag{13}
$$

$$
S_2(t) \triangleq d_2 \cdot [p_3(t) - p_2(t)]\,. \tag{14}
$$

The final values $q_1(T), q_2(T)$ are free according to the problem statement, thus according to the transversality conditions for the costates, it holds that

$$
p_1(T) = 0 \Rightarrow P_1(T) = 0\,, \qquad p_2(T) = 0\,. \tag{15}
$$

Hence, from (11) and (10), $P_1(t)$ and $p_2(t)$ are respectively changing according to

$$
P_1(t) = [B + \tilde{\lambda}_{1U}(t) - \tilde{\lambda}_{1L}(t)] \cdot t - [B + \tilde{\lambda}_{1U}(T) - \tilde{\lambda}_{1L}(T)] \cdot T\,, \tag{16}
$$

$$
p_2(t) = [M + \lambda_{2U}(t) - \lambda_{2L}(t)] \cdot t - [M + \lambda_{2U}(T) - \lambda_{2L}(T)] \cdot T\,. \tag{17}
$$

Case A solution summary

The time interval is divided into different time periods according to how the costate variables change with the control variables. Therefore, the problem has to be solved in order to determine the various periods.

It is important to note that the switching time notations is as follows:

- t_{1L}: The time instant when the constraint $q_1 \geq 0$ becomes active
- t_{s1}: The time instant when the the car pre-signal control is forced to go to a minimum, i.e. $u_2 = \underline{u}_2$, in order to get a feasible solution
- t_s: The time instant when the constraint $q \leq \bar{q}$ becomes active
- t_{s2}: The time instant when the constraint $q_2 \leq \bar{q}_2$ becomes active

In conclusion;

$$P_1(t) = \begin{cases} -B \cdot [t_{1L} - t] + P_1(t_s) & \forall\, 0 \leq t < t_{1L} \\ P_1(t_s) & \forall\, t_{1L} \leq t < t_s \\ -B \cdot [T - t] & \forall\, t_s \leq t \leq T, \end{cases} \tag{18}$$

$$p_2(t) = \begin{cases} -M \cdot [t_{s2} - t] + P_1(t_{s_2}) & \forall\, 0 \leq t < t_{s2} \\ -B \cdot [T - t] & \forall\, t_{s2} \leq t \leq T, \end{cases} \tag{19}$$

$$p_3(t) = \begin{cases} P_1(t_s) & \forall\, 0 \leq t < t_s \\ -B \cdot [T - t] & \forall\, t_s \leq t \leq T, \end{cases} \tag{20}$$

Accordingly, the control input has the following form

$$\begin{aligned} &\text{Period I:} \quad u_1(t) = \overline{u}_1\,, \quad u_2(t) = \overline{u}_2\,, \quad \forall 0 \leq t < t_{1L}\,, \\ &\text{Period II:} \quad u_1(t) = u_1^{**}\,, \quad u_2(t) = \overline{u}_2\,, \quad \forall t_{1L} \leq t < t_{s1}\,, \\ &\text{Period III:} \quad u_1(t) = u_1^{**}\,, \quad u_2(t) = \underline{u}_2\,, \quad \forall t_{s1} \leq t < t_s\,, \\ &\text{Period IV:} \quad u_1(t) = u_1^{*}\,, \quad u_2(t) = \underline{u}_2\,, \quad \forall t_s \leq t < t_{s2}\,, \\ &\text{Period V:} \quad u_1(t) = u_1^{c*}\,, \quad u_2(t) = u_2^{c*}\,, \quad \forall t_{s2} \leq t \leq T\,. \end{aligned} \tag{21}$$

The optimal control solution for Case A is given by (21). The control synthesis for Case A is also depicted in Fig. 2. The costate variables $P_1(t)$, $p_2(t)$, and $p_3(t)$, change over time are respectively given by (18), (19), and (20), as also shown in Fig. 3.

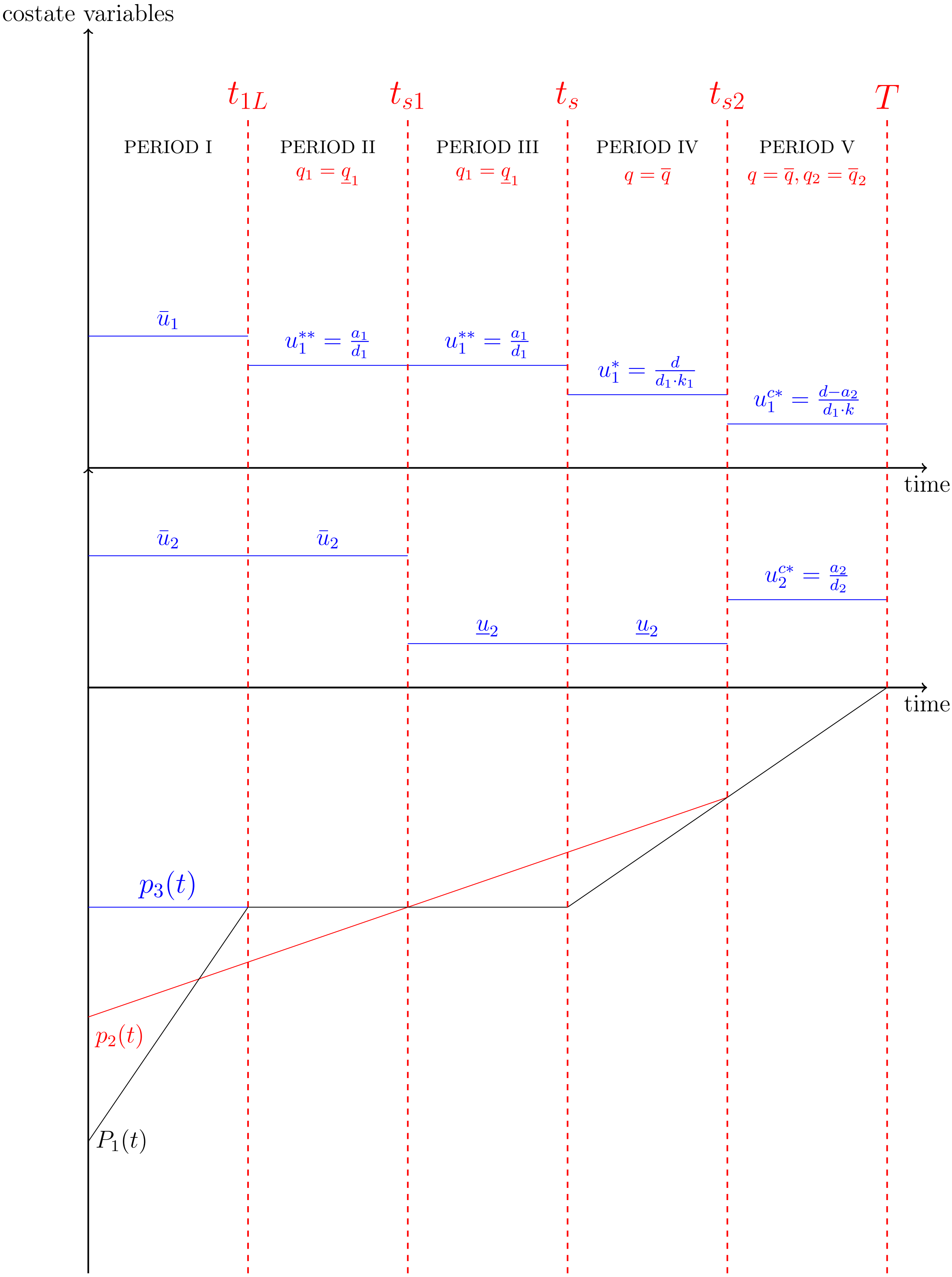

Figure 2: Control synthesis for Case A solution.

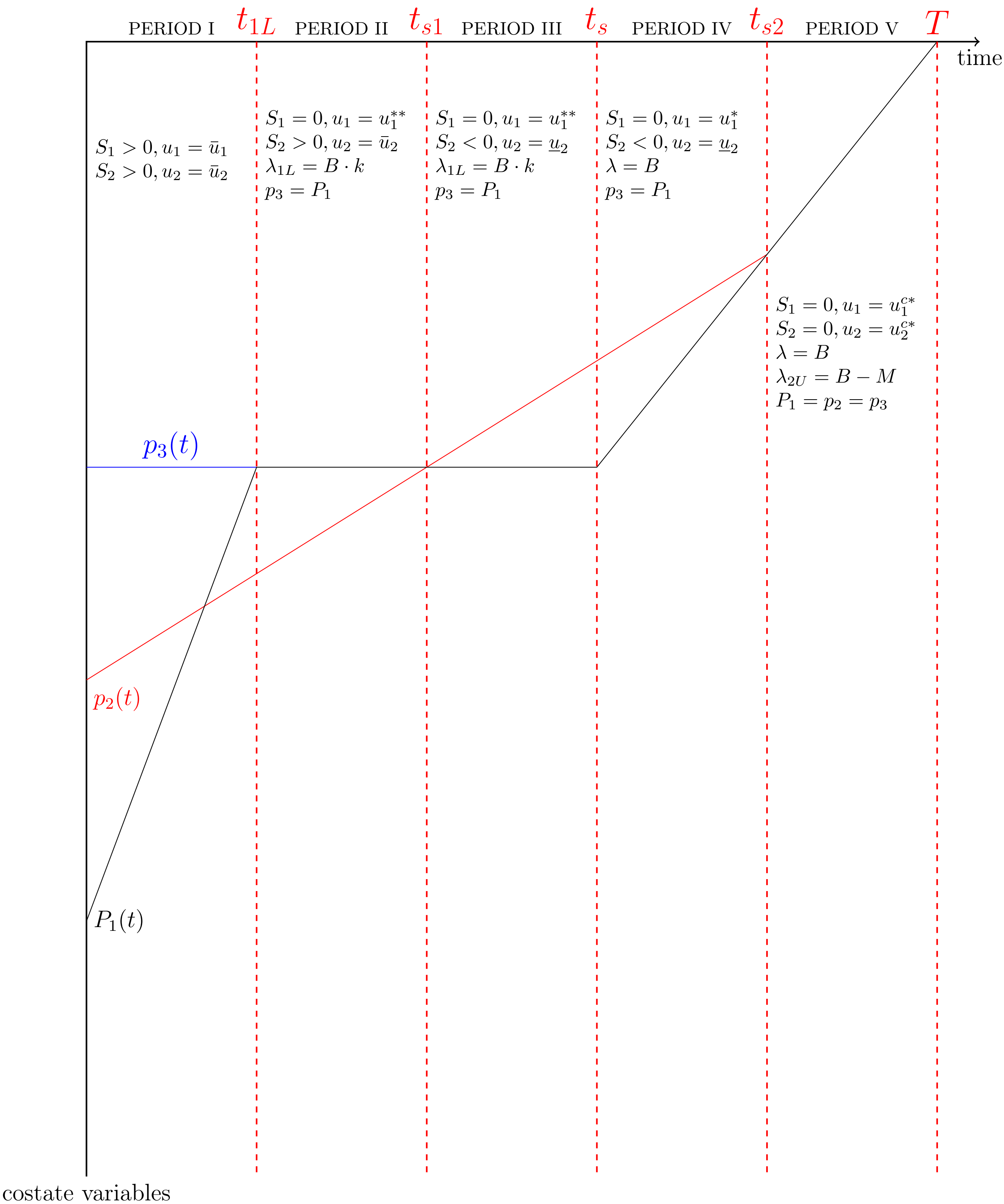

Figure 3: Costates' evolution for Case A solution.

Verification of Case A

Denote this verification of Case A as example 1. Verification was done through five variables of interest,

1. Pre-signal states.
2. Lagrange multipliers.
3. Pre-signal costates.
4. Pre-signal control inputs.
5. Hamiltonian.

This example represents Case A solution. The chosen parameters for this example are shown in Table 1.

Table 1: Set of parameters for Example 1.

Example 1 Solution Set								
Given (determined) Values			Knowns			Bounds		
parameter	value	unit	variable	value	unit	variable	lower	upper
d_1	610	veh/hr	$q_{1,0}$	2	bus	q	-	70
d_2	950	veh/hr	$q_{2,0}$	20	car	q_1	0	-
d	1000	veh/hr	q_0	25	car	q_2	-	25
a_1	430	veh/hr	t_0	0	s	u_1	0	1
a_2	850	veh/hr	t_f	200	s	u_2	0	1
k	2.5					t	0	200
M	1.5							
B	15							

This example considers lower and upper constraints on car queue and upper constraint on equivalent car queues, i.e. $q_2(t) \geq 0$, $q_2(t) \leq \overline{q}_2$ and $q(t) \leq \overline{q}$.

In this example only the lower constraint on the bus queue, the upper constraint on the car queue and the upper constraint on the equivalent car queue become active. This is shown in Fig. 4, where the function $q_1(t)$ approach it's minimum value at some point and remains at it for some time, whereas the functions $q(t)$ and $q_2(t)$ approach a maximum value at some point in time and remain limited by them until the end of the time horizon.

In this example, the lower constraint on the bus queue, $q_1(t)$, becomes active in the second period, at $t = t_{1L} = 40$ seconds where $q_1(t) = q_{1,\min} = 0$. Whereas the upper constraint on the equivalent car queue, $q(t)$, becomes active in the fourth period, at $t = t_s = 172$ seconds where $q(t) = q_{\max} = 70$. And the upper constraint on the car queue, $q_2(t)$, becomes active in the fifth period, at $t = t_{s2} = 176$ seconds where $q_2(t) = q_{2_{\max}} = 25$, as illustrated in Fig. 4.

Additionally, there are three Lagrange multipliers in this example, each one is associated with either a lower or an upper state constraint,

1. $\lambda_{1L}(t)$ for constraint $q_1(t) \geq 0$ on bus queue,
2. $\lambda_{2U}(t)$ for constraint $q_2(t) \leq \overline{q}_2$ on car queue,
3. $\lambda(t)$ for constraint $q(t) \leq \overline{q}$ on equivalent car queue.

The lower constraint on the bus queue, $q_1(t) \geq 0$, becomes active when $\lambda_{1L} > 0$, whereas whenever the constraint inactive $\lambda_{1L} = 0$. The same concept applies to the constraint on the equivalent car queue, $q(t) \leq \overline{q}(t)$ and to the upper constraint on the car queue, $q_2(t) \leq \overline{q}_2(t)$. Whereas, the other constraints never become active and therefore their corresponding Lagrange multipliers are zero through all the time horizon. All these conditions are illustrated in Fig. 5.

In this example, the bus queue, $q_1(t)$, becomes active in the second period, at $t = t_{1L} = 40$ seconds where $\lambda_{1L}(t) = B \cdot k = 37.5$. Whereas, the equivalent car queue, $q(t)$, becomes active in the fourth period, at $t = t_s = 172$ seconds where $\lambda(t) = B = 15$. And, the upper constraint on the car queue, $q_2(t)$, becomes active in the fifth period, at $t = t_{s2} = 176$ seconds where $\lambda_{2U}(t) = B - M = 13.5$, as illustrated in Fig. 5.

Furthermore, the costates in this example change over time according to (10). On one hand, B and M are positive values, therefore the costate variables $P_1(t)$, $p_2(t)$ are always increasing. Additionally, it is logical to assume that the average number of persons in bus is larger than the average number of persons in car, i.e. it assumed that $B > M$, where B is the slope of the line of $P_1(t)$ and M is the slope of the line of $p_2(t)$. Hence, it holds that $P_1(t) < p_2(t)$ for the first period of the solution, refer to (10). Finally, according to the transversality condition the end points for the costates, i.e. $P_1(T)$ and $p_2(T)$ are zero, see subsection .

On the other hand, since $\lambda(t) > 0$ if the upper constraint on equivalent car queue, $q(t) \leq \overline{q}$, is active and $\lambda(t) = 0$ otherwise. Therefore, the costate $p_3(t)$ is either constant when the upper constraint is not active, or increasing otherwise.

This example follows the analytical derived solution for the costates, shown in Fig. 3, and numerically in the bottom of Fig. 6. The corresponding numerical solution for the costates is shown in the top of Fig. 6. The conditions described above are almost perfectly compatible in both figures.

Moreover, the aim is to find the control inputs that maximize the Hamiltonian, according to the optimality condition, are determined according to (21). Correspondingly, the derived analytical solution for the control inputs for example 1 is illustrated in Fig. 2. This behavior is met by both the analytical solution and the numerical DIDO optimization method solution, as shown in Fig. 7, where both graphs are compatible.

Lastly, the Hamiltonian (H) can be thought of implicitly as a function of time (t). Because the costates and the Lagrange multipliers are continuous, and the control inputs are piecewise continuous, it follows that H is a piecewise continuous function of t. Our problem is an autonomous problem, since there is no explicit dependence on t, $\partial H/\partial t = 0$, see (9). And when an optimal control is autonomous, then the Hamiltonian is a constant function of time along the optimal trajectory, as shown in Fig. 8.

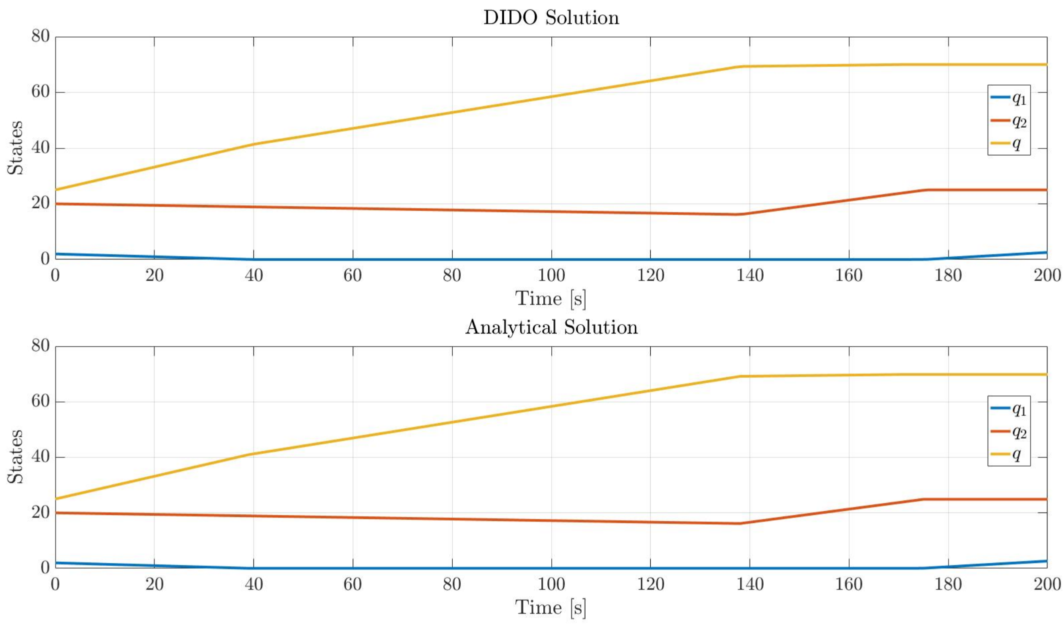

Figure 4: Example 1: States evolution over time obtained numerically and analytically.

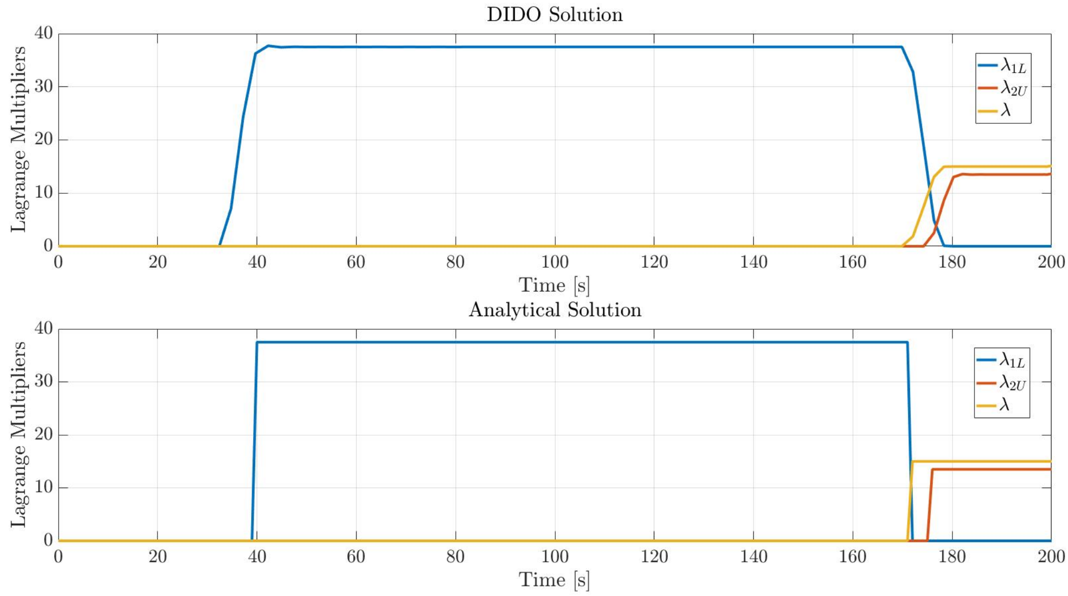

Figure 5: Example 1: Lagrange multipliers evolution over time.

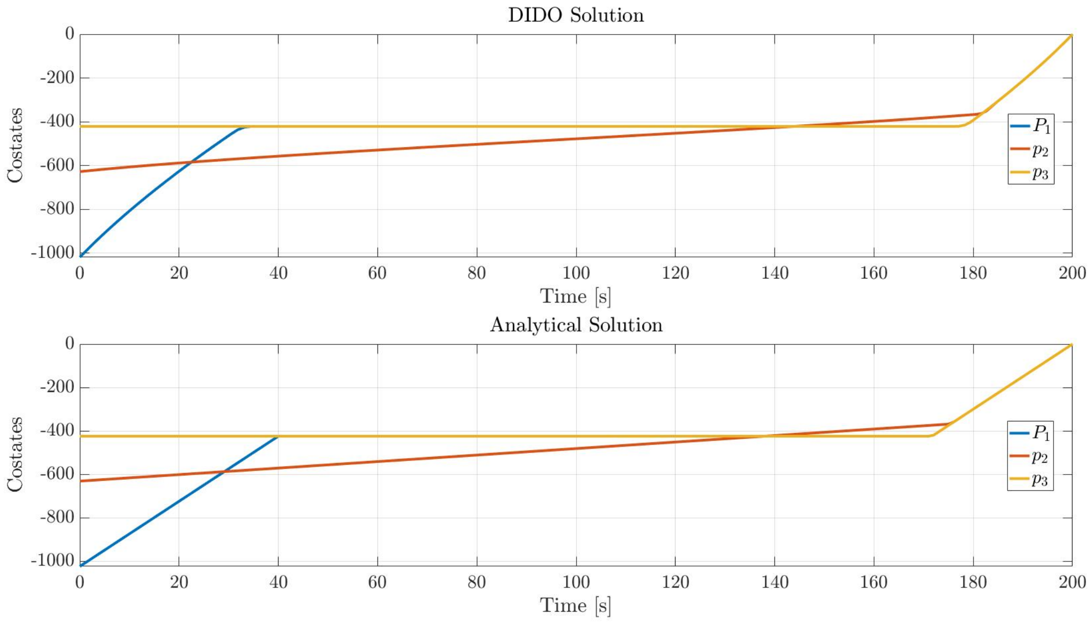

Figure 6: Example 1: Costates evolution over time obtained numerically and analytically.

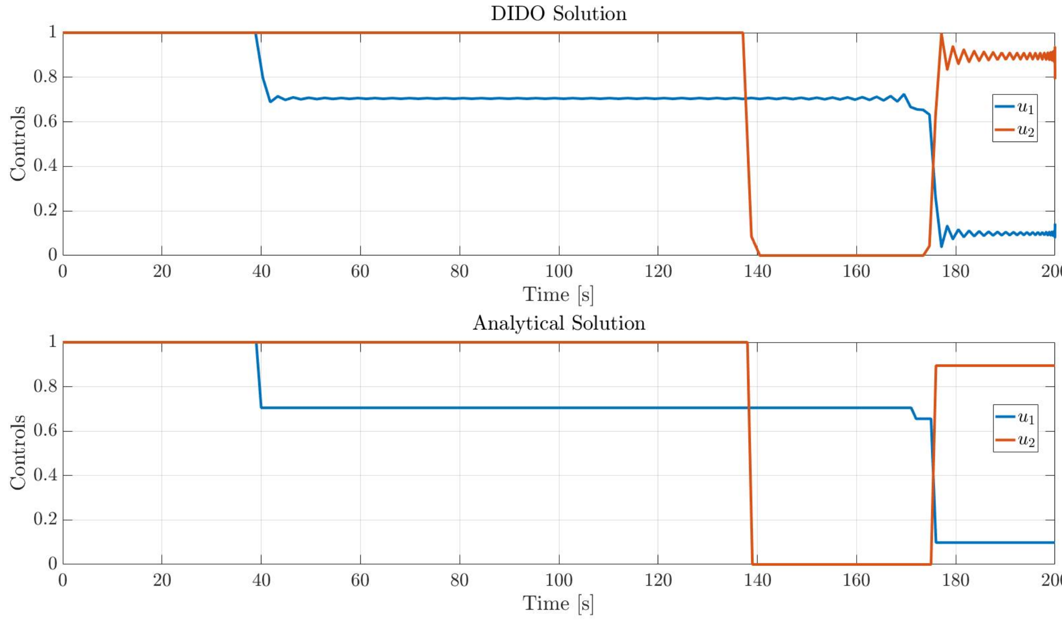

Figure 7: Example 1: Control inputs evolution over time obtained numerically and analytically.

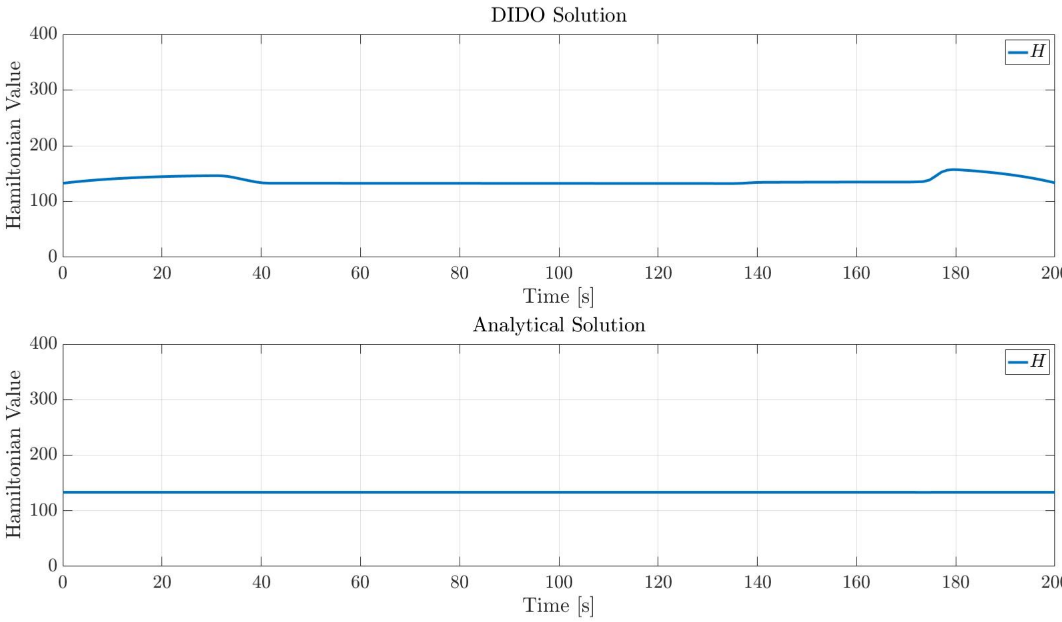

Figure 8: Example 1: Hamiltonian evolution over time.

The derived optimal analytical solutions were verified via numerical solutions. Comparison results show that analytical solutions are very close to the numerical results obtained from the DIDO optimization method.

References

[1] M. Eichler and C. F. Daganzo. Bus lanes with intermittent priority: Strategy formulae and an evaluation. *Transportation Research Part B: Methodological*, 40(9):731–744, 2006.

[2] W. Gu, M. J. Cassidy, V. V. Gayah, and Y. Ouyang. Mitigating negative impacts of near-side bus stops on cars. *Transportation Research Part B: Methodological*, 47:42–56, 2013.

[3] W. Gu, V. V. Gayah, M. J. Cassidy, and N. Saade. On the impacts of bus stops near signalized intersections: Models of car and bus delays. *Transportation research part B: methodological*, 68:123–140, 2014.

[4] S. I. Guler, V. V. Gayah, and M. Menendez. Bus priority at signalized intersections with single-lane approaches: A novel pre-signal strategy. *Transportation Research Part C: Emerging Technologies*, 63:51–70, 2016.

[5] S. I. Guler and M. Menendez. Analytical formulation and empirical evaluation of pre-signals for bus priority. *Transportation Research Part B: Methodological*, 64:41–53, 2014.

[6] R. F. Hartl, S. P. Sethi, and R. G. Vickson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM review*, 37(2):181–218, 1995.

[7] H. He, M. Menendez, and I. Guler. Analytical evaluation of flexible sharing strategies on multi-modal arterials. *Transportation research procedia*, 23:980–999, 2017.

[8] M. Khwais and J. Haddad. Optimal pre-signal control for buses and cars at isolated signalized intersections. *IFAC-PapersOnLine*, 50(1):9706–9711, 2017.

[9] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko. *The mathematical theory of optimal processes*. Wiley-Interscience, NY, 1962.

[10] H. Rakha and Y. Zhang. Sensitivity analysis of transit signal priority impacts on operation of a signalized intersection. *Journal of transportation engineering*, 130(6):796–804, 2004.

[11] H. R. Smith, B. Hemily, and M. Ivanovic. Transit signal priority (tsp): A planning and implementation handbook. 2005.

[12] S. R. Sunkari, P. S. Beasley, T. Urbanik, and D. B. Fambro. Model to evaluate the impacts of bus priority on signalized intersections. *Transportation Research Record*, (1494):117–123, 1995.