Distributed Control for an Urban Traffic Network

Viet Hoang Pham, Kazunori Sakurama, *Member, IEEE*, Shaoshuai Mou, *Member, IEEE*, and Hyo-Sung Ahn, *Senior Member, IEEE*

***Abstract*— In this paper, we develop a distributed control method for an urban traffic network in order to improve traffic conditions and guarantee a smooth operation for all roads and intersections. An optimal coordination among traffic flows is determined by a model predictive control idea. The control objective is formulated as a constrained optimization problem in which the cost function to be minimized is a combination of some traffic network-related performance indexes and the constraints are derived from capacities of roads and intersections. The green time assigned to each road link is computed from its optimal downstream traffic flow. In the proposed algorithms, every road link uses only its local information to determine its control decision which corresponds to a global optimal solution. The algorithms are developed by using a gradient projection-based method and properties of the minimal polynomial of a matrix pair. The effectiveness of the proposed algorithms is validated via numerical simulations with MATLAB and VISSIM.**

***Index Terms*— Urban traffic control, model predictive control, distributed method.**

I. Introduction

THOUGH traffic congestion impacts negatively to economy, human health and environment [1], [2], it becomes more and more complicated and inevitable recently, particularly in urban areas. This is because urban areas are centers of population and economy. The traffic demands increase dramatically while it is difficult to extend the road infrastructures. Even if the capacity of a road is high, traffic congestion may still occur when the number of vehicles entering into this road exceeds its capacity. So traffic flows from roads to roads need to be coordinated effectively to avoid traffic congestion. To control traffic flows, regulating traffic signals at intersections is a popular and convenient way. The downstream traffic flow of one road, which is the number of vehicles departing from this road and entering to neighboring roads, in a cycle is directly proportional to its assigned green time duration.

Motivated by recent advancement in sensing and communication techniques, intelligent traffic control systems have attracted huge attention from both the transportation research community and the control systems society [3], [4]. Applying control system theories, various traffic signal control strategies for urban traffic networks have been developed [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. Some strategies are implemented for isolated intersection control [5], [6], [7], [8], showing a good performance with low-level traffic demand. However, these schemes do not guarantee a globally optimal operation for urban traffic networks because a best control decision for an intersection may influence negatively to other ones. Based on macroscopic model and historical data, TRANSYT-7F [9] formulates coordinated control decisions off-line for intersections to maximize the total throughput and minimize delay. Aiming to improve the control performance, coordinated traffic-responsive strategies including SCOOT [10], SCATS [11], PRODYN [12], RHODES [13] and OPAC [14] use the traffic states measured by road sensors in real time and optimize traffic signals timing with the near forecasted traffic information. Though these methods show nice performance, they suffer exponential complexities.

Relying on a vehicle conservation to estimate future traffic behaviors, model predictive control (MPC) is shown to be a powerful tool in controlling traffic networks [15], [16], [17], [18], [19], [20], [21], [22], [23]. This approach aims to determine a traffic signals timing plan for optimizing some interested performance indexes over a given time horizon. Traffic models such as store-and-forward model [24] and cell transmission model [25], [26] are employed because they are simple but accurate enough to describe the traffic evolution over time and space. According to the communication structure among controllers and sensors, MPC-based traffic network control strategies are grouped into centralized methods [15], [16], [17], [18] or distributed methods [19], [20], [21], [22], [23]. In the centralized setup, one powerful controller gathers information of all roads and makes all control decisions for all intersections. On the contrary, in a distributed setup, an urban traffic network is divided into multiple subnetworks and each of them is assigned to one controller. Then these local controllers cooperate to solve the control problem of the overall network. The distributed strategies have several superior features over the centralized ones:

- *Moderateness:* Since each local controller requires only its local information, the expense for collecting information is low and the risk of executing huge size data is avoided. The computation load of one local controller may not increase dramatically although the size of the network increases.
- *Adaptation:* The communication cost is small because each local controller exchanges information only with its

Manuscript received 25 July 2020; revised 14 June 2021, 10 January 2022, and 3 May 2022; accepted 14 September 2022. Date of publication 28 September 2022; date of current version 5 December 2022. This work was supported by the National Research Foundation of Korea (NRF) under Grant NRF-2022R1A2B5B03001459. The Associate Editor for this article was S. Sacone. *(Corresponding author: Hyo-Sung Ahn.)*

Viet Hoang Pham and Hyo-Sung Ahn are with the School of Mechanical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea (e-mail: vietph@gist.ac.kr; hyosung@gist.ac.kr).

Kazunori Sakurama is with the Graduate School of Informatics, Kyoto University, Sakyoku, Kyoto 606-8501, Japan (e-mail: sakurama@i.kyoto-u.ac.jp).

Shaoshuai Mou is with the School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47906 USA (e-mail: mous@purdue.edu).

Digital Object Identifier 10.1109/TITS.2022.3208565

neighboring controllers. Consequently, the time delay is reduced significantly when collecting information. The local controllers make control decisions in real-time using on-going traffic data.

- *Robustness:* It is not necessary to reprogram all local controllers even though the traffic network structure varies. For example, when one local controller fails or one subnetwork cannot be used due to accident, others are able to work independently. This may enable us to do an automatic plug-and-play operation.

In our conference paper [27], a distributed traffic control strategy is developed for a special traffic network, which consists of many consecutive intersections in a main arterial road. We formulated the traffic problem as a constrained Economic Dispatch Problem (EDP) in order to have an optimal distribution of vehicles. The consensus-based computation technique proposed in [28] was applied with the assumption that every road link knows the minimal polynomial of the weighted matrix associated with the communication graph. This paper is to generalize the work in [27] by considering a general network and relaxing the requirement on global knowledge of the communication graph. We consider an urban traffic network whose physical graph is required to be weakly connected (i.e., its underlying graph is connected). To assess traffic conditions, an objective function that depends on the numbers of contained vehicles and the downstream traffic flows of road links is considered. The control variables we choose are downstream traffic flows because they are suitable to compute the length of active (green) time assigned to road links. To have a smooth operation of roads and intersections, some safety constraints on roads occupancy ratios, downstream traffic flows and traffic signal splits are required to be satisfied. Consequently, one of the main purposes of this paper is to formulate the traffic control problem for an urban traffic network in detailed form as a constrained optimization problem. After providing an analytic solution for the traffic control problem in a centralized manner, we design algorithms to allow each road link to determine the solution in a distributed way using only its local information. Taking the advantage of an accelerated gradient-based projection method, the optimal solution is asymptotically achieved and the convergence rate is $O\left(\frac{1}{s^2}\right)$ where s is the number of the iterations. Based on the minimal time consensus technique developed in [28], [29], and [30], an algorithm is proposed to compute the final converged values of a time-invariant linear nonhomogeneous system in a minimum time.

The remainder of this paper is organized as follows. We describe an urban traffic network as a multi-agent system in Section II and formulate traffic control problem via MPC approach in Section III. By analyzing the primal and dual problems, we design an iterative update law to find the optimal solution of traffic control problem in Section IV. In Section V, we propose an algorithm for minimum-time final value computation and then provide our main distributed algorithm for urban traffic network control. Section VI includes numerical simulations to validate our proposed algorithms and Section VII contains concluding remarks.

TABLE I
TWO COMMON SEQUENCES OF TRAFFIC SIGNAL PHASES

	Intersection	1^{st} phase	2^{nd} phase	3^{rd} phase	4^{th} phase
Type 1	I_1				
Type 2	I_2				

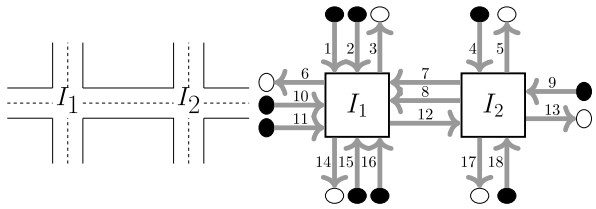


Fig. 1. The traffic network consisting of two intersections I_1 and I_2, and its graph representation.

II. URBAN TRAFFIC MODEL

A. Graph Representation

In an urban traffic network, signalized intersections are used to regulate traffic flows for the aim of avoiding the conflicts among vehicle movements. Traffic flows crossing an intersection are grouped into several traffic signal phases, and the traffic flows assigned to the same phase are activated simultaneously. A *road* connecting two consecutive intersections has several lanes, and a group of lanes (of one road) that are activated in the same traffic signal phase is called a *road link*. So, there could be several different road links even in the same road. To have a simple representation for an urban traffic network, we consider an intersection as a node and a road link as an edge. Then, an urban traffic network is illustrated by a directed graph $\mathcal{P} = (\mathcal{J}, \mathcal{R})$ where $\mathcal{J} = \{I_1, \cdots, I_M\}$ is the set of nodes and $\mathcal{R} = \{1, 2, \ldots, N\}$ is the set of road links. For simplicity, we assume that traffic signal phases and their sequences are fixed in all intersections. This assumption is similar to the ones in many well-known traffic models which are widely employed in MPC traffic signal control approaches; for examples, store-and-forward model [15], [16], [18], [19], [21], and cell transmission model [22].

The traffic model using a directed graph concept could represent a general traffic network. For an illustration purpose, we consider a network of two intersections where they have different setup of traffic signal phases as given in Table I. This table shows two common sequences of traffic signal phases for typical four-arm intersections. Type 1 is suitable for arterial roads when the number of turning-left vehicles is significantly smaller than the number of going-straight and turning-right vehicles. Type 2 enhances the safety of movements when the turning ratios are arbitrary. Fig. 1 depicts the network of two intersections. For each incoming road to the intersection I_1, one road link represents some lanes for turning left movements and another road link corresponds to some lanes for going straight and turning right movements. Each incoming road to the intersection I_2 is represented by one road link. The road from the intersection I_1 to the intersection I_2 is represented by the road link 12 while the road from the intersection I_2 to the intersection I_1 is represented

TABLE II
ACTIVE ROAD LINKS IN EACH PHASE

Intersection	1^{st} phase	2^{nd} phase	3^{rd} phase	4^{th} phase
I_1	$1, 16$	$2, 15$	$7, 11$	$8, 10$
I_2	4	12	18	9

by two road links 7 and 8, in which the road link 7 corresponds to turning-right and going-straight movements, and the road link 8 represents turning-left movements. Table II shows the links which are active in each phase for both two intersections.

Let $\sigma(i)$ and $\tau(i)$ represent respectively the source node and the sink node of the road link $i \in \mathcal{R}$, i.e., the direction of vehicles moving in this road link is from $\sigma(i)$ to $\tau(i)$. For convenience, we use $\mathcal{O}$ to denote the set of all source nodes (from which the vehicles move into the network) and destination nodes (into which the vehicles exit the network). A node in $\mathcal{O}$ is a virtual node representing the outside of the considered traffic network. So, $\mathcal{O} \cap \mathcal{J} = \emptyset$ and $\sigma(i), \tau(i) \in \mathcal{J} \cup \mathcal{O}$ for all $i \in \mathcal{R}$. We call $i \in \mathcal{R}$ a source road link if $\sigma(i) \in \mathcal{O}$ and a destination road link if $\tau(i) \in \mathcal{O}$. For any two road links $i, j \in \mathcal{R}$, j is reachable from i or i can reach to j if there exists at least one directed sequence of road links $\{z_1, z_2, \ldots, z_n\}$ such that $z_1 = i, z_n = j$ and $\tau(z_k) = \sigma(z_{k+1}) \in \mathcal{J}, k = 1, \ldots, n-1$. A road link $i \in \mathcal{R}$ is reachable from and can reach to itself. In this paper, we pay our attention to an urban traffic network whose graph representation satisfies the following assumption.

Assumption 1: For every road link $i \in \mathcal{R}$, it is reachable from at least one source road link $j \in \mathcal{R}$, where $\sigma(j) \in \mathcal{O}$, and it can reach to at least one destination road link $k \in \mathcal{R}$, where $\tau(k) \in \mathcal{O}$.

We use $\mathcal{N}_i^-$ and $\mathcal{N}_i^+$ to denote the set of downstream neighbors and the set of upstream neighbors of the road link i:

$$\mathcal{N}_i^- = \{j \in \mathcal{R} : \sigma(j) = \tau(i)\},\ \mathcal{N}_i^+ = \{j \in \mathcal{R} : \tau(j) = \sigma(i)\}.$$

Thus, vehicles departing from the road link i will enter into the road links in $\mathcal{N}_i^-$, and vehicles departing from the road links in $\mathcal{N}_i^+$ will enter into the road link i. We use $\mathcal{I}_\sigma$ to denote the set of road links in $\mathcal{R}$ which have the same ending node as σ:

$$\mathcal{I}_\sigma = \{j \in \mathcal{R} : \tau(j) = \sigma \in \mathcal{J}\}.$$

For better understanding of the aforementioned notations, we consider the traffic network in Fig. 1 where $\mathcal{R} = \{1, 2, \ldots, 18\}$, $\mathcal{J} = \{I_1, I_2\}$ and $\mathcal{O} = \big\{\sigma(1), \sigma(2), \tau(3), \sigma(4), \tau(5), \tau(6), \sigma(9), \sigma(10), \sigma(11), \tau(13), \tau(14), \sigma(15), \sigma(16), \tau(17), \sigma(18)\big\}$. In this figure, the gray arrows illustrate road links in $\mathcal{R}$, the squares represent internal intersections in $\mathcal{J}$ and the black/white ellipses represent external source/destination nodes in the set $\mathcal{O}$. For the road links 7, 8 and 12, we have $\sigma(7) = \sigma(8) = I_2$, $\tau(7) = \tau(8) = I_1$, $\mathcal{N}_7^+ = \mathcal{N}_8^+ = \{4, 9, 18\}$, $\mathcal{N}_7^- = \{3, 6\}$, $\mathcal{N}_8^- = \{14\}$, $\sigma(12) = I_1, \tau(12) = I_2, \mathcal{N}_{12}^+ = \{2, 11, 16\}, \mathcal{N}_{12}^- = \{5, 13, 17\}$. For the intersections, we have $\mathcal{I}_{I_1} = \{1, 2, 7, 8, 10, 11, 15, 16\}$ and $\mathcal{I}_{I_2} = \{4, 9, 12, 18\}$.

B. Dynamic Model

The discrete-time dynamic model of a road link $i \in \mathcal{R}$ is given by the following conservation law:

$$n_i(t+1) = n_i(t) + \sum_{j \in \mathcal{N}_i^+} f_{ji}(t) - f_i(t) + \mu_i(t) - \xi_i(t) \quad (1)$$

where t is the time instant index; $n_i(t)$ denotes the number of vehicles contained in the road link i at time tT; $f_i(t)$ is the downstream traffic flow departing from the road link i to its downstream neighbors or a node in $\mathcal{O}$, $f_{ji}(t)$ is the traffic flow exiting $j \in \mathcal{N}_i^+$ and entering into the road link i, $\xi_i(t)$ is the total outflow exiting the network at the road link i (to some places such as buildings, parking lots, etc.) and $\mu_i(t)$ is the exogenous cumulative inflow entering into the network from the source node $\sigma(i)$ if $\sigma(i) \in \mathcal{O}$ and the places in the road link i, all these traffic flows are measured by the numbers of corresponding vehicles in the time interval $[tT, (t+1)T]$; T is the time length between two consecutive instants t and $t+1$. In the traffic model (1), we consider $n_i(t)$ as traffic state, $f_i(t)$ and $f_{ji}(t), \forall j \in \mathcal{N}_i^+$, as control inputs, and $\xi_i(t)$ and $\mu_i(t)$ as disturbances. The equation (1) is a kind of vehicles conservation law, i.e., the number of total vehicles contained in the road link i at the time $(t+1)T$ is equal to the one at the time tT after adding the number of total vehicles traveling into this road link and subtracting the number of total vehicles exiting from this road link in the interval time $[tT, (t+1)T]$.

When we consider the traffic flows departing from the road link i and entering to its downstream neighbors, the turning ratio $r_{ij}(t)$ from this road link to one of its downstream neighbor $j \in \mathcal{N}_i^-$ is defined by

$$f_{ij}(t) = r_{ij}(t) f_i(t), \text{ where } \tau(i) \in \mathcal{J}. \quad (2)$$

Then we have $\sum_{j \in \mathcal{N}_i^-} r_{ij}(t) = 1$ and $r_{ij}(t) = \begin{cases} \geq 0, & \tau(i) = \sigma(j) \\ = 0, & otherwise \end{cases}$ for all t. Let $\mathbf{R}(t) = [r_{ij}(t)] \in \mathbb{R}^{N \times N}$ be the turning-ratio matrix. The following lemma shows that the matrix $\mathbf{R}(t)$ is Schur stable in general case. Its proof is given in APPENDX A.

Lemma 1: The turning-ratio matrix $\mathbf{R}(t)$ has the spectral radius less than 1 for all t.

From equations (1) and (2), we have a state model for the urban traffic network as follows:

$$\mathbf{n}(t+1) = \mathbf{n}(t) - (\mathbf{G}(t))^T \mathbf{f}(t) + \mathbf{e} \quad (3)$$

where $\mathbf{G}(t) = \mathbf{I} - \mathbf{R}(t)$ contains topological characteristics of the network and the turning ratios, $\mathbf{n}(t) = \big[n_1(t) \cdots n_N(t)\big]^T$, $\mathbf{f}(t) = \big[f_1(t) \cdots f_N(t)\big]^T$ and $\mathbf{e}(t) = \big[\mu_1(t) - \xi_1(t) \cdots \mu_N(t) - \xi_N(t)\big]^T$. Besides the equality constraints associated with the conservation law, we need to consider further constraints for a smooth operation of the traffic network.

1) Constraints From Cell Transmission Model (CTM)-Based Approximation: In cell transmission model [25], [26], a road cell c is a short-length segment with the same vehicle densities at all points in this cell. Let ρ_c be its density; then, the demand function $D_c(\rho_c)$ is the upper boundary of the flow rate of vehicles that want to exit c and the supply function $S_c(\rho_c)$ is the upper boundary of the flow rate that can be received by c.

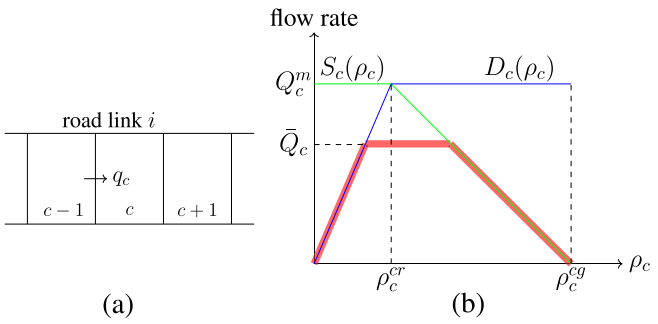


Fig. 2. (a) The road link i consists of multiple cells. (b) The supply and demand functions (i.e., $S_c(\rho_c)$ and $D_c(\rho_c)$) of a road cell c.

The demand function D_c is assumed to be non-decreasing and the supply function S_c is supposed to be non-increasing. In this paper, we consider the linear-piecewise form [20], [22] of these functions as $D_c(\rho_c) = \min\{u_c^{free}\rho_c, Q_c^m\}$ and $S_c(\rho_c) = \min\{Q_c^m, u_c^{back}(\rho_c^{cg} - \rho_c)\}$ (depicted in Fig.2b) where u_c^{free} is the free flow speed and u_c^{back} is the speed of the backward shock wave. The congestion density ρ_c^{cg} is defined as the point satisfying the condition $S_c(\rho_c) = 0, \forall \rho_c \geq \rho_c^{cg}$, and the critical density ρ_c^{cr} is the point satisfying the condition $S_c(\rho_c^{cr}) = D_c(\rho_c^{cr})$. Based on CTM theory [25], the allowable flow (denoted as $q_c(\Delta t)$) between two neighboring segments c and $c+1$ (the red line in Fig. 2b) is computed by

$$q_c(\Delta t) = \min\{n_{c-1}(\Delta t), \bar{Q}_c, \frac{u_c^{back}}{u_c^{free}}(N_c - n_c(\Delta t))$$

where $n_c(\Delta t)$, $\bar{Q}_c$ and N_c denote, respectively, the number of vehicles, the maximum allowable inflow, and the holding capacity. Since each road link consists of multiple road cells, its demand and supply can be computed as the corresponding summations of all its road cells. For simplicity, we assume that the free flow speed and the backward shock wave (i.e., u_i^{free} and u_i^{back}) of all cells in the road link i are the same and define its ratio as $u_i = \frac{u_i^{back}}{u_i^{free}}$. Let $\bar{n}_i = \sum_c N_c$ be the maximum number of vehicles that can be contained in the road link i. The number of contained vehicles and the downstream traffic flow of every road link $i \in \mathcal{R}$ are required to satisfy the constraints of their capacities at all time $t > 0$ as follows:

$$0 \leq n_i(t) \leq \bar{n}_i \tag{4a}$$

$$\sum_{j \in \mathcal{N}_i^+} f_{ji}(t) + \mu_i(t) \leq u_i(\bar{n}_i - n_i(t)) \tag{4b}$$

$$0 \leq f_i(t) \leq n_i(t) \tag{4c}$$

Since the outflows of the considered traffic network, i.e., the downstream traffic flows departing from the destination road links, are the exogenous inflows of other traffic networks, there may be some restrictions for these flows. Assume that the upper boundary of the downstream traffic flow of a destination road link i, where $\tau(i) \in \mathcal{O}$, is given by $f_i^m(t)$. We also set $f_i^m(t) = S_i T$ in the case of $\tau(i) \in \mathcal{J}$ where S_i is the saturated flow rate of vehicles departing from the road link i to the intersection $\tau(i)$. Then we have an additional constraint

$$f_i(t) \leq f_i^m(t)$$

for every road link $i \in \mathcal{R}$ with known $f_i^m(t)$.

2) Constraints for Green Time: Consider the boundary cell of the road link i connecting to the signalized intersection $\tau(i)$ when the signal switches to green. At the beginning, vehicles enter the intersection at the saturation flow rate. If the constraints (4b) are satisfied for all downstream neighbors of the road link i, i.e., there is enough space in the intersection $\tau(i)$ to receive vehicles departing from the road link i, the downstream traffic flow is maintained at the saturation rate in all active time duration. Then, the length of green (active) time $G_i(t)$ assigned to the road link i in the time interval $[tT, (t+1)T]$ is required to be larger than the necessary one for its downstream traffic flow, i.e., $G_i(t) \geq \frac{1}{S_i} f_i(t)$. To avoid the waste of green time, we let

$$G_i(t) = \frac{1}{S_i} f_i(t)$$

Consider two road links y and z belonging to the same traffic signal phase, which are activated simultaneously. Then, the relation between the green times assigned to these links can be represented by one or both of the following two inequalities:

$$\frac{1}{S_y} f_y(t) \leq \frac{1}{S_z} f_z(t) \text{ and } \frac{1}{S_y} f_y(t) \geq \frac{1}{S_z} f_z(t) \tag{5}$$

This relation can be determined based on current queue lengths or historical traffic demand. For example, the equality constraint $\frac{1}{S_y} f_y(t) = \frac{1}{S_z} f_z(t)$, which is equivalent to both the inequalities in (5), is used in the saturated scenario when there always exist vehicles waiting in these two road links. In another situation where the traffic demand of road link z is higher than the demand of road link y, the relation should be $\frac{1}{S_y} f_y(t) \leq \frac{1}{S_z} f_z(t)$. Let $\bar{\mathcal{I}}_\sigma \subseteq \mathcal{I}_\sigma$ be the set of road links which have the biggest assigned green time in their traffic signal phases in the intersection σ. Then the following inequality is required to be satisfied:

$$\sum_{j \in \bar{\mathcal{I}}_\sigma} \frac{1}{S_j} f_j(t) \leq T \tag{6}$$

For a simplicity, we stack the relation (5) and the constraint for green time of all traffic signal phases (6) into the following matrix form:

$$\mathbf{D}_\sigma(t)\mathbf{f}_\sigma(t) \leq \mathbf{d}_\sigma(t) \tag{7}$$

where $\mathbf{f}_\sigma(t)$ is the column vector whose elements are downstream traffic flows of the incoming road links of the intersection σ, i.e., $f_j(t)$ where $j \in \mathcal{I}_\sigma$. The matrix $\mathbf{D}_\sigma(t)$ and the vector $\mathbf{d}_\sigma(t)$ are assumed to be predetermined at time t. For the traffic network depicted in Fig. 1, we have $\mathbf{f}_{I_1}(t) = \left[f_1(t)\ f_2(t)\ f_7(t)\ f_8(t)\ f_{10}(t)\ f_{11}(t)\ f_{15}(t)\ f_{16}(t)\right]^T$ and $\mathbf{f}_{I_2}(t) = \left[f_4(t)\ f_9(t)\ f_{12}(t)\ f_{18}(t)\right]^T$. If we assume that the numbers of vehicles going through the road links 1, 2, 10, 11 are more than the ones through 16, 15, 8, 7, respectively,

then the matrix $\mathbf{D}_{I_1}(t)$ and the vector $\mathbf{d}_{I_1}(t)$ can be chosen as follows:

$$\mathbf{D}_{I_1}(t)=\begin{bmatrix}-1 & 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0\\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0\end{bmatrix}, \mathbf{d}_{I_1}(t)=\begin{bmatrix}0\\0\\0\\0\\T\end{bmatrix}.$$

For the intersection I_2, we have $\mathbf{D}_{I_2}(t) = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}$ and $\mathbf{d}_{I_2}(t) = T$.

3) Constraints From Uncertainties of Model Parameters: The turning-ratio matrix $\mathbf{R}(t)$ is usually assumed to be given [8], [15], [16], [17], [18], [19], [20], [21], [22], [23]. This assumption is motivated by the fact that drivers usually tend to choose the lane suitable for their turning choices; so $r_{ij}(t)$ could be estimated empirically or measured by road sensors. The simplest estimation model for turning ratios is based on the numbers of vehicles counted at the entrance and exit of lanes. They can be also measured by turning-ratio sensors (counting sensor, video sensors, Bluetooth and Wifi detector) installed on intersections [31], [32], [33]. Other strategies estimate the turning proportions relying on prior historical knowledge of arrival flows [34], [35]. Besides that, they also use data recollection campaigns to construct a database of average routing behavior, or these measurements are assumed to depend on a given probability distribution. Although recent advancement in sensing and information technology enhances the accuracy of collected data, it is impossible to predetermine the turning ratios exactly. Due to unpredictable situations and events such as weather conditions, accidents or drivers' decision changes, uncertainties in the estimation of the turning ratios cannot be avoided. Consequently, in this paper, we assume that these traffic model parameters are given with small uncertainties. Let $\underline{r}_{ij}(t)$ and $\overline{r}_{ij}(t)$ be the lower and upper bounds of possible values of turning split ratio $r_{ij}(t)$ and let $r^*_{ij}(t)$ be the nominal value measured or estimated by the road link i. Thus, $r^*_{ij}(t)$ needs to satisfy $\sum_{j\in\mathcal{N}_i^-} r^*_{ij}(t) = 1$ but $\sum_{j\in\mathcal{N}_i^-} \underline{r}_{ij}(t) < 1$, $\sum_{j\in\mathcal{N}_i^-} \overline{r}_{ij}(t) > 1$. In the case of there is no uncertainty or the measurement is absolutely accurate, we can write $\underline{r}_{ij}(t) = \overline{r}_{ij}(t) = r^*_{ij}(t)$. Assume that the nominal values $r^*_{ij}(t)$, $\forall j \in \mathcal{N}_i^-, i \in \mathcal{R}$, may be incorrect but highly reliable and the correct ones are bounded in small intervals $[\underline{r}_{ij}(t), \overline{r}_{ij}(t)]$. Mathematically, this assumption is given by

$$F_P(r_{ij}(t), r^*_{ij}(t)) \gg F_P(r_{ij}(t), r'_{ij}(t)), \forall r'_{ij}(t) \neq r^*_{ij}(t)$$
$$P\left(\underline{r}_{ij}(t) \le r_{ij}(t) \le \overline{r}_{ij}(t)\right) = 1$$

where $\overline{r}_{ij}(t) - \underline{r}_{ij}(t)$ is small for all $i \in \mathcal{R}, j \in \mathcal{N}_i^-$, $P(x)$ is the the probability of x and $F_P(x, y) = P(y - \Delta < x < y + \Delta)$ for any $\Delta > 0$.

In an urban traffic network, exogenous inflows and outflows are provided and received by not only source/destination nodes in $\mathcal{O}$ but also by some places (for example parking lot, building, etc) in its road links. This fact makes it difficult to measure $\mu_i(t)$ and $\xi_i(t)$. With similar manner as for turning ratios, we define $\underline{\mu}_i(t)$, $\overline{\mu}_i(t)$, $\mu^*_i(t)$, $\underline{\xi}_i(t)$, $\overline{\xi}_i(t)$, $\xi^*_i(t)$ as their lower, upper and nominal values and assume that:

$$F_P(\mu_i(t), \mu^*_i(t)) \gg F_P(\mu_i(t), \mu'_i(t)), \quad \forall \mu'_i(t) \neq \mu^*_i(t)$$
$$\times P\left(\underline{\mu}_i(t) \le \mu_i(t) \le \overline{\mu}_i(t)\right) = 1$$
$$F_P(\xi_i(t), \xi^*_i(t)) \gg F_P(\xi_i(t), \xi'_i(t)), \quad \forall \xi'_i(t) \neq \xi^*_i(t)$$
$$\times P\left(\underline{\xi}_i(t) \le \xi_i(t) \le \overline{\xi}_i(t)\right) = 1$$

where $\overline{\mu}_i(t) - \underline{\mu}_i(t)$ and $\overline{\xi}_i(t) - \underline{\xi}_i(t)$ are small for all $i \in \mathcal{R}$. Finally, Assumption 2 is required to solve the traffic control problem.

Assumption 2: The values of $r^*_{ij}(t)$, $\underline{r}_{ij}(t)$, $\overline{r}_{ij}(t)$, $\mu^*_i(t)$, $\underline{\mu}_i(t)$, $\overline{\mu}_i(t)$, $\xi^*_i(t)$, $\underline{\xi}_i(t)$, $\overline{\xi}_i(t)$ are given for every road link $i \in \mathcal{R}$ at time t and they are considered to be constant in next K cycles.

Consequently, with the given interval uncertainties, we have

$$\sum_{j\in\mathcal{N}_i^+} \underline{r}_{ji}(t) f_j(t) \le \sum_{j\in\mathcal{N}_i^+} r_{ji}(t) f_j(t) \le \sum_{j\in\mathcal{N}_i^+} \overline{r}_{ji}(t) f_j(t)$$
$$n_i(t+1) \ge n_i(t) + \underline{\mu}_i(t) - \overline{\xi}_i(t) + \sum_{j\in\mathcal{N}_i^+} \underline{r}_{ji}(t) f_j(t) - f_i(t)$$
$$n_i(t+1) \le n_i(t) + \overline{\mu}_i(t) - \underline{\xi}_i(t) + \sum_{j\in\mathcal{N}_i^+} \overline{r}_{ji}(t) f_j(t) - f_i(t)$$

Therefore, the road link capacity constraints (4a) are guaranteed if

$$n_i(t) + \underline{\mu}_i(t) - \overline{\xi}_i(t) + \sum_{j\in\mathcal{N}_i^+} \underline{r}_{ji}(t) f_j(t) - f_i(t) \ge 0 \quad (8a)$$
$$n_i(t) + \overline{\mu}_i(t) - \underline{\xi}_i(t) + \sum_{j\in\mathcal{N}_i^+} \overline{r}_{ji}(t) f_j(t) - f_i(t) \le \overline{n}_i \quad (8b)$$

III. Problem Formulation

A. Objective of Traffic Control

The main objective of urban traffic control is to determine the green time lengths assigned to road links in each traffic signal cycle in order to improve traffic conditions and guarantee a smooth operation for all roads and intersections. The green time lengths are computed based on the optimal downstream traffic flows. Assume that all signalized intersections in the considered urban traffic network have a common cycle time $C = cT$ where $c \in \mathbb{Z}^+$ is a positive integer. At the beginning of a new cycle, the following cost function needs to be minimized:

$$\Phi(t) = \sum_{k=1}^{K}\sum_{i=1}^{N} \Phi_i(t+k) \quad (9)$$

where K is the number of time steps ahead (here K needs to satisfy that K/c is an integer larger than 1) and $\Phi_i(t+k)$ is a function depending on the traffic state of the road link i at time $(t+k)T$ and its downstream traffic flow in the time interval $[(t+k-1)T, (t+k)T]$. The cost function $\Phi(t)$ corresponds to one or some performance indexes, which are used to evaluate the effectiveness of traffic control methods. It can be a quadratic function [15], [16] or can have a linear form [20]. Since there are some situations in which some road links are congested but the overall cost function $\Phi(t)$ achieves

a minimum value, the safety traffic constraints we formulated in the previous section need to be satisfied to avoid serious aftereffects. Let $f_i^{opt}(t+k), i \in \mathcal{R}, k = 1, \cdots, K$ be the obtained optimal downstream traffic flows. Then the green time lengths applied in this cycle correspond to only the first c time steps, i.e., $G_i(t+\tilde{k}) = \frac{1}{S_i} f_i^{opt}(t+\tilde{k}), \forall i : \tau(i) \in \mathcal{J}$, $\tilde{k} = 1, \cdots, c$.

Aiming to minimize the risk of traffic congestion and enhance the operational efficiency of the urban traffic network, we are interested in the following three performance indexes:

- The total time spent

$$TTS(t) = T \sum_{k=1}^{K} \sum_{i=1}^{N} n_i(t+k);$$

- The relative occupancy balance

$$ROB(t) = \sum_{k=1}^{K} \sum_{i=1}^{N} \frac{(n_i(t+k))^2}{\bar{n}_i};$$

- The total downstream traffic flow

$$TDF(t) = \sum_{k=0}^{K-1} \sum_{i=1}^{N} f_i(t+k)$$

Combining these performance indexes with weights, we consider the local objective function corresponding to a road link i, i.e., $\Phi_i(t+k)$ in (9), as the following equation:

$$\Phi_i(t+k) = a_i(n_i(t+k))^2 + b_i n_i(t+k) - w_i f_i(t+k-1) \tag{10}$$

where $a_i, b_i, w_i > 0$ are positive coefficients known by the road link i.

B. MPC Traffic Control Problem

We use $\hat{n}_{i,k} = n_i(t+k|t)$ and $\hat{f}_{i,k} = f_i(t+k-1|t)$ to denote the predicted number of vehicles of the road link $i \in \mathcal{R}$ at time $t+k$ and the predicted downstream traffic flow departing from the road link i in the time interval $[(t+k-1)T, (t+k)T]$ given the current states at time t. The cost function for the traffic control problem is given by

$$\Phi = \sum_{k=1}^{K} \sum_{i=1}^{N} \left(a_i \hat{n}_{i,k}^2 + b_i \hat{n}_{i,k} - w_i \hat{f}_{i,k} \right) \tag{11}$$

and the constraints described in Subsection II-B are rewritten as:

$$\hat{n}_{i,k} = \hat{n}_{i,k-1} + e_{i,k} + \sum_{j \in \mathcal{N}_i^+} r_{ji} \hat{f}_{j,k} - \hat{f}_{i,k} \tag{12a}$$

$$-\hat{f}_{i,k} \le 0 \tag{12b}$$

$$\hat{f}_{i,k} - \hat{n}_{i,k-1} \le 0 \tag{12c}$$

$$\hat{f}_{i,k} - \overline{f}_{i,k} \le 0 \tag{12d}$$

$$\sum_{j \in \mathcal{N}_i^+} r_{ji} \hat{f}_{j,k} + u_i \hat{n}_{i,k-1} - \bar{e}_{i,k} \le 0 \tag{12e}$$

$$\mathbf{D}_{\sigma,k} \hat{\mathbf{f}}_{\sigma,k} \le \mathbf{d}_{\sigma,k} \tag{12f}$$

$$-\hat{n}_{i,k-1} + \hat{f}_{i,k} - \sum_{j \in \mathcal{N}_i^+} \underline{r}_{ji} \hat{f}_{j,k} - \underline{n}_{i,k} \le 0 \tag{12g}$$

$$\hat{n}_{i,k-1} + \sum_{j \in \mathcal{N}_i^+} \overline{r}_{ji} \hat{f}_{j,k} - \hat{f}_{i,k} - \overline{n}_{i,k} \le 0 \tag{12h}$$

where $e_{i,k} = \mu_i^*(t+k|t) - \xi_i^*(t+k|t)$, $\underline{r}_{ij} = \underline{r}_{ij}(t)$, $\overline{r}_{ij} = \overline{r}_{ij}(t)$, $r_{ij} = r_{ij}^*(t)$, $\bar{e}_{i,k} = u_i(\bar{n}_i - \overline{\mu}_i(t+k|t) + \underline{\xi}_i(t+k|t))$, $\underline{n}_{i,k} = \underline{\mu}_i(t+k|t) - \overline{\xi}_i(t+k|t)$, $\overline{n}_{i,k} = \bar{n}_i + \underline{\xi}_i(t+k|t) - \overline{\mu}_i(t+k|t)$ and $\overline{f}_{i,k} = f_i^m(t+k|t)$ for all $i \in \mathcal{R}, \sigma \in \mathcal{J}, k = 1, \ldots, K$. We assume that $n_{i,0} = n_i(t)$ is known. The vectors and matrix in (12f) are defined as $\hat{\mathbf{f}}_{\sigma,k} = [\hat{f}_{i_1,k}, \cdots, \hat{f}_{i_\sigma,k}]^T$ where $\{i_1, \cdots, i_\sigma\} = \mathcal{I}_\sigma$, $\mathbf{D}_{\sigma,k} = \mathbf{D}_\sigma(t+k|t)$ and $\mathbf{d}_{\sigma,k} = \mathbf{d}_\sigma(t+k|t)$. For simplicity, we can choose $\mathbf{D}_\sigma(t+k|t) = \mathbf{D}_\sigma(t)$ and $\mathbf{d}_\sigma(t+k|t) = \mathbf{d}_\sigma(t)$ for all $k = 1, \ldots, K$. The parameters $n_{i,0}, e_{i,k}, r_{ij}, \overline{r}_{ij}, \underline{r}_{ij}, \overline{e}_{i,k}, \underline{e}_{i,k}, \bar{n}_i, v_i, \overline{f}_{i,k}, \mathbf{D}_{\sigma,k}, \mathbf{d}_{\sigma,k}$ are supposed to be known by the road link i at the beginning of cycle. Formally, the MPC traffic control problem is formulated by the constrained optimization problem (13) as follows:

$$\min_{\hat{n}_{i,k}, \hat{f}_{i,k}, i \in \mathcal{R}, k=1,\ldots,K} \Phi \text{ in (11) s.t. constraints in (12)} \tag{13}$$

Denoting $\hat{\mathbf{n}}_k = [\hat{n}_{1,k}, \cdots, \hat{n}_{N,k}]^T$, $\hat{\mathbf{f}}_k = [\hat{f}_{1,k}, \cdots, \hat{f}_{N,k}]^T$ and $\mathbf{e}_k = [e_{1,k}, \cdots, e_{N,k}]^T$, the conservation law (12a) could be written by the following matrix form:

$$\begin{aligned} \hat{\mathbf{n}}_k &= \hat{\mathbf{n}}_{k-1} + \mathbf{e}_k - \mathbf{G}^T \hat{\mathbf{f}}_k \\ &= \mathbf{n}_0 + \mathbf{e}_1 + \cdots + \mathbf{e}_k - \mathbf{G}^T \hat{\mathbf{f}}_1 - \cdots - \mathbf{G}^T \hat{\mathbf{f}}_k \end{aligned}$$

We also define two stack vectors $\hat{\mathbf{n}} = [\hat{\mathbf{n}}_1^T, \cdots, \hat{\mathbf{n}}_K^T]^T$ and $\hat{\mathbf{f}} = [\hat{\mathbf{f}}_1^T, \cdots, \hat{\mathbf{f}}_K^T]^T$. Combining with the equality constraint (12a), the objective function can be rewritten as

$$\Phi = 0.5 \hat{\mathbf{f}}^T \mathbf{Q} \hat{\mathbf{f}} + \mathbf{q}^T \hat{\mathbf{f}} + const \tag{14}$$

where $\mathbf{Q} = 2 \begin{bmatrix} K & K-1 & \cdots & 1 \\ K-1 & K-1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \otimes (\mathbf{G}\mathbf{A}\mathbf{G}^T)$,

$$\begin{aligned} \mathbf{q} = &2 \begin{bmatrix} K & K-1 & \cdots & 1 \end{bmatrix} \otimes (\mathbf{x}_0^T \mathbf{A} \mathbf{G}^T + \mathbf{b}^T \mathbf{G}^T) \\ &+2 \begin{bmatrix} K & K-1 & \cdots & 1 \end{bmatrix} \otimes (\mathbf{e}_1^T \mathbf{A} \mathbf{G}^T) \\ &+2 \begin{bmatrix} 0 & K-1 & \cdots & 1 \end{bmatrix} \otimes (\mathbf{e}_2^T \mathbf{A} \mathbf{G}^T) + \cdots \\ &+2 \begin{bmatrix} 0 & \cdots & 1 \end{bmatrix} \otimes (\mathbf{e}_K^T \mathbf{A} \mathbf{G}^T) - \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \otimes \mathbf{w} \end{aligned}$$

in which $\mathbf{A} = Diag(a_1, \ldots, a_N)$, $\mathbf{b} = [b_1, \ldots, b_N]^T$, $\mathbf{w} = [w_1, \ldots, w_N]^T$. The matrix $\mathbf{G}$ is a positive definite matrix due to Lemma 1. It implies $\mathbf{Q}$ is a positive definite matrix or $\Phi(t)$ is a strictly convex quadratic function of downstream traffic flows. Thus, the MPC traffic control problem (13) has a unique optimal solution if there exists at least one set of $\{\hat{f}_{i,k} : i \in \mathcal{R}, k = 1, \ldots, K\}$ satisfying the inequalities (12b-12h) strictly. It is easy to verify that the constraints in (12b-12h) can be described by the following affine form:

$$\mathbf{P}\hat{\mathbf{f}} + \mathbf{p} \le \mathbf{0} \tag{15}$$

We make the following assumption for the feasibility of the problem (13).

Assumption 3: The set $\mathcal{D} = \left\{ \hat{\mathbf{f}} : \mathbf{P}\hat{\mathbf{f}} + \mathbf{p} < \mathbf{0} \right\}$ is nonempty.

C. Distributed Control Problem Statement

Since an urban traffic network usually consists of many roads and intersections, effective traffic control strategies are required not only to be able to improve traffic conditions but also to smartly react to changes in the local regions. Compared to a centralized control system, a distributed approach achieves a better scalability and robustness. In this setup, an urban traffic network is divided into many subnetworks and each of them is controlled by one local controller. To optimize the operation of the whole urban traffic network, controllers cooperate to solve the constrained optimization problem (13). From the constraints in (12), it is easy to see that one road link $i \in \mathcal{R}$ requires information belonging to itself, its sink node $\tau(i)$ and its neighbors (in the set $\mathcal{N}_i^- \cup \mathcal{N}_i^+$) to check the feasibility of decision control variables $\hat{n}_{i,k}$, $\hat{f}_{i,k}$, $k = 1, \ldots, K$. So we make an assumption on information topology of the urban traffic network.

Assumption 4: Each road link i can receive information of its upstream and downstream neighbors, i.e., the road links in $\mathcal{N}_i^+ \cup \mathcal{N}_i^-$, and the information of its sink node $\tau(i)$. Each intersection σ could receive information of its incoming road links in the set $\mathcal{I}_\sigma$.

With the above assumption, we could state the main control problem as follows:

Problem 1: Design update laws for every road link and every intersection to solve the MPC traffic control problem (13) of an urban traffic network under Assumptions 1, 2, 3, 4.

In the next section, we first give optimality conditions for the constrained optimization problem (13) and develop an analytic solution for this problem by applying the dual ascent method (described in Appendix B) to an equivalent problem. Then we show how to implement the proposed method in a distributed manner in Section V. Although we aim to develop an update rule applied for each road link and intersection, we do not restrict our work to the assumption that each road link or intersection has exactly one controller. We will show in Remark 2 that our proposed method is applicable for various setups of distributed traffic control strategies.

IV. Centralized Method for MPC Traffic Control Problem

A. Optimality Condition

The Lagrangian function of the constrained optimization problem (13) is given by the equation (16), shown at the bottom of the page, where $\zeta_{i,k}$, $\alpha_{i,k}$, $\beta_{i,k}$, $\varrho_{i,k}$, $\nu_{i,k}$, $\lambda_{i,k}$, $\theta_{i,k}$, with $i \in \mathcal{R}$, and $\boldsymbol{\gamma}_{\sigma,k}$, with $\sigma \in \mathcal{J}$, are Lagrange multipliers.

We stack these multipliers into vectors as $\boldsymbol{\zeta} = [\boldsymbol{\zeta}_1^T, \boldsymbol{\zeta}_2^T, \cdots, \boldsymbol{\zeta}_K^T]^T$ where $\boldsymbol{\zeta}_k = [\zeta_{1,k}, \cdots, \zeta_{N,k}]^T$, $\boldsymbol{\chi}_{i,k} = [\lambda_{i,k}, \theta_{i,k}, \alpha_{i,k}, \beta_{i,k}, \varrho_{i,k}, \nu_{i,k}]^T$ for $i = 1, \ldots, N$, $k = 1, \ldots, K$, and $\boldsymbol{\eta} = \big[\boldsymbol{\chi}_{1,1}^T, \cdots, \boldsymbol{\chi}_{N,1}^T, \boldsymbol{\gamma}_{I_1,1}^T, \cdots, \boldsymbol{\gamma}_{I_M,1}^T, \boldsymbol{\chi}_{1,2}^T, \cdots, \boldsymbol{\chi}_{N,2}^T, \boldsymbol{\gamma}_{I_1,2}^T, \cdots, \boldsymbol{\gamma}_{I_M,2}^T, \cdots, \boldsymbol{\chi}_{1,K}^T, \cdots, \boldsymbol{\chi}_{N,K}^T, \boldsymbol{\gamma}_{I_1,K}^T, \cdots, \boldsymbol{\gamma}_{I_M,K}^T\big]^T$. So the Lagrangian function is in the form of $\mathcal{L} = \mathcal{L}(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\zeta}, \boldsymbol{\eta})$. It is noteworthy that the partial derivative $\nabla_{\boldsymbol{\zeta}} \mathcal{L}$ coincides with the equality constraint (12a) and does not depend on $\boldsymbol{\eta}$, the partial derivative $\nabla_{\boldsymbol{\eta}} \mathcal{L}$ corresponds to the inequality constraints (12b-12h) and does not depend on $\boldsymbol{\zeta}$, which means

$$\mathcal{L}(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \boldsymbol{\Phi}(\hat{\mathbf{n}}, \hat{\mathbf{f}}) + \boldsymbol{\zeta}^T \nabla_{\boldsymbol{\zeta}} \mathcal{L}(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\zeta}) + \boldsymbol{\eta}^T \nabla_{\boldsymbol{\eta}} \mathcal{L}(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\eta})$$

Let $\hat{\mathbf{n}}^{opt}, \hat{\mathbf{f}}^{opt}, \boldsymbol{\zeta}^{opt}$ $\boldsymbol{\eta}^{opt}$ be corresponding to the optimal primal and dual solutions of the constrained optimization problem (13). From the necessary and sufficient conditions for the convex optimization problem, we have

$$\nabla_{\hat{\mathbf{n}}} \mathcal{L}(\hat{\mathbf{n}}^{opt}, \hat{\mathbf{n}}^{opt}, \boldsymbol{\zeta}^{opt}, \boldsymbol{\eta}^{opt}) = \mathbf{0}, \tag{17a}$$

$$\nabla_{\hat{\mathbf{f}}} \mathcal{L}(\hat{\mathbf{n}}^{opt}, \hat{\mathbf{f}}^{opt}, \boldsymbol{\zeta}^{opt}, \boldsymbol{\eta}^{opt}) = \mathbf{0}, \tag{17b}$$

$$\nabla_{\boldsymbol{\zeta}} \mathcal{L}(\hat{\mathbf{n}}^{opt}, \hat{\mathbf{f}}^{opt}, \boldsymbol{\zeta}^{opt}) = \mathbf{0}, \tag{17c}$$

$$\nabla_{\boldsymbol{\eta}} \mathcal{L}(\hat{\mathbf{n}}^{opt}, \hat{\mathbf{f}}^{opt}, \boldsymbol{\eta}^{opt}) \leq \mathbf{0}, \tag{17d}$$

$$\boldsymbol{\eta}^{opt} \geq \mathbf{0}, \tag{17e}$$

$$Diag\left(\boldsymbol{\eta}^{opt}\right) \nabla_{\boldsymbol{\eta}} \mathcal{L}(\hat{\mathbf{n}}^{opt}, \hat{\mathbf{f}}^{opt}, \boldsymbol{\eta}^{opt}) = \mathbf{0} \tag{17f}$$

The above equations are KKT conditions of the problem (13). In which, the equations (17a) and (17b) are stationary conditions. The derivative of the Lagrangian function with respect to variables $\hat{n}_{i,k}$ (respectively, $\hat{f}_{i,k}$) is required to be

$$\begin{aligned}
\mathcal{L} = \sum_{k=1}^{K} \Bigg\{ & \sum_{i=1}^{N} \left(a_i \hat{n}_{i,k}^2 + b_i \hat{n}_{i,k} - w_i \hat{f}_{i,k} \right) + \sum_{i=1}^{N} \zeta_{i,k} \left(\hat{n}_{i,k-1} + e_{i,k} + \sum_{j \in \mathcal{N}_i^+} r_{ji} \hat{f}_{i,k} - \hat{f}_{i,k} - \hat{n}_{i,k} \right) \\
& + \sum_{i=1}^{N} \alpha_{i,k} \left(-\hat{f}_{i,k} \right) \\
& + \sum_{i=1}^{N} \beta_{i,k} \left(\hat{f}_{i,k} - \hat{n}_{i,k-1} \right) + \sum_{i=1}^{N} \varrho_{i,k} \left(\hat{f}_{i,k} - \overline{f}_{i,k} \right) + \sum_{i=1}^{N} \nu_{i,k} \left(\sum_{j \in \mathcal{N}_i^+} r_{ji} \hat{f}_{j,k} + u_i \hat{n}_{i,k-1} - \bar{e}_{i,k} \right) \\
& + \sum_{\sigma \in \mathcal{J}} \boldsymbol{\gamma}_{\sigma,k}^T \left(\mathbf{D}_{\sigma,k} \hat{\mathbf{f}}_{\sigma,k} - \mathbf{d}_{\sigma,k} \right) \\
& + \sum_{i=1}^{N} \lambda_{i,k} \left(-\hat{n}_{i,k-1} + \hat{f}_{i,k} - \sum_{j \in \mathcal{N}_i^+} \underline{r}_{ji} \hat{f}_{j,k} - \underline{n}_{i,k} \right) + \sum_{i=1}^{N} \theta_{i,k} \left(\hat{n}_{i,k-1} - \hat{f}_{i,k} + \sum_{j \in \mathcal{N}_i^+} \overline{r}_{ji} \hat{f}_{j,k} - \overline{n}_{i,k} \right) \Bigg\}
\end{aligned} \tag{16}$$

zero at the optimal solution. The equations (17c) and (17d) are equality and inequality constraints. For the multipliers corresponding to inequality constraints, they need to be nonnegative as shown in (17e) and need to satisfy the complementary slackness condition (17f). In (17f), $Diag\left(\boldsymbol{\eta}^{opt}\right)$ is a diagonal matrix whose main diagonal consists of the same elements as in the vector $\boldsymbol{\eta}^{opt}$.

B. Dual Ascent Method for an Equivalent Problem

1) Dual Problem: Let $\mathcal{F}$ be a subset of $\mathcal{R}^{2NK}$ where the equality constraint (12a) holds for all $i = 1, \ldots, N$. That is

$$\mathcal{F} = \{(\hat{\mathbf{n}}, \hat{\mathbf{f}}) : \hat{\mathbf{n}}_k = \hat{\mathbf{n}}_{k-1} + \mathbf{e}_k - \mathbf{G}^T \hat{\mathbf{f}}_k, \forall k = 1, \cdots, K\}.$$

The set $\mathcal{F}$ is convex and the following constrained optimization problem is equivalent to (13).

$$\inf_{(\hat{\mathbf{n}}, \hat{\mathbf{f}}) \in \mathcal{F}} \Phi(\hat{\mathbf{n}}, \hat{\mathbf{f}}) \text{ in (11) s.t. (12}b - 12h) \tag{18}$$

The Lagrangian function of the problem (18) is

$$\mathcal{L}^e(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\eta}) = \Phi(\hat{\mathbf{n}}, \hat{\mathbf{f}}) + \boldsymbol{\eta}^T \nabla_{\boldsymbol{\eta}} \mathcal{L}(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\eta}).$$

We define the dual function as $\Psi(\boldsymbol{\eta}) = \inf_{(\hat{\mathbf{n}}, \hat{\mathbf{f}}) \in \mathcal{F}} \mathcal{L}^e(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\eta})$; then the dual problem corresponging to the constrained optimization problem (18) is

$$\max_{\boldsymbol{\eta} \geq \mathbf{0}} \Psi(\boldsymbol{\eta}) \tag{19}$$

Under Assumption 3, the minimization problem (18) satisfies Assumption 5. According to Proposition 1 given in the Appendix-B, there exists at least one Lagrange multiplier $\boldsymbol{\eta}^*$ such that

$$\boldsymbol{\eta}^* = \arg\max_{\boldsymbol{\eta} \geq \mathbf{0}} \Psi(\boldsymbol{\eta}) \geq \mathbf{0} \tag{20}$$

and $\Phi^* = \inf_{(\hat{\mathbf{n}}, \hat{\mathbf{f}}) \in \mathcal{F}} \mathcal{L}^e(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\eta}^*)$ where Φ^* is the optimal value of (18). Moreover, if

$$Diag\left(\boldsymbol{\eta}^*\right) \nabla_{\boldsymbol{\eta}} \mathcal{L}^e(\hat{\mathbf{n}}^*(\boldsymbol{\eta}^*), \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*), \boldsymbol{\eta}^*) = \mathbf{0} \tag{21}$$

where $(\hat{\mathbf{n}}^*(\boldsymbol{\eta}^*), \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*)) = \arg\min_{(\hat{\mathbf{n}}, \hat{\mathbf{f}}) \in \mathcal{F}} \mathcal{L}^e(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\eta}^*)$, then $(\hat{\mathbf{n}}^*(\boldsymbol{\eta}^*), \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*))$ is the optimal solution of the primal problem (18) as stated in Proposition 2 (see Appendix-B). That means $(\hat{\mathbf{n}}^*(\boldsymbol{\eta}^*), \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*)) = (\hat{\mathbf{n}}^{opt}, \hat{\mathbf{f}}^{opt})$ because the problem (13) has a unique optimal solution.

Fixing $\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}}$, consider the problem $\inf_{(\mathbf{x}, \mathbf{f}) \in \mathcal{F}} \mathcal{L}^e(\mathbf{x}, \mathbf{f}, \tilde{\boldsymbol{\eta}})$ and its Lagrangian function is given by

$$\tilde{\mathcal{L}} = \boldsymbol{\Phi}(\hat{\mathbf{n}}, \hat{\mathbf{f}}) + \tilde{\boldsymbol{\eta}}^T \nabla_{\boldsymbol{\eta}} \mathcal{L}(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\eta}) + \boldsymbol{\zeta}^T \nabla_{\boldsymbol{\zeta}} \mathcal{L}(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\zeta})$$

Furthermore, we define two functions $h^n_{i,k}(\tilde{\boldsymbol{\eta}}) = b_i - \tilde{\beta}_{i,k+1} - \tilde{\lambda}_{i,k+1} + \tilde{\theta}_{i,k+1} + u_i \tilde{\nu}_{i,k+1}$ and $h^f_{i,k}(\tilde{\boldsymbol{\eta}}) = -w_i + \tilde{\lambda}_{i,k} - \sum_{j \in \mathcal{N}_i^-} \underline{r}_{ij} \tilde{\lambda}_{j,k} - \tilde{\theta}_{i,k} + \sum_{j \in \mathcal{N}_i^-} \overline{r}_{ij} \tilde{\theta}_{j,k} + \tilde{\varrho}_{i,k} - \tilde{\alpha}_{i,k} + \tilde{\beta}_{i,k} + \sum_{j \in \mathcal{N}_i^-} r_{ij} \tilde{\nu}_{j,k} + \tilde{\boldsymbol{\gamma}}^T_{\tau(i),k} \mathbf{D}^i_{\tau(i),k}$ where $\mathbf{D}^i_{\tau(i),k}$ is the column in the matrix $\mathbf{D}_{\tau(i),k}$ corresponding to the element $\hat{f}_{i,k}$ in the vector $\hat{\mathbf{f}}_{\tau(i),k}$. Set $\tilde{\beta}_{i,K+1} = \tilde{\nu}_{i,K+1} = \tilde{\lambda}_{i,K+1} = \tilde{\theta}_{i,K+1} = 0$ for all $i = 1, \cdots, N$. Then, taking the derivative of the Lagrangian function with respect to traffic states $\hat{\mathbf{n}}$, traffic downstream flows $\hat{\mathbf{f}}$, and Lagrangian multipliers corresponding to equality constraints $\boldsymbol{\zeta}$, we have

$$\frac{\partial \tilde{\mathcal{L}}}{\partial \hat{n}_{i,k}} = 2a_i \hat{n}_{i,k} - \zeta_{i,k} + \zeta_{i,k+1} + h^n_{i,k}(\tilde{\boldsymbol{\eta}}) \tag{22a}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial \hat{f}_{i,k}} = -\zeta_{i,k} + \sum_{j \in \mathcal{N}_i^-} r_{ij} \zeta_{j,k} + h^f_{i,k}(\tilde{\boldsymbol{\eta}}) \tag{22b}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial \zeta_{i,k}} = e_{i,k} + \hat{n}_{i,k-1} + \sum_{j \in \mathcal{N}_i^+} r_{ji} \hat{f}_{j,k} - \hat{f}_{i,k} - \hat{n}_{i,k} \tag{22c}$$

Let $\hat{\mathbf{n}}^*(\tilde{\boldsymbol{\eta}})$, $\hat{\mathbf{f}}^*(\tilde{\boldsymbol{\eta}})$ and $\boldsymbol{\zeta}^*(\tilde{\boldsymbol{\eta}})$ be the optimal solution of the problem $\inf_{(\hat{\mathbf{n}}, \hat{\mathbf{f}}) \in \mathcal{F}} \mathcal{L}^e(\hat{\mathbf{n}}, \hat{\mathbf{f}}, \boldsymbol{\zeta}, \tilde{\boldsymbol{\eta}})$ for a given $\tilde{\boldsymbol{\eta}}$. The optimality conditions are obtained by equalizing the right-hand sides of (22) to zero. They are equivalent to the following equations:

$$\mathbf{G}\boldsymbol{\zeta}^*_k(\tilde{\boldsymbol{\eta}}) = \mathbf{h}^f_k(\tilde{\boldsymbol{\eta}}) \tag{23a}$$

$$2\mathbf{A}\hat{\mathbf{n}}^*_k(\tilde{\boldsymbol{\eta}}) = \boldsymbol{\zeta}^*_k(\tilde{\boldsymbol{\eta}}) - \boldsymbol{\zeta}^*_{k+1}(\tilde{\boldsymbol{\eta}}) - \mathbf{h}^n_k(\tilde{\boldsymbol{\eta}}) \tag{23b}$$

$$\mathbf{G}^T \hat{\mathbf{f}}^*_k(\tilde{\boldsymbol{\eta}}) = \mathbf{n}^*_{k-1}(\tilde{\boldsymbol{\eta}}) - \mathbf{n}^*_k(\tilde{\boldsymbol{\eta}}) + \mathbf{e}_k \tag{23c}$$

where $\mathbf{h}^f_k(\tilde{\boldsymbol{\eta}}) = [h^f_{1,k}(\tilde{\boldsymbol{\eta}}), \cdots, h^f_{N,k}(\tilde{\boldsymbol{\eta}})]^T \in \mathbb{R}^N$ and $\mathbf{h}^n_k(\tilde{\boldsymbol{\eta}}) = [h^n_{1,k}(\tilde{\boldsymbol{\eta}}), \cdots, h^n_{N,k}(\tilde{\boldsymbol{\eta}})]^T \in \mathbb{R}^N$.

Solving (23), we obtain

$$\boldsymbol{\zeta}^*_k(\tilde{\boldsymbol{\eta}}) = \mathbf{G}^{-1} \mathbf{h}^f_k(\tilde{\boldsymbol{\eta}}) \tag{24a}$$

$$\hat{n}^*_{i,k}(\tilde{\boldsymbol{\eta}}) = \frac{\zeta^*_{i,k} - \zeta^*_{i,k+1} - h^n_{i,k}(\tilde{\boldsymbol{\eta}})}{2a_i} \tag{24b}$$

$$\hat{\mathbf{f}}^*_k(\tilde{\boldsymbol{\eta}}) = \mathbf{G}^{-T} \left(\mathbf{n}^*_{k-1}(\tilde{\boldsymbol{\eta}}) - \mathbf{n}^*_k(\tilde{\boldsymbol{\eta}}) + \mathbf{e}_k\right) \tag{24c}$$

From (24), it is easy to verify that the optimal traffic flow vector can be described in the following linear form:

$$\hat{\mathbf{f}}^*(\tilde{\boldsymbol{\eta}}) = \mathbf{Q}^{-1} \left(-\mathbf{P}^T \tilde{\boldsymbol{\eta}} - \mathbf{q}\right). \tag{25}$$

2) Iterative Update for Solving the Dual Problem: Since $\hat{\mathbf{n}}^*(\boldsymbol{\eta})$, $\hat{\mathbf{f}}^*(\boldsymbol{\eta})$ and $\boldsymbol{\zeta}^*(\boldsymbol{\eta})$ are continuous function with respect to $\boldsymbol{\eta}$, we have $\Psi(\boldsymbol{\eta}) = \mathcal{L}(\hat{\mathbf{n}}^*(\boldsymbol{\eta}), \hat{\mathbf{f}}^*(\boldsymbol{\eta}), \boldsymbol{\zeta}^*(\boldsymbol{\eta}), \boldsymbol{\eta})$, which is also continuous. According to Proposition 3, we have $\nabla_{\boldsymbol{\eta}} \Psi(\boldsymbol{\eta}) = \nabla_{\boldsymbol{\eta}} \mathcal{L}(\boldsymbol{\eta})$. The partial derivatives of the Lagrangian function $\mathcal{L}(\boldsymbol{\eta})$ with respect to Lagrange multipliers $\boldsymbol{\eta}$ coincide with the inequality constraints (12b-12h). So, $\nabla \Psi(\boldsymbol{\eta})$ has the linear form of $\nabla \Psi(\boldsymbol{\eta}) = \mathbf{P}\mathbf{f}^*(\boldsymbol{\eta}) + \mathbf{p}$. Then, we have

$$\nabla \Psi(\boldsymbol{\eta}^{(1)}) - \nabla \Psi(\boldsymbol{\eta}^{(2)}) = -\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T \left(\boldsymbol{\eta}^{(1)} - \boldsymbol{\eta}^{(2)}\right)$$

which implies

$$||\nabla \Psi(\boldsymbol{\eta}^{(1)}) - \nabla \Psi(\boldsymbol{\eta}^{(2)})|| \leq ||\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T|| \cdot ||\mathbf{f}^*(\boldsymbol{\eta}^{(1)}) - \mathbf{f}^*(\boldsymbol{\eta}^{(2)})||$$

Since the function $\Psi(\boldsymbol{\eta})$ is concave or $-\Psi(\boldsymbol{\eta})$ is convex, the optimal solution of the dual problme (19) is asymptotically achieved under the following gradient-based projection update law

$$\boldsymbol{\eta}(s+1) = [\mathbf{0}, \boldsymbol{\eta}(s) + \epsilon \nabla \Psi(\boldsymbol{\eta}(s))]^+$$

with a step size $\epsilon < \left(||\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T||\right)^{-1}$, where s is an estimation step index, and the function $\mathbf{v} = [\mathbf{u}]^+ \in \mathbb{R}^p$ is defined as $v_i = u_i$ if $u_i \geq 0$ and $v_i = 0$ otherwise. A limitation of classical gradient-based methods is a slow rate of convergence, which

is $O(\frac{1}{s})$. In order to enhance the convergence rate, we apply accelerated gradient method as follows:

$$\tilde{\boldsymbol{\eta}}(s) = \boldsymbol{\eta}(s) + \frac{s-1}{s+2}\left(\boldsymbol{\eta}(s) - \boldsymbol{\eta}(s-1)\right) \tag{26a}$$

$$\boldsymbol{\eta}(s+1) = \left[\tilde{\boldsymbol{\eta}}(s) + \epsilon \nabla \Psi(\tilde{\boldsymbol{\eta}}(s))\right]^{+} \tag{26b}$$

with $\boldsymbol{\eta}(0) = \mathbf{0}$. According to Proposition 4, we have

$$-\Psi(\boldsymbol{\eta}(s)) + \Psi(\boldsymbol{\eta}^*) \leq \frac{2||\boldsymbol{\eta}(0) - \boldsymbol{\eta}^*||_2^2}{\epsilon(s+1)^2}, \forall s \geq 1, \tag{27}$$

where $\boldsymbol{\eta}^* = \arg\max_{\boldsymbol{\eta} \geq \mathbf{0}} \Psi(\boldsymbol{\eta})$.

C. Convergence to the Optimal Solution of the Problem (13)

Following the proof of Theorem 3 in [36], we prove that $\mathbf{f}^*(\boldsymbol{\eta}(s))$ asymptotically converges to the optimal solution of the MPC traffic control problem (13) and the convergence rate is $O\left(\frac{1}{(s+1)^2}\right)$. This fact is stated in the following theorem.

Theorem 1: By applying the update rules (24) and (26) with a sufficiently small ϵ, the optimal solution of the constrained minimization problem (13) is asymptotically achieved in the following sense:

$$||\hat{\mathbf{f}}^*(\boldsymbol{\eta}(s)) - \mathbf{f}^{opt}||_2^2 \leq \frac{4||\mathbf{Q}^{-1}||}{\epsilon(s+1)^2}||\boldsymbol{\eta}^* - \boldsymbol{\eta}(0)||_2^2 \tag{28}$$

Proof: We have $2\Psi(\boldsymbol{\eta}) = -\left(\mathbf{P}^T\boldsymbol{\eta} + \mathbf{q}\right)^T \mathbf{Q}^{-1}\left(\mathbf{P}^T\boldsymbol{\eta} + \mathbf{q}\right) - 2\mathbf{p}^T\boldsymbol{\eta} = -\mathbf{q}^T\mathbf{Q}^{-1}\mathbf{q} - 2\mathbf{p}^T\boldsymbol{\eta} - \boldsymbol{\eta}^T\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T\boldsymbol{\eta} - 2\mathbf{q}^T\mathbf{Q}^{-1}\mathbf{P}^T\boldsymbol{\eta}$ and $\nabla\Psi(\boldsymbol{\eta}) = -\mathbf{P}\mathbf{Q}^{-1}\left(\mathbf{P}^T\boldsymbol{\eta} + \mathbf{q}\right) - \mathbf{p}$. They lead to

$$\begin{aligned}
&||\mathbf{P}^T\boldsymbol{\eta}(s) - \mathbf{P}^T\boldsymbol{\eta}^*||_{\mathbf{Q}^{-1}}^2 = \left(\boldsymbol{\eta}(s) - \boldsymbol{\eta}^*\right)^T \mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T\left(\boldsymbol{\eta}(s) - \boldsymbol{\eta}^*\right) \\
&= (\boldsymbol{\eta}(s))^T\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T\boldsymbol{\eta}(s) - (\boldsymbol{\eta}^*)^T\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T\boldsymbol{\eta}^* \\
&\quad -2(\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T\boldsymbol{\eta}^*)^T\boldsymbol{\eta}(s) + 2(\boldsymbol{\eta}^*)^T\mathbf{P}\mathbf{Q}^{-1}\mathbf{P}^T\boldsymbol{\eta}^* \\
&\quad +2\left(\mathbf{p} + \mathbf{P}\mathbf{Q}^{-1}\mathbf{q}\right)^T\left(\boldsymbol{\eta}(s) - \boldsymbol{\eta}^* + \boldsymbol{\eta}^* - \boldsymbol{\eta}(s)\right) \\
&= 2\Psi(\boldsymbol{\eta}^*) - 2\Psi(\boldsymbol{\eta}(s)) - \left(\mathbf{P}\mathbf{Q}^{-1}\left(\mathbf{P}^T\boldsymbol{\eta}^* + \mathbf{q}\right) + \mathbf{p}\right)^T(\boldsymbol{\eta}(s) \\
&\quad -\boldsymbol{\eta}^*) \\
&= 2\Psi(\boldsymbol{\eta}^*) - 2\Psi(\boldsymbol{\eta}(s)) + (\nabla\Psi(\boldsymbol{\eta}^*))^T\left(\boldsymbol{\eta}(s) - \boldsymbol{\eta}^*\right).
\end{aligned}$$

Considering the point $\boldsymbol{\eta}^* = \arg\max_{\boldsymbol{\eta} \geq \mathbf{0}} \Psi(\boldsymbol{\eta})$, we have

$$\left(\boldsymbol{\eta} - \boldsymbol{\eta}^*\right)^T \nabla\Psi(\boldsymbol{\eta}^*) \leq 0, \forall \boldsymbol{\eta} \geq \mathbf{0}$$

Since $||\hat{\mathbf{f}}^*(\boldsymbol{\eta}(s)) - \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*)||_2^2 = ||\mathbf{Q}^{-1}\left(\mathbf{P}^T\boldsymbol{\eta}(s) - \mathbf{P}^T\boldsymbol{\eta}^*\right)||_2^2 \leq ||\mathbf{Q}^{-1}|| \cdot ||\mathbf{P}^T\boldsymbol{\eta}(s) - \mathbf{P}^T\boldsymbol{\eta}^*||_{\mathbf{Q}^{-1}}^2$, from (27), we have $||\hat{\mathbf{f}}^*(\boldsymbol{\eta}(s)) - \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*)||_2^2 \leq \frac{4||\mathbf{Q}^{-1}||}{\epsilon(s+1)^2}||\boldsymbol{\eta}^* - \boldsymbol{\eta}(0)||_2^2$.

Let η_j and η_j^* be the j-th elements of the vectors $\boldsymbol{\eta}$ and $\boldsymbol{\eta}^*$ respectively, and let ∇_j be the j-th element of the vector $\nabla\Psi(\boldsymbol{\eta}^*)$; then we have

$$\left(\boldsymbol{\eta} - \boldsymbol{\eta}^*\right)^T \nabla\Psi(\boldsymbol{\eta}^*) = \sum_j \left(\eta_j - \eta_j^*\right)\nabla_j \leq 0, \forall \eta_j \geq 0. \tag{29}$$

Consider j^* where $\eta_j^* > 0$, the equation (29) holds if and only if $\nabla_j = 0$. In another case of $\eta_j^* = 0$, the equation (29) holds if and only if $\nabla_j \leq 0$. These facts imply that $\eta_j^*\nabla_j = 0$ and $\nabla_j \leq 0$ for all j. It means that $(\hat{\mathbf{n}}^*(\boldsymbol{\eta}^*), \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*), \boldsymbol{\zeta}^*(\boldsymbol{\eta}^*), \boldsymbol{\eta}^*)$ satisfies (17d), (17e) and (17f). In addition, the equation (23) guarantees the conditions (17a), (17b), (17c), respectively. Thus, all KKT conditions are satisfied or $(\hat{\mathbf{n}}^*(\boldsymbol{\eta}^*), \hat{\mathbf{f}}^*(\boldsymbol{\eta}^*), \boldsymbol{\zeta}^*(\boldsymbol{\eta}^*), \boldsymbol{\eta}^*)$ is the optimal solution of the constrained optimization problem (13). ■

Summarizing the analysis in this section, we provide Algorithm 1 as a centralized method to find the optimal solution of the problem (13) in which the stopping criterion is chosen as

$$||\boldsymbol{\eta}(s+1) - \boldsymbol{\eta}(s)||_\infty < \delta \tag{30}$$

for a sufficiently small δ.

Algorithm 1 Centralized Algorithm to Solve Problem (13)

1: **Input**: Parameters of road link $i \in \mathcal{R}$: $a_i, b_i, w_i, e_{i,k}$, $\overline{n}_{i,k}, \underline{n}_{i,k}, \overline{f}_{i,k}, r_{ij}, \underline{r}_{ij}, \overline{r}_{ij}$, $j \in \mathcal{N}_i^-$, small number ϵ.
2: **Output**: Optimal solution $\mathbf{f}^{opt} = \mathbf{f}^*(\boldsymbol{\eta}(s))$.
3: *Initialization*: $s \leftarrow 0$, $\boldsymbol{\eta}(s) \leftarrow \mathbf{0}$.
4: *Iterative update*:
5: **while** the stopping criterion is not satisfied **do**
6: $\tilde{\boldsymbol{\eta}}(s) \leftarrow (26a)$
7: $\mathbf{f}^*(\tilde{\boldsymbol{\eta}}(s)) \leftarrow (25)$
8: $\boldsymbol{\eta}(s+1) \leftarrow (26b)$
9: $s \leftarrow s+1$
10: **end while**

V. Distributed Algorithm for MPC Traffic Control Problem

The detailed formulation of $\boldsymbol{\chi}_{i,k}(s+1) = [\lambda_{i,k}(s+1), \theta_{i,k}(s+1), \alpha_{i,k}(s+1), \varrho_{i,k}(s+1), \beta_{i,k}(s+1), \nu_{i,k}(s+1)]^T$ for the road link $i \in \mathcal{R}$ and $\boldsymbol{\gamma}_{\sigma,k}(s+1)$ for the intersection $\sigma \in \mathcal{J}$ in (26b) are given in the equation (31), shown at the bottom of the next page.

It is easy to see that only local information is required to update these dual variables. In order to design a distributed version of Algorithm 1, we first provide a distributed method to find $\boldsymbol{\zeta}_k^*(\tilde{\boldsymbol{\eta}}), \hat{\mathbf{f}}_k{}^*(\tilde{\boldsymbol{\eta}})$ given in (24) in Subsection V-A. Then the main algorithm and discussions about implementation issues are provided in Subsection V-B.

A. Minimum-Time Final Value Computation

1) Computing the Final Value: The analysis in this subsection is inspired by the minimal time consensus method presented in [28], [29], and [30]. The main difference from these works is that we do not restrict to a consensus problem where the discrete state matrix is a stochastic matrix. Instead, we consider the following discrete-time model

$$\mathbf{x}(l+1) = \mathbf{M}\mathbf{x}(l) + \mathbf{m}, l \geq 1 \tag{32a}$$

$$y_r(l) = \mathbf{e}_n^T[r]\mathbf{x}(l). \tag{32b}$$

where $\mathbf{M} \in \mathbb{R}^{n \times n}$ is strictly stable, $\mathbf{m} \in \mathbb{R}^n$ is a constant vector and $\mathbf{e}_n[r]$ is an n-dimensional vector with all elements being zero except the r-th element being one. In (32), $\mathbf{x} \in \mathbb{R}^n$ is a variable vector and y_r is an observation corresponding to the r-th element in $\mathbf{x}$. Let $\mathbf{z}(l) = \mathbf{x}(l+1) - \mathbf{x}(l)$ be the difference

between two consecutive iterations. From (32), we can verify that

$$\mathbf{z}(l+1) = \mathbf{M}\mathbf{z}(l)$$

Notice that $z_r(l+i) = \mathbf{e}_n^T[r]\mathbf{M}^i\mathbf{z}(l) = y_r(l+i+1) - y_r(l+i)$. Since $\mathbf{M}$ has only stable eigenvalues, we have $\lim_{l\to\infty}\mathbf{z}(l) = \mathbf{0}_n$, which implies a convergence of $y_r(l)$, i.e., there exists $y_r^\infty = \lim_{l\to\infty} y_r(l)$.

For a matrix pair $(\mathbf{M}, \mathbf{e}_n^T[r])$, there is a minimal polynomial $q_r(t) = t^{m_r+1} + \sum_{i=0}^{m_r}\Theta_i t^i$ with minimum degree $m_r + 1 \le n$ that satisfies $\mathbf{e}_n^T[r]q_r(\mathbf{M}) = \mathbf{0}_n^T$. So, we have

$$\mathbf{e}_n^T[r]q_r(\mathbf{M})\mathbf{z}(l) = 0 = \mathbf{e}_n^T[r]\sum_{i=0}^{m_r+1}\Theta_i\mathbf{M}^i\mathbf{z}(l),\ \ \Theta_{m_r+1} = 1.$$

The above equation is equivalent to $0 = \Theta_0[y_r(l+1) - y_r(l)] + \Theta_1[y_r(l+2) - y_r(l+1)] + \cdots + \Theta_{m_r}[y_r(l+m_r+1) - y_r(l+m_r)] + \Theta_{m_r+1}[y_r(l+m_r+2) - y_r(l+m_r+1)]$ or

$$y_r(l)\Theta_0 + y_r(l+1)\Theta_1 + \cdots + y_r(l+m_r+1)\Theta_{m_r+1} = const$$

for all $l \ge 0$. Thus, the final value of (32)b is computed by

$$y_r^\infty = \frac{y_r(1)\Theta_0 + y_r(2)\Theta_1 + \cdots + y_r(m_r+2)\Theta_{m_r+1}}{\Theta_0 + \Theta_1 + \cdots + \Theta_{m_r+1}}. \quad (33)$$

Let $q(t)$ be the polynomial of the matrix $\mathbf{M}$, then $q(t) = \prod_{i=1}^n(t-\kappa_i)$ where κ_i is an eigenvalue of $\mathbf{M}$. So, 1 is not a root of $q(t)$. As shown in [28], $q_r(t)$ divides the minimal polynomial of $\mathbf{M}$. That means all roots of $q_r(t)$ are also roots of $q(t)$. It guarantees that all roots of $q_r(t)$ are different from 1 or the denominator of (33) is non-zero, i.e., $q_r(1) = \Theta_0 + \Theta_1 + \cdots + \Theta_{m_r+1} \ne 0$.

2) Distributed Method to Estimate Θ: In [30], Yuan *et al.* developed an algorithm to calculate the coefficients $\Theta_0, \Theta_1, \ldots, \Theta_{m_r}$ using only observations $y_r(l)$. The key idea is to find the first defective Hankel matrix given as follows:

$$\mathbf{H}_{z_r}[l] = \begin{bmatrix} z_r(0) & z_r(1) & \cdots & z_r(l) \\ z_r(1) & z_r(2) & \ldots & z_r(l+1) \\ \vdots & \vdots & \ddots & \vdots \\ z_r(l) & z_r(l+1) & \ldots & z_r(2l) \end{bmatrix} \quad (34)$$

Let $l = l^*$ be the first index where the matrix $\mathbf{H}_{z_r}[l]$ looses its rank, which means

$$Rank(\mathbf{H}_{z_r}[l^*]) = Rank(\mathbf{H}_{z_r}[l^*+h]),\ h = 1, 2, \ldots \quad (35a)$$

$$\mathbf{H}_{z_r}[l^*]\mathbf{\Theta} = \mathbf{0} \quad (35b)$$

where $\mathbf{\Theta}$ is the the corresponding normalized kernel. Hence, we have $l^* = m_r + 1$ and $\begin{bmatrix}\Theta_0\ \Theta_1 \cdots \Theta_{l^*-1}\ 1\end{bmatrix}^T \equiv \begin{bmatrix}\Theta_0\ \Theta_1 \cdots \Theta_{m_r}\ \Theta_{m_r+1}\end{bmatrix}^T$.

Next, we provide Algorithm 2 to determine l^* and $\mathbf{\Theta}$. Denoting $\mathbf{c}_l = \begin{bmatrix} z_r(l)\ z_r(l+1) \cdots z_r(2l-1)\end{bmatrix}^T$ and $\mathbf{Z}_l = \mathbf{H}_{z_r}[l]$, we have

$$\mathbf{Z}_l = \begin{bmatrix} \mathbf{Z}_{l-1} & \mathbf{c}_l \\ \mathbf{c}_l^T & z_r(2l) \end{bmatrix} \text{ for } l > 1 \quad (36)$$

Consider the Schur complement $\Lambda_l = z_r(2l) - \mathbf{c}_l^T\mathbf{Z}_{l-1}^{-1}\mathbf{c}_l$, we could verify that if $\Lambda_l \ne 0$ and $\mathbf{Z}_{l-1}$ is full rank, the matrix $\mathbf{Z}_l$ is nonsingular and its inverse matrix is given as

$$\mathbf{Z}_l^{-1} = \begin{bmatrix} \mathbf{Z}_{l-1}^{-1}(\mathbf{I}_l + \mathbf{c}_l\Lambda_l^{-1}\mathbf{c}_l^T\mathbf{Z}_{l-1}^{-1}) & -\mathbf{Z}_{l-1}^{-1}\mathbf{c}_l\Lambda_l^{-1} \\ -\Lambda_l^{-1}\mathbf{c}_l^T\mathbf{Z}_{l-1}^{-1} & \Lambda_l^{-1} \end{bmatrix} \quad (37)$$

It implies that $\mathbf{Z}_l$ will not loose the rank until $\Lambda_l = 0$. Or, the minimum index l^* that satisfies $\Lambda_{l^*} = 0$ corresponds to the first defective Hankel matrix. From (36), we have

$$\begin{bmatrix}\Theta_0\ \Theta_1 \cdots \Theta_{l^*-1}\end{bmatrix}^T = -\mathbf{Z}_{l^*-1}^{-1}\mathbf{c}_{l^*} \quad (38)$$

since $\mathbf{Z}_{l^*}\mathbf{\Theta} = \mathbf{0}$ and $\Theta_{l^*} = 1$. Since Algorithm 2 requires only observations $y_r(l)$, it can run in parallel with the update process (32).

$$\alpha_{i,k}(s+1) = \left[\tilde{\alpha}_{i,k}(s) + \epsilon(-\hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)))\right]^+ \quad (31a)$$

$$\beta_{i,k}(s+1) = \left[\tilde{\beta}_{i,k}(s) + \epsilon\left(\hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) - \hat{n}_{i,k-1}^*(\tilde{\boldsymbol{\eta}}(s))\right)\right]^+ \quad (31b)$$

$$\varrho_{i,k}(s+1) = \left[\tilde{\varrho}_{i,k}(s) + \epsilon(\hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) - \overline{f}_{i,k})\right]^+ \quad (31c)$$

$$\nu_{i,k}(s+1) = \left[\tilde{\nu}_{i,k}(s) + \epsilon\left(\sum_{j\in\mathcal{N}_i^+} r_{ji}\hat{f}_{j,k}^*(\tilde{\boldsymbol{\eta}}(s)) + u_i\hat{n}_{i,k-1}^*(\tilde{\boldsymbol{\eta}}(s)) - \bar{e}_{i,k}\right)\right]^+ \quad (31d)$$

$$\lambda_{i,k}(s+1) = \left[\tilde{\lambda}_{i,k}(s) + \epsilon\left(-\hat{n}_{i,k-1}^*(\tilde{\boldsymbol{\eta}}(s)) + \hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) - \sum_{j\in\mathcal{N}_i^+}\underline{r}_{ji}\hat{f}_{j,k}^*(\tilde{\boldsymbol{\eta}}(s)) - \underline{n}_{i,k}\right)\right]^+ \quad (31e)$$

$$\theta_{i,k}(s+1) = \left[\tilde{\theta}_{i,k}(s) + \epsilon\left(\hat{n}_{i,k-1}^*(\tilde{\boldsymbol{\eta}}(s)) - \hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) + \sum_{j\in\mathcal{N}_i^+}\overline{r}_{ji}\hat{f}_{j,k}^*(\tilde{\boldsymbol{\eta}}(s)) - \overline{n}_{i,k}\right)\right]^+ \quad (31f)$$

$$\boldsymbol{\gamma}_{\sigma,k}(s+1) = \left[\tilde{\boldsymbol{\gamma}}_{\sigma,k}(s) + \epsilon\left(\mathbf{D}_{\sigma,k}\hat{\mathbf{f}}_{\sigma,k}^*(\tilde{\boldsymbol{\eta}}(s)) - \mathbf{d}_{\sigma,k}\right)\right]^+ \quad (31g)$$

Algorithm 2 Minimum-Time Computation of the Final Value

Input: Successive observations of $y_r(i), l = 0, 1, \ldots$.
Output: Final value y_r^{∞} and vector of coefficients Θ.
Initialization: $l = 1$

$$\mathbf{Y} \leftarrow \frac{1}{z_r(0)}, \mathbf{c} \leftarrow z_r(1), \Lambda \leftarrow z_r(2) - \frac{(z_r(1))^2}{z_r(0)}.$$

First defective Hankel detection loop:
while $\Lambda \neq 0$ **do**
  $l \leftarrow l + 1$
  $\mathbf{Y} \leftarrow \begin{bmatrix} \mathbf{Y} + \mathbf{Y}\mathbf{c}\Lambda^{-1}\mathbf{c}^T\mathbf{Y} & -\mathbf{Y}\mathbf{c}\Lambda^{-1} \\ -\Lambda^{-1}\mathbf{c}^T\mathbf{Y} & \Lambda^{-1} \end{bmatrix}$
  $\mathbf{c} \leftarrow \begin{bmatrix} z_r(l) & \cdots & z_r(2l-1) \end{bmatrix}^T$
  $\Lambda \leftarrow z_r(2l) - \mathbf{c}^T\mathbf{Y}\mathbf{c}$
end while
Computation $\Theta \leftarrow \begin{bmatrix} -\mathbf{Y}\mathbf{c} \\ 1 \end{bmatrix}$, $y_r^{\infty} \leftarrow$ (33)

B. Main Algorithm

Since the matrix $\mathbf{R}$ is strictly stable, the solutions $\boldsymbol{\zeta}_k^*(\tilde{\boldsymbol{\eta}})$ in (24a) and $\hat{\mathbf{f}}_k^*(\tilde{\boldsymbol{\eta}})$ in (24c) can be found by applying the Jacobi method as follows:

$$\tilde{\boldsymbol{\zeta}}_k(l+1) = \mathbf{R}\tilde{\boldsymbol{\zeta}}_k(l) + \mathbf{h}_k^f(\tilde{\boldsymbol{\eta}}(s)),$$
$$\tilde{\mathbf{f}}_k(l+1) = \mathbf{R}^T\tilde{\mathbf{f}}_k(l) + \hat{\mathbf{n}}_{k-1}^*(\tilde{\boldsymbol{\eta}}(s)) - \hat{\mathbf{n}}_k^*(\tilde{\boldsymbol{\eta}}(s)) + \mathbf{e}_k.$$

As stated in the previous subsection, we have $\lim_{l\to\infty} \tilde{\boldsymbol{\zeta}}_k(l) = \boldsymbol{\zeta}_k^*(\tilde{\boldsymbol{\eta}}(s))$, $\lim_{l\to\infty} \tilde{\mathbf{f}}_k(l) = \hat{\mathbf{f}}_k^*(\tilde{\boldsymbol{\eta}}(s))$ and the Jacobi method allows road links to choose initial estimations $\tilde{\boldsymbol{\zeta}}_k(0)$ and $\tilde{\mathbf{f}}_k(0)$ arbitrarily. The detailed formulations of this application for each road link $i \in \mathcal{R}$ are given as

$$\tilde{\zeta}_{i,k}(l+1) = \sum_{j\in\mathcal{N}_i^-} r_{ij}\tilde{\zeta}_{j,k}(l) + h_{i,k}^f(\tilde{\boldsymbol{\eta}}(s)) \tag{39a}$$

$$\tilde{f}_{i,k}(l+1) = \sum_{j\in\mathcal{N}_i^+} r_{ji}\tilde{f}_{j,k}(l) + \hat{n}_{i,k-1}^*(\tilde{\boldsymbol{\eta}}(s)) - \hat{n}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) + e_{i,k} \tag{39b}$$

By applying Algorithm 2, each road link i is able to estimate coefficient vectors $\boldsymbol{\Theta}^{(\zeta_i)} = [\Theta_0^{(\zeta_i)}, \Theta_1^{(\zeta_i)}, \cdots, \Theta_{m_i}^{(\zeta_i)}, \Theta_{m_i+1}^{(\zeta_i)}]^T$ and $\boldsymbol{\Theta}^{(f_i)} = [\Theta_0^{(f_i)}, \Theta_1^{(f_i)}, \cdots, \Theta_{m_i'}^{(f_i)}, \Theta_{m_i'+1}^{(f_i)}]^T$ where $\Theta_{m_i+1}^{(\zeta_i)} = \Theta_{m_i'+1}^{(f_i)} = 1$, such that

$$\zeta_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) = \frac{\sum_{j=0}^{m_i+1} \Theta_j^{(\zeta_i)}\tilde{\zeta}_{i,k}(1+j)}{\sum_{j=0}^{m_i+1} \Theta_j^{(\zeta_i)}} \tag{40a}$$

$$\hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) = \frac{\sum_{j=0}^{m_i'+1} \Theta_j^{(f_i)}\tilde{f}_{i,k}(1+j)}{\sum_{j=0}^{m_i'+1} \Theta_j^{(f_i)}} \tag{40b}$$

with the data of $\tilde{\zeta}_{i,k}(l)$ obtained in (39a) (resp., $\tilde{f}_{i,k}(l)$ obtained in (39b)). In [37], Charalambous *et al.* show that there exists a set of initial states ($\tilde{\zeta}_{i,k}(0)$'s, $\tilde{f}_{i,k}(0)$'s) of measure zero for which the determined vector $\boldsymbol{\Theta}$ is different from $\boldsymbol{\Theta}^{(\zeta_i)}$ (resp., $\boldsymbol{\Theta}^{(f_i)}$). The main reason for which the algorithm fails is because the Hankel matrix (34) loses its rank too early. Although characterizing such a set of initial states is a hard problem, several practical techniques to cope this issue are available; e.g. see Remark 7 and Remark 8 in [37]. We propose here a simple strategy to avoid the failure. That is, once a road link has finished Algorithm 2, it uses the determined vector $\boldsymbol{\Theta}$ to compute the final state as in (33) and sends this value to its neighbors with a special flag. By this way, every road link knows the final values belonging to itself and its neighbors. So, it can check whether its local equation ((39a) or (39b)) is satisfied. If there is any violation determined by a road link i, it sends a special message to notify its neighbors to run Algorithm 2 again with other initial states.

Finally, we summarize our analysis by Algorithm 3 which is our main traffic control strategy.

Algorithm 3 Proposed Control Strategy for Each Road Link $i \in \mathcal{R}$ and Its Sink Node $\tau(i) \in \mathcal{J}$

Initialize: Apply Algorithm 2
  Determine $\boldsymbol{\Theta}^{(\zeta_i)}$, $\boldsymbol{\Theta}^{(f_i)}$ and number m_i, m_i'.
  $m_{max} = \max\left\{\{m_i\}_{i=1,\ldots,N}, \{m_i'\}_{i=1,\ldots,N}\right\}$
Iterative update:
  Initialization $s \leftarrow 0$, $\boldsymbol{\eta}_{i,k}(0) \leftarrow \mathbf{0}$, $k = 1, \ldots, K$
  while the stop criterion is not satisfied **do**
    $\tilde{\boldsymbol{\eta}}(s) \leftarrow$ (26a)
    Run (39a) in m_{max} times, then $\zeta_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) \leftarrow$ (40a),
    $\hat{n}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) \leftarrow$ (24b),
    Run (39b) in m_{max} times, then $\hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s)) \leftarrow$ (40b).
    $\boldsymbol{\chi}_{i,k}(s+1) \leftarrow$ (31a-31f)
    $\boldsymbol{\gamma}_{\tau(i),k}(s+1) \leftarrow$ (31g) if $\tau(i) \in \mathcal{J}$
    $s \leftarrow s + 1$
  end while
Finish algorithm $\hat{f}_{i,k}^{opt} \leftarrow f_{i,k}^*(\boldsymbol{\eta}(s))$

It is noteworthy that only local information is required in the update rules (24b), (26a), (31), (39), (40) (see more detailed explanation in Remark 1 and Remark 2). Algorithm 3 is a distributed version of Algorithm 1 and its correctness is guaranteed by Theorem 1. Thus we have the following theorem.

Theorem 2: By applying Algorithm 3, each road link $i \in \mathcal{R}$ is able to determine its downstream traffic flows $\hat{f}_{i,k}^{opt}, k = 1, \ldots, K$ in the optimal solution of the constrained minimization problem (13) while using only its local information.

Remark 1: In Algorithm 3, we consider the traffic model parameters $(e_{i,k}, \overline{e}_{i,k}, \overline{f}_{i,k}, \overline{n}_{i,k}, \underline{n}_{i,k}, r_{ij}, \underline{r}_{ij}, \overline{r}_{ij}$ where $j \in \mathcal{N}_i^-)$, the updated multipliers $\boldsymbol{\chi}_{i,k}(s)$ in (31), $\tilde{\boldsymbol{\chi}}_{i,k}(s)$ in (26), the Jacobi estimation $\tilde{\zeta}_{i,k}(l)$, $\tilde{f}_{i,k}(l)$ in (39), and the computed values $\hat{n}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s))$ in (24b), $\zeta_{i,k}^*(\tilde{\boldsymbol{\eta}}(s))$, $\hat{f}_{i,k}^*(\tilde{\boldsymbol{\eta}}(s))$ in (40) as the information of the road link $i \in \mathcal{R}$ for all $k = 1, \ldots, K$. It is clear that the road link $i \in \mathcal{R}$ uses only its own information in the equations (24b), (26a), (40) and the updates (31a)-(31c) of the multipliers $\alpha_{i,k}, \beta_{i,k}, \varrho_{i,k}$. For the updates of other multipliers and the estimated downstream traffic flow, the road link i needs additional information from its neighboring road links and its sink node. It includes the estimated solution $\hat{f}_{j,k}^*(\tilde{\boldsymbol{\eta}}(s))$ and traffic model parameters $r_{ji}, \overline{r}_{ji}, \underline{r}_{ji}$ from every upstream road link $j \in \mathcal{N}_i^+$ for the update of the multipliers

$v_{i,k}, \lambda_{i,k}, \theta_{i,k}$ in (31d)-(31f); the values of $\tilde{\lambda}_{j,k}(s), \tilde{\theta}_{j,k}(s)$ from its downstream road link $j \in \mathcal{N}_i^-$ and the values of $\tilde{\gamma}_{\tau(i),k}(s)$ from its sink node (if $\tau(i) \in \mathcal{J}$) for the update of the function $h_{i,k}^f(\tilde{\boldsymbol{\eta}}(s))$; and the computed values $\tilde{\zeta}_{j,k}(l), l = 1, 2, \ldots,$ from its downstream neighbor $j \in \mathcal{N}_i^-$ and $r_{ji}\tilde{f}_{j,k}(l), l = 1, 2, \ldots,$ from its upstream neighbor $j \in \mathcal{N}_i^+$ in the Jacobi update (39).

In the case of an intersection $\sigma \in \mathcal{J}$, its own information includes the matrices $\mathbf{D}_{\sigma,k}$, the vectors $\mathbf{d}_{\sigma,k}$ and the multipliers $\gamma_{\sigma,k}(s), \tilde{\gamma}_{\sigma,k}(s)$ for all $k = 1, \ldots, K$. The intersection σ uses only its own information in the update (26a) and it requires additional information from the incoming road links in the set $\mathcal{I}_\sigma$, i.e., $\hat{\mathbf{f}}_{\sigma,k}^*(\tilde{\boldsymbol{\eta}}(s))$, to update the multipliers $\gamma_{\sigma,k}(s), k = 1, \ldots, K$, as in (31g).

Remark 2: Since Algorithm 3 is applicable for every road link and every intersection, it is reasonable to expect a significant reduction of execution time when there are many computation units working in parallel. We present here a suggestion for the implementation of our proposed distributed traffic control strategy. Assume the considered urban traffic network is divided into L subnetworks, denoted by $\mathcal{P}_i, i = 1, \ldots, L$, and each subnetwork $\mathcal{P}_i$ is controlled by one local controller $\mathcal{C}_{\mathcal{P}_i}$. Let $\mathcal{J}_{\mathcal{P}_i}$ be the set of internal signalized intersections and $\mathcal{R}_{\mathcal{P}_i} = \{z \in \mathcal{R} : \tau(z) \in \mathcal{J}_{\mathcal{P}_i}\} \cup \{z \in \mathcal{R} : \sigma(z) \in \mathcal{J}_{\mathcal{P}_i}, \tau(z) \in \mathcal{O}\}$ be the set of internal road links of the subnetwork $\mathcal{P}_i$. The controller $\mathcal{C}_{\mathcal{P}_i}$ needs to measure/estimate traffic model parameters and decide the green time duration for road links in the set $\mathcal{R}_{\mathcal{P}_i}$. Its own information consists of the one belonging to all road links in the set $\mathcal{R}_{\mathcal{P}_i}$ and all signalized intersections in the set $\mathcal{J}_{\mathcal{P}_i}$. Let $\mathcal{R}_{\mathcal{P}_j\mathcal{P}_i} = \{z \in \mathcal{R}_{\mathcal{P}_i} : \sigma(z) \in \mathcal{J}_{\mathcal{P}_j}, \tau(z) \in \mathcal{J}_{\mathcal{P}_i}\}$ be the set of road links connecting $\mathcal{P}_j$ to $\mathcal{P}_i$. Since the upstream neighbors of a road link in $\mathcal{R}_{\mathcal{P}_j\mathcal{P}_i}$ are contained in $\mathcal{R}_{\mathcal{P}_j}$ while its downstream neighbors are contained in $\mathcal{R}_{\mathcal{P}_i}$, controllers $\mathcal{C}_{\mathcal{P}_i}$ and $\mathcal{C}_{\mathcal{P}_j}$ need to exchange information when applying Algorithm 3 if $\mathcal{R}_{\mathcal{P}_j\mathcal{P}_i} \neq \emptyset$.

At the beginning of every cycle, the local controller $\mathcal{C}_{\mathcal{P}_i}$ runs Algorithm 3 for all road links in the set $\mathcal{R}_{\mathcal{P}_i}$ and all intersections in the set $\mathcal{J}_{\mathcal{P}_i}$. It uses the information of itself and the one received from the controller $\mathcal{C}_{\mathcal{P}_j}$ if $\mathcal{R}_{\mathcal{P}_j\mathcal{P}_i} \neq \emptyset$ or $\mathcal{R}_{\mathcal{P}_i\mathcal{P}_j} \neq \emptyset$. For each road link $z \in \mathcal{R}_{\mathcal{P}_j\mathcal{P}_i}$, $\mathcal{C}_{\mathcal{P}_i}$ needs to receive information of $y \in \mathcal{N}_z^+$ from $\mathcal{C}_{\mathcal{P}_j}$ (including $r_{yz}, \underline{r}_{yz}, \overline{r}_{yz}$, $\hat{f}_y^*(\tilde{\boldsymbol{\eta}}(s))$ for the update (31) and $\tilde{f}_{y,k}(l)$ for the update (39b) of the road link z) and needs to send information of z to $\mathcal{C}_{\mathcal{P}_j}$ (including $\tilde{\zeta}_{z,k}(l)$ for the update (39a) of the road link $u \in \mathcal{N}_z^+$).

Remark 3: It is remarkable that Algorithm 3 has some key properties of distributed strategies. At the beginning of each cycle, the current traffic states are used to determine the optimal downstream traffic flows which satisfy all safety constraints. The updates of every road link $i \in \mathcal{R}$ and every intersection $\sigma \in \mathcal{J}$ require only local information. If there is any structural change (for example, one road link is added or removed), only controllers corresponding to this road link and its neighbors need to be reprogrammed. In the case of any road link j cannot be operated due to accident or other reasons, Algorithm 3 is still applicable for remaining road links with the setup of $r_{ij} = 0$ for all $i \in \mathcal{R}$ and $\bar{f}_{j,k} = 0$ for all k.

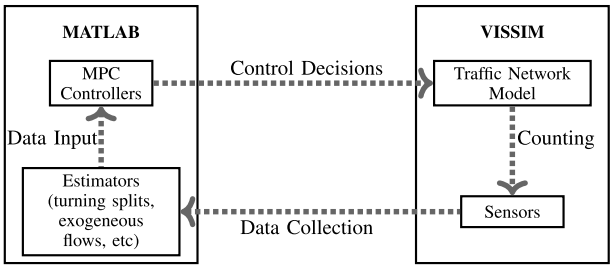


Fig. 3. Information exchange between VISSIM and MATALB.

Fig. 4. The mainstream road network of Gangnam district in Seoul is simulated (Image is captured from VISSIM software).

VI. Simulation

A. Microscopic Simulation

1) Simulation Setup: To verify the effectiveness of our proposed traffic control strategy (Algorithm 3), we conduct simulations using VISSIM.[1] This software allows us to implement the traffic signal timing plan, which is obtained by customized traffic management strategies, into traffic simulation. Our control algorithm is written in MATLAB. Fig. 3 illustrates the information flow between VISSIM and MATALB. MATLAB program employs COM interface provided by VISSIM to transfer simulation data from VISSIM to MATLAB and transfer control decisions from MATLAB to VISSIM. The traffic network used in the simulation is a hypothesis traffic network, which is based on the mainstream road network of Gangnam district in Seoul, and is illustrated in Fig. 4. In this figure, internal intersections are marked by red and yellow squares while the source/destination nodes are marked by black circles. The red squares correspond to four-arms intersections whose traffic signal phases are set as Type 1, and the yellow squares correspond to three-arms intersections and four-arms intersections whose traffic signal phases are set as Type 2. The simulated traffic network consists of 654 road links and 131 intersections, in which there are 35 four-arms intersections with Type 1, and 50 four-arms intersections and

[1] VISSIM is a microscopic traffic simulation software that is widely used to evaluate various traffic management methods. Vissim PTV AG is available at www.vissim.de.

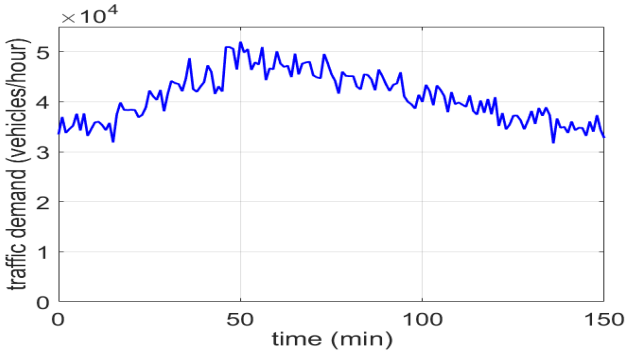


Fig. 5. Medium level traffic demand of the traffic network.

46 three-arms intersections with Type 2. The capacities of these road links depend on their lengths, which range from 120 to 750 meters. The saturated flow of a road link is set by 0.53 vehicle/lane/second. In this simulation, we assume that every destination road link connects to another traffic region with infinite free space.

We consider two different scenarios in traffic demand: (1) the case of medium level demand and (2) the case of high level demand. In the first case, the flow rate of the exogenous inflow entering into the network is given in Fig. 5. The traffic demand in the high level case is 120 percents of the one in the medium level case. Let $r^l_{\sigma\tau}(t), r^r_{\sigma\tau}(t)$ and $r^s_{\sigma\tau}(t)$ be respectively the nominal ratios of turning left, turning right and going straight directions of a road connecting two consecutive intersections σ and τ from time t minute to time $t+1$ minute. We set $r^s_{\sigma\tau}(t)$ randomly as $r^s_{\sigma\tau}(t) \in \{0.6, 0.8\}$ when there is a straight direction and $r^s_{\sigma\tau}(t) = 0$, otherwise (for example in some three-arms intersections), and $r^l_{\sigma\tau}(t) = r^r_{\sigma\tau}(t) = 0.5 - 0.5 r^s_{\sigma\tau}(t)$. Assume that these nominal turning ratios are estimated by a local controller while the correct ones are $\hat{r}^l_{\sigma\tau}(t) = r^l_{\sigma\tau}(t) + \delta^l_{\sigma\tau}(t)$, $\hat{r}^r_{\sigma\tau}(t) = r^r_{\sigma\tau}(t) + \delta^r_{\sigma\tau}(t)$ and $\hat{r}^s_{\sigma\tau}(t) = 1 - \hat{r}^l_{\sigma\tau}(t) - \hat{r}^r_{\sigma\tau}(t)$ where $\delta^l_{\sigma\tau}(t), \delta^r_{\sigma\tau}(t)$ are random numbers in the range $[-0.05, 0.05]$.

2) Simulation Results: In order to employ Algorithm 3, we need to choose the weights a_i, b_i and w_i for the local cost function (10) where $i \in \mathcal{R}$. From the simulation results, we observe that a_i should be chosen such that $a_i \bar{n}_i = const$ for all i in order to balance vehicles distribution over the traffic network. When setting $a_i = \frac{100}{\bar{n}_i}$ for all road links, the bigger coefficient b_i plays the role of driving vehicles to their corresponding destination road links and exiting the considered traffic network faster. However, if b_i is set to too large, it will increase the pressure of destination road links and their neighbors frequently. So, b_i should be chosen such that $b_i n_i(t) > a_i (n_i(t))^2$ when the road link i has low occupancy and $b_i n_i(t) < a_i (n_i(t))^2$ otherwise. The coefficient w_i is to render more vehicles to travel from roads to roads through intersections. Similarly as the weight b_i, the weight w_i should be larger than $a_i n_i(t)$ when the downstream neighbors of the road link i have low occupancy and should be smaller otherwise. Thus, the local cost function (10) is designed as

$$\Phi_i(k) = \frac{100}{\bar{n}_i} (n_i(k+1))^2 + 15 n_i(k+1) - 15 f_i(k) \quad (41)$$

Five evaluation criteria $N_{total}, T_{ave}, T_{eff}, N_{wait}$ and N_{high} are calculated to compare the performance of our proposed control algorithm with other existing methods. These criteria are described in what follows.

- The total number of vehicles served by the network is

$$N_{total} = \sum_{t=0}^{N_C - 1} \sum_{i \in \mathcal{R}} \Delta \mu_i(t) \text{ (vehicles)}$$

  where N_C is the simulation time (in minutes) and $\Delta \mu_i(t)$ is the number of all vehicles entering into the road link i in $(t+1)^{th}$-minute from the outside environment.
- The total time spent is a summation of the times all vehicles spent inside the network (i.e., summed over all severed vehicles). The average total time spent per vehicle is

$$T_{ave} = \frac{1}{N_{total}} \sum_{t=1}^{N_C} \sum_{i \in \mathcal{R}} \Delta n_i(t) \text{ (minutes)}$$

  where $\Delta n_i(t)$ is the traffic state of the road link i at time t minute. By reducing this index, overall travel time is reduced.
- The total number of vehicles through one intersection (or the considered network) reflects its operational effectiveness. Thus, we are interested in the following performance index

$$T_{eff} = \sum_{t=0}^{N_C - 1} \sum_{i \in \mathcal{R}} \Delta f_i(t) \text{ (vehicles)}$$

  where $\Delta f_i(t)$ is the total downstream traffic flow departing from the road link i in $(t+1)^{th}$-minute.
- The number of vehicles which have to wait in the road link i during the t^{th}-minute is $N_{i,wait}(t) = max\{0, \Delta n_i(t) - \Delta f_i(t)\}$. The average number of stops per vehicle is

$$N_{wait} = \frac{1}{N_{total}} \sum_{t=0}^{N_C - 1} \sum_{i \in \mathcal{R}} N_{i,wait}(t)$$

- To measure the risk of congestion at the beginning of every minute, we consider the number of road links whose occupancy is larger than the number δ_{high}. Summing these numbers over all simulation time, we have

$$N_{high} = \sum_{t=1}^{N_C} \left| \left\{ i \in \mathcal{R} : \frac{n_i(t)}{\bar{n}_i} \geq \delta_{high} \right\} \right|$$

  where $|(\cdot)|$ denotes the cardinality of the set in $(\cdot)$.

To validate the effectiveness of our proposed control method, we compare the above performance indexes with two existing control strategies. In the simulations, the traffic signal cycle time is set as $C = 60$ seconds for all intersections and the simulation time is set as $N_C = 150$ minutes.

- (Existing strategy 1) The pretimed signal control: Every intersection $\sigma \in \mathcal{J}$ has predetermined and fixed green timing plan regardless of changes in traffic states of road links. The green time length of a traffic signal phase in a cycle is determined as $\frac{w_i}{\sum_{j \in \overline{\mathcal{I}}_\sigma} w_j} C$ where w_j is the weight of the road link $j \in \mathcal{R}$ and i is the road link with biggest weight in this phase. The weights

TABLE III
SIMULATION RESULTS OF THREE CONTROL STRATEGIES

	Strategies	N_{total}	T_{ave}	N_{wait}	T_{eff}	N_{high}
Med. level	Pretimed	123296	16.8	5.3	1299200	1539
	BP-based	123307	10.3	2.3	1465227	113
	Algorithm 3	123307	10.2	1.8	1466672	0
High level	Pretimed	126426	24.4	14.8	1199010	8595
	BP-based	135341	20.6	11.3	1246350	8235
	Algorithm 3	144066	13.9	4.5	1676247	127

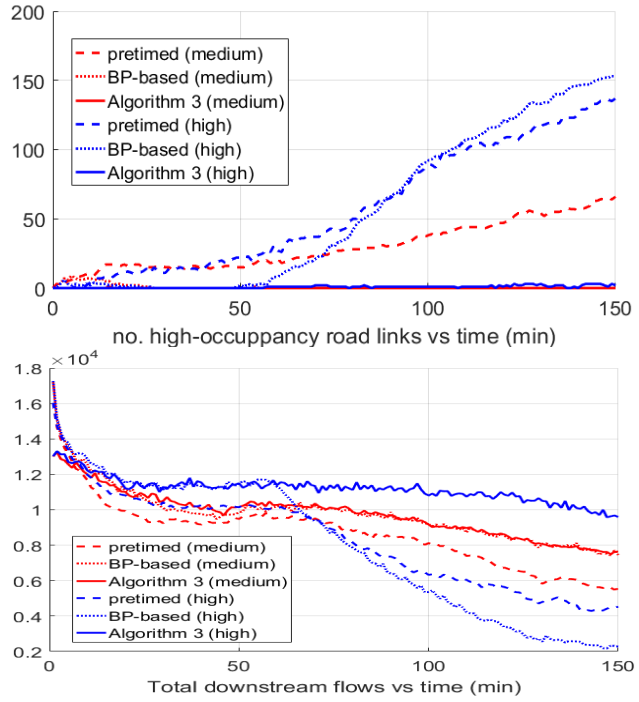


Fig. 6. The number of road links having high occupancy and the total flows vs time. The red lines represent the cases of medium level demand and the blue lines represent the cases of high level demand.

are proportional to the demand of traffic movements based on historical data. Here, we set ϖ_i as the i^{th}-element in the vector of the planed total traffic flows $\mathbf{f}^{plan} = \sum_{t=0}^{N_c-1} \mathbf{G}(t)^{-T} \Delta\boldsymbol{\mu}(t)$ where $\mathbf{G}(t)$ is constructed from the turning ratios given in the simulation setup and $\Delta\boldsymbol{\mu}(t) = [\Delta\mu_1(t), \cdots, \Delta\mu_N(t)]^T$ with $\Delta\mu_i(t)$ being the exogenous flow of the road link $i \in \mathcal{R}$.

- (Existing strategy 2) The back-pressure (BP)-based control: This method is firstly proposed in [38] and is modified in [39]. The control step is chosen by 10 second.
- (Our proposed method) Algorithm 3: The time interval between two time instants is chosen as $T = 30$ seconds; so $c = \frac{C}{T} = 2$. The employed local cost function is (41).

TABLE III shows the simulation results for two scenarios (i.e., medium level demand and high level demand). In the first one corresponding to the case of medium level demand, we set $\delta_{high} = 0.65$. For the second case corresponding to high level demand, $\delta_{high} = 0.85$. In Fig. 6, for the three methods (i.e., pretimed signal control, BP-based control and our proposed control strategies), we show the number of road links having high occupancy and the total downstream traffic flows of all road links in every minutes. It is observed that our proposed control has the advantage of 1) reducing the number of the road links having high occupancy, and 2) maintaining a high level of operating efficiency for intersections in longer time. Moreover, the proposed method is successful for reducing the risk of traffic congestion (N_{high}) and the travel time delay (T_{ave}, N_{wait}) while enhancing the total downstream traffic flows (T_{eff}).

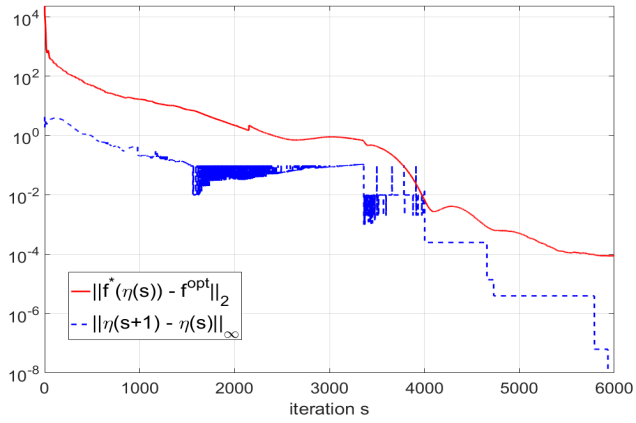


Fig. 7. The evolution of two functions $U(s)$ and $V(s)$ when applying Algorithm 3 with $K = 4$. The y-axis is in log-scale.

B. Computational Effectiveness of Proposed Algorithms

Let the function $U(s) = ||\boldsymbol{\eta}(s+1) - \boldsymbol{\eta}(s)||_\infty$ represent the difference between two consecutive dual variables and the function $V(s) = ||\mathbf{f}^*(\boldsymbol{\eta}(s)) - \mathbf{f}^{opt}||_2$ be defined as the distance from the estimated solution at iteration s of Algorithm 3 to the true optimal solution of the problem (13). Fig. 7 illustrates the evolution of two function $U(s)$ and $V(s)$ in one run of Algorithm 3. From this figure, when $U(s) < 10^{-6}$, we can also observe $V(s) < 10^{-4}$. So, the stopping criteria (30) with $\delta = 10^{-6}$ is chosen for the *while loop* in Algorithm 3. With $K = 4$, it takes about 21 seconds for Algorithm 1 and about 120 seconds for Algorithm 3 to execute 6000 iterations. However, they correspond to the total time required by all road links and all intersections. We expect significant reduction in the case when there are more computation units working in parallel.

In order to test the effectiveness of Algorithm 3, we try to implement the distributed setup outlined in Remark 2. It is assumed that one local controller is assigned to one intersection and run Algorithm 3 for this intersection and the corresponding road links. We use functions *tic* and *toc* in MATLAB to measure the computation time taken by each local controller. For every parallel process which can start at the same time, we add the maximum value of the time length taken by local controllers to the total execution time. Under this setup, we could observe that the averaged execution time of Algorithm 3 is only about 0.6 second in the case of $K = 4$. To find the number of *while* loops and the execution time of the proposed algorithms required until the stopping criterion is satisfied, we run Algorithm 1 and Algorithm 3 for the tested traffic network (described in Subsection VI. A) with 1000 different initial traffic states and exogenous traffic flows. TABLE IV shows the average and maximum values of these indexes with respect to various K. Note that, the numbers of necessary *while* loops for the centralized setup

TABLE IV

THE NUMBER OF *While* LOOPS AND TOTAL COMPUTATION TIME REQUIRED FOR RUNNING ALGORITHM 3

$K =$	2	4	6	8
ave. loops	3479	4900	6097	6907
max. loops	4592	6102	7604	8832
ave. time (centralized)	7.124	15.364	31.34	45.682
max. time (centralized)	11.594	21.412	38.707	59.493
ave. time (distributed)	0.214	0.607	1.301	1.991
max. time (distributed)	0.297	0.975	2.062	2.88

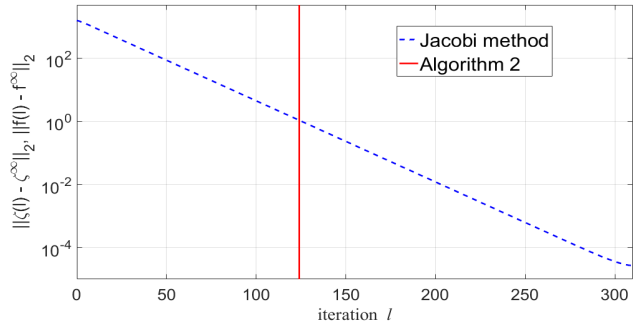


Fig. 8. The convergence of Jacobi updates (the y-axis is in log-scale) and the number of updates required by Algorithm 2.

(i.e., Algorithm 1) and the distributed setup (i.e., Algorithm 3) are the same. However, the execution time in distributed setup (i.e., Algorithm 3) is significantly smaller than the centralized setup. Since the execution time is small compared to the time interval between two consecutive control steps, our proposed control method can be considered promising when a good communication among local controllers is provided. Our experiments are conducted in a computer with chip Intel Core I5 8500 and 16 GB RAM.

Let $N^{Alg.2}$ be the maximum number of the updates (i.e., l) computed from Algorithm 2, which is required for one road link to use (40a) or (40b). To check the advantage of Algorithm 2, we compare $N^{Alg.2}$ with the necessary number of Jacobi updates (39a) (or (39b)). In Fig. 8, the dotted blue curve represents the function $log\left(||\tilde{\boldsymbol{\zeta}}_k(l) - \boldsymbol{\zeta}_k^*(\tilde{\boldsymbol{\eta}}(s))||_2\right)$, which is the evolution of the difference between $\tilde{\boldsymbol{\zeta}}_k(l)$ and the final converged value $\boldsymbol{\zeta}_k^*(\tilde{\boldsymbol{\eta}}(s))$ in log scale, and the red vertical line indicates the number $N^{Alg.2} = 128$. Note that regardless of choice of the numbers k and s, the number $N^{Alg.2}$ was computed as $N^{Alg.2} = 128$, and it was the same on both the update of $\tilde{\boldsymbol{\zeta}}_k(l)$ and the update of $\tilde{\mathbf{f}}_k(l)$ for all $k = 0, \cdots, K-1$. So, Algorithm 2 requires significantly less numbers of the updates (39a) and (39b) than the ones of traditional Jacobi method.

VII. CONCLUSION

In this paper, a distributed control method for an urban traffic network is proposed. First, the problem of improving traffic conditions while guaranteeing a smooth operation of the overall traffic network is formulated as a constrained optimization problem. Then two distributed algorithms are designed in order to allow every road link and every intersection to use only local information. Based on the accelerated-gradient method and the properties of the minimal polynomial of a matrix pair, it was shown that the proposed algorithms could be applied to reduce the execution time significantly in a distributed setup. The VISSIM-based microscopic simulation results verify that our proposed traffic control method is successful for reducing the risk of traffic congestion and enhancing the total flows through intersections and the whole traffic network.

APPENDIX

A. Proof of Lemma 1

According to the Gerschgorin circles theorem, we have all eigenvalues of $\mathbf{R}(t)$ lying within the union of the N circles defined by

$$|z - r_{ii}(t)| \leq -|r_{ii}(t)| + \sum_{j=1}^{N} |r_{ij}(t)|, i = 1, \ldots, N.$$

From the definition of the turning-ratio matrix, we have $r_{ii}(t) = 0$ for all $i \in \mathcal{R}$ and the sum $\sum_{j=1}^{N} |r_{ij}(t)|$ equals to one if $\tau(i) \in \mathcal{J}$ or to zero if $\tau(i) \in \mathcal{O}$. This means $|\kappa| \leq 1$ for every eigenvalue κ of the matrix $\mathbf{R}(t)$. Let $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_N \end{bmatrix}^T$ be the eigenvector corresponding to eigenvalue κ, we have

$$r_{i1}(t)y_1 + r_{i2}(t)y_2 + \cdots + r_{iN}(t)y_N = \kappa y_i$$

for all $i \in \mathcal{R}$. Consider the case of $|\kappa| = 1$, then

$$\begin{aligned} |\kappa y_i| = |y_i| &= |r_{i1}(t)y_1 + r_{i2}(t)y_2 + \cdots + r_{iN}(t)y_N| \\ &\leq |r_{i1}(t)||y_1| + |r_{i2}(t)||y_2| + \cdots + |r_{iN}(t)||y_N| \end{aligned}$$

Since $r_{ij}(t) \geq 0$ for all $i, j \in \mathcal{R}$, we have

$$|y_i| \leq r_{i1}(t)|y_1| + r_{i2}(t)|y_2| + \cdots + r_{iN}(t)|y_N| \quad (42)$$

Let i^* be the index where $|y_{i^*}| = \max\left\{|y_j|, j \in \mathcal{R}\right\}$ and $|y_{i^*}| > 0$. We have

$$|y_{i^*}| = r_{i^*1}(t)|y_{i^*}| + r_{i^*2}(t)|y_{i^*}| + \cdots + r_{i^*N}(t)|y_{i^*}|.$$

because $\sum_{j=1}^{N} r_{i^*j}(t) = 1$. It implies $\sum_{j=1}^{N} r_{i^*j}(t)(|y_j| - |y_{i^*}|) \leq 0$. Because $r_{i^*j}(t) \geq 0$ if $j \in \mathcal{N}_{i^*}^-$ and $r_{i^*j}(t) = 0$ if $j \notin \mathcal{N}_{i^*}^-$, the inequality (42) holds for i^* if and only if $|y_j| = |y_{i^*}|$ for all $j \in \mathcal{N}_{i^*}^-$. Then we generalize that for all j which is reachable from i^*, we have $|y_j| = |y_{i^*}|$. By Assumption 1, there exists j^* where $\tau(j^*) \in \mathcal{O}$ and $|u_{j^*}| = |u_{i^*}|$. Since $r_{j^*k}(t) = 0$ for all $k \in \mathcal{R}$, (42) implies $|y_{j^*}| \leq 0$ or $y_{j^*} = 0$. That means $\max\{|y_i| : i \in \mathcal{R}\} = 0$ or $y_i = 0$ for all $i = 1, \ldots, N$. This contradicts with the assumption that $\mathbf{y}$ is an eigenvector of $\mathbf{R}(t)$. Thus, $|\kappa| < 1$ for all eigenvalues of $\mathbf{R}(t)$.

B. Convex Optimization

Consider the nonlinear constrained problem

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}} \quad & \phi(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, i = 1, \ldots, p \end{aligned} \quad (43)$$

under convexity and interior point assumptions as follows.

Assumption 5 (Assumption 5.3.1 [40]): The set $\mathcal{X}$ is a convex subset of $\mathbb{R}^n$ and the functions $\phi : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$ are convex over $\mathcal{X}$. In addition, there exists a vector $\bar{\mathbf{x}} \in \mathcal{X}$ such that $g_i(\bar{\mathbf{x}}) < 0, i = 1, \ldots, p$.

1) Primal Problem and Dual Problem: The Lagrangian function corresponding to the problem (43) is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = \phi(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x})$$

where $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_p(\mathbf{x})]^T$. Define the dual function by $\varphi(\boldsymbol{\mu}) = \inf_{\mathbf{x}\in\mathcal{X}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$, then the associated dual problem of (43) is given as:

$$\max_{\boldsymbol{\mu}\geq\mathbf{0}} \varphi(\boldsymbol{\mu}). \tag{44}$$

Let ϕ^* and φ^* be the optimal values of the primal problem (43) and the dual problem (44), respectively. From the definition, we have $\varphi^* \leq \phi^*$ in general. Moreover, if the primal problem (43) satisfies the conditions stated in Assumption 5, the existence of Lagrange multiplier and the optimal solution of the problem (43) is guaranteed by the following propositions.

Proposition 1 (Proposition 5.3.1 [40]): Let Assumption 5 hold for the problem (43). Then $\phi^* = \varphi^*$ and there exists at least one Lagrange multiplier $\boldsymbol{\mu}^*$ where

$$\boldsymbol{\mu}^* \geq \mathbf{0} \text{ and } \phi^* = \inf_{\mathbf{x}\in\mathcal{X}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}^*)$$

Proposition 2 (Proposition 5.1.1 [40]): Let $\boldsymbol{\mu}^*$ be a Lagrange multiplier having the properties given in Proposition 1. Then $\mathbf{x}^*$ is a global minimum of the primal problem if and only if $\mathbf{x}^*$ is feasible and

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}^*) \tag{45a}$$

$$\mu_j^* g_j(\mathbf{x}^*) = 0, \, j = 1, \ldots, p \tag{45b}$$

According to Proposition 1 and Proposition 2, the optimization problem (43) can be solved by the following two steps:

- Find the optimal solution of the dual problem (44);
- The optimal solution of the primal problem is (45a).

2) Accelerated Proximal Gradient Method for Finding Optimal Dual Solution: In this subsection, we present a fast iterative algorithm for solving the dual problem (44) when the primal problem (43) satisfies Assumption 5. It is easy to verify that $\varphi(\boldsymbol{\mu})$ is a concave function. If it is continuously differentiable with Lipschitz continuous gradient L, i.e.,

$$||\nabla\varphi(\mathbf{x}) - \nabla\varphi(\mathbf{y})|| \leq L||\mathbf{x} - \mathbf{y}||, \forall \mathbf{x}, \mathbf{y} \geq \mathbf{0},$$

then the optimal solution of the dual problem (44) can be found by applying gradient-based methods. The following proposition gives a condition under which the dual function $\varphi(\boldsymbol{\mu})$ is differentiable.

Proposition 3 (Proposition 6.1.1 [40]): Let $\mathcal{X}$ be a convex set, and let ϕ and $\mathbf{g}$ be continuous over $\mathcal{X}$. Assume also that ϕ is strictly convex, g_i's are linear and for every $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$ is minimized over $\mathbf{x} \in \mathcal{X}$ at a unique point $\mathbf{x}_{\boldsymbol{\mu}}$. Then, $\varphi(\boldsymbol{\mu})$ is differentiable and

$$\nabla\varphi(\boldsymbol{\mu}) = \mathbf{g}(\mathbf{x}_{\boldsymbol{\mu}}).$$

It is known that gradient-based optimization methods are of low complexity but have a slow rate of convergence $O(\frac{1}{s})$ where s is the iteration number. This limitation is overcome by the following iteration

$$\tilde{\boldsymbol{\mu}}(s) = \boldsymbol{\mu}(s) + \frac{s-1}{s+2}(\boldsymbol{\mu}(s) - \boldsymbol{\mu}(s-1)) \tag{46a}$$

$$\boldsymbol{\mu}(s+1) = \left[\tilde{\boldsymbol{\mu}}(s) + \epsilon\nabla\varphi(\tilde{\boldsymbol{\mu}}(s))\right]^+ \tag{46b}$$

where $\epsilon \leq \frac{1}{L}$ is a step size, $[\mathbf{u}]^+ = \mathbf{v} \in \mathbb{R}^p$ is defined as $v_i = u_i$ if $u_i \geq 0$ and $v_i = 0$ otherwise. The accelerated proximal gradient method (46) is derived by the iterations in Algorithm 2 in [41] and equations (4.1)-(4.3) in [42]. We remark here that $\max_{\boldsymbol{\mu}\geq\mathbf{0}} \varphi(\boldsymbol{\mu})$ is equivalent to the problem $\min_{\boldsymbol{\mu}\geq\mathbf{0}} -\varphi(\boldsymbol{\mu})$ and the function $-\varphi(\boldsymbol{\mu})$ is convex. The following proposition is the result according to Proposition 2 in [41] and Theorem 4.4 in [42]

Proposition 4: Let $\boldsymbol{\mu}^* = \arg\max_{\boldsymbol{\mu}\geq\mathbf{0}} \varphi(\boldsymbol{\mu})$ and $\boldsymbol{\mu}(s)$ be generated by (46). Then for any $s \geq 1$

$$\varphi(\boldsymbol{\mu}^*) - \varphi(\boldsymbol{\mu}(s)) \leq \frac{2||\boldsymbol{\mu}(0) - \boldsymbol{\mu}^*||_2^2}{\epsilon(s+1)^2} \tag{47}$$

REFERENCES

[1] G. Treyz, G. E. Weisbrod, and D. Vary, *Economic Implications of Congestion*, National Cooperative Highway Research Program and American Association of State Highway and Transportation Officials and National Research Council, Transportation Research Board, Washington, DC, USA, 2001.

[2] K. Zhang and S. Batterman, "Air pollution and health risks due to vehicle traffic," *Sci. Total Environ.*, vols. 450–451, pp. 307–316, Apr. 2013.

[3] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proc. IEEE*, vol. 91, no. 12, pp. 4416–4426, Dec. 2003.

[4] L. Chen and C. Englund, "Cooperative intersection management: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 570–586, Feb. 2016.

[5] R. B. Allsop, "SIGSET: A computer program for calculating traffic capacity of signal-controlled road junctions," *Traffic Eng. Control*, vol. 12, pp. 58–60, Jan. 1971.

[6] R. B. Allsop, "SIGCAP: A computer program for assessing the traffic capacity of signal-controlled road junctions," *Traffic Eng. Control*, vol. 12, pp. 338–341, Jan. 1976.

[7] B. D. Schutter and B. D. Moor, "Optimal traffic light control for a single intersection," *Eur. J. Control*, vol. 4, no. 3, pp. 260–276, 1998.

[8] J. Haddad, B. D. Schutter, D. Mahalel, I. Ioslovich, and P.-O. Gutman, "Optimal steady-state control for isolated traffic intersections," *IEEE Trans. Autom. Control*, vol. 55, no. 11, pp. 2612–2617, Nov. 2010.

[9] K. Courage and C. E. Wallace, *Transyt-7F User's Guide*. Washington, DC, USA: Federal Highway Administration, 1991.

[10] P. B. Hunt, D. L. Robertson, and R. D. Bretherton, "The SCOOT online traffic signal optimization technique," *Traffic Eng. Control*, vol. 23, no. 4, pp. 190–192, 1982.

[11] P. R. Lowrie, "Scats: The Sydney co-ordinated adaptive traffic system-principles, methodology, algorithms," in *Proc. IEE Int. Conf. Road Traffic Signalling*, 1982, pp. 67–70.

[12] J. L. Farges, J. J. Henry, and J. Tufal, "The PRODYN real-time traffic algorithm," in *Proc. 4th IFAC Symp. Trasport. Syst.* Baden-Baden, Germany, 1983, pp. 307–312.

[13] P. Mirchandani and L. Head, "Rhodes—A real-time traffic signal control system: Architecture, algorithms, and analysis," in *Proc. TRISTAN III Triennial Symp. Transport. Anal.*, vol. 2. San Juan, Puerto Rico, Jun. 1998, pp. 17–23.

[14] N. H. Gartner, "OPAC: A demand-responsive strategy for traffic signal control," *Transp. Res. Rec.*, vol. 906, pp. 75–84, Jan. 1983.

[15] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 5, pp. 680–694, 2010.

[16] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "A rolling-horizon quadratic programming approach to the signal control problem in large scale congested urban road networks," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 2, pp. 163–174, 2009.

[17] S. Lin, B. D. Schutter, Y. Xi, and H. Hellendoorn, "Fast model predictive control for urban road networks via MILP," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 846–856, Sep. 2011.

[18] Z. Zhou, B. D. Schutter, S. Lin, and Y. Xi, "Two-level hierarchical model-based predictive control for large-scale urban traffic networks," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 2, pp. 496–508, Mar. 2017.

[19] T. Tettamanti, T. Luspay, B. Kulcsar, T. Peni, and I. Varga, "Robust control for urban road traffic network," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 385–398, Feb. 2014.

[20] S. Timotheou, C. G. Panayiotou, and M. M. Polycarpou, "Distributed traffic signal control using the cell transmission model via the alternating direction method of multipliers," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 919–933, Apr. 2015.

[21] B. L. Ye, W. Wu, L. Li, and W. Mao, "A Hierarchical model predictive control approach for signal splits optimization in large-scale urban road networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2182–2192, Aug. 2016.

[22] P. Grandinetti, C. C. de Wit, and F. Garin, "Distributed optimal traffic lights design for large-scale urban networks," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 950–963, May 2019.

[23] N. Wu, D. Li, and Y. Xi, "Distributed weighted balanced control of traffic signals for urban traffic congestion," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3710–3720, Oct. 2019.

[24] D. C. Gazis, "Optimal control of oversaturated store-and-forward transportation networks," *Transp. Sci.*, vol. 10, pp. 1–9, Feb. 1976.

[25] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. B, Methodol.*, vol. 28, no. 4, pp. 269–287, Aug. 1994.

[26] C. F. Daganzo, "The cell transmission model—Part II: Network traffic," *Transp. Res. B, Methodol.*, vol. 29, no. 2, pp. 79–93, Apr. 1995.

[27] V. H. Pham, K. Sakurama, and H.-S. Ahn, "A decentralized control strategy for urban traffic network," in *Proc. 12th Asian Control Conf. (ASCC)*, Jun. 2019, pp. 1–6.

[28] S. Sundaram and C. N. Hadjicostis, "Finite-time distributed consensus in graphs with time-invariant topologies," in *Proc. Amer. Control Conf.*, Jul. 2007, pp. 711–716.

[29] Y. Yuan, G.-B. Stan, L. Shi, and J. Goncalves, "Decentralised final value theorem for discrete-time LTI systems with application to minimal-time distributed consensus," in *Proc. 48h IEEE Conf. Decis. Control (CDC) Held Jointly With 28th Chin. Control Conf.*, Dec. 2009, pp. 2664–2669.

[30] Y. Yuan, G.-B. Stan, L. Shi, M. Barahona, and J. Goncalves, "Decentralised minimum-time consensus," *Automatica*, vol. 49, no. 5, pp. 1227–1235, 2013.

[31] C.-J. Lan and G. A. Davis, "Real-time estimation of turning movement proportions from partial counts on urban networks," *Transp. Res. C, Emerg. Technol.*, vol. 7, no. 5, pp. 305–327, Oct. 1999.

[32] M. Rodriguez-Vega, C. Canudas-de-Wit, and H. Fourati, "Location of turning ratio and flow sensors for flow reconstruction in large traffic networks," *Transp. Res. B, Methodol.*, vol. 121, pp. 21–40, Mar. 2019.

[33] A. Bhaskar, M. Qu, and E. Chung, "Bluetooth vehicle trajectory by fusing Bluetooth and loops: Motorway travel time statistics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 113–122, Feb. 2015.

[34] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with limited traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 653–662, Aug. 2015.

[35] S. Lee, S. C. Wong, C. C. C. Pang, and K. Choi, "Real-time estimation of lane-to-lane turning flows at isolated signalized junctions," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1549–1558, Jun. 2015.

[36] P. Giselsson, M. D. Doan, T. Keviczky, B. D. Schutter, and A. Rantzer, "Accelerated gradient methods and dual decomposition in distributed model predictive control," *Automatica*, vol. 49, no. 3, pp. 829–833, 2013.

[37] T. Charalambous, Y. Yuan, T. Yang, W. Pan, C. N. Hadjicostis, and M. Johansson, "Distributed finite-time average consensus in digraphs in the presence of time delays," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 4, pp. 370–381, Dec. 2015.

[38] T. Wongpiromsarn, T. Uthaicharoenpong, Y. Wang, E. Frazzoli, and D. Wang, "Distributed traffic signal control for maximum network throughput," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2012, pp. 588–595.

[39] J. Gregoire, X. Qian, E. Frazzoli, A. D. L. Fortelle, and T. Wongpiromsarn, "Capacity-aware backpressure traffic signal control," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 2, pp. 164–173, Jun. 2015.

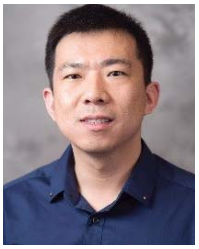
[40] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. New York, NY, USA: Academic, 1999.

[41] P. Tseng. (2008). *On Accelerated Proximal Gradient Methods for Convex-Concave Optimization*. [Online]. Available: https://www.mit.edu/~dimitrib/PTseng/papers/apgm.pdf

[42] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

Viet Hoang Pham received the B.E. degree in electrical engineering from the Hanoi University of Science and Technology (HUST), Hanoi, Vietnam, in 2013, and the M.S. degree in mechatronics from the Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, in 2017, where he is currently pursuing the Ph.D. degree in mechanical engineering. His work focuses on distributed optimization methods and urban traffic control systems designed by numerical methods.

Kazunori Sakurama (Member, IEEE) received the bachelor's degree in engineering and the master's and Ph.D. degrees in informatics from Kyoto University, Kyoto, Japan, in 1999, 2001, and 2004, respectively. He was a Research Fellow at the Japan Society for the Promotion of Science from 2003 to 2004, an Assistant Professor at The University of Electro-Communications, Tokyo, Japan, from 2004 to 2011, and an Associate Professor at the Graduate School of Engineering, Tottori University, Tottori, Japan, from 2011 to 2018. He is currently an Associate Professor at the Graduate School of Informatics, Kyoto University. His research interests include control of multi-agent systems, networked systems, and nonlinear systems.



Shaoshuai Mou (Member, IEEE) received the Ph.D. degree in electrical engineering from Yale University in 2014. He worked as a Post-Doctoral Researcher at MIT for a year, and then joined Purdue University as a Tenure-Track Assistant Professor in August 2015. He is currently an Associate Professor with the School of Aeronautics and Astronautics, Purdue University. He co-directs Purdue's research Center for Innovation in Control, Optimization and Networks (ICON), which consists of 63 faculty from 12 departments across the university with the aim of enhancing research and teaching collaboration in autonomous and connected systems. The Center for ICON was launched in September 2020 as a university unit to serve Purdue's Autonomous and Connected Systems Initiative (ACSI). He also starts to co-chair this strategic initiative ACSI in Spring 2022. His research has been focusing on advancing control theories with recent progress in optimization, networks, and learning to address fundamental challenges in autonomous systems, with particular research interests in multi-agent systems, control of autonomous systems, learning and adaptive systems, cybersecurity, and resilience.

Hyo-Sung Ahn (Senior Member, IEEE) received the B.S. and M.S. degrees in astronomy from Yonsei University, Seoul, South Korea, in 1998 and 2000, respectively, the M.S. degree in electrical engineering from the University of North Dakota, Grand Forks, ND, USA, in 2003, and the Ph.D. degree in electrical engineering from Utah State University, Logan, UT, USA, in 2006. Since July 2007, he has been with the School of Mechatronics and the School of Mechanical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea. He was a Dasan Distinguished Professor (Dasan Professor) from 2013 to 2018. Before joining GIST, he was a Senior Researcher at the Electronics and Telecommunications Research Institute, Daejeon, South Korea. He was a Visiting Scholar at the Colorado School of Mines in 2019. He is currently a Professor at the School of Mechanical Engineering, GIST. He is the author of the books *Iterative Learning Control: Robustness and Monotonic Convergence for Interval Systems* (Springer, 2007) and *Formation Control–Approaches for Distributed Agents* (Springer, 2020). His research interests include distributed control, aerospace navigation and control, network localization, and learning control.