2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Data-Driven Optimal Control of Traffic Signals for Urban Road Networks

Tong Liu[1], Hong Wang[2] and Zhong-Ping Jiang[1]

***Abstract*— This paper studies the issue of data-driven optimal control design for traffic signals of oversaturated urban road networks. The signal control system based on the store and forward model is generally uncontrollable for which the controllable decomposition is needed. Instead of identifying the unknown parameters like saturation rates and turning ratios, a finite number of measured trajectories can be used to parametrize the system and help directly construct a transformation matrix for Kalman controllable decomposition through the fundamental lemma of J. C. Willems. On top of that, an infinite-horizon linear quadratic regulator (LQR) problem is formulated considering the constraints of green times for traffic signals. The problem can be solved through a two-phase data-driven learning process, where one solves an infinite-horizon unconstrained LQR problem and the other solves a finite-horizon constrained LQR problem. The simulation result shows the theoretical analysis is effective and the proposed data-driven controller can yield desired performance for reducing traffic congestion.**

I. INTRODUCTION

Traffic congestion in urban road can cause a strong degradation in performance of network infrastructure, increase the travel delay and decrease the traffic throughput [1]. Traffic signal control aims at reducing the traffic congestion by adjusting the signal splits in each intersection to smooth the traffic flow and improve the traffic efficiency.

Decreasing the queue length is a straightforward way to mitigate traffic congestion, and a stabilizing controller is defined to make the queue length finite in average [2]. TUC (Traffic-Responsive Urban Control) utilized a store and forward model to describe the traffic flow of urban networks and introduced linear quadratic regulator to design a dynamic signal split law such that the number of vehicles can be minimized [3]. This strategy was extended to handle the constraints with multiple states and controls combined with nonlinear programming [4]. Furthermore, a multi-agent control strategy was derived based on this model and, together with a model predictive control technique, had improved the traffic conditions [5]. Similarly, decentralized controller design was considered in [6] and the obtained controller had comparable performance with the centralized TUC controller. Besides, TUC has been extensively studied and experimentally implemented in many places, and had shown around 15% improvement in average network speed compared with a pre-existing signal plan in a mid-sized Brazilian city [7].

However, most of the research on TUC assumes the system model is known, and the model parameters are identified from experimental data offline [3]. In practice, the traffic conditions at intersections can change due to the randomness of vehicle movements [8]. Therefore, the offline parameters may not be applicable for the controller design. On the other hand, data-driven control is a promising approach as it derives desirable control laws directly from real-time data collected from the traffic systems. Adaptive linear quadratic regulator was proposed to online identify the traffic dynamics and adjust the controller to decrease travel delay [9]. Function approximation based reinforcement learning was proposed to dynamically react to different traffic conditions and had been shown to outperform other pretimed and longest queue methods [10]. Besides, multi-agent deep reinforcement learning was employed to control local traffic signals using neighboring information and had shown the superior optimality, robustness, and sample efficiency over other decentralized algorithms [11].

The learning or adaptation process can be complicated and the optimality of the controlled process is generally not guaranteed. By contrast, data-driven control by adaptive dynamic programming and Willems' fundamental lemma has been developed with rigorous theoretical guarantee on optimality, stability and robustness [12][13]. The purpose of this paper is to integrate the store and forward model for network-wide traffic signals with data-driven control methods. Besides, the controllability and controllable decomposition are analyzed and data-driven methods are proposed to solve the constrained infinite-horizon linear quadratic regulator problem with theoretical convergence and optimality analysis.

The rest of the paper is organized as follows. Section II describes the store and forward model for traffic networks. In Section III, the controllability of the system is analyzed and the transformation matrix for Kalman controllable decomposition is obtained. A linear quadratic regulator problem with control constraints is formulated and solved by a two-phase data-driven algorithm in Section IV. The simulation results are included in Section V and some concluding remarks are contained in Section VI.

Notations. Throughout this paper, I_n stands for the identity matrix with size n. $\mathbb{N}$ denotes the natural number and $\mathbb{N}_+$ denotes the positive integers. $\otimes$ stands for the Kronecker product. The image space of an $m \times n$ matrix A is $\mathrm{Im}[A] =$

*This work has been supported in part by the U.S. Department of Energy under Contract DE-AC05-00OR22725 and in part by the National Science Foundation under Grant EPCN-1903781.

[1]T. Liu and Z.-P. Jiang are with the Control and Networks Lab, Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201 USA (e-mail: tl3049@nyu.edu; zjiang@nyu.edu).

[2]H. Wang is with the Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (e-mail: wangh6@ornl.gov).

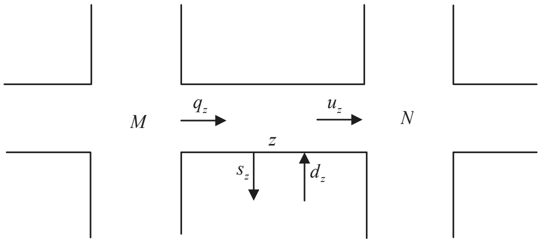

Fig. 1. An urban road link

$\{y \in \mathbb{R}^m : \exists x \in \mathbb{R}^n, y = Ax\}$; $\mathrm{rank}(A)$ is the rank of matrix A and $\mathrm{col}(A) \in \mathbb{R}^{mn}$ is the vector which stacks all columns of A. For a linear space $\mathcal{V}$, $\mathcal{V}^{\{} = \{x \in \mathbb{R}^n : x^T z = 0, \forall z \in \mathcal{V}\}$ is its orthogonal complement. Given a discrete-time signal $x : \mathbb{N} \to \mathbb{R}^q$ and $i, j \in \mathbb{N}$ with $i \leq j$, $f_{[i,j]} = \begin{bmatrix} f_i^T & f_{i+1}^T & \cdots & f_j^T \end{bmatrix}^T$ and the Hankel matrix of depth d ($d \leq j - i + 1$) associated with $f_{[i,j]}$ is

$$H_d(f_{[i,j]}) = \begin{bmatrix} f_i & f_{i+1} & \cdots & f_{j\ d+1} \\ f_{i+1} & f_{i+2} & \cdots & f_{j\ d+2} \\ \vdots & \vdots & \ddots & \vdots \\ f_{i+d\ 1} & f_{i+d} & \cdots & f_j \end{bmatrix}.$$

II. System Modelling

In this section, we consider the store and forward model in [3] for the network-wide traffic signal control of oversaturated urban roads.

The traffic network can be defined as a directed graph where the intersections are regarded as vertices and links are regarded as directed edges. Consider a typical unban road link as shown in Fig. 1, where vehicles run through link z from intersection M to intersection N. For a fixed sampling time T, the vehicle flow satisfies

$$x_z(k+1) = x_z(k) + T[q_z(k) - s_z(k) + d_z(k) - u_z(k)], \quad (1)$$

where for moment k, $x_z(k)$ denotes the number of vehicles at link z; $q_z(k)$ denotes the inflows from upstream intersection M; $u_z(k)$ denotes the outflows towards downstream intersection N; $s_z(k)$ is the exit flow and $d_z(k)$ is the demand flow. The exit flow can be defined as $s_z(k) = t_{z,0} q_z(k)$ where $t_{z,0}$ denotes the exit rate from inflow $q_z(k)$. Besides, the inflow $q_z(k)$ can be expressed as

$$q_z(k) = \sum_{w \perp I(M)} t_{w,z} u_w(k) \quad (2)$$

where $I(M)$ denotes the approaching links of intersection M and $t_{w,z}$ denotes the turning rate from link w to link z. The outflow $u_z(k)$ has a close relation with the signal timing plan as

$$u_z(k) = \frac{S_z}{C} \sum_{i \perp v_z} g_{N,i}(k) \quad (3)$$

where S_z denotes the saturation rates of link z; C is the cycle time for all intersections; $g_{N,i}(k)$ is the green time for phase i in intersection N and v_z denotes the phases when link z has right of way. Notice that (3) holds only when T is an integer multiple of C. Combining the formulas (1), (2) and (3), we have

$$\begin{aligned} x_z(k+1) = x_z(k) + T[(1 - t_{z,0}) \\ \sum_{w \perp I(M)} \frac{t_{w,z} S_w(\sum_{i \perp v_w} g_{M,i}(k))}{C} + \\ d_z(k) - \frac{S_z(\sum_{i \perp v_z} g_{N,i}(k))}{C}]. \end{aligned} \quad (4)$$

For intersection $j \in \{1, 2, ..., S\}$ where S is the number of intersections in the network, F_j denotes the number of phases for intersection j. Therefore $g_j(k) = \mathrm{col}(g_{j,1}(k), g_{j,2}(k), .., g_{j,F_j}(k)) \in \mathbb{R}^{F_j}$ denotes the time of all phases for intersection j at moment k. We only consider green time for each phase, while, for the sake of simplicity, the yellow time and the red time are assumed to be fixed constants. Considering all incoming links $z \in \{1, 2, ..., Z\}$, where Z is the number of incoming links in the network, the system model is described by

$$x(k+1) = Ax(k) + Bg(k) + Td(k) \quad (5)$$

where $x(k) = \mathrm{col}(x_1(k), x_2(k), ..., x_Z(k)) \in \mathbb{R}^Z$; $d(k) = \mathrm{col}(d_1(k), d_2(k), ..., d_Z(k)) \in \mathbb{R}^Z$; $g(k) = \mathrm{col}(g_1(k), g_2(k), ..., g_S(k)) \in \mathbb{R}^m$ with $m = \sum_{j=1}^{S} F_j$; $A = I_Z \in \mathbb{R}^{Z* Z}$ and $B \in \mathbb{R}^{Z* m}$.

Assume there exist a nominal control $\bar{g}(k) \in \mathbb{R}^m$ and historical demand $\bar{d}(k) \in \mathbb{R}^Z$ such that $B\bar{g}(k) + T\bar{d}(k) = 0$ [4]. Then, using this equation, (5) can be rewritten as

$$x(k+1) = Ax(k) + B\Delta g(k) + T\Delta d(k) \quad (6)$$

where $\Delta g(k) = g(k) - \bar{g}(k)$ and $\Delta d(k) = d(k) - \bar{d}(k)$. By selecting an appropriate nominal control $\bar{g}(k)$, $\|\Delta d(k)\|$ can be minimized [6] and removed from (5) [3]. Thus the dynamic model becomes

$$x(k+1) = Ax(k) + B\Delta g(k). \quad (7)$$

Some constraints on the phases of traffic signals are described as follows. For each $j \in \{1, 2, ..., S\}$ and $i \in \{1, 2, ..., F_j\}$,

$$g_{j,i,\min} \leq g_{j,i}(k) \leq g_{j,i,\max} \quad (8)$$

where $g_{j,i,\min}$ and $g_{j,i,\max}$ are the minimum and maximum green times for phase i in intersection j, respectively. In addition, for each $j \in \{1, 2, ..., S\}$,

$$\sum_{i=1}^{F_j} g_{j,i}(k) + L_j = C \quad (9)$$

where L_j is the sum of yellow and red times of intersection j. Notice that (9) still holds for the nominal control $\bar{g}(k)$, and, as a result, for each $j \in \{1, 2, ..., S\}$,

$$\sum_{i=1}^{F_j} \Delta g_{j,i}(k) = 0. \quad (10)$$

Therefore, control variable $\Delta g_{j,1}(k)$ can be discarded from $\Delta g(k)$ by reorganizing the matrix B as follows. Define m_j as the index of the first column for intersection j in matrix B, and by (10),

$$\sum_{i=1}^{F_j} \Delta g_{j,i} B_{[\times m_j+i \;\; 1]} = \sum_{i=2}^{F_j} \Delta g_{j,i} (B_{[\times m_j+i \;\; 1]} - B_{[\times m_j]})$$

where $B_{[\times k]}$ denotes the kth column of B. Therefore, define $\Delta g_j^r(k) = \mathrm{col}(\Delta g_{j,2}(k), ..., \Delta g_{j,F_j}(k))$ and corresponding columns $B_j^r = [B_{[.,m_j+1]} - B_{[.,m_j]}, ..., B_{[.,m_j+F_j \;\; 1]} - B_{[.,m_j]}]$. Then

$$\Delta g^r(k) = \mathrm{col}(\Delta g_1^r(k), \Delta g_2^r(k), ..., \Delta g_S^r(k)) \in \mathbb{R}^{m \;\; S}$$

and $B^r = [B_1^r, B_2^r, ..., B_S^r] \in \mathbb{R}^{Z*(m \;\; S)}$. Thus the system (7) can be rewritten as

$$x(k+1) = Ax(k) + B^r \Delta g^r(k). \tag{11}$$

III. Controllability Analysis and Decomposition

In this section, we investigate the controllability of (11) and obtain a data-driven transformation matrix for Kalman controllable decomposition through the fundamental lemma of Willems [14].

We have the following result on system controllability.

Lemma 1: The pair (A, B) in (7) is not controllable when $m < Z$, and the pair (A, B^r) in (11) is not controllable when $m < Z + S$.

Proof: Since $A = I_Z$, the controllability matrix $\mathcal{C} = [B, B, ..., B]$ and $\mathrm{rank}(\mathcal{C}) = \mathrm{rank}(B) \leq m < Z$ is not of full row rank. Thus (A, B) in (7) is not controllable [15]. The second argument can be proved similarly. ■

Remark 1: It is proved in [6] that B in (7) has full column rank assuming open traffic network and minimum complete stage strategy. For simplicity, we assume the assumptions still hold and $\mathrm{rank}(B) = m$ implies $\mathrm{rank}(B^r) = m - S$. Thus when $m = Z$, the pair (A, B) in (7) is controllable. However, in general, the number of phases is fewer than that of links, i.e. $m < Z$, thus the pairs (A, B) and (A, B^r) will be uncontrollable.

Furthermore, since A is an identity matrix, the uncontrollable mode $\lambda = 1$. Therefore, the pairs (A, B) and (A, B^r) are not stabilizable [16] and the controllable decomposition is needed for the control design.

To obtain a data-driven transformation matrix for controllable decomposition, we first introduce some necessary preliminaries.

Definition 1 ([17]): $\{u_{[0,T^i \;\; 1]}\}_{i=1}^{\tau}$ is collectively persistently exciting of order $d \in \mathbb{N}_+$ if $d \leq T^i$ for all $i = 1, 2, ..., \tau$, and the mosaic-Hankel matrix,

$$[H_d(u^1_{[0,T^1 \;\; 1]}), ..., H_d(u^{\tau}_{[0,T^{\tau} \;\; 1]})],$$

has full row rank.

Lemma 2 ([18]): Let $\{u^i_{[0,T^i \;\; 1]}, x^i_{[0,T^i \;\; 1]}\}_{i=1}^{\tau}$ be a set of input-state trajectories generated by the system $x_{k+1} = Ax_k + Bu_k$ with $x_k \in \mathbb{R}^n$ and $u_k \in \mathbb{R}^m$. If $\{u^i_{[0,T^i \;\; 1]}\}_{i=1}^{\tau}$ is collectively persistently exciting of order $\delta + L$ where δ is no more than the degree of the minimal polynomial of A, then

$$\begin{aligned} &\mathrm{Im}\begin{bmatrix} H_1(x^1_{[0,T^1 \;\; L]}) & \cdots & H_1(x^{\tau}_{[0,T^1 \;\; L]}) \\ H_L(u^1_{[0,T^1 \;\; 1]}) & \cdots & H_L(x^{\tau}_{[0,T^1 \;\; 1]}) \end{bmatrix} \\ &= (\mathcal{R} + \mathcal{K}[x_0^1, x_0^2, ..., x_0^{\tau}]) \times \mathbb{R}^{mL} \end{aligned} \tag{12}$$

where $\mathcal{R} = \mathrm{Im}[B, AB, ..., A^{n \;\; 1}B]$ and $\mathcal{K}[x_0^1, x_0^2, ..., x_0^{\tau}] = \mathrm{Im}[X_0, AX_0, ..., A^{n \;\; 1}X_0]$ with $X_0 = [x_0^1, x_0^2, ..., x_0^{\tau}]$ and x_0^i is the first state in the trajectory $x^i_{[0,T^i \;\; 1]}$ for all $1 \leq i \leq \tau$.

The key of controllable decomposition is to find a basis of the controllable subspace $\mathcal{R}$, which stays in the span of trajectories $\{x^i_{[0,T^i \;\; L]}\}_{i=1}^{\tau}$. With appropriate initial states x_0^i, the controllable subspace can be fully identified in our case as follows.

Proposition 1: Let $u^1_{[0,T^1 \;\; 1]}, x^1_{[0,T^1 \;\; 1]}$ be an input-state trajectory generated by system (11). If $u^1_{[0,T^1 \;\; 1]}$ is collectively persistently exciting of order 2 and $x_0^1 \in \mathcal{R}$, then

$$\mathcal{R} = \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})]. \tag{13}$$

Proof: Suppose $\tau = 1$ and $L = 1$, since $A = I_Z$, the degree of the minimal polynomial of A is 1. Since $u^1_{[0,T^1 \;\; 1]}$ is collectively persistently exciting of order 2, by Lemma 2, $\mathcal{R} + \mathrm{Im}[x_0^1] = \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})]$. When $x_0^1 \in \mathcal{R}$, $\mathcal{R} = \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})]$. ■

The main difficulty of applying Proposition 1 is that we do not know B^r matrix in advance, and we can only check it afterwards through $\mathrm{rank}[H_1(x^1_{[0,T^1 \;\; 1]})] = m - S$. If the condition does not hold or equivalently $x_0^1 \notin \mathcal{R}$, we have the following proposition to find $\mathcal{R}$ by collecting another trajectory.

Proposition 2: Let $\{u^i_{[0,T^i \;\; 1]}, x^i_{[0,T^i \;\; 1]}\}_{i=1}^{2}$ be two input-state trajectories generated by system (11). If $u^1_{[0,T^1 \;\; 1]}$ and $u^2_{[0,T^2 \;\; 1]}$ are both collectively persistently exciting of order 2, then if $x_0^1 \notin \mathcal{R}$ and $x_0^2 \notin \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})]$,

$$\mathcal{R} = \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})] \cap \mathrm{Im}[H_1(x^2_{[0,T^2 \;\; 1]})]. \tag{14}$$

Proof: Suppose $\tau = 1$ and $L = 1$, then by Lemma 2, $\mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})] = \mathcal{R} + \mathrm{Im}[x_0^1]$. Thus $\mathcal{R} \subseteq \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})]$. Similarly $\mathcal{R} \subseteq \mathrm{Im}[H_1(x^2_{[0,T^2 \;\; 1]})]$, so $\mathcal{R} \subseteq \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})] \cap \mathrm{Im}[H_1(x^2_{[0,T^2 \;\; 1]})]$. Conversely, when $x_0^1 \notin \mathcal{R}$, $\mathrm{rank}[H_1(x^1_{[0,T^1 \;\; 1]})] = m - S + 1$ and there exists a matrix $H_1^{\{} \in \mathbb{R}^{Z*(Z \;\; m+S \;\; 1)}$ such that $\mathrm{rank}[H_1, H_1^{\{}] = Z$ and $\mathrm{Im}[H_1^{\{}]$ is the orthogonal complement of $\mathrm{Im}[H_1]$ (H_1 refers to $H_1(x^1_{[0,T^1 \;\; 1]})$). Besides, there exists a matrix $R \in \mathbb{R}^{Z*(m \;\; S)}$ such that $\mathcal{R} = \mathrm{Im}[R]$. If $x \in \mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})] \cap \mathrm{Im}[H_1(x^2_{[0,T^2 \;\; 1]})]$, there exist $a_0 \in \mathbb{R}^{m \;\; S}$ and $a_1 \in \mathbb{R}$ such that $x = Ra_0 + x_0^1 a_1$, and there exist $b_0 \in \mathbb{R}^{m \;\; S}$ and $b_1 \in \mathbb{R}$ such that $x = Rb_0 + x_0^2 b_1$. Since $x_0^2 \notin \mathrm{Im}[H_1]$, there exist $m_0 \in \mathbb{R}^{T^1 \;\; 1}$ and $m_1 \in \mathbb{R}^{Z \;\; m+S \;\; 1}$ such that $x_0^2 = H_1 m_0 + H_1^{\{} m_1$ with $H_1^{\{} m_1 \neq 0$. Arranging these equations, we have $R(a_0 - b_0) + x_0^1 a_1 - H_1 m_0 b_1 = H_1^{\{} m_1 b_1$, where the left side term lies in $\mathrm{Im}[H_1]$ and the right side term lies in $\mathrm{Im}[H_1^{\{}]$. Since $\mathrm{Im}[H_1] \cap \mathrm{Im}[H_1^{\{}] = 0$ and $H_1^{\{} m_1 \neq 0$, $b_1 = 0$ and $x = Rb_0 \in \mathcal{R}$. Thus $\mathrm{Im}[H_1(x^1_{[0,T^1 \;\; 1]})] \cap \mathrm{Im}[H_2(x^2_{[0,T^2 \;\; 1]})] \subseteq \mathcal{R}$

and the proof is completed considering the other inclusion relation. ■

To get the basis of $\mathcal{R}$ from (14), the following equation

$$\begin{bmatrix} H_1(x^1_{[0,T^1-1]}) & -H_1(x^2_{[0,T^2-1]}) \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = 0$$

can be solved with unknown variables $X \in \mathbb{R}^{T^1-1}$ and $Y \in \mathbb{R}^{T^2-1}$. The basic solutions $X = \{X_i\}_{i=1}^{I}$ define a set of vectors $\{H_1(x^1_{[0,T^1-1]})X_i\}_{i=1}^{I}$, where we can find the maximum number of linearly independent vectors $\{r_i\}_{i=1}^{m-S}$ through the Gaussian Elimination method [19] as the basis of $\mathcal{R}$. Similarly, when $x_0^1 \in \mathcal{R}$, the basis $\{r_i\}_{i=1}^{m-S}$ can be directly constructed from $H_1(x^1_{[0,T^1-1]})$ based on Proposition 1. Therefore we have matrix $R = [r_1, r_2, ..., r_{m-S}] \in \mathbb{R}^{Z*(m-S)}$ with $\mathcal{R} = \text{Im}[R]$.

On top of that, there exists a matrix $R^{\{} \in \mathbb{R}^{Z*(Z-m+S)}$ with full column rank such that $R^T R^{\{} = 0$. As a result, $T = [R, R^{\{}]$ is the transformation matrix for Kalman controllable decomposition. Under $x(k) = Tz(k)$, for system (11), we have

$$\begin{cases} z^c(k+1) = z^c(k) + B_c \Delta g^r(k) \\ z^u(k+1) = z^u(k) \end{cases} \tag{15}$$

where $z(k) = \text{col}[z^c(k), z^u(k)]$; $z^c(k) \in \mathbb{R}^{m-S}$ and $T^{-1}B^r = \begin{bmatrix} B_c \\ 0 \end{bmatrix}$ with $B_c \in \mathbb{R}^{(m-S)*(m-S)}$, and the pair (I_{m-S}, B_c) is controllable. From (15), the controllable subsystem and uncontrollable system are decoupled and we only need to design a controller for the controllable subsystem.

IV. Data-driven Optimal Controller Design

In this section, we consider a constrained infinite-horizon linear quadratic regulator (LQR) problem for system (11). Specially, based on the controllable decomposition, we employ a data-driven algorithm to solve the problem with only controllable subsystems without relying on a prior knowledge of system parameters.

To decrease the number of vehicles and mitigate traffic congestion with moderate control efforts, the following problem is first considered,

$$\begin{aligned} \min_{\mathbf{u}} \; & \sum_{k=0}^{\in} x_k^T Q x_k + u_k^T R u_k \\ \text{s.t.} \; & x_{k+1} = x_k + B^r u_k, \forall k \geq 0; \\ & x_0 = x^0; \mathbf{u} \in \mathbb{U} \end{aligned} \tag{16}$$

where $Q = Q^T > 0$ and $R = R^T > 0$ are weighting matrices; u_k refers to $\Delta g^r(k)$; x_k refers to $x(k)$; $\mathbf{u} = \text{col}[u_0, u_1, ...]$ and $\mathbb{U} = \{\mathbf{u} | u_k \in U, \forall k \geq 0\}$ denotes the constraints of (8) with $U = \{u_k | M u_k \leq \bar{u}\}$ where $\bar{u} > 0 \in \mathbb{R}^{2m}$ and $M \in \mathbb{R}^{2m*(m-S)}$. We assume the nominal control $\bar{g}(k)$ satisfies (8) strictly for each component, thus $\mathbf{0}$ is an interior point of $\mathbb{U}$. Under $x_k = Tz_k$, problem (16) can be transformed into an equivalent form where the relevant terms of uncontrollable variables z_k^u can be removed to get a well-defined problem as follows,

$$\begin{aligned} \min_{\mathbf{u}} \; & \sum_{k=0}^{\in} (z_k^c)^T Q_c z_k^c + u_k^T R u_k \\ \text{s.t.} \; & z_{k+1}^c = z_k^c + B_c u_k, \forall k \geq 0; \\ & z_0^c = z^c(0); \mathbf{u} \in \mathbb{U} \end{aligned} \tag{17}$$

where $z_k^c = z^c(k)$; $Q_c > 0$ is the upper left block matrix of $T^T Q T$ with size $(m-S) \times (m-S)$ and z_k^c is the first $m-S$ components of $T^{-1}x_k$. If there exists a feasible solution to the problem (17), a two-phase data-driven algorithm can be employed to solve the problem.

A. Phase 1. Unconstrained Infinite-horizon LQR

In this stage, the aim is to find the optimal solution of the following problem,

$$\begin{aligned} \min_{\mathbf{u}} \; & \sum_{k=0}^{\in} (z_k^c)^T Q_c z_k^c + u_k^T R u_k \\ \text{s.t.} \; & z_{k+1}^c = z_k^c + B_c u_k, \forall k \geq 0; \\ & z_0^c = z^c(0) \end{aligned} \tag{18}$$

where the control constraints are removed from (17). The optimal solution can be found through the following value iteration process.

Lemma 3: [20] For $i \in \mathbb{N}$, consider the iteration

$$\begin{aligned} P_{i+1} &= P_i - P_i B_c K_i + Q_c \\ K_i &= (B_c^T P_i B_c + R)^{-1} B_c^T P_i \end{aligned} \tag{19}$$

with $P_0 \geq 0$, then

$$\lim_{i \infty \in} P_i = P^{\to}, \quad \lim_{i \infty \in} K_i = K^{\to}.$$

In Lemma 3, $P^{\to}$ defines the optimal cost function $J^{\to}(z_0^c) = (z_0^c)^T P^{\to} z_0^c$ and $K^{\to}$ defines the optimal control $u_k^{\to} = -K^{\to} z_k^c$ for problem (18). But the iteration process needs the explicit form of B_c matrix, which can be bypassed by adaptive dynamic programming techniques as follows. Consider that

$$\begin{aligned} (z_{k+1}^c)^T P_i z_{k+1}^c &= (z_k^c + B_c u_k)^T P_i (z_k^c + B_c u_k) \\ &= (z_k^c)^T P_i z_k^c + 2u_k^T B_c^T P_i z_k^c + u_k^T B_c^T P_i B_c u_k, \end{aligned}$$

and by using the Kronecker product,

$$\begin{aligned} &((\tilde{z}_{k+1}^c)^T - (\tilde{z}_k^c)^T)\text{vecs}(P_i) \\ &= 2(z_k^c \otimes u_k)^T \text{vec}(B_c^T P_i) + \tilde{u}_k^T \text{vecs}(B_c^T P_i B_c) \end{aligned} \tag{20}$$

where the notations of $\tilde{z}_k^c$ and $\tilde{u}_k$ can be found in [21]. Let $\{u_{[0,M]}, z_{[0,M]}^c\}$ be a long input-state trajectory which leads to a matrix equation as

$$\Phi \begin{bmatrix} \text{vec}(B_c^T P_i) \\ \text{vecs}(B_c^T P_i B_c) \end{bmatrix} = \Psi \text{vecs}(P_i) \tag{21}$$

where

$$\begin{aligned} \Psi &= \begin{bmatrix} \tilde{z}_1^c - \tilde{z}_0^c & \tilde{z}_2^c - \tilde{z}_1^c & \cdots & \tilde{z}_M^c - \tilde{z}_{M-1}^c \end{bmatrix}^T, \\ \Phi &= \begin{bmatrix} \delta_0 & \delta_1 & \cdots & \delta_{M-1} \end{bmatrix}^T \end{aligned}$$

with $\delta_j^T = \begin{bmatrix} 2(z_j^c \otimes u_j)^T & \tilde{u}_j^T \end{bmatrix}$ for $j = 0, 1, ..., M-1$. Therefore we can solve (21) iteratively to execute the value iteration process. To get a unique solution, the following assumption is imposed.

Assumption 1: Given $\{u_{[0,M]}, z^c_{[0,M]}\}$, Φ has full column rank.

Now we are ready to give our algorithm and its convergence result.

Algorithm 1 Data-driven Value Iteration for Phase 1

1: Choose a large enough $I \in \mathbb{N}$. Collect an input-state trajectory $\{u_{[0,M]}, z^c_{[0,M]}\}$ such that Assumption 1 holds. Let $i = 0$ and $P_0 = 0$.
2: Solve $B_c^T P_i$ and $B_c^T P_i B_c$ from (21), and get K_i and P_{i+1} from (19).
3: If $i \geq I$, go to step 4; Otherwise, let $i \leftarrow i + 1$ and return to step 2.
4: Output P_I and K_I.

Theorem 1: Under Assumption 1, for Algorithm 1,

$$\lim_{I \infty \in} P_i = P^{\rightarrow}; \quad \lim_{I \infty \in} K_i = K^{\rightarrow}$$

Proof: Since $P_0 = 0$, under Assumption 1, solving (21) gives the unique solution of (19) for each iteration, and the convergence is obtained by Lemma 3. ■

B. Phase 2. Constrained Finite-horizon LQR

After we get P_I and K_I from Phase 1, the following problem is considered,

$$\begin{aligned} \min_{\}u_k \ {}_0^N \ {}_1} \ & (z^c_N)^T P_I z^c_N + \sum_{k=0}^{N-1} (z^c_k)^T Q_c z^c_k + u_k^T R u_k \\ \text{s.t.} \ & z^c_{k+1} = z^c_k + B_c u_k, u_k \in U, \forall 0 \leq k < N; \\ & z^c_0 = z^c(0), \end{aligned} \tag{22}$$

where the time horizon N is determined based on whether z^c_N is in the admissible set [21]. To solve (22) needs the information of B_c matrix, based on the following Lemma, sampled trajectories can be used to implement the parametrization.

Lemma 4: [14][18] For system $z^c_{k+1} = z^c_k + B_c u_k$, if the input $u^d_{[0,T-1]}$ is collectively persistently exciting of order $1 + L$, and $x^d_{[0,T-1]}$ is the state trajectory, then any L-long input-state trajectory of system can be expressed as

$$\begin{bmatrix} u_{[0,L-1]} \\ z^c_{[0,L-1]} \end{bmatrix} = \begin{bmatrix} H_L(u^d_{[0,T-1]}) \\ H_L(x^d_{[0,T-1]}) \end{bmatrix} g$$

with $g \in \mathbb{R}^{T-L+1}$. Conversely, for any $g \in \mathbb{R}^{T-L+1}$,

$$\begin{bmatrix} H_L(u^d_{[0,T-1]}) \\ H_L(x^d_{[0,T-1]}) \end{bmatrix} g$$

is an L-long input-state trajectory.

Based on this Lemma, for $L = 2$, we have the following equivalent form of problem (22) [21].

$$\begin{aligned} \min_{\}g_k \ {}_0^N \ {}_1} \ & (z^c_N)^T P_I z^c_N + \sum_{k=0}^{N-1} (z^c_k)^T Q_c z^c_k + u_k^T R u_k \\ \text{s.t.} \ & \begin{bmatrix} u_{[k,k+1]} \\ z^c_{[k,k+1]} \end{bmatrix} = \begin{bmatrix} H_2(u^d_{[0,T-1]}) \\ H_2(x^d_{[0,T-1]}) \end{bmatrix} g_k, \\ & H_1(x^d_{[1,T-1]}) g_k = H_1(x^d_{[0,T-2]}) g_{k+1}, \\ & u_k \in U, \forall 0 \leq k < N; z^c_0 = z^c(0) \end{aligned} \tag{23}$$

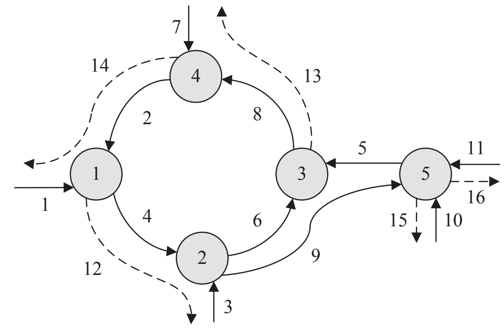

Fig. 2. Traffic network for simulation

It can be proved that this problem is a convex quadratic programming problem and can be solved efficiently. If the solution for (23) is $u_k^{\rightarrow} = H_1(u^d_{[0,T-2]}) g_k$ with $k < N$, then the near optimal control law for the whole problem is

$$\tilde{u}_k = \begin{cases} u_k^{\rightarrow}, & 0 \leq k < N \\ -K_I z^c_k, & k \geq N \end{cases}.$$

The whole algorithm is as follows, which is data-driven and converges to the optimal solution of (17) as I tends to infinity.

Algorithm 2 Data-driven Algorithm for the Whole Problem

1: Find the transformation matrix T based on Section III.
2: Run Algorithm 1 to find P_I and K_I.
3: Solve (23) to get $u_k^{\rightarrow}$ for $k < N$.
4: Output the near optimal control law $\tilde{u}_k$.

V. Simulation Results

In this section, we illustrate our proposed data-driven control method for a benchmark traffic signal network.

We use the traffic network in Section 2.3 of [6] as Fig. 2 shows, where there are 5 intersections, 11 incoming links and 9 phases in total, and the solid lines refer to the incoming links and the dashed lines are outgoing links of the network. The parameters including phase setting, turing rates, saturation rates and exit rates are same as those in [6]. The cycle time and sampling time are both 60 seconds. For all intersections, the fixed yellow and red time are 4 seconds in total. And the minimum and maximum green time can be obtained by shifting the nominal green time by 5 seconds for all phases.

In the first stage, the number of control variables is reduced to get $\Delta g^r(k)$. It is noted that intersection 3 only has 1 phase and it occupies the whole available time, which means it is open all the time. In this case, we can directly discard the 5th column in B matrix since $\Delta g_{3,1} = 0$. Through the elimination method in Section II, we only have 4 control variables $\Delta g_{1,2}, \Delta g_{2,2}, \Delta g_{4,2}$ and $\Delta g_{5,2}$ and all the other phases' time can be derived according to (10). To get a transformation matrix T, two persistently exciting input-state trajectories with order 2 are collected

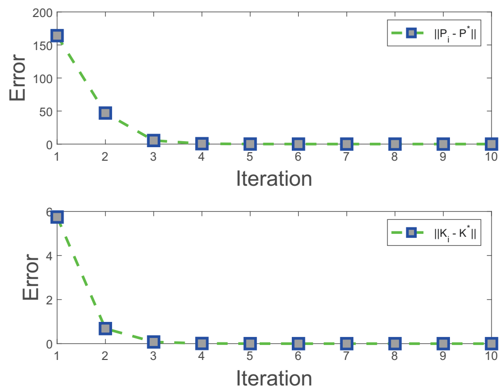

Fig. 3. Convergence of P_i and K_i

with initial states $[11,5,27,18,23,16,15,13,18,15,16]^T$ and $[10,28,12,34,28,13,20,13,26,12,15]^T$. And the persistently exciting inputs are chosen through random numbers between $(0,1)$ for 30 data points in each trajectory. The conditions of Proposition 2 are satisfied and T matrix is found by the proposed data driven method. B_c matrix is with full rank under this T matrix. Notice that the orthogonal basis can give more numerically stable solutions compared with the regular basis in the solving process.

After that, we apply the two-phase learning process to compute the optimal control law. In the first unconstrained case, Q and R are both identity matrix and another trajectory with 30 data points is collected based on the random signals as above to satisfy the Assumption 1. Algorithm 1 is implemented and $P^{\rightarrow}$and $K^{\rightarrow}$are computed for comparison. It can be seen from Fig. 3 that, P^i and K^i converge to $P^{\rightarrow}$ and $K^{\rightarrow}$within 5 steps. After solving P_I, the admissible set is calculated and the planning horizon N is selected as 2 to guarantee z_N^c can reach the admissible set [21], which implies that after 2 steps, the trajectory is generated by the controller from the unconstrained linear quadratic regulator. A new persistent exciting input-state trajectory is collected with order 3 and length $T = 15$ to parametrize the system. The optimal cost is 917.43, higher than 844.11, which refers to the optimal cost for the unconstrained case.

VI. Conclusions

In this paper, the optimal traffic signal control problem in urban road network based on the store and forward model is considered. The controllability of the traffic system is investigated and the controllable decomposition process is constructed by collecting at most two persistently exciting input-state trajectories. A constrained infinite-horizon linear quadratic regulator problem is proposed and solved through a two-phase data-driven learning process based on the controllable decomposition result. Finally, the simulation results have validated the effectiveness of our proposed method.

References

[1] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043–2067, 2003.

[2] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 177–195, 2013.

[3] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network-wide signal control," *Control Engineering Practice*, vol. 10, no. 2, pp. 183–195, 2002.

[4] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 2, pp. 163–174, 2009.

[5] L. B. De Oliveira and E. Camponogara, "Multi-agent model predictive control of signaling split in urban traffic networks," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 1, pp. 120–139, 2010.

[6] L. Pedroso and P. Batista, "Decentralized store-and-forward based strategies for the signal control problem in large-scale congested urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 132, p. 103412, 2021.

[7] W. Kraus, F. A. de Souza, R. C. Carlson, M. Papageorgiou, L. D. Dantas, E. Camponogara, E. Kosmatopoulos, and K. Aboudolas, "Cost effective real-time traffic signal control using the TUC strategy," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 6–17, 2010.

[8] H. Wang, S. V. Patil, H. A. Aziz, and S. Young, "Modeling and control using stochastic distribution control theory for intersection traffic flow," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[9] H. Wang, M. Zhu, W. Hong, C. Wang, G. Tao, and Y. Wang, "Optimizing signal timing control for large urban traffic networks using an adaptive linear quadratic regulator control strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 333–343, 2020.

[10] L. Prashanth and S. Bhatnagar, "Reinforcement learning with function approximation for traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 412–421, 2010.

[11] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1086–1095, 2019.

[12] Z.-P. Jiang, T. Bian, and W. Gao, "Learning-based control: A tutorial and some recent results," *Foundations and Trends in Systems and Control*, vol. 8, no. 3, pp. 176–284, 2020.

[13] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2019.

[14] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.

[15] C. T. Chen, *Linear System Theory and Design*. Oxford University Press, 1998.

[16] J. P. Hespanha, *Linear systems theory*. Princeton university press, 2018.

[17] H. J. van Waarde, C. De Persis, M. K. Camlibel, and P. Tesi, "Willems' fundamental lemma for state-space systems and its extension to multiple datasets," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 602–607, 2020.

[18] Y. Yu, S. Talebi, H. J. van Waarde, U. Topcu, M. Mesbahi, and B. Akmee, "On controllability and persistency of excitation in data-driven control: extensions of Willems' fundamental lemma," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 6485–6490.

[19] G. Strang, *Linear algebra and its applications.* Belmont, CA: Thomson, Brooks/Cole, 2006.

[20] D. Bertsekas, *Dynamic programming and optimal control: Volume I.* Athena scientific, 2012, vol. 1.

[21] B. Pang and Z.-P. Jiang, "A data-driven approach for constrained infinite-horizon linear quadratic regulation," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 6010–6015.