A Unified Framework for Data-Driven Optimal Control of Connected Vehicles in Mixed Traffic

Tong Liu, *Student Member, IEEE*, Leilei Cui, Bo Pang and Zhong-Ping Jiang, *Fellow, IEEE*

***Abstract*—This paper presents a unified approach to the problem of learning-based optimal control of connected human-driven and autonomous vehicles in mixed-traffic environments including both the freeway and ring road settings. The stabilizability of a string of connected vehicles including multiple autonomous vehicles (AVs) and heterogeneous human-driven vehicles (HDVs) is studied by a model reduction technique and the Popov-Belevitch-Hautus (PBH) test. For this problem setup, a linear quadratic regulator (LQR) problem is formulated and a solution based on adaptive dynamic programming (ADP) techniques is proposed without a priori knowledge on model parameters. To start the learning process, an initial stabilizing control law is obtained using the small-gain theorem for the ring road case. It is shown that the obtained stabilizing control law can achieve general $\mathcal{L}_p$ string stability under appropriate conditions. Besides, to minimize the impact of external disturbance, a linear quadratic zero-sum game is introduced and solved by an iterative learning-based algorithm. Finally, the simulation results verify the theoretical analysis and the proposed methods achieve desirable performance for control of a mixed-vehicular network.**

***Index Terms*—Connected and autonomous vehicles (CAVs), stabilizability, adaptive dynamic programming, optimal control, disturbance attenuation.**

I. Introduction

INTELLIGENT transportation is aimed at enhancing the safety, throughput, and energy efficiency through emergent communication techniques, e.g., vehicle-to-vehicle communication, by which vehicles can be virtually connected and controlled, leading to significant performance improvements [1]–[3].

One emerging important topic in connected vehicles is cooperative driving where connected vehicles communicate with each other to coordinate the motions of the vehicular network. The longitudinal motion control of connected vehicles has a huge impact on the energy efficiency of the whole system [4]. In this topic, considerable theoretical research has been carried out on the system modeling and control of CAVs with guaranteed internal stability and string stability under desirable communication topologies [5]. Naus et al. [6] designed a cooperative adaptive cruise control (CACC) system and analyzed the frequency-domain condition for string stability considering a velocity-dependent intervehicle spacing policy. Zheng et al. [7] discussed the impact of different information flows on the internal stability and stability margin of a platoon of AVs. Hu et al. [8] proposed a distributed coordinated control strategy for the longitudinal collision avoidance of CAVs. Zhang et al. [9] considered the unknown input delays for the connected vehicles and proposed adaptive switching control methods with guaranteed stability. Besides, Xiao et al. [10] presented a distributed cooperative platooning control of CAVs with an event-triggered communication mechanism to achieve efficient utilization of communication resources.

With the deployment of AVs, there will be a transition stage when both AVs and HDVs exist in the connected vehicles [2]. This mixed-traffic environment provides new challenges to control synthesis and closed-loop vehicular system analysis as HDVs cannot be directly controlled. In this direction, there have been plenty of studies on the control of CAVs in the mixed-traffic setting. When full HDVs running on a ring road, a traffic jam can occur without any bottleneck [11], which inspired the control of a CAV in the string to maintain the expected states of the vehicles [12]. The optimal control design of a single CAV in the freeway was studied in [13] based on LQR considering communication delay, driver reaction time, and string stability of the vehicular system. A robust control strategy was introduced for the CAV considering the uncertain driving reaction times of the preceding HDVs [14]. An eco-driving strategy for CAVs in mixed platoons was developed to reduce the total fuel consumption of the platoons [15]. Control of a single CAV and multiple heterogeneous HDVs in the ring road through $\mathcal{H}_2$ optimal control was proposed in [16], and the controllability and reachability conditions were established to achieve the internal stability of the mixed-traffic platoon. Similarly, multiple CAVs and homogeneous HDVs were considered in [17] to derive the conditions on system stabilizability and an $\mathcal{H}_2$ optimal control problem was formulated to decrease the impact of external disturbance.

In the mixed-traffic environment, the behaviors of HDVs can be different from the predicted behaviors using the models with empirical parameters, causing the degradation of performance of model-based control methods. Reinforcement learning (RL) and ADP are learning-based optimal control methods, where the actions or controls are learned from data through the interaction between agents and environments in an online adaptive process [18]. Input and state data represent the past experience and data-driven controllers can be gradually learned as the experience increases, e.g., a Gaussian process model was utilized to learn car-following behaviors [19], and a model-free approach was applied to CACC using a function approximation RL algorithm [20]. Besides, deep RL methods were applied to design stabilizing and platooning policies for

This work has been supported in part by the National Science Foundation under Grants EPCN-1903781 and ECCS-2210320.

The authors are with Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, 370 Jay Street, Brooklyn, NY 11201, USA (e-mail: tl3049@nyu.edu; l.cui@nyu.edu; bo.pang@nyu.edu; zjiang@nyu.edu).

mixed vehicular systems [21], and validated in the freeway and ring road environments [22], [23]. However, the convergence and optimality of the algorithms are not guaranteed with the learning mechanisms. By contrast, ADP techniques exploit the structure of the optimal control problems and provide learning-based iterative algorithms with rigorous theoretical analysis, which have been applied to the longitudinal control of one CAV in a platoon with heterogeneous HDVs on a freeway [24], with further consideration of driver reaction time and lateral control of AVs [25], [26]. In addition, Gao et al. [27] proposed a nonlinear model for the CAV in the platoon and employed global ADP to design learning-based suboptimal controllers for the mixed vehicular system with robustness to nonvanishing disturbances.

External disturbance can cause stop-and-go waves along the platoon and the designed controller should be able to attenuate the disturbance [28], [29]. $\mathcal{L}_p$ string stability and head-to-tail string stability are proposed to describe the propagation of the disturbance in the freeway [13], [30]. However, in the ring road case, it is difficult to recognize the head vehicle or the tail vehicle and we need to consider the impact of disturbance on all the vehicles. Thus, a system-level performance that describes the impact of the disturbance on the output related to all states was proposed [16]. The minimization of the performance comes down to an $\mathcal{H}_\infty$ control problem and can be transformed into a linear quadratic zero-sum game [31], [32]. A data-driven algorithm is proposed to solve the optimal controller for the zero-sum game in order to reduce the impact of the external disturbance in this paper.

This paper is aimed at proposing a unified approach for the optimal control of a general class of mixed vehicular systems including CAVs and heterogeneous HDVs in both the freeway and ring road cases. This paper has combined the results in our previous work [33], [34], and presented more systematic tools on this issue. Specifically, the stabilizability of the mixed-traffic system is established under mild conditions. On top of that, instead of using a model-based method, ADP techniques are employed to design a data-driven optimal controller without relying on the precise knowledge of human driver parameters. Besides, the initial stabilizing controllers are constructed explicitly to start the learning algorithm. The convergence analysis of the algorithm and the $\mathcal{L}_p$ string stability analysis of the vehicular network in closed-loop with the obtained learning-based controllers are also given under appropriate conditions. Furthermore, a robust controller which achieves the maximum level of attenuation for external disturbance is constructed by solving a linear-quadratic zero-sum game and implemented by a learning-based value iteration algorithm. The proposed methods are further validated through SUMO simulation, a microscopic traffic simulation [35].

The main contributions are summarized as follows:

1) We introduce a novel way to study the stabilizability of the connected vehicles, independent of the order in which HDVs and AVs are arranged in the platoon, and the environments of freeways and ring roads. The first independence relies on the fact that the stabilizability of the system does not change after applying appropriate state feedback transformation, and the second one depends on a model reduction technique for the ring road environment, from which the similarity of systems in the freeways and ring roads will be clear. Most of the previous studies focus on the control of connected vehicles on either the freeway case or the ring road case, e.g. [13]–[17], [24], [25], [27], but after using the stabilizability results in this paper, the control approach can be synthesized in a unified way without specifying the environment as shown in Sections IV and V.

2) We propose a complete data-driven solution based on ADP techniques for the linear quadratic optimal control of the mixed vehicular platoons, by optimizing the performance of operational costs and enhancing the ability of disturbance attenuation. Compared with the model-based control [13], [16], [17], our approach does not rely on any prior information about system parameters, which usually come from the offline system identification process [36]. By contrast, our method is an off-policy learning process and optimal controllers can be learned either online or offline by collecting enough input and state data along the trajectories of the vehicles [18]. And this feature is extremely useful for the control of mixed vehicular platoons, as human drivers' behavior can vary significantly in different situations, and this online learning process can capture the real-time behavior of human drivers and adaptively change the output of the controllers to improve the driving performance.

3) We provide rigorous theoretical analysis of the learning algorithms for the mixed vehicular platoons. In most applications of RL and ADP, an initial stabilizing controller is not easy to find. But for the control of connected vehicles, the initial stabilizing controllers are constructed to guarantee the asymptotic stability of the closed-loop systems based on the small-gain theorem [37], which can also be used for data collection. And the resulting optimal controllers can guarantee the general $\mathcal{L}_p$ string stability of the platoon [30]. Sufficient data, which is characterized as the persistent excitation condition, is utilized to guarantee the convergence of the learning algorithms combined with the policy iteration process. Furthermore, for the robust optimal control aiming at reducing the impact of external disturbance, the original $\mathcal{H}_\infty$ optimization problem is translated into a linear quadratic zero-sum game [32], and solved by a learning-based value iteration algorithm without any initial stabilizing controller.

Although we consider the system models with linear dynamics, the techniques in this paper can be extended to nonlinear system models by using similar model reduction techniques, nonlinear optimal control theory, small-gain techniques, nonlinear $\mathcal{H}_\infty$ control theory, and ADP techniques [18], [38], [39]. In this sense, we have provided a unified framework for the data-driven optimal control of connected vehicles in mixed platoons either in the freeway or the ring road environments.

The rest of the paper is organized as follows. Section II describes the mathematical model of HDVs and CAVs on the freeway and ring roads. In Section III, we analyze the stabilizability of the connected vehicles by the classical PBH test [40, Theorem 14.2]. In Section IV, we formulate an LQR problem and propose a learning-based algorithm with constructed initial stabilizing controllers by ADP techniques, and the general $\mathcal{L}_p$ string stability is guaranteed with the obtained controllers. Section V presents an iterative data-driven algorithm that

yields a robust optimal controller with guaranteed disturbance attenuation ability. Section VI demonstrates the effectiveness of the theoretical analysis by simulation results and the paper is concluded in Section VII.

Notations. $\mathbb{R}_+$ denotes the set of non-negative real numbers. I denotes the identity matrix of the appropriate size. $\mathbb{C}_-$ denotes the set of complex numbers with negative real parts. $\mathbb{C}_+$ denotes the set of complex numbers with non-negative real parts. A^H denotes the conjugate transpose of matrix A. $\mathbb{Z}_{\geq 0}$ denotes the set of non-negative integers. $\|\cdot\|$ denotes the Euclidean norm of a vector and the induced 2-norm of a matrix. $\otimes$ denotes the Kronecker product. $\sigma(\cdot)$ denotes the set of eigenvalues of a matrix. $\text{vec}(A) = [A_{[\cdot,1]}^T, A_{[\cdot,2]}^T, ..., A_{[\cdot,n]}^T]^T$ where $A_{[\cdot,i]} \in \mathbb{R}^m$ denotes the ith column of $A \in \mathbb{R}^{m\times n}$.

II. Mathematical Modelling

This section describes the system models of the mixed vehicular platoons with CAVs and heterogeneous HDVs in the freeway and ring road cases.

A. Car Following Models

We consider the longitudinal control of a platoon of n connected vehicles, in which m vehicles are CAVs with $1 \leq m \leq n$, and let $S = \{c_1, c_2, ..., c_m\}$ denote the order set of CAVs in the platoon. The longitudinal car-following models for HDVs and CAVs are presented as follows.

The car-following model for HDV i is,

$$\dot{v}_i = f_i(\Delta p_i, \Delta \dot{p}_i, v_i) \tag{1}$$

where $v_i \in \mathbb{R}_+$ denotes the vehicle speed and $\Delta p_i \in \mathbb{R}$ denotes the space headway between vehicle i and $i-1$, as shown in Fig. 1 and Fig. 2 for the freeway and ring road cases, respectively. Let $(v^*, \Delta p_i^*)$ be the equilibrium of (1), which satisfies

$$0 = f_i(\Delta p_i^*, 0, v^*). \tag{2}$$

The objective is to keep vehicle i in the platoon achieving the speed v^* and headway Δp_i^*. Denote the speed error and headway error by $\tilde{v}_i = v_i - v^*$ and $\tilde{p}_i = \Delta p_i - \Delta p_i^*$, respectively, and $x_i = [\tilde{p}_i, \tilde{v}_i]^T$ denotes the state of vehicle i. After linearizing (1) around $(\Delta p_i^*, v^*)$, for HDV i,

$$\dot{x}_i = A_i^h x_i + E_i^h x_{i-1} \tag{3}$$

with $A_i^h = \begin{bmatrix} 0 & -1 \\ a_i & -b_i \end{bmatrix}$ and $E_i^h = \begin{bmatrix} 0 & 1 \\ 0 & c_i \end{bmatrix}$ where a_i, b_i, c_i are three positive constants that describes the regular diving behavior of HDVs, and can vary from driver to driver [16]. Similarily, for CAV i [24],

$$\dot{x}_i = A_i^c x_i + E_i^c x_{i-1} + B_i^c u_i \tag{4}$$

with $A_i^c = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}, E_i^c = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, and $B_i^c = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Notice that u_i can have access to the information of all connected vehicles, not limited to the neighboring vehicles [41]. Here we use superscript h and c to denote the matrices about HDVs and CAVs, respectively. The problem is to design control input $u_i \in \mathbb{R}$ for each CAV i such that the whole platoon can achieve different control objectives. Besides, we employ a linear system to describe the dynamics of car-following models since the behavior of a nonlinear system can be approximated by the linearized system around the equilibrium, and the analysis and control of the linearized system is effective in a small neighborhood of the equilibrium for the original nonlinear system [42].

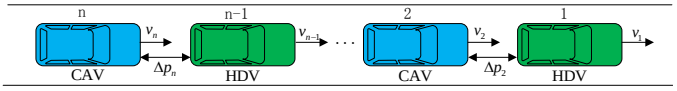


Fig. 1. A platoon of n vehicles on the freeway

B. Freeway Case

In the freeway case as shown in Fig. 1, we assume a virtual vehicle 0 in front of vehicle 1 and let $x_0 = [\tilde{p}_0, \tilde{v}_0]^T$ denote the state of virtual vehicle 0. Then the state-space representation for the platoon is

$$\dot{x} = A_f x + B_f u + E\tilde{v}_0 \tag{5}$$

where $x = [x_1^T, x_2^T, ..., x_n^T]^T$ denotes the state of the platoon and $u = [u_1, ..., u_m]^T$ denotes the control input of the platoon. $B_f = [e_{2c_1}, e_{2c_2}, ..., e_{2c_m}]$ with $e_j \in \mathbb{R}^{2n}$ being the standard coordinate vector where its jth component is 1 and the others are 0. Besides, when the head vehicle is an HDV, $E = [1, c_1, 0, ..., 0]^T$, otherwise, $E = e_1$. In addition, $A_f \in \mathbb{R}^{2n\times 2n}$ can be arranged into n^2 submatrices $A_{i,j} \in \mathbb{R}^{2\times 2}$ with $i \in \{1, ..., n\}$ and $j \in \{1, ..., n\}$. When $i \in S$, $A_{i,i} = A_i^c$ and $A_{i,i-1} = E_i^c$, otherwise, $A_{i,i} = A_i^h$ and $A_{i,i-1} = E_i^h$, and the other elements in A_f are zero. For example, consider the platoon with the sequence of vehicles "CAV-HDV-HDV" where the tail vehicle is a CAV, then $S = \{3\}$ and

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} A_1^h & 0 & 0 \\ E_2^h & A_2^h & 0 \\ 0 & E_2^c & A_i^c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + e_6 u + E\tilde{v}_0$$

with $x = [x_1^T, x_2^T, x_3^T]^T \in \mathbb{R}^6$ and $u \in \mathbb{R}$.

C. Ring Road Case

The ring road case provides a closed environment for connected vehicles and removes other effects like boundary conditions and intersections on the traffic flows [12]. In this case, vehicle n is in front of vehicle 1 as shown in Fig. 2, and the state-space representation for the platoon is

$$\dot{x} = A_r x + B_r u \tag{6}$$

where $A_r \in \mathbb{R}^{2n\times 2n}$ and $B_r \in \mathbb{R}^{2n\times m}$. A_r and B_r have the same expressions as A_f and B_f, respectively, except that if $1 \in S$, $A_{1,n} = E_1^c$, otherwise $A_{1,n} = E_1^h$. Considering the same example in the freeway case, we have

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} A_1^h & 0 & E_1^h \\ E_2^h & A_2^h & 0 \\ 0 & E_2^c & A_i^c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + e_6 u$$

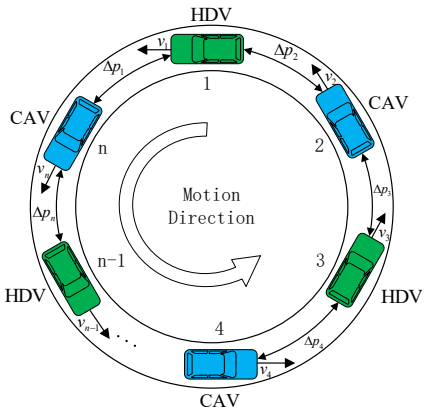


Fig. 2. A platoon of n vehicles on the ring road

with $x = [x_1^T, x_2^T, x_3^T]^T \in \mathbb{R}^6$ and $u \in \mathbb{R}$, where E_1^h appears in the top right corner of A_r, and this extra term implies that the state of the head vehicle is influenced by the tail vehicle.

The ring road case is more challenging than the freeway case in two aspects: firstly, the vehicle subsystems are more coupled, since in the ring road case, the behaviors of the head vehicle and the tail vehicle can affect each other, while the behavior of the tail vehicle cannot influence the head vehicle in the freeway case. Secondly, the circumference of the ring road is fixed, thus partial states in the ring road case are constrained in a hyperplane and system (6) cannot be controllable.

To bypass the uncontrollable situation, we exploit the fact that the length of the ring road, denoted as L, is fixed, to consider a reduced-dimension state space. And the following assumption is imposed, i.e.,

Assumption 1. $\sum_{i=1}^{n} \Delta p_i^* = L$.

Recall that Δp_i^* is the expected relative distance between vehicle i and $i-1$, and this assumption implies that the sum of expected relative distance for all vehicles should be the circumference of the ring road (vehicle length is ignored without loss of generality), otherwise the expected states can not be reached physically.

Under Assumption 1,

$$\sum_{i=1}^{n} \tilde{p}_i = \sum_{i=1}^{n} (\Delta p_i - \Delta p_i^*) = 0$$

and $\tilde{p}_n = -\sum_{i=1}^{n-1} \tilde{p}_i$. Thus, the state $\tilde{p}_n$ can be discarded from (6) using $\dot{\tilde{v}}_n = -(\sum_{i=1}^{n-1} a_n \tilde{p}_i) - b_n \tilde{v}_n + c_n \tilde{v}_{n-1}$ if vehicle n is an HDV. To represent the reduced-order model in this case, let us first assume that the first $n-1$ vehicles are running on a freeway, and correspondingly, (5) is denoted as $\dot{\hat{x}} = \bar{A}_f \hat{x} + \bar{B}_f \hat{u} + \bar{E} \tilde{v}_0$. Then, defining state $x = [x_1^T, x_2^T, ..., x_{n-1}^T, \tilde{v}_n]^T \in \mathbb{R}^{2n-1}$ and if vehicle n is a CAV, we have

$$\dot{x} = \begin{bmatrix} \bar{A}_f & \bar{E} \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} \bar{B}_f & 0 \\ 0 & 1 \end{bmatrix} u,$$

otherwise,

$$\dot{x} = \begin{bmatrix} \bar{A}_f & \bar{E} \\ F^T & -b_n \end{bmatrix} x + \begin{bmatrix} \bar{B}_f \\ 0 \end{bmatrix} u$$

with $F = [-a_n, 0, -a_n, 0, ..., -a_n, c_n]^T \in \mathbb{R}^{2n-2}$. This reduced-order system is denoted as

$$\dot{x} = A_{rr} x + B_{rr} u \tag{7}$$

with $A_{rr} \in \mathbb{R}^{(2n-1)\times(2n-1)}$ and $B_{rr} \in \mathbb{R}^{(2n-1)\times m}$. Considering the same example as before, the corresponding reduced-order system is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{\tilde{v}}_3 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 & 0 & 1 \\ a_1 & -b_1 & 0 & 0 & c_1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & c_2 & a_2 & -b_2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \tilde{v}_3 \end{bmatrix} + e_5 u$$

with $x = [x_1^T, x_2^T, \tilde{v}_3]^T \in \mathbb{R}^5$ and $u \in \mathbb{R}$.

Remark 1. It turns out that using (7) instead of (6) greatly simplifies the stabilizability analysis as shown in section III.

III. Stabilizability Analysis

In this section, the stabilizability of the mixed vehicular systems (5) and (7) is studied by the PBH test. To make the paper self-contained, we introduce the following necessary preliminaries.

Definition 1. The pair (A, B) is stabilizable if there exists a matrix K such that $A - BK$ has all eigenvalues in $\mathbb{C}_-$.

Lemma 1 (PBH test [39]). The pair (A, B) is stabilizable if and only if the matrix $[A - \lambda I, B]$ is of full row rank for any $\lambda \in \mathbb{C}_+$.

The following result gives a useful characterization of stabilizability under state feedback.

Lemma 2. The pair (A, B) is stabilizable if and only if for any given K, the pair $(A - BK, B)$ is stabilizable.

Proof. Let $V(A) = \{\eta | \eta^H (A - \lambda I) = 0, \eta^H B = 0, \forall \lambda \in \mathbb{C}_+\}$. Then, by Lemma 1, the pair (A, B) is stabilizable if and only if $V(A) = \{0\}$. Similarly, the pair $(A - BK, B)$ is stabilizable if and only if $V(A - BK) = \{0\}$. Since $V(A) = V(A - BK)$ for any given K, the proof is completed. □

By Lemma 2, the stabilizability does not change after state feedback, which is employed to study the stabilizability of (A, B) as follows.

Proposition 1. In the freeway case, the pair (A_f, B_f) is stabilizable.

Proof. Let $u_i = a_i \tilde{p}_i - b_i \tilde{v}_i + c_i \tilde{v}_{i-1}$ with $a_i, b_i > 0$ for $i \in S$ and $c_1 = 0$ if $1 \in S$, which defines a matrix K_f such that $u = -K_f x$. Consequently, $\hat{A}_f = A_f - B_f K_f$ and

$$\hat{A}_f = \begin{bmatrix} A_1^h & 0 & \cdots & 0 \\ E_2^h & A_2^h & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & E_n^h & A_n^h \end{bmatrix},$$

where A_i^h is Hurwitz since $a_i, b_i > 0$ for $i \in \{1, ..., n\}$. Hence $\hat{A}_f$ is Hurwitz and the pair $(\hat{A}_f, B_f)$ is stabilizable by Lemma 1. Then, by Lemma 2, the pair (A_f, B_f) is stabilizable. □

The same conclusion can be established for the ring road case using similar proof techniques as follows.

Proposition 2. In the ring road case, the pair (A_{rr}, B_{rr}) is stabilizable.

Proof. Let $u_i = a_i\tilde{p}_i - b_i\tilde{v}_i + c_i\tilde{v}_{i-1}$ with $a_i, b_i > 0$ for $i \in S$ and $\tilde{v}_0 = \tilde{v}_n$ if $1 \in S$, which defines a matrix K_r such that $u = -K_r x$. Consequently, $\hat{A}_{rr} = A_{rr} - B_{rr}K_r$ and

$$\hat{A}_{rr} = \begin{bmatrix} \hat{A}_f^c & E^h \\ F^T & -b_n \end{bmatrix}$$

where $E^h = [1, c_1, 0, ..., 0]^T \in \mathbb{R}^{2n-2}$, F has been defined in section II.C, and

$$\hat{A}_f^c = \begin{bmatrix} A_1^h & 0 & \cdots & 0 \\ E_2^h & A_2^h & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & E_{n-1}^h & A_{n-1}^h \end{bmatrix}$$

is Hurwitz since A_i^h is Hurwitz for $i \in \{1, 2, ..., n-1\}$. Without loss of generality, we assume the last vehicle is a CAV and the last column in B_{rr} is e_{2n-1}. Suppose there exists a vector $\rho = [\rho_1, \rho_2]^T \in \mathbb{C}^{2n-1}$ with $\rho_2 \in \mathbb{C}$, such that $\rho^H B_{rr} = 0$, then $\rho_2 = 0$. Thus, $\rho^H[\hat{A}_{rr} - \lambda I, B_{rr}] = 0$ is equivalent to $\rho_1^H[\hat{A}_f^c - \lambda I, E^h] = 0$. Since $\hat{A}_f^c$ is Hurwitz, $\hat{A}_f^c - \lambda I$ is of full row rank for any $\lambda \in \mathbb{C}_+$, which implies that $\rho_1 = 0$ and $\rho = 0$. By Lemma 1, the pair $(\hat{A}_{rr}, B_{rr})$ is stabilizable. Then, by Lemma 2, the pair (A_{rr}, B_{rr}) is stabilizable. □

Remark 2. The stabilizability result in [17] assuming homogeneous HDVs is a special case of Proposition 2. If we use system (6) for the stabilizability analysis, there exists a non-zero vector $\rho_0 = [1, 0, 1, 0, ..., 1, 0]^T \in \mathbb{R}^{2n}$ such that $\rho_0^T A_r = 0$ and $\rho_0^T B_r = 0$, which implies $[A_r - \lambda I, B_r]$ is not of full row rank when $\lambda = 0$. Thus the pair (A_r, B_r) is not stabilizable by Lemma 1. In fact, what hinders the stabilizability is exactly the constraint imposed by the fixed ring road circumference, and this constraint has been explicitly exploited by establishing Assumption 1 and the system model (7). Furthermore, for the original system (6), Assumption 1 is equivalent to the condition $x(0) \in \Pi$ with

$$\Pi = \{x \in \mathbb{R}^{2n} | \sum_{i=1}^{n} \tilde{p}_i(0) = 0\}.$$

Besides, by the facts of $\sum_{i=1}^{n} \tilde{p}_i = \sum_{i=1}^{n} \tilde{p}_i(0)$ and Proposition 2, there exists a control u in (6) such that $x(t)$ can converge to 0 as $t \to \infty$ if and only if $x(0) \in \Pi$, which relaxes the condition required in Corollary 1 of [34] and provides another characterization of Assumption 1.

The stabilizability results for both freeway and ring-road environments, provide a sound foundation for further control synthesis including stabilization and disturbance attenuation in Section IV and Section V, respectively.

IV. Learning-Based Optimal Control

This section presents an LQR problem for mixed vehicular systems with disturbance, and a learning-based method by ADP techniques is proposed to obtain the optimal control in the absence of prior knowledge of system dynamics. Besides, we construct the initial stabilizing controllers to start the data-driven control algorithm. Finally, we investigate the general $\mathcal{L}_p$ string stability of the system in the closed-loop with the obtained optimal controllers.

A. Data-Driven Algorithm

To unify the two cases, let us consider a general uncertain system

$$\dot{x} = Ax + Bu + Hw \tag{8}$$

where x refers to the distance and velocity errors of all vehicles relative to the expected states; u is the control input of the CAVs; w is an exogenous disturbance and constant matrix H denotes the gain of the disturbance. Besides, the pair (A, B) is stabilizable as shown in Section III. Let us consider the following LQR problem:

$$\begin{aligned} &\min_u \int_0^\infty (x^T Q x + u^T R u) dt \\ &\text{subject to } \ \dot{x} = Ax + Bu \end{aligned} \tag{9}$$

where Q and R are two real symmetric and positive definite matrices. The cost function penalizes the state and control, and Q and R can be manually chosen to guarantee satisfactory closed-loop system performance. Besides, by selecting appropriate Q and R matrices, some constraints, e.g., the saturation of control, can be incorporated into consideration [24].

It is well-known that solving problem (9) reduces to solving the following algebraic Riccati equation [38],

$$A^T P + PA + Q - PBR^{-1}B^T P = 0. \tag{10}$$

The solution P^* defines the optimal control $u = -K^* x$ with $K^* = R^{-1}B^T P^*$ for the problem (9), and can be obtained by a policy iteration process.

Lemma 3 ([43]). Select a stabilizing gain K^0, i.e., $\sigma(A - BK^0) \subseteq \mathbb{C}_-$. Solve P^j from the Lyapunov equation

$$(A - BK^j)^T P^j + P^j(A - BK^j) + Q + (K^j)^T R K^j = 0 \tag{11}$$

where

$$K^{j+1} = R^{-1}B^T P^j \tag{12}$$

for $j \in \mathbb{Z}_{\geq 0}$. Then, $\sigma(A - BK^j) \subseteq \mathbb{C}_-$ for $j \in \mathbb{Z}_{\geq 0}$, and $\lim_{j\to\infty} P^j = P^*$, $\lim_{j\to\infty} K^j = K^*$ where P^* is the solution of (10) and $K^* = R^{-1}B^T P^*$.

Equation (11) evaluates current policy $u = -K^j x$ and (12) improves the policy. ADP techniques build a bridge between the iterative equations in Lemma 3 and the collected input and state data from system (8). Then the optimal controller can be learned with the collected data iteratively without resorting to the specific parameters of A and B in the computation of (11) and (12). The process is described as follows.

For any $P^j \geq 0$, along the solution curve $x(t)$ of (8) for an interval $[T_0, T_s]$,

$$x^T P^j x|_{T_0}^{T_s} = \int_{T_0}^{T_s} x^T(-Q-(K^j)^T RK^j)x + 2w^T H^T P^j x + 2(u+K^j x)^T RK^{j+1} x dt \tag{13}$$

where (8), (11) and (12) were used. Since $\text{vec}(ABD) = (D^T \otimes A)\text{vec}(B)$ for any matrices A, B, D with compatible sizes [38], $x^T P^j x\big|_{T_0}^{T_1} = (x^T \otimes x^T\big|_{T_0}^{T_1})\text{vec}(P^j)$. Similarly, the right-side terms of (13) can be represented as the multiplication of matrices about state and input, and the other matrices.

For vectors $r \in \mathbb{R}^{n_r}$, and a symmetric matrix $M \in \mathbb{R}^{N\times N}$ with positive integers n_r and N, respectively, and time-varying signals $g(t)$ and $q(t)$, let us define

$$\bar{M} = [M_{1,1}, 2M_{1,2}, ..., 2M_{1,N}, M_{2,2}, ..., 2M_{N-1,N}, M_{N,N}]^T,$$

$$\bar{r} = [r_1^2, r_1 r_2, ..., r_1 r_{n_r}, r_2^2, r_2 r_3, ...r_{n_r-1} r_{n_r}, r_{n_r}^2]^T,$$

$$I_{g,q} = \left[\int_{T_0}^{T_1} g\otimes q dt, \int_{T_1}^{T_2} g\otimes q dt, ..., \int_{T_{s-1}}^{T_s} g\otimes q dt\right]^T,$$

and

$$D_{g,g} = [\bar{g}(T_1)-\bar{g}(T_0), \bar{g}(T_2)-\bar{g}(T_1), ..., \bar{g}(T_s)-\bar{g}(T_{s-1})]^T,$$

where $T_0 < T_1 < T_2 < ... < T_s$ are several selected times. Then, considering the corresponding data of the trajectory at the times,

$$\Omega_j \begin{bmatrix} \bar{P}^j \\ \text{vec}(K^{j+1}) \\ \text{vec}(H^T P^j) \end{bmatrix} = \Psi_j \tag{14}$$

with $\Omega_j =$

$$\left[D_{x,x}, -2I_{x,u}(I\otimes R) - 2I_{x,x}(I\otimes (K^j)^T R), -2I_{x,w}\right]$$

and $\Psi_j = -I_{x,x}\text{vec}(Q+(K^j)^T RK^j)$. It can be seen that Ω_j depends on the input and state data, and known matrices R and K^j in the iteration. Therefore, solving (14) provides a way of solving (11) and (12) iteratively with only state and input data. To guarantee the solvability, the following assumption is imposed.

Assumption 2. Ω_j in (14) is of full column rank for $j \in \mathbb{Z}_{\geq 0}$.

Therefore, the main theorem for this section is summarized as follows.

Theorem 1. Assume there exists a matrix K^0 such that $\sigma(A-BK^0) \subseteq \mathbb{C}_-$, and Assumption 2 holds. Then, starting from K^0, P^j and K^{j+1} can be uniquely solved from (14) for $j \in \mathbb{Z}_{\geq 0}$. Besides, $\lim_{j\to\infty} P^j = P^*$, $\lim_{j\to\infty} K^j = K^*$, and $u = -K^* x$ is the optimal control of problem (9).

Proof. For $j \in \mathbb{Z}_{\geq 0}$, let $V_1(j)$ denote the set of pairs (P^j, K^{j+1}) satisfying (11) and (12), and $V_2(j)$ be the set of pairs (P^j, K^{j+1}) satisfying (14). When $j = 0$, $V_1(0)$ contains only one pair, because there exists a unique solution P^0 of (11) since $\sigma(A-BK^0) \subseteq \mathbb{C}_-$, and K^1 is uniquely defined by (12). $V_2(0)$ contains this pair by (13) and only this pair by Assumption 2, thus $V_1(0) = V_2(0)$. Assuming $V_1(j) = V_2(j)$ when $j = k$, let us consider the case when $j = k+1$. Since $\sigma(A-BK^{k+1}) \subseteq \mathbb{C}_-$ by Lemma 3, P^{k+1} can be uniquely solved by (11) and K^{k+2} is uniquely decided by (12), thus $V_1(j)$ contains only one pair. Similarly, this pair is the only pair of $V_2(j)$ by (13) and Assumption 2, which implies that $V_1(j) = V_2(j)$ when $j = k+1$. Therefore, by mathematical induction, $V_1(j) = V_2(j)$ for $j \in \mathbb{Z}_{\geq 0}$. By Lemma 3, $\lim_{j\to\infty} P^j = P^*$ and $\lim_{j\to\infty} K^j = K^*$ where P^* is the solution of (10) and $K^* = R^{-1}B^T P^*$, and $u = -K^* x$ is the optimal control of problem (9) [38]. □

Based on Theorem 1, Algorithm 1 is constructed. There are two unsolved issues in Algorithm 1: one is the initial stabilizing gain K^0 and the other is the check of Assumption 2, since Ω_j relies on the iteration j. The first issue is thoroughly solved in Section IV-B and this section focuses on the second issue. We provide sufficient conditions to guarantee Assumption 2 that are independent of j, using only $I_{x,u}, I_{x,x}$ and $I_{x,w}$. We consider the freeway and ring road cases separately to provide accurate descriptions.

Algorithm 1 Learning-Based Optimal Control of AVs

1: Choose a stabilizing gain K^0, i.e., $\sigma(A-BK^0) \subseteq \mathbb{C}_-$, a small threshold value ε_0, and the selected times of exploration: $T_0, T_1, ..., T_s$.
2: For system (8), let $u = -K_0 x + \xi$ with exploration noise $\xi(t)$ and gather trajectories of $x(t), u(t)$, and $w(t)$ for $t \in [T_0, T_s]$ such that Assumption 2 holds.
3: Obtain P^0, K^1, P^1, and K_2 by solving (14). Let $j = 1$.
4: **while** $\left\|P^j - P^{j-1}\right\| > \varepsilon_0$ **do**
5: Let $j = j+1$.
6: Obtain P^j and K^{j+1} by solving (14).
7: **end while**
8: Save P^j and K^{j+1}, and $u = -K^{j+1}x$ is the control output.

1) Freeway Case: In this case, system (8) is replaced by system (5) with $A = A_f$, $B = B_f$, $H = E$, and $w = \tilde{v}_0$, where the speed error of virtual vehicle 0 is regarded as the disturbance [24]. Besides, Assumption 2 is assured as follows.

Lemma 4. In the freeway case, when

$$\text{rank}([I_{x,x}, I_{x,u}, I_{x,\tilde{v}_0}]) = n(2n+2m+3), \tag{15}$$

Assumption 2 holds.

Proof. See [44]. □

Therefore in Algorithm 1, we can directly check condition (15) with collected data to know if Assumption 2 holds. Hence, the condition in step 2 of Algorithm 1 can be replaced by "such that condition (15) holds". Similarly, a revision of Theorem 1 is as follows, which provides an extension of the result in [24] in the sense of uncertain number of CAVs in the platoon.

Corollary 1. In the freeway case, assume there exists a matrix K^0 such that $\sigma(A_f - B_f K^0) \subseteq \mathbb{C}_-$, and condition (15) holds. Then, starting from K^0, P^j and K^{j+1} can be uniquely solved from (14) for $j \in \mathbb{Z}_{\geq 0}$. Besides, $\lim_{j\to\infty} P^j = P^*$, $\lim_{j\to\infty} K^j = K^*$, and $u = -K^* x$ is the optimal control of problem (9).

2) Ring Road Case: In this case, we consider extra disturbance $w = [w_1, w_2, ..., w_p]^T \in \mathbb{R}^p$ for system (7) and $p \leq n$ [16] (see an example of noise signals in [34]). Thus $A = A_{rr}$, $B = B_{rr}$ and the disturbed system is

$$\dot{x} = A_{rr}x + B_{rr}u + Hw. \tag{16}$$

with some matrix $H \in \mathbb{R}^{(2n-1)\times p}$. Similarly, to guarantee Assumption 2, the following condition can be obtained.

Lemma 5. In the ring road case, when

$$\text{rank}([I_{x,x}, I_{x,u}, I_{x,w}]) = (m+n+p)(2n-1), \tag{17}$$

Assumption 2 holds.

Proof. See [44]. □

Therefore, the condition in step 2 of Algorithm 1 can be replaced by "such that condition (17) holds". A revision of Theorem 1 is as follows.

Corollary 2. In the ring road case, assume there exists a matrix K^0 such that $\sigma(A_{rr} - B_{rr}K^0) \subseteq \mathbb{C}_-$, and condition (17) holds. Then, starting from K^0, P^j and K^{j+1} can be uniquely solved from (14) for $j \in \mathbb{Z}_{\geq 0}$. Besides, $\lim_{j\to\infty} P^j = P^*$, $\lim_{j\to\infty} K^j = K^*$, and $u = -K^*x$ is the optimal control of problem (9).

B. Initial Stabilizing Controllers

To complete Algorithm 1, initial stabilizing controllers are constructed for both cases as follows.

For the freeway case, from the proof of Proposition 1, it is straightforward to construct a stabilizing controller. For each CAV $i \in S$ in the platoon, $u_i = a_i\tilde{p}_i - b_i\tilde{v}_i + c_i\tilde{v}_{i-1}$ with $a_i, b_i > 0$ and $c_1 = 0$ if $i = 1$. This control is denoted as the regular HDV control. Since $\hat{A}_f$ is Hurwitz, the controller is stabilizing.

For the ring road case, the stabilizing controller is not obvious. In fact, if we still apply the regular HDV control laws as the freeway case, the controller can be destabilizing. Let us consider a platoon of two HDVs in the ring road case with $a_1 = b_1 = a_2 = b_2 = 1, c_1 = c_2 = 2$ and in this case

$$\hat{A}_{rr} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & -1 & 2 \\ -1 & 2 & -1 \end{bmatrix},$$

and $\det(\lambda I - \hat{A}_{rr}) = (\lambda-1)(\lambda+1)(\lambda+2)$. Thus an unstable mode $\lambda = 1$ will appear when we apply the regular HDV control laws, which may explain the reason why stop-to-go waves can occur when all HDVs were running on a ring road without bottleneck [11], which shows the necessity of introducing CAVs into the platoon in the ring road environment.

However, the stabilizing controller can be constructed with a minor revision to the regular HDV control laws. The problem is to ask that the eigenvalues of

$$\hat{A}_{rr} = \begin{bmatrix} \hat{A}_f^c & E^h \\ F^T & -b_n \end{bmatrix}$$

all have negative real parts and this fact can be achieved by the well-known small-gain theorem [37].

Proposition 3. There exists a constant $c > 0$ such that when $b_n > 0$ and

$$\frac{\sqrt{(n-1)a_n^2 + c_n^2}}{b_n} < c, \tag{18}$$

$\hat{A}_{rr}$ is Hurwitz.

Proof. Consider the system $\dot{x} = \hat{A}_{rr}x$ with $x = [x_1^T, x_2]^T$ where $x_1 \in \mathbb{R}^{2n-2}$ and $x_2 \in \mathbb{R}$, and the system can be decomposed into two interconnected systems:

$$\begin{cases} \dot{x}_1 = \hat{A}_f^c x_1 + E^h x_2 \\ \dot{x}_2 = -b_n x_2 + F^T x_1. \end{cases} \tag{19}$$

Since $\hat{A}_f^c$ is Hurwitz, $\left\| e^{\hat{A}_f^c t} \right\| \leq k_1 e^{-\lambda_1 t}$ with constants $k_1, \lambda_1 > 0$ [40], and

$$x_1(t) = e^{\hat{A}_f^c t}x_1(0) + \int_0^t e^{\hat{A}_f^c (t-\tau)}E^h x_2(\tau)d\tau,$$

which implies that

$$\begin{aligned} \|x_1(t)\| \leq &\max\{2k_1 e^{-\lambda_1 t}\|x_1(0)\|, \\ &\frac{2k_1\sqrt{1+c_1^2}}{\lambda_1}\max_{0\leq\tau\leq t}\|x_2(\tau)\|\} \end{aligned}$$

and x_1 subsystem is input-to-state stable regarding x_2 as the input [45]. Similarly, since $b_n > 0$,

$$\begin{aligned} \|x_2(t)\| \leq &\max\{2e^{-b_n t}\|x_2(0)\|, \\ &\frac{2\sqrt{(n-1)a_n^2 + c_n^2}}{b_n}\max_{0\leq\tau\leq t}\|x_1(\tau)\|\}, \end{aligned}$$

which implies x_2 subsystem is input-to-state stable regarding x_1 as the input. Thus, if the following small-gain condition

$$\frac{2k_1\sqrt{1+c_1^2}}{\lambda_1} \cdot \frac{2\sqrt{(n-1)a_n^2 + c_n^2}}{b_n} < 1$$

holds [39] or equivalently (18) holds with $c = \frac{\lambda_1}{4k_1\sqrt{1+c_1^2}}$, the system $\dot{x} = \hat{A}_{rr}x$ is asymptotically stable and $\hat{A}_{rr}$ is Hurwitz. □

Remark 3. Any CAV can be chosen as the tail vehicle n, and its HDV control law can be designed by (18) to achieve the asymptotic stability of the vehicular system, since permutation transformation does not change the stability properties of systems. When $a_n = c_n = 0, b_n > 0$, condition (18) always holds. In this case, $F^T = 0$ and $\sigma(\hat{A}_{rr}) = -b_n \cup \sigma(\hat{A}_f^c)$ have negative real parts. In general, the constant c can be small and condition (18) implies that the stabilizing control of a CAV in the platoon should be more sensitive to current speed variations, compared with other factors. It is worth pointing out that, due to the existence of uncontrollable mode 0, the full-order system (6) cannot be rendered asymptotically stable from any initial conditions. Thus this small-gain technique can not be applied to system (6) to get an initial stabilizing policy, which shows the necessity of introducing the reduced-order system (7).

Therefore, a stabilizing control law in the ring road case can be obtained as follows: all CAVs use the regular HDV control laws with one CAV taking the control parameters based

on Proposition 3. And it should be mentioned that, the initial controllers in both cases do not require the specific information of parameters of the heterogenous HDVs, which is necessary for data-driven control methods.

C. String Stability

String stability is proposed to study the impact of disturbance when propagated along the platoon [46]. The general $\mathcal{L}_p$ string stability is considered with any positive integer p [30]. The definition is introduced as follows.

Definition 2 ([30]). The platoon system (8) is $\mathcal{L}_p$ string stable if there exist class $\mathcal{K}$ functions α_1, α_2 such that, for any initial state $\hat{x}(0) = [w(0), x(0)^T]^T$, and any $\zeta(t) \in \mathcal{L}_p$,

$$\|x_i\|_{\mathcal{L}_p} \leq \alpha_1(\|\zeta\|_{\mathcal{L}_p}) + \alpha_2(\|\hat{x}(0)\|) \tag{20}$$

for $i \in \{1, 2, ..., n\}$, assuming w is governed by an exosystem $\dot{w} = f(w, \zeta)$.

The definitions of class $\mathcal{K}$ functions and $\mathcal{L}_p$ space are in [42]. The definition of $\mathcal{L}_p$ string stability considers $y_i = x_i$ as the output for each vehicle i in the platoon, and applies to both freeway and ring road cases.

Proposition 4. The platoon system (8) with the obtained controller by Algorithm 1 is $\mathcal{L}_p$ string stable if there exist class $\mathcal{K}$ functions α_3, α_4 such that for any $w(0)$ and $\zeta(t) \in \mathcal{L}_p$,

$$\|w\|_{\mathcal{L}_p} \leq \alpha_3(\|\zeta\|_{\mathcal{L}_p}) + \alpha_4(\|w(0)\|). \tag{21}$$

Proof. The closed-loop system is $\dot{x} = (A - BK^{j+1})x + Hw$ with $u = -K^{j+1}x$ from Algorithm 1, and by Lemma 3, $\sigma(A - BK^{j+1}) \subseteq \mathbb{C}_-$. Thus there exist class $\mathcal{K}$ functions α_5, α_6 such that, for any $x(0)$ and $w \in \mathcal{L}_p$, $\|x\|_{\mathcal{L}_p} \leq \alpha_5(\|w\|_{\mathcal{L}_p}) + \alpha_6(\|x(0)\|)$ [42]. With (21), $\|x\|_{\mathcal{L}_p} \leq \alpha_1(\|\zeta\|_{\mathcal{L}_p}) + \alpha_2(\|\hat{x}(0)\|)$ where $\alpha_1 = \alpha_5 \circ \alpha_3$ and $\alpha_2 = \alpha_5 \circ \alpha_4 + \alpha_6$. Since $\|x_i\|_{\mathcal{L}_p} \leq \|x\|_{\mathcal{L}_p}$ for $i \in \{1, ..., n\}$, the proof is completed. □

Condition (21) implies that the exosystem should behave well in the sense of input-output stability, and can represent a large class of signals, e.g., constant signals and sinusoidal signals. Proposition 4 extends the result in [24] for $p = 2$. Similar to the discussions in Section IV-A, different implementations of A, B and H can provide more accurate descriptions in the freeway and ring road cases. Furthermore, when $p = \infty$, the string stability gives an upper bound of amplitude of the states and can be utilized to study the safety constraints [47], [48].

V. Learning-Based Robust Optimal Control

In the previous contents, disturbance attenuation is not considered, which is critical for the suppression of stop-and-go waves along the platoon. In this section, by invoking the robust and optimal control theory, a model-based controller design method is first introduced for the disturbance attenuation, which highly relies on the parameters of human behavior. Then, based on the model-based results, a learning-based controller design approach is proposed such that the robust and optimal controller can be directly obtained from input-state data in the absence of the accurate parameters of the human behavior.

A. Model-Based Robust Optimal Control

This section presents model-based robust and optimal control for the disturbance attenuation. Let us consider a system-level output [16]

$$z(t) = \begin{bmatrix} Q^{\frac{1}{2}} \\ 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ R^{\frac{1}{2}} \end{bmatrix} u(t) \tag{22}$$

where $Q^{\frac{1}{2}}$ and $R^{\frac{1}{2}}$ are the square roots of two real, symmetric, and positive definite matrices Q and R[1], respectively, denoting the weights on the states and controls (see Appendix A.3 in [38]). Since undesired perturbation w exists in the vehicular system (8), the state x may have a large fluctuation causing stop-and-go waves, and control u can be amplified causing additional energy assumption. By suppressing the impact of w on the output z, it is expected that stop-and-go waves can be dampened and energy consumption can be reduced.

Denote the transfer function matrix from input w to output z as G_{zw} whose $\mathcal{H}_\infty$ norm is

$$\|G_{zw}\|_\infty = \operatorname{ess\,sup}_{w \in \mathbb{R}} \bar{\sigma}\{G_{zw}(jw)\}$$

where $\bar{\sigma}(\cdot)$ denotes the maximum singular value of a given matrix [31]. The disturbance attenuation problem is to minimize $\|G_{zw}\|_\infty$ as

$$\begin{aligned} &\min_u \|G_{zw}\|_\infty \\ &\text{s.t. } \dot{x} = Ax + Bu + Hw \\ &\qquad z = \begin{bmatrix} Q^{\frac{1}{2}} \\ 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ R^{\frac{1}{2}} \end{bmatrix} u. \end{aligned} \tag{23}$$

whose solution is related to the linear quadratic zero-sum differential game

$$\min_u \max_w J_\gamma(x_0, u, w) = \int_0^\infty z^T(t)z(t) - \gamma^2 w^T(t)w(t)dt$$

where x_0 is the initial state and $\gamma \in \mathbb{R}$ is a positive constant [32].

Proposition 5 ([32]). Let (A, B) be stabilizable and $(A, Q^{\frac{1}{2}})$ be observable. Define $\gamma^\infty = \inf\{\gamma > 0 | \min_u \max_w J_\gamma(0, u, w) \leq 0\}$. If $\gamma > \gamma^\infty$, the game has a finite upper value, and there exists a unique symmetric positive definite solution P^* of the Riccati equation

$$A^T P + PA - P(BR^{-1}B^T - \gamma^{-2}HH^T)P + Q = 0 \tag{24}$$

with the property that $A - (BR^{-1}B^T - \gamma^{-2}HH^T)P^*$ is Hurwitz. Besides, the optimal control of the game is $u = -K^*x$ with $K^* = R^{-1}B^T P^*$.

Since the $\mathcal{H}_\infty$ norm is an induced gain from the $\mathcal{L}_2$ space of the inputs to the $\mathcal{L}_2$ space of the outputs and

$$\|G_{zw}\|_\infty = \sup_{w \in \mathcal{L}_2} \left(\frac{\int_0^\infty z^T(t)z(t)dt}{\int_0^\infty w^T(t)w(t)dt}\right)^{\frac{1}{2}},$$

by solving the linear quadratic zero-sum game with $\gamma > \gamma^\infty$, the suboptimal control will approach the optimal solution of (23) when γ decreases towards γ^∞.

[1]This section studies a different control problem and some notations, i.e., Q, R, P, K, P^*, K^*, N, are reused but irrelevant to the content of section IV.

Since the pair (A, B) is stabilizable as shown in Section III, and the pair $(A, Q^{\frac{1}{2}})$ is observable as Q is positive definite, the central problem is to solve (24) with a given feasible γ. Different from the Riccati equation (10) considered in section IV, (24) has an indefinite quadratic term $P(BR^{-1}B^T - \gamma^{-2}HH^T)P$, thus the iterative algorithm based on Lemma 3 cannot be applied. The policy iteration process [49] is one way of solving (24) with an initial admissible policy, but the initial policy is not easy to obtain considering general heterogeneous HDVs. Instead, the following way is taken, known as a value iteration process, which does not rely on any initial admissible policy, but guarantees the convergence to the solution of (24). Denote $B_2 = BR^{-\frac{1}{2}}$ and $B_1 = \gamma^{-1}H$.

Lemma 6 ([50]). Given real matrices $A, B_1, B_2, Q^{\frac{1}{2}}$ with compatible dimensions such that $(A, Q^{\frac{1}{2}})$ is observable, and (A, B_2) is stabilizable. Define a mapping F as $F(P) = PA + A^TP - P(B_2B_2^T - B_1B_1^T)P + Q$. Suppose there exists a positive definite stabilizing solution P^* of the algebraic Riccati equation (24). Then, two square matrix series Z_k and P_k can be defined for all $k \in \mathbb{Z}_{\geq 0}$ recursively as follows:

$$\begin{aligned} &P_0 = 0 \\ &A_k = A + B_1B_1^TP_k - B_2B_2^TP_k \end{aligned} \tag{25}$$

and $Z_k \geq 0$ is the unique stabilizing solution of

$$0 = Z_kA_k + A_k^TZ_k - Z_kB_2B_2^TZ_k + F(P_k) \tag{26}$$

and $P_{k+1} = P_k + Z_k$. Then, $\lim_{k\to\infty} P_k = P^*$.

The algebraic Riccati equation (26) with a negative semidefinite quadratic term needs to be recursively solved in Lemma 6. Applying Lemma 3 is difficult because A_k is updated in each iteration, making it generally impossible to construct a universal stabilizing controller based on A_k. Instead, a value iteration process is proposed without any initial stabilizing controller as follows.

Lemma 7 ([51]). Assume (A_k, B_2) is stabilizable and $(A_k, F(P_k)^{\frac{1}{2}})$ is detectable. Define the differential Riccati equation

$$\dot{Z}_k = Z_kA_k + A_k^TZ_k - Z_kB_2B_2^TZ_k + F(P_k) \tag{27}$$

with a real symmetric matrix $Z_k(0) \geq 0$, then $\lim_{t\to\infty} Z_k(t) = Z_k^\infty$ where Z_k^∞ is the unique positive semidefinite solution of (26).

In [50], it was proved that (A_k, B_2) is stabilizable and $(A_k, F(P_k)^{\frac{1}{2}})$ is detectable for $k \in \mathbb{Z}_{\geq 0}$. Thus, for the Riccati equation (26), Z_k can be obtained by solving the corresponding differential Riccati equation (27) from any initial real symmetric and positive semidefinite matrix. We propose the Euler method to solve (27) numerically as follows [52]. Let $h > 0$ be a small constant, then,

(1) $Z_k^0 = 0$;
(2) For $i = 0, ..., N-1$,

$$\begin{aligned} Z_k^{i+1} =& Z_k^i + h(Z_k^iA_k + A_k^TZ_k^i \\ &- Z_k^iB_2B_2^TZ_k^i + F(P_k)). \end{aligned} \tag{28}$$

For this process, by selecting a small enough step size h, Z_k^∞ can be approximated by Z_k^N with any accuracy. Indeed, for any $\frac{\varepsilon}{2} > 0$, there exists a $T > 0$ such that when $t > T$, $\|Z_k(t) - Z_k^\infty\| < \frac{\varepsilon}{2}$. For any $t > T$, there exist $h > 0$ and $N = \lfloor \frac{t}{h} \rfloor$ such that $\left\|Z_k(t) - Z_k^N\right\| < \frac{\varepsilon}{2}$ [52], from which $\left\|Z_k^N - Z_k^\infty\right\| < \varepsilon$.

In summary, the robust optimal control algorithm relies on the iteration in Lemma 6, which is implemented through the value iteration process (28). The following section develops learning-based methods by ADP techniques without requiring the parameters of A, B, and H.

B. Learning-Based Robust Optimal Control

This section presents learning-based methods based on Lemma 6 and value iteration (28) with fully unknown system dynamics by ADP techniques. Let $\hat{u} = R^{\frac{1}{2}}u$ and $\hat{w} = \gamma w$, then (8) can be rewritten as

$$\dot{x} = Ax + B_1\hat{w} + B_2\hat{u}. \tag{29}$$

For any $P_k \geq 0$, along the solution curve $x(t)$ of (29),

$$\begin{aligned} \frac{d}{dt}(x^TP_kx) =& (Ax + B_1\hat{w} + B_2\hat{u})^TP_kx \\ &+ x^TP_k(Ax + B_1\hat{w} + B_2\hat{u}) \\ =& x^T(A^TP_k + P_kA)x + 2\hat{w}^TB_1^TP_kx \\ &+ 2\hat{u}^TB_2^TP_kx \\ =& x^TH_kx + 2\hat{w}^TL_kx + 2\hat{u}^TM_kx \end{aligned}$$

where $H_k = A^TP_k + P_kA$, $L_k = B_1^TP_k$, and $M_k = B_2^TP_k$. Integrating the above equation from T_0 to T_s,

$$\begin{aligned} x^TP_kx|_{T_0}^{T_s} =& \int_{T_0}^{T_s} x^T\otimes x^Tdt\mathrm{vec}(H_k) \\ &+ 2\int_{T_0}^{T_s} x^T\otimes\hat{w}^Tdt\mathrm{vec}(L_k) \\ &+ 2\int_{T_0}^{T_s} x^T\otimes\hat{u}^Tdt\mathrm{vec}(M_k). \end{aligned} \tag{30}$$

For selected times $\{T_0, T_1, ..., T_s\}$, let

$$I_{\bar{x}} = \left[\int_{T_0}^{T_1} \bar{x}dt, \int_{T_1}^{T_2} \bar{x}dt, ..., \int_{T_{s-1}}^{T_s} \bar{x}dt\right]^T$$

where $\bar{x}$ has been defined in Section IV. Then, using the corresponding sampling data consistent with (30),

$$\Theta \begin{bmatrix} \bar{H}_k \\ \mathrm{vec}(L_k) \\ \mathrm{vec}(M_k) \end{bmatrix} = D_{xx}\bar{P}_k \tag{31}$$

with $\Theta = [I_{\bar{x}}, 2I_{x\hat{w}}, 2I_{x\hat{u}}]$. When Θ is of full column rank,

$$\begin{bmatrix} \bar{H}_k \\ \mathrm{vec}(L_k) \\ \mathrm{vec}(M_k) \end{bmatrix} = (\Theta^T\Theta)^{-1}\Theta^TD_{xx}\bar{P}_k,$$

and $F(P_k)$ can be solved accordingly as $F(P_k) = H_k + L_k^TL_k - M_k^TM_k + Q$. Similarly, let $\tilde{w} = \hat{w} - L_kx$ and $\tilde{u} = \hat{u} + M_kx$, and (8) can be rewritten as

$$\dot{x} = A_kx + B_1\tilde{w} + B_2\tilde{u}. \tag{32}$$

For any $Z_k^i \geq 0$, along the solution curve $x(t)$ of (32),

$$\begin{aligned}\frac{d}{dt}(x^T Z_k^i x) &= (A_k x + B_1\tilde{w} + B_2\tilde{u})^T Z_k^i x \\ &\quad + x^T Z_k^i (A_k x + B_1\tilde{w} + B_2\tilde{u}) \\ &= x^T (A_k^T Z_k^i + Z_k^i A_k)x + 2\tilde{w}^T B_1^T Z_k^i x \\ &\quad + 2\tilde{u}^T B_2^T Z_k^i x \\ &= x^T H_k^i x + 2\tilde{w}^T L_k^i x + 2\tilde{u}^T M_k^i x\end{aligned}$$

where $H_k^i = A_k^T Z_k^i + Z_k^i A_k$, $L_k^i = B_1^T Z_k^i$ and $M_k^i = B_2^T Z_k^i$. Integrating the above equation from T_0 to T_s,

$$\begin{aligned}x^T Z_k^i x|_{T_0}^{T_s} &= \int_{T_0}^{T_s} x^T \otimes x^T dt \mathrm{vec}(H_k^i) \\ &\quad + 2\int_{T_0}^{T_s} x^T \otimes \tilde{w}^T dt \mathrm{vec}(L_k^i) \\ &\quad + 2\int_{T_0}^{T_s} x^T \otimes \tilde{u}^T dt \mathrm{vec}(M_k^i).\end{aligned} \tag{33}$$

Using the data consistent with (33) based on the selected times $\{T_0, T_1, ..., T_s\}$,

$$\Theta_k \begin{bmatrix} \bar{H}_k^i \\ \mathrm{vec}(L_k^i) \\ \mathrm{vec}(M_k^i) \end{bmatrix} = D_{xx}\bar{Z}_k^i \tag{34}$$

with $\Theta_k = [I_{\bar{x}}, 2I_{x\tilde{w}}, 2I_{x\tilde{u}}]$. When Θ_k is of full column rank,

$$\begin{bmatrix} \bar{H}_k^i \\ \mathrm{vec}(L_k^i) \\ \mathrm{vec}(M_k^i) \end{bmatrix} = (\Theta_k^T\Theta_k)^{-1}\Theta_k^T D_{xx}\mathrm{vecs}(Z_k^i),$$

and the increment in (28) can be solved as $\Delta Z_k^i = H_k^i - (M_k^i)^T M_k^i + F(P_k)$. Thus the value iteration step (28) can be achieved by data-driven methods. Regarding the data needed to make Θ and Θ_k full column rank, the following Lemma shows that Θ_k is of full column rank when Θ is of full column rank. Therefore, there is no need to repeatedly collect data as long as Θ is of full rank.

Lemma 8. If Θ has full column rank, so does Θ_k, $\forall k \in \mathbb{Z}_{\geq 0}$.

Proof. Since $\tilde{w} = \hat{w} - L_k x$ and $\tilde{u} = \hat{u} + M_k x$,

$$I_{x\tilde{w}} = I_{x\hat{w}} - I_{xx}(I \otimes L_k^T),\ I_{x\tilde{u}} = I_{x\hat{u}} + I_{xx}(I \otimes M_k^T).$$

Besides, there exits a matrix Y such that $I_{xx} = I_{\bar{x}}Y$. Using these equations, when Θ has full column rank, so does Θ_k, $\forall k \in \mathbb{Z}_{\geq 0}$. □

Algorithm 2 is given based on the above analysis. Similar to step 2 in Algorithm 1, exploration noise is injected into input u to guarantee that Θ has full column rank. This algorithm starts with a predefined large constant γ and the value of γ is gradually decreased until there is no feasible solution for Algorithm 2, and the final value of γ is the approximately optimal cost of (23).

VI. Simulation and Results

In this section, we validate our theoretical analysis and algorithms by simulations using SUMO [35].

Algorithm 2 Learning-Based Robust Optimal Control of AVs

1: Given $\gamma > 0$ and let $k = 0$ and $P_0 = 0$. Collect data of $x, \hat{u}, \hat{w}$ such that Θ has full column rank.
2: If $k = 0$, $F(P_k) = Q$; Otherwise, solve H_k, L_k and M_k from (31) and let $F(P_k) = H_k + L_k^T L_k - M_k^T M_k + Q$.
3: Let $i = 0$ and $Z_k^i = 0$.
4: Solve H_k^i, L_k^i and M_k^i from (34) and let $\Delta Z_k^i = H_k^i - (M_k^i)^T M_k^i + F(P_k)$.
5: If $\left\|\Delta Z_k^i\right\| < \varepsilon_1$ where ε_1 is a predefined small constant, go to step 6; Otherwise, $Z_k^{i+1} = Z_k^i + h\Delta Z_k^i$ where h is a predefined small constant. Let $i = i + 1$ and return to step 4.
6: If $\left\|Z_k^i\right\| < \varepsilon_2$ where ε_2 is a predefined small constant, stop the procedure and $u = -R^{-\frac{1}{2}} M_k x$ is the approximate optimal control law; Otherwise, $P_{k+1} = P_k + Z_k^i$. Let $k = k + 1$ and return to step 2.

A. Parameter Settings for the Platoons

For the freeway case, a four-vehicle platoon is considered with 2 CAVs and 2 HDVs, and the sequence of the vehicles is CAV-HDV-CAV-HDV as shown in Fig. 3. For the ring road case, an eight-vehicle platoon is considered with 2 CAVs and 6 HDVs, and the sequence of the vehicles is CAV-HDV-HDV-HDV-CAV-HDV-HDV-HDV where the leading vehicle is an HDV as shown in Fig. 4. For each HDV i, optimal velocity model in [25] is taken as the car-following model and the desired velocity is

$$v^*(\Delta p_i) = \begin{cases} 0, & \Delta p_i \leq \Delta p_l \\ \frac{v_m}{2}(1 - \cos(\pi \frac{\Delta p_i - \Delta p_l}{\Delta p_h - \Delta p_l})), & \Delta p_l \leq \Delta p_i \leq \Delta p_h \\ v_m, & \Delta p_i \geq \Delta p_h \end{cases}$$

where Δp_l and Δp_h denote the lower and upper bounds of spacing headway, respectively. On top of that, the dynamics is

$$\begin{cases} \Delta\dot{p}_i = v_{i-1} - v_i, \\ \dot{v}_i = a_i^*[v^*(\Delta p_i) - v_i] + b_i^* \Delta\dot{p}_i, \end{cases} \tag{35}$$

where a_i^* and b_i^* denote the relative velocity gain and spacing headway gain of vehicle i, respectively. For HDVs, $a_i^* = 0.15$ and $b_i^* = 0.25$ when i is an odd number, otherwise $a_i^* = 0.25$, and $b_i^* = 0.25$. Each vehicle's length is 4.8 m [12]. For the freeway case, the platoon runs in a sufficiently long straight line. Besides, for HDVs, the desired velocity is $v^* = 28$ m/s and the desired headway is $\Delta p^* = 30.02$ m based on the freeway traffic data [13]. While desired headways decrease to 16 m for AVs under the same desired velocities [53]. In addition, the velocity of the virtual leading vehicle 0 is $v^* + 2e^{-t}$. For the ring road case, the total length of the ring road is 99.2 m. Besides, we set the desired velocity $v^* = 7.5$ m/s and headway $\Delta p^* = 7.6$ m for all the vehicles based on Experiment A in [12], and it can be checked that Assumption 1 holds. Moreover, the disturbance w in (16) is $2e^{-t}$ and $H = e_2 \in \mathbb{R}^7$.

B. Learning-Based Optimal Control

In this section, we implement Algorithm 1 for both freeway and ring road cases. For the freeway case, based on Proposition

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3287131

Fig. 3. Vehicular network for the freeway case in the SUMO simulation.

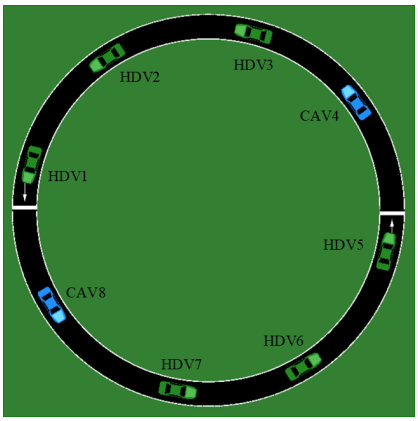


Fig. 4. Vehicular network for the ring road case in the SUMO simulation.

1, let $u_2 = a_2\tilde{p}_2 - b_2\tilde{v}_2 + c_2\tilde{v}_1 + \xi_2(t)$ and $u_4 = a_4\tilde{p}_4 - b_4\tilde{v}_4 + c_4\tilde{v}_3 + \xi_4(t)$ be the initial control controllers for CAV 2 and CAV 4, where $a_2 = a_4 = 0.3927$, $b_2 = b_4 = 0.5$, $c_2 = c_4 = 0.25$, and $\xi_2(t)$ and $\xi_4(t)$ are exploration noise. Then, the initial state feedback gain is

$$K_0 = \begin{bmatrix} 0 & -c_2 & -a_2 & b_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -c_4 & -a_4 & b_4 \end{bmatrix}.$$

To collect the training data, the initial state $x(0) = [0, -1, 1, 1.5, 0.1, 0.2, 0.3, -0.1]^T$. Exploration noise is injected into each input with

$$\xi_i(t) = \frac{1}{M}\sum_{k=1}^{M} \sin f_k t$$

where $i \in \{2, 4\}$, $M = 100$ and frequency $f_k \sim U[-250, 250]$ with U denoting the uniform distribution. The sampling time is 0.01 s and rank condition (15) is satisfied by collecting 800 data points. Q and R are identity matrices. P^* and K^* are solved by the algebraic Riccati equation (10) using A_f and B_f as the baseline for comparison. The errors in the iteration process are shown in Fig. 5, and it can be seen that P^j and K^j converge to P^* and K^* within 6 iteration steps, respectively, which validates the effectiveness of Corollary 1.

For the ring road case, based on Proposition 3, let $u_4 = a_4\tilde{p}_4 - b_4\tilde{v}_4 + c_4\tilde{v}_3 + \xi_4(t)$ and $u_8 = -b_8\tilde{v}_8 + \xi_4(t)$ where $a_4 = 0.3927$, $b_4 = b_8 = 0.5$, $c_4 = 0.25$, and $\xi_4(t)$ and $\xi_8(t)$ are exploration noise. The initial state feedback gain can be obtained similarly. To collect the training data, the initial state $x(0) = [1, -1, 1, 1.5, 0.1, 0.2, 0.3, 0.5, -0.5, 1, 0.4, 0.5, -0.5, 1, -1]^T$. Exploration noise signals ξ_4 and ξ_8 have the same forms as the freeway case and rank condition (17) holds with 3300 data points. $Q = 2I$ and R is the identity matrix. P^* and K^* are obtained by solving the algebraic Riccati equation (10) using A_{rr} and B_{rr}. The errors in the iteration process are shown

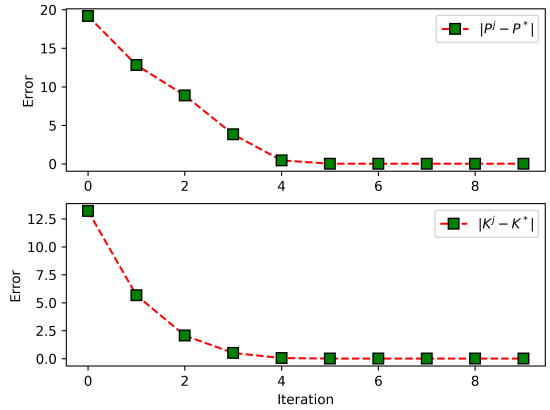


Fig. 5. Convergence of P^j and K^j for the freeway case using Algorithm 1.

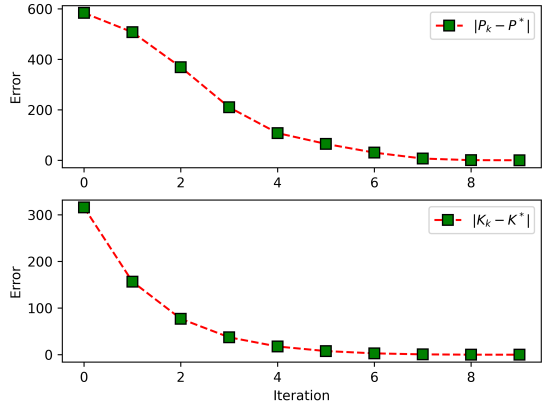


Fig. 6. Convergence of P^j and K^j for the ring road case using Algorithm 1.

in Fig. 6. It can be observed that P^j and K^j converge to P^* and K^* within 8 iteration steps, respectively, which validates the effectiveness of Corollary 2. Besides, we carried out an experiment to illustrate the efficiency of the training process. We fix 2 CAVs and gradually increase the number of HDVs in the platoon, and observe the required data points to satisfy condition (17), and compute the training time for different platoon sizes. The result is shown in Fig. 7. During this process, we do not change any settings except the collected data points, and it can be seen that condition (17) can be satisfied by collecting large enough data. In addition, the training time is within 4 seconds for the platoons with sizes less than 10, which is due to the quadratic convergence speed of the policy iteration algorithm [43], and the training time is significantly lower than the training time of deep-RL-based methods, e.g., the training process takes a few hours in [23].

C. Learning-Based Robust Optimal Control

To apply Algorithm 2, for the freeway case, exploration noise is injected using the same signal as Section VI-B. Θ in (31) is of full column rank by collecting 200 data points.

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2023.3287131

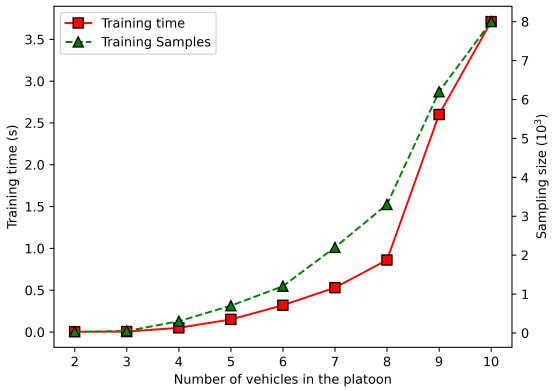


Fig. 7. Training time and training samples needed for condition (17).

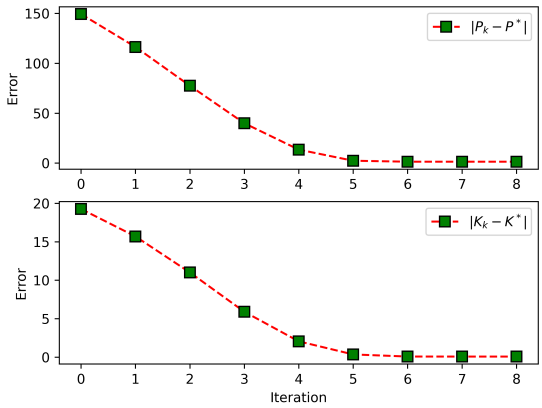


Fig. 9. Convergence of P_k and K_k for the ring road case using Algorithm 2.

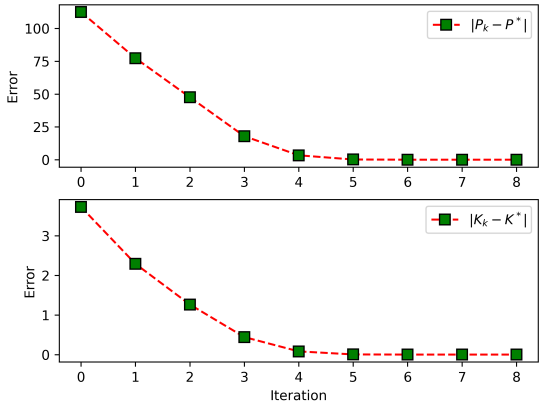


Fig. 8. Convergence of P_k and K_k for the freeway case using Algorithm 2.

The step size h is 0.01. Q and R matrices remain the same as Section VI-B. P^* and K^* are obtained by solving the algebraic Riccati equation (24) using A_f, B_f, and E for comparison. The errors in the learning process are shown in Fig. 8, and it can be seen that P_k and K_k converge to P^* and K^* within 8 steps. The minimum γ found by the learning algorithm is 3.91, which is close to the $\mathcal{H}_\infty$ gain 3.901 from input $\tilde{v}_0$ to output z, and this value is smaller than 4.075, the gain computed by the previous learning-based optimal controller.

For the ring road case, Q and R remain the same matrices as Section VI-B, and the other parameters are based on the freeway case. Similarly, P^* and K^* are solved from (24) based on A_{rr}, B_{rr}, and H for comparison. The errors in the learning process are shown in Fig. 9, and it can be observed that P_k and K_k converge to P^* and K^* within 8 steps. The minimum γ found by the learning algorithm is 6.45 which is close to the $\mathcal{H}_\infty$ gain 6.443 from input w to output z, and this value is smaller than 8.17, the gain computed by the previous learning-based optimal controller.

D. Performance Comparison

This section presents the comparison results of the performance of different controllers. Three performance indices are first considered, i.e. J_0, J_1 and J_2, where

$$J_0 = \int_0^\infty x^T Q x + u^T R u \; dt$$

denoting the overall cost; J_1 represents the overshoot described by the percentage of maximum offsets compared to the equilibrium states and J_2 denotes the entering time when the state errors stay within the 2% range. To illustrate the effectiveness of the learning process, for both cases, we compute the corresponding performance indices of the learned controllers during iterations under the same initial conditions. The result is shown in Fig. 10, where the normalized value refers to the original value divided by the minimum value over the iteration process. It turns out that, in the freeway case, compared with the initial controller, J_0, J_1 and J_2 have improved by $31.6\%, 5.8\%$ and 55.5%, respectively, and in the ring road case, compared with the initial controller, J_0, J_1 and J_2 have improved by $433.7\%, 22.1\%$ and 395.5%, respectively. Therefore, the performance of the learning-based controllers has improved during the learning process.

In [12], FollowerStopper (FS) and PI with saturation (PI) controllers are two non-model-based controllers that have been validated experimentally to stabilize the mixed traffic flow. Both controllers need to track the command speeds where the FS controller uses a proportional control and the latter applies integral control additionally. The control parameters are chosen from [12] and well-tuned to achieve good performance. Let J_3 denote the average speed variation per vehicle per second [54], and J_4 denote the fuel consumption [55] (unit: mililiter). Table I and Table II show the comparison results for different controllers with the same initial conditions, where ADPs refers to the optimal controller in Section VI-B and ADPr refers to the robust controller in Section VI-C. It can be seen that ADPs outperforms the other three controllers on J_0, J_1 and J_4, and has a similar performance to that of ADPr on J_2 and J_3 for both freeway and ring road cases. Finally, we obtain the

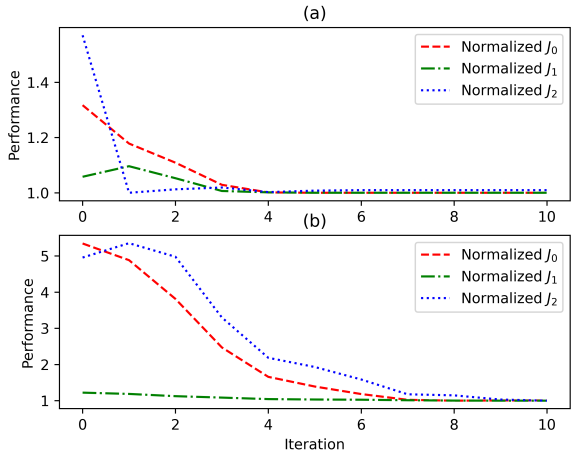


Fig. 10. (a) Normalized performance of the learned controllers in the freeway case. (b) Normalized performance of the learned controllers in the ring road case.

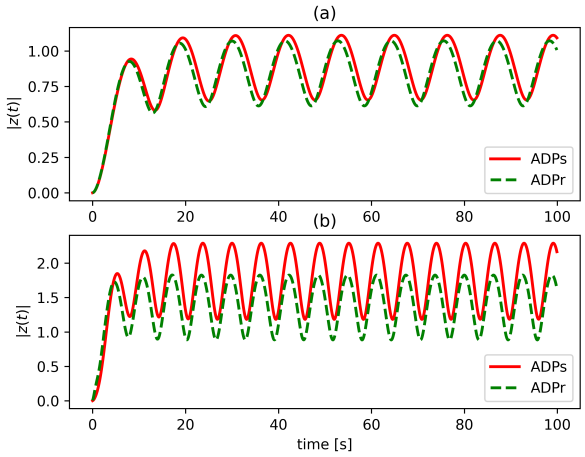


Fig. 11. (a) Trajectory of $\|z(t)\|$ in the freeway case. (b) Trajectory of $\|z(t)\|$ in the ring road case.

execution times of the controllers for the control process, and the times are 0.001, 0.004, 0.014 and 0.014 seconds for ADPs, ADPr, FS, and PI, respectively. Considering the training times in Section VI-B, the proposed methods have relatively short computation times. Also, after training, an optimal controller can be utilized for real-time control without frequent retraining until the performance deteriorates.

TABLE I
COMPARISON OF PERFORMANCE FOR DIFFERENT CONTROLLERS IN THE FREEWAY CASE

Perf. / Met.	J_0	J_1	J_2 [s]	J_3	J_4
ADPs	2422.8	26.9%	15.4	1.7	52.2
ADPr	2943.0	33.7%	13.1	1.5	57.3
FS	3984.3	29.3%	18.0	1.9	55.3
PI	4303.7	30.1%	19.4	2.3	59.5

TABLE II
COMPARISON OF PERFORMANCE FOR DIFFERENT CONTROLLERS IN THE RING ROAD CASE

Perf. / Met.	J_0	J_1	J_2 [s]	J_3	J_4
ADPs	72.2	18.6%	6.7	0.14	11.9
ADPr	226.9	21.4%	8.5	0.18	12.1
FS	1134.0	19.7%	14.7	0.20	13.1
PI	1126.4	20.1%	15.2	0.20	13.1

The robust controller ADPr is not proposed to optimize the cost function J_0, but to attenuate the impact of the external disturbance. Here the disturbance $2e^{-t}$ vanishes fast and its impact is not prominent. Therefore, to show the effectiveness of ADPr, other forms of disturbance signals are selected. For the freeway case, let the disturbance $\tilde{v}_0$ be $\sin(0.2759t)$. The trajectory of $\|z(t)\|$ is shown in Fig. 11 with zero initial conditions. The output of the robust controller has a smaller fluctuation. The truncated costs J_0, J_3 and J_4 decrease from 776, 0.76, and 41.2 to 719, 0.71, and 40.8, respectively, compared with the optimal controller in Section VI-B. Similarly, for the ring road case, when the disturbance w is $\sin(0.5t)$, the trajectory of $\|z(t)\|$ is shown in Fig. 11 with zero initial conditions. The attenuation of the disturbance can be observed. The truncated costs J_0, J_3 and J_4 decrease from 3185, 1.3, and 38.2 to 2020, 0.91, and 33.7, respectively, which shows the effectiveness of the proposed robust controller.

VII. CONCLUSIONS

This paper has presented a unified framework for learning-based optimal control of mixed vehicular systems with CAVs and heterogenous HDVs in the freeway and ring road environments. By using a model reduction technique and the PBH test, it is shown that the vehicular systems in both cases are stabilizable, and this fact is independent of the formation of HDVs and CAVs in the platoon. Based on ADP techniques, a data-driven algorithm with guaranteed convergence has been employed to solve an LQR problem for the mixed vehicular system without prior knowledge of system parameters, and small-gain techniques are utilized to construct the initial stabilizing control laws. The obtained optimal controllers can achieve the general $\mathcal{L}_p$ string stability. To attenuate the effects of the disturbance, a learning-based value iteration process has been proposed to solve the corresponding linear quadratic zero-sum game. SUMO simulation has been used to demonstrate the effectiveness of the proposed methods. Our future work will investigate the robustness of our proposed learning algorithms with respect to model uncertainties [56], [57], and consider the coordination of CAVs in particular several platoons at intersections [58] and the integration of the proposed methods with data-driven traffic signal control [59], [60].

REFERENCES

[1] D. Cao, X. Wang, L. Li, C. Lv, X. Na, Y. Xing, X. Li, Y. Li, Y. Chen, and F. Y. Wang, "Future directions of intelligent vehicles: Potentials, possibilities, and perspectives," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 1, pp. 7–10, 2022.

[2] A. Vahidi and A. Sciarretta, "Energy saving potentials of connected and automated vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 822–843, 2018.

[3] L. Chen, Y. Zhang, B. Tian, Y. Ai, D. Cao, and F.-Y. Wang, "Parallel driving os: A ubiquitous operating system for autonomous driving in cpss," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 886–895, 2022.

[4] Y. Zhang, Z. Ai, J. Chen, T. You, C. Du, and L. Deng, "Energy-saving optimization and control of autonomous electric vehicles with considering multiconstraints," *IEEE Transactions on Cybernetics*, vol. 52, no. 10, pp. 10869–10881, 2021.

[5] S. E. Li, Y. Zheng, K. Li, Y. Wu, J. K. Hedrick, F. Gao, and H. Zhang, "Dynamical modeling and distributed control of connected and automated vehicles: Challenges and opportunities," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 3, pp. 46–58, 2017.

[6] G. J. Naus, R. P. Vugts, J. Ploeg, M. J. van De Molengraft, and M. Steinbuch, "String-stable CACC design and experimental validation: A frequency-domain approach," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4268–4279, 2010.

[7] Y. Zheng, S. E. Li, J. Wang, D. Cao, and K. Li, "Stability and scalability of homogeneous vehicular platoon: Study on the influence of information flow topologies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 14–26, 2015.

[8] M. Hu, J. Li, Y. Bian, J. Wang, B. Xu, and Y. Zhu, "Distributed coordinated brake control for longitudinal collision avoidance of multiple connected automated vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 745–755, 2022.

[9] H. Zhang, J. Liu, Z. Wang, C. Huang, and H. Yan, "Adaptive switched control for connected vehicle platoon with unknown input delays," *IEEE Transactions on Cybernetics*, vol. 53, no. 3, pp. 1511–1521, 2021.

[10] S. Xiao, X. Ge, Q.-L. Han, and Y. Zhang, "Resource-efficient platooning control of connected automated vehicles over vanets," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 579–589, 2022.

[11] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S. Tadaki, and S. Yukawa, "Traffic jams without bottlenecks–experimental evidence for the physical mechanism of the formation of a jam," *New Journal of Physics*, vol. 10, no. 3, p. 033001, 2008.

[12] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, H. Pohlmann, F. Wu, B. Piccoli, *et al.*, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205–221, 2018.

[13] I. G. Jin and G. Orosz, "Optimal control of connected vehicle systems with communication delay and driver reaction time," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 2056–2070, 2016.

[14] Z. Xu and X. Jiao, "Robust control of connected cruise vehicle platoon with uncertain human driving reaction time," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 368–376, 2021.

[15] J. Yang, D. Zhao, J. Lan, S. Xue, W. Zhao, D. Tian, Q. Zhou, and K. Song, "Eco-driving of general mixed platoons with CAVs and HDVs," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1190–1203, 2022.

[16] J. Wang, Y. Zheng, Q. Xu, J. Wang, and K. Li, "Controllability analysis and optimal control of mixed traffic flow with human-driven and autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7445–7459, 2020.

[17] Y. Zheng, J. Wang, and K. Li, "Smoothing traffic flow via control of autonomous vehicles," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 3882–3896, 2020.

[18] 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.

[19] Y. Wang, Z. Wang, K. Han, P. Tiwari, and D. B. Work, "Gaussian process-based personalized adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21178–21189, 2022.

[20] C. Desjardins and B. Chaib-Draa, "Cooperative adaptive cruise control: A reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1248–1260, 2011.

[21] C. Wu, A. Kreidieh, E. Vinitsky, and A. M. Bayen, "Emergent behaviors in mixed-autonomy traffic," in *Conference on Robot Learning (CoRL)*, pp. 398–407, 2017.

[22] A. R. Kreidieh, C. Wu, and A. M. Bayen, "Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1475–1480, 2018.

[23] C. Wu, A. R. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: A modular learning framework for mixed autonomy traffic," *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 1270–1286, 2022.

[24] W. Gao, Z. P. Jiang, and K. Ozbay, "Data-driven adaptive optimal control of connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1122–1133, 2016.

[25] M. Huang, Z. P. Jiang, and K. Ozbay, "Learning-based adaptive optimal control for connected vehicles in mixed traffic: robustness to driver reaction time," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 5267–5277, 2020.

[26] L. Cui, K. Ozbay, and Z. P. Jiang, "Combined longitudinal and lateral control of autonomous vehicles based on reinforcement learning," in *2021 American Control Conference (ACC)*, pp. 1929–1934, IEEE, 2021.

[27] W. Gao and Z. P. Jiang, "Nonlinear and adaptive suboptimal control of connected vehicles: A global adaptive dynamic programming approach," *Journal of Intelligent & Robotic Systems*, vol. 85, no. 3-4, pp. 597–611, 2017.

[28] G. Gunter, C. Janssen, W. Barbour, R. E. Stern, and D. B. Work, "Model-based string stability of adaptive cruise control systems using field data," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 1, pp. 90–99, 2019.

[29] E. Kayacan, "Multiobjective H_∞ control for string stability of cooperative adaptive cruise control systems," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 1, pp. 52–61, 2017.

[30] J. Ploeg, N. Van De Wouw, and H. Nijmeijer, "$\mathcal{L}_p$ string stability of cascaded systems: Application to vehicle platooning," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 2, pp. 786–793, 2013.

[31] K. Zhou and J. C. Doyle, *Essentials of Robust Control*, vol. 104. Prentice Hall Upper Saddle River, NJ, 1998.

[32] T. Başar and P. Bernhard, *H-Infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Springer Science & Business Media, 2008.

[33] T. Liu, L. Cui, B. Pang, and Z. P. Jiang, "Learning-based control of multiple connected vehicles in the mixed traffic by adaptive dynamic programming," *IFAC-PapersOnLine*, vol. 54, no. 14, pp. 370–375, 2021.

[34] T. Liu, L. Cui, B. Pang, and Z. P. Jiang, "Data-driven adaptive optimal control of mixed-traffic connected vehicles in a ring road," in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 77–82, IEEE, 2021.

[35] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y. P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using SUMO," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582, 2018.

[36] Y. Wang, M. L. Delle Monache, and D. B. Work, "Identifiability of car-following dynamics," *Physica D: Nonlinear Phenomena*, no. 0167-2789, p. 133090, 2021.

[37] Z. P. Jiang, A. R. Teel, and L. Praly, "Small-gain theorem for ISS systems and applications," *Mathematics of Control, Signals and Systems*, vol. 7, no. 2, pp. 95–120, 1994.

[38] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. John Wiley & Sons, 2012.

[39] A. Isidori, *Lectures in Feedback Design for Multivariable Systems*. Springer, 2017.

[40] J. P. Hespanha, *Linear Systems Theory*. Princeton University Press, 2018.

[41] I. G. Jin and G. Orosz, "Dynamics of connected vehicle systems with delayed acceleration feedback," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 46–64, 2014.

[42] H. K. Khalil, *Nonlinear Systems*. Prentice-Hall, NJ, 2002.

[43] D. Kleinman, "On an iterative technique for Riccati equation computations," *IEEE Transactions on Automatic Control*, vol. 13, no. 1, pp. 114–115, 1968.

[44] Y. Jiang and Z. P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699–2704, 2012.

[45] E. D. Sontag, "Smooth stabilization implies coprime factorization," *IEEE Transactions on Automatic Control*, vol. 34, no. 4, pp. 435–443, 1989.

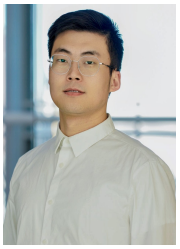
[46] S. Feng, Y. Zhang, S. E. Li, Z. Cao, H. X. Liu, and L. Li, "String stability for vehicular platoon control: Definitions and analysis methods," *Annual Reviews in Control*, vol. 47, pp. 81–97, 2019.

[47] Z. Ju, H. Zhang, X. Li, X. Chen, J. Han, and M. Yang, "A survey on attack detection and resilience for connected and automated vehicles: From vehicle dynamics and control perspective," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 815–837, 2022.

[48] G. Sidorenko, A. Fedorov, J. Thunberg, and A. Vinel, "Towards a complete safety framework for longitudinal driving," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 809–814, 2022.

[49] L. Cui and Z. P. Jiang, "A reinforcement learning look at risk-sensitive linear quadratic Gaussian control," *ArXiv Preprint arXiv:2212.02072*, 2022.
[50] A. Lanzon, Y. Feng, B. D. Anderson, and M. Rotkowitz, "Computing the positive stabilizing solution to algebraic Riccati equations with an indefinite quadratic term via a recursive method," *IEEE Transactions on Automatic Control*, vol. 53, no. 10, pp. 2280–2291, 2008.
[51] V. Kučera, "A review of the matrix Riccati equation," *Kybernetika*, vol. 9, no. 1, pp. 42–61, 1973.
[52] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical Analysis*. Cengage Learning, 2015.
[53] Y. Zhu, D. Zhao, and Z. Zhong, "Adaptive optimal control of heterogeneous CACC system with uncertain dynamics," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 4, pp. 1772–1779, 2018.
[54] S. Wang, M. Shang, M. W. Levin, and R. Stern, "A general approach to smoothing nonlinear mixed traffic via control of autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 146, p. 103967, 2023.
[55] N. Wan, A. Vahidi, and A. Luckow, "Optimal speed advisory for connected vehicles in arterial roads and the impact on mixed traffic," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 548–563, 2016.
[56] B. Pang, T. Bian, and Z. P. Jiang, "Robust policy iteration for continuous-time linear quadratic regulation," *IEEE Transactions on Automatic Control*, vol. 67, no. 1, pp. 504–511, 2021.
[57] X. Tang, K. Yang, H. Wang, J. Wu, Y. Qin, W. Yu, and D. Cao, "Prediction-uncertainty-aware decision-making for autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 849–862, 2022.
[58] S. D. Kumaravel, A. A. Malikopoulos, and R. Ayyagari, "Optimal coordination of platoons of connected and automated vehicles at signal-free intersections," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 186–197, 2021.
[59] T. Liu, H. Wang, and Z. P. Jiang, "Data-driven optimal control of traffic signals for urban road networks," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 844–849, IEEE, 2022.
[60] Z. Yin, T. Liu, C. Wang, H. Wang, and Z. P. Jiang, "Reducing urban traffic congestion using deep learning and model predictive control," *IEEE Transactions on Neural Networks and Learning Systems*, 2023. doi:10.1109/TNNLS.2023.3264709.

Tong Liu received the B.Sc. and M.Sc. degrees from Beijing Jiaotong University, Beijing, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree at the Control and Networks Lab, Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY, USA. His research interests include adaptive dynamic programming, cooperative control of connected vehicles, and adaptive traffic signal control.



Leilei Cui received the B.Eng. degree in automation from Northwestern Polytechnical University, Xian, China, in 2016, and the M.S. degree in control engineering at Shanghai Jiao Tong University, Shanghai, China, in 2019. He is currently a Ph.D. candidate in the Control and Networks Lab, Tandon School of Engineering, New York University. His research interests include robot control, reinforcement learning, adaptive dynamic programming (ADP), and optimal control.

Bo Pang received the B.Sc. degree in Automation from the Beihang University, Beijing, China, in 2014, and the M.Sc. degree in Control Science and Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2017 and the Ph.D. degree in Electrical Engineering from New York University, NY, U.S.A, in 2021. His research interests include optimal/stochastic control, approximate/adaptive dynamic programming, and reinforcement learning.

Zhong-Ping Jiang received the M.Sc. degree in statistics from the University of Paris XI, France, in 1989, and the Ph.D. degree in automatic control and mathematics from the Ecole des Mines de Paris (now, called ParisTech-Mines), France, in 1993, under the direction of Prof. Laurent Praly.

Currently, he is a Professor of Electrical and Computer Engineering at the Tandon School of Engineering, New York University. His main research interests include stability theory, robust/adaptive/distributed nonlinear control, robust adaptive dynamic programming, reinforcement learning and their applications to information, mechanical and biological systems. In these fields, he has written six books and is author/co-author of over 500 peer-reviewed journal and conference papers.

Prof. Jiang is a recipient of the prestigious Queen Elizabeth II Fellowship Award from the Australian Research Council, CAREER Award from the U.S. National Science Foundation, JSPS Invitation Fellowship from the Japan Society for the Promotion of Science, Distinguished Overseas Chinese Scholar Award from the NSF of China, and several best paper awards. He has served as Deputy Editor-in-Chief, Senior Editor and Associate Editor for numerous journals. Prof. Jiang is a Fellow of the IEEE, IFAC, CAA, and AAIA, a foreign member of the Academia Europaea (Academy of Europe), and is among the Clarivate Analytics Highly Cited Researchers. In 2022, he received the Excellence in Research Award from the NYU Tandon School of Engineering.