2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Learning Optimal Robust Control of Connected Vehicles in Mixed Traffic Flow

Jie Li, Jiawei Wang, Shengbo Eben Li, Keqiang Li

Abstract— **Connected and automated vehicles (CAVs) technologies promise to attenuate undesired traffic disturbances. However, in mixed traffic where human-driven vehicles (HDVs) also exist, the nonlinear human-driving behavior has brought critical challenges for effective CAV control. This paper employs the policy iteration method to learn the optimal robust controller for nonlinear mixed traffic systems. Precisely, we consider the H_∞ control framework and formulate it as a zero-sum game, the equivalent condition for whose solution is converted into a Hamilton–Jacobi inequality with a Hamiltonian constraint. Then, a policy iteration algorithm is designed to generate stabilizing controllers with desired attenuation performance. Based on the updated robust controller, the attenuation level is further optimized in sum of squares program by leveraging the gap of the Hamiltonian constraint. Simulation studies verify that the obtained controller enables the CAVs to dampen traffic perturbations and smooth traffic flow.**

I. Introduction

Undesired traffic disturbances may easily lead to the occurrence of traffic waves, where the involved vehicles periodically accelerate and decelerate, resulting in decreased travel efficiency, fuel economy and driving safety [1]. The emergence of connected and automated vehicles (CAVs) promises efficient attenuation of traffic disturbances [2]. Recent research has either theoretically or empirically revealed that in mixed traffic, where human-driven vehicles (HDVs) also exist, CAVs could mitigate traffic waves and stabilize traffic flow even in a low penetration rate [3]–[5].

Regarding the specific control methods of CAVs, existing model-based research mostly relies on a linearized dynamics model for the mixed traffic system. Common modeling frameworks include Lagrangian control [6], [7], connected cruise control (CCC) [8], and leading cruise control (LCC) [9]. To obtain such a model, these research usually needs to linearize a car-following model of HDVs, *e.g.*, optimal velocity model (OVM) [10], around certain traffic equilibrium state. In practice, however, the performance of these methods may easily be compromised due to the nonlinear human-driving behaviors and the time-varying traffic equilibrium states. To address these issues, some model-free learning policies have been recently proposed via reinforcement learning (RL) [11], [12] or data-driven predictive control [13], [14], but the dependence on large-scale traffic data has limited its practical deployment.

This study is supported by National Key R&D Program of China with 2022YFB2502901. It is also partially supported by the National Natural Science Foundation of China under grant number 52221005. All correspondence should be sent to S. Li.

J. Li, J. Wang, S. Li and K. Li are with the School of Vehicle and Mobility, Tsinghua University, Beijing, China. ({jie-li18,wang-jw18}@mails.tsinghua.edu.cn, {lishbo,likq}@tsinghua.edu.cn).

To our best knowledge, very few studies have addressed the disturbance attenuation problem in nonlinear mixed traffic systems, with a very recent exception in [15], where Lyapunov methods are employed for stability analysis. Besides, existing methods have not considered optimizing the disturbance attenuation performance of CAVs in mixed traffic flow. To optimize and control the nonlinear traffic system via CAVs, approximate dynamic programming (ADP) provides a promising technique through solving the nonlinear H_∞ optimal control problem of the mixed traffic system. Particularly, this work utilizes the tool of policy iteration (PI) from ADP to learn a robust optimal controller for mixed traffic systems with explicit consideration of nonlinear human-driving behaviors.

PI is a class of effective numerical method for nonlinear robust control [16], [17], and has been recently applied to a wide range of diverse fields; see, *e.g.*, robot manipulator [18] and vehicle platooning [19]. Compared with RL, PI enjoys complete theoretical foundations, including algorithm convergence and closed-loop stability [20], which play a critical role in connected vehicle control. Indeed, benefiting from the advantages of this method, an adaptive optimal controller has been recently designed in [21] with respect to unknown and heterogeneous HDV behaviors in mixed traffic. However, how to achieve an optimal disturbance attenuation performance remains an open question. To address this issue, this paper develops a model-based learning algorithm to optimize the disturbance attenuation level and derive an H_∞ optimal controller for the CAVs from the nonlinear mixed traffic dynamics. Precisely, the main contributions of this paper are as follows:

- An affine nonlinear model is established for the mixed traffic system based on the LCC framework. Compared with existing work where linearized dynamics around equilibrium states are under consideration [7]–[9], we directly focus on the nonlinear dynamics to design model-based learning control policies for the CAVs.
- The H_∞ control problem of the nonlinear mixed traffic system is formulated as a zero-sum game, whose control policy can be obtained by solving the converted Hamilton-Jacobi (HJ) inequality. The obtained state-feedback controller is proved to achieve the given attenuation level for the mixed traffic system.
- The HJ inequality reserves the optimization space for attenuation level. Accordingly, we further develop a model-based learning algorithm, which optimizes the attenuation performance in outer-loop iterations through

2023 62nd IEEE Conference on Decision and Control (CDC) | 979-8-3503-0124-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/CDC49753.2023.10383277

sum of squares programs, and generates stabilizing controllers with attenuation performance guarantees at every inner-loop iteration in a PI paradigm.

The rest of this paper is organized as follows. Section II establishes the nonlinear mixed traffic model. The model-based learning algorithm and its theoretical analysis are presented in Section III. Section IV shows the simulation results, and Section V concludes this work.

II. Nonlinear Modeling of Mixed Traffic

Consider the mixed traffic system shown in Fig. 1, where there is one head vehicle (indexed as 0), m CAVs and $n-m$ HDVs following behind (indexed from 1 to n). Without loss of generality, we assume that the first vehicle behind the head vehicle is CAV. Denote $S = \{l_1, l_2, \ldots, l_m\}$ as the set of all the CAV indexes, where $1 = l_1 < l_2 < \cdots < l_m \leq n$. Such a multi-vehicle system in mixed traffic is named as a special form of LCC [9], which incorporates both upstream and downstream traffic information, and allows the CAV to attenuate the disturbances from the head vehicle, whilst actively leading the motion of the HDVs behind.

Denote $p_i(t)$ and $v_i(t)$ as the position and velocity of vehicle i, respectively. Then, $s_i(t) = p_{i-1}(t) - p_i(t)$ and $\dot{s}_i(t) = v_{i-1}(t) - v_i(t)$ represent the spacing (relative distance) and relative velocity of vehicle i with respect to its predecessor. Motivated by recent research on mixed traffic [4], [6], [8], we consider the OVM model for the HDVs, a typical nonlinear car-following model, to represent its longitudinal driving behavior, which is given by [10]

$$\dot{v}_i(t) = \alpha_i \left(v^{\mathrm{d}}(s_i(t)) - v_i(t)\right) + \beta_i \dot{s}_i(t),\ i \notin S, \quad (1)$$

where α_i, β_i denote the sensitivity coefficients for vehicle i, and the spacing-dependent desired velocity $v^{\mathrm{d}}(s_i(t))$ is

$$v^{\mathrm{d}}(s_i(t)) = \begin{cases} 0, & s_i(t) \leq s_{\mathrm{st}}, \\ \bar{v}^{\mathrm{d}}\left(s_i(t)\right), & s_{\mathrm{st}} < s_i(t) < s_{\mathrm{go}}, \\ v_{\max}, & s_i(t) \geq s_{\mathrm{go}}, \end{cases} \quad (2)$$

with $\bar{v}^{\mathrm{d}}$ given by

$$\bar{v}^{\mathrm{d}}\left(s_i(t)\right) = \frac{v_{\max}}{2}\left(1 - \cos\left(\frac{s_i(t) - s_{\mathrm{st}}}{s_{\mathrm{go}} - s_{\mathrm{st}}}\pi\right)\right).$$

Define the deviation state of each vehicle from the equilibrium state as $\tilde{s}_i(t) = s_i(t) - s^*$, $\tilde{v}_i(t) = v_i(t) - v^*$, where s^*, v^* denote the equilibrium spacing and velocity, respectively. For simplicity, a homogeneous setup for s^* is under consideration, but all the results can be generalized to the heterogeneous case. Denote $x_i(t) = [\tilde{s}_i(t), \tilde{v}_i(t)]^\top$ as the states of vehicle i. Then, the nonlinear dynamics model for each HDV around the equilibrium state is obtained as follows

$$\begin{aligned} \dot{\tilde{s}}_i(t) &= \tilde{v}_{i-1}(t) - \tilde{v}_i(t), \\ \dot{\tilde{v}}_i(t) &= h_i\left(\tilde{s}_i(t),\ \tilde{v}_i(t),\ \tilde{v}_{i-1}(t)\right), \end{aligned} \quad (3)$$

where

$$\begin{aligned} h_i(\cdot) = \alpha_i &\left(v^{\mathrm{d}}(\tilde{s}_i(t) + s^*) - (\tilde{v}_i(t) + v^*)\right) \\ &+ \beta_i\left(\tilde{v}_{i-1}(t) - \tilde{v}_i(t)\right). \end{aligned} \quad (4)$$

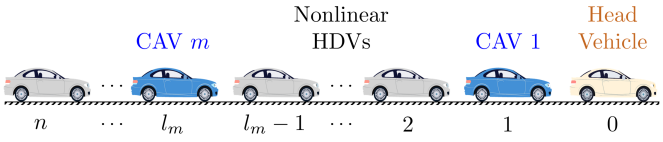


Fig. 1. Schematic of the mixed traffic system in the LCC framework. There are multiple CAVs (colored in blue) and HDVs (colored in gray, with a nonlinear car-following model) following behind the head vehicle. The first vehicle behind the head vehicle is CAV.

For the CAV, its acceleration signal $\dot{\tilde{v}}_{l_i}(t)$ is regarded as the control input $u_i(t)$, $i \in \mathbb{N}_1^m$, where $\mathbb{N}_1^m$ denotes all the natural numbers within $[1, m]$. Then, the longitudinal control model of the CAVs is given by [9]

$$\dot{\tilde{v}}_{l_i}(t) = u_i(t),\ i \in \mathbb{N}_1^m. \quad (5)$$

Lumping the dynamics of the CAV and the HDVs, a nonlinear model for the LCC system can be established as follows

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) + k(x(t))w(t), \quad (6)$$

where the lumped state and control input are defined as

$$\begin{aligned} x(t) &= \left[x_1^\top(t), x_2^\top(t), \cdots, x_n^\top(t)\right]^\top \in \mathbb{R}^{2n}, \\ u(t) &= \left[u_1^\top(t), u_2^\top(t), \cdots, u_m^\top(t)\right]^\top \in \mathbb{R}^{m}, \end{aligned}$$

respectively, and the disturbance $w(t) = \tilde{v}_0(t) \in \mathbb{R}$ represents the velocity deviation of the head vehicle. The functions $f: \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $g: \mathbb{R}^{2n} \to \mathbb{R}^{2n\times m}$ and $k: \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ are known vector-valued functions. In the system model (6), the dynamics of the rear $n-1$ vehicles satisfy (3). In the case of $n = 3$, $m = 1$, for example, the specific expression of the dynamic model (6) is as follows

$$f(x(t)) = \begin{bmatrix} -\tilde{v}_1(t) \\ 0 \\ \tilde{v}_1(t) - \tilde{v}_2(t) \\ h\left(\tilde{s}_2(t),\ \tilde{v}_2(t),\ \tilde{v}_1(t)\right) \\ \tilde{v}_2(t) - \tilde{v}_3(t) \\ h\left(\tilde{s}_3(t),\ \tilde{v}_3(t),\ \tilde{v}_2(t)\right) \end{bmatrix},$$

$$g(x(t)) = [0,1,0,0,0,0]^\top,\ k(x(t)) = [1,0,0,0,0,0]^\top.$$

To describe the performance of the mixed traffic system, we define $z(t) \in \mathbb{R}^{2n+m}$ as the output

$$z(t) \triangleq \begin{bmatrix} \sqrt{Q}x(t) \\ \sqrt{R}u(t) \end{bmatrix}, \quad (7)$$

and the square of its norm is given by

$$\|z(t)\|^2 = z^\top(t)z(t) = x^\top(t)Qx(t) + u^\top(t)Ru(t), \quad (8)$$

where $\sqrt{Q} = \mathrm{diag}(\theta_s, \theta_v, \ldots, \theta_s, \theta_v) \in \mathbb{R}^{2n\times 2n}$ and $\sqrt{R} = \mathrm{diag}(\theta_u, \ldots, \theta_u) \in \mathbb{R}^{m\times m}$ are positive definite matrices, with $\theta_s, \theta_v, \theta_u$ denoting the weight coefficients for penalizing spacing deviations, velocity deviations and control inputs. Note that the system (6)-(7) is zero-state observable.

Remark 1: Existing research mostly considers the linearized dynamics of the mixed traffic system to design the

control policies for the CAVs. In this paper, we directly focus on the affine model (6), and aim at developing model-based learning method with theoretical guarantees and ability to process nonlinear mixed traffic systems. Note that although the OVM model (1) is utilized for describing the HDVs' dynamics and the resulting expression of $g(x)$ is a constant function, our control method is applicable to any mixed traffic system in the general nonlinear affine form of (6).

III. H_∞ Optimal Control by Policy Iteration

This section first formulates the H_∞ control problem of the mixed traffic system as a zero-sum game. Based on the PI framework, a model-based learning algorithm is then developed to solve the equivalent HJ inequality.

A. Problem Formulation

To characterize the disturbance attenuation performance of the closed-loop system, we first give the following definition. For the convenience of writing, the time t will be omitted in the subsequent content.

Definition 1 (Disturbance Attenuation): For all disturbance $w \in L_2[0,\infty)$, the closed-loop system (6)-(7) with the initial state $x(0)=0$ is said to have an L_2-gain $\leq \gamma$, if

$$\int_0^\infty \|z\|^2 \, dt \leq \gamma^2 \int_0^\infty \|w\|^2 \, dt.$$

In other words, the system satisfies the disturbance attenuation performance with attenuation level $\gamma > 0$.

The attenuation level γ captures the influence of the external disturbance w on the performance output z. Precisely, a smaller value of γ indicates a better capability of CAVs in dissipating traffic waves. Then, given the nonlinear mixed traffic system (6), define the value function of the initial state $x = x(0)$ as

$$V(x) \triangleq \int_0^\infty (l(x(\tau), u(\tau), w(\tau)))d\tau, \tag{9}$$

where the cost function is defined as

$$l(x,u,w) \triangleq x^\top Q x + u^\top R u - \gamma^2 w^\top w. \tag{10}$$

From the point of view of game theory, disturbance aims at deteriorating control performance, while control policy optimizes the worst-case performance in H_∞ control [22]. Given a suitable attenuation level $\gamma > 0$, the H_∞ control problem can be formulated as the following zero-sum game [23]

$$V^*(x) = \min_{u(\cdot)} \max_{w(\cdot)} \int_0^\infty (l(x(\tau), u(\tau), w(\tau)))d\tau, \tag{11}$$

where $V^*(x)$ is the Nash value, control $u(\cdot)$ and disturbance $w(\cdot)$ are two sides of the game. Moreover, the controller at the Nash equilibrium should stabilize the system at $w \equiv 0$, and allow the closed-loop system to have an L_2-gain $\leq \gamma$ for all $w \in L_2[0,\infty)$. Further, the H_∞ optimal control problem explores the lowest attenuation level $\gamma^* > 0$ and resolves the corresponding zero-sum game (11). The existence of the lowest attenuation level of nonlinear affine systems is guaranteed by [22]. The following assumption declares the existence of the desired controller in mixed traffic flow.

Assumption 1: Given an attenuation level $\gamma \geq \gamma^*$, there exits a robust controller $u = \pi(x)$ with $\pi(0) = 0$ such that the system (6)-(7) is stabilized at $w \equiv 0$ and that the closed-loop system has an L_2-gain $\leq \gamma$ for all $w \in L_2[0,\infty)$.

On the premise of zero-state observability, the solution to the following Hamilton–Jacobi–Isaacs (HJI) equation solves the zero-sum game [23]

$$\begin{aligned} &x^\top Q x + (\nabla V^*(x))^\top f(x) \\ &\quad - \frac{1}{4}(\nabla V^*(x))^\top g(x) R^{-1} g^\top(x) \nabla V^*(x) \\ &\quad + \frac{1}{4\gamma^2}(\nabla V^*(x))^\top k(x) k^\top(x) \nabla V^*(x) = 0, \end{aligned} \tag{12}$$

which is a nonlinear partial differential equation about the optimal value function $V^*(x)$ with the boundary condition $V^*(0) = 0$. If the HJI equation has a smooth positive semi-definite solution $V^*(x)$, then the controller is derived as

$$u^*(x) = -\frac{1}{2} R^{-1} g^\top(x) \nabla V^*(x).$$

Traditional PI algorithms usually focus on providing numerical methods for solving the HJI equation (12) to generate robust controllers. Through the following lemma, or as shown in Theorem 1, we can also derive a robust controller from the associated HJ inequality (13). Compared with solving HJI equation directly, the gap of HJ inequality allows for further optimizing the attenuation level.

Lemma 1 ([24, Theorem 16 & Corollary 17]): Consider the nonlinear system (6)-(7) with an attenuation level γ. Suppose that there is a smooth positive semi-definite solution $V(x)$ to the HJI equation (12) or the HJ inequality

$$\begin{aligned} &x^\top Q x + (\nabla V(x))^\top f(x) \\ &\quad - \frac{1}{4}(\nabla V(x))^\top g(x) R^{-1} g^\top(x) \nabla V(x) \\ &\quad + \frac{1}{4\gamma^2}(\nabla V(x))^\top k(x) k^\top(x) \nabla V(x) \leq 0, \end{aligned} \tag{13}$$

with the boundary condition $V(0) = 0$, then the closed-loop system with the state feedback controller

$$u(x) = -\frac{1}{2} R^{-1} g^\top(x) \nabla V(x),$$

is asymptotically stable at $w \equiv 0$, and has an L_2-gain $\leq \gamma$ for all disturbance $w \in L_2[0,\infty)$.

Besides, the existence of the solution to the HJ inequality is guaranteed by the following lemma.

Lemma 2 ([24, Theorem 18]): Consider the nonlinear system (6)-(7) and an attenuation level γ. If there is a controller $u = \pi(x)$ satisfying Assumption 1, there exists a smooth positive semi-definite solution $V_a(x)$ to the HJ inequality (13).

Similar to the HJI equation (12), the HJ inequality (13) contains two nonlinear terms about the differential of value function, which are related to control input function $g(x)$ and disturbance input function $k(x)$, respectively. This makes it non-trivial to get the problem solutions.

B. Inequality Conversion

We proceed to provide a concrete procedure to solve the HJ inequality (13). To begin with, the inequality is transformed to eliminate the nonlinear differential term about disturbance input function $k(x)$ while retaining the characteristics of the control policy.

In order to facilitate the solution through conversion, add the following square term

$$-\gamma^2 \left\| w - \frac{1}{2\gamma^2} k^\top(x) \nabla V(x) \right\|^2 \leq 0,$$

to both sides of the HJ inequality (13) and get the following converted inequality for all disturbance signal w

$$\begin{aligned} & x^\top Q x + (\nabla V(x))^\top f(x) \\ & \quad - \frac{1}{4} (\nabla V(x))^\top g(x) R^{-1} g^\top(x) \nabla V(x) \\ & \quad + (\nabla V(x))^\top k(x) w - \gamma^2 w^\top w \leq 0. \end{aligned} \tag{14}$$

This transformed inequality (14) is exactly the problem that we aim to solve in this work. It can be proved that the controller derived from the feasible solution of the converted inequality (14) reserves the stability and disturbance attenuation performance. With Lemma 2 in place, it follows that the inequality (14) admits a feasible solution $V_a(x)$.

Theorem 1 (Stability and Robustness): Suppose that the converted inequality (14) admits a feasible solution $V(x)$. Then, the closed-loop system with the controller

$$u(x) = -\frac{1}{2} R^{-1} g^\top(x) \nabla V(x),$$

is asymptotically stable at $w \equiv 0$, and has an L_2-gain $\leq \gamma$ for all $w \in L_2[0, \infty)$.

Proof: Substituting the expression of $u(x)$ into the transformed inequality (14) yields

$$\begin{aligned} & (\nabla V(x))^\top \big(f(x) + g(x) u(x) + k(x) w \big) \\ & \leq - x^\top Q x - u^\top(x) R u(x) + \gamma^2 w^\top w. \end{aligned} \tag{15}$$

When $w \equiv 0$, one has

$$(\nabla V(x))^\top \big(f(x) + g(x) u(x) \big) \leq -x^\top Q x - u^\top(x) R u(x) \leq 0.$$

Therefore, the asymptotic stability is obtained by Lyapunov's direct method, where $V(x)$ is a Lyapunov function candidate. For all $w \in L_2[0, \infty)$, by integrating the derived inequality (15), it can be directly obtained by [24, Theorem 16] that the closed-loop system has an L_2-gain $\leq \gamma$. ■

C. Model-based Learning Algorithm

With stability and robustness results shown in Theorem 1, we are ready to design a model-based learning algorithm to solve the converted inequality (14). Precisely, given a desired attenuation level γ, the inner-loop iteration of the algorithm employs the policy iteration method to derive stabilizing controllers. When the iterative process converges, the converted inequality (14) can be restored by substituting the improved control policy (18) into the Hamiltonian constraint (17b). In outer-loop iteration, the attenuation level is optimized by using the gap of the Hamiltonian constraint. The pseudocode of the developed method is shown in Algorithm 1. In the following, we present further elaborations and theoretical guarantees on the developed algorithm.

Algorithm 1: Model-based Learning Algorithm

Input: initial control policy $u^{(0)}(x)$.

for $i = 1, 2, \cdots$ **do**

  Attenuation Level Optimization:

$$\gamma^{(i)} = \arg\min_{\gamma > 0} \gamma \tag{16a}$$

$$\text{s.t.} \quad \mathcal{L}\Big(V(x), u^{(i-1)}(x), \gamma \Big) \geq 0 \tag{16b}$$

$$V(x) \geq 0. \tag{16c}$$

  Let $V_0^{(i)}(x) \leftarrow V(x)$ and $u_0^{(i)}(x) \leftarrow u^{(i-1)}(x)$.

  for $k = 1, 2, \cdots$ **do**

    Policy Evaluation:

$$V_k^{(i)}(x) = \arg\min_{V} \int_\Omega V(x) dx \tag{17a}$$

$$\text{s.t.} \quad \mathcal{L}\Big(V(x), u_{k-1}^{(i)}(x), \gamma^{(i)} \Big) \geq 0 \tag{17b}$$

$$V_{k-1}^{(i)}(x) - V(x) \geq 0 \tag{17c}$$

$$V(x) \geq 0. \tag{17d}$$

    Policy Improvement:

$$u_k^{(i)}(x) = -\frac{1}{2} R^{-1} g^\top(x) \nabla V_k^{(i)}(x). \tag{18}$$

  end

  Let $u^{(i)}(x) \leftarrow u_\infty^{(i)}(x)$.

end

(Inner-loop) Policy Iteration: For an attenuation level $\gamma^{(i)} \geq \gamma^*$, a stabilizing controller is designed in the inner-loop iteration to allow the closed-loop system to have an L_2-gain smaller than $\gamma^{(i)}$. Enlightened by the existing PI framework [20], the step of policy improvement (18) allows the nonlinear differential term about control input function $g(x)$ to be simplified to a linear term (17b) in the step of policy evaluation (17a). Consider the negative Hamiltonian as

$$\begin{aligned} \mathcal{L}\left(V(x), u(x), \gamma\right) \triangleq & -(\nabla V(x))^\top \big(f(x) + g(x) u(x) + k(x) w \big) \\ & - x^\top Q x - u^\top(x) R u(x) + \gamma^2 w^\top w. \end{aligned} \tag{19}$$

Given an improved controller $u_{k-1}^{(i)}(x)$ at the beginning of the k-th iteration, the value function $V_k^{(i)}(x)$ is updated by imposing a constraint on Hamiltonian (17b) in policy evaluation, which only contains the linear term of the differential of value function. The following lemma illustrates the existence of the initial feasible solution of the PI framework.

Lemma 3 (Feasibility of Hamiltonian Constraint): There exists a controller $u_a(x)$ such that the Hamiltonian constraint $\mathcal{L}\left(V(x), u_a(x), \gamma^{(i)}\right) \geq 0$ has a non-empty feasible set about the value function $V(x)$.

Proof: According to Lemma 2, the value function $V_a(x)$

satisfies the HJ inequality (13). Construct the controller as $u_a(x) = -\frac{1}{2}R^{-1}g^\top(x)\nabla V_a(x)$. It is straightforward that $\mathcal{L}\left(V_a(x), u_a(x), \gamma^{(i)}\right) \geq 0$. So, $V_a(x)$ is a feasible solution to the Hamiltonian constraint $\mathcal{L}\left(V(x), u_a(x), \gamma^{(i)}\right) \geq 0$. ■

Therefore, the value function and control policy can be initialized as $V_0^{(i)}(x) = V_a(x)$ and $u_0^{(i)}(x) = u_a(x)$ such that the first iteration of policy evaluation has a feasible solution. Besides, it can be proved recursively that the subsequent iterations of policy evaluation (17) have a feasible solution.

Theorem 2 (Recursive Feasibility): Consider the PI paradigm with the policy evaluation (17) and the policy improvement (18). If policy evaluation is feasible at the k-th iteration, it will also be feasible at the $(k+1)$-th iteration.

Proof: Assume that for $u_{k-1}^{(i)}(x)$, the Hamiltonian constraint (17b) in policy evaluation has a feasible solution $V_k^{(i)}(x)$, *i.e.*, $\mathcal{L}\left(V_k^{(i)}(x), u_{k-1}^{(i)}(x), \gamma^{(i)}\right) \geq 0$. After updating the control policy $u_k^{(i)}(x)$ at the policy improvement step (18), we have

$$\begin{aligned}
&\mathcal{L}\left(V_k^{(i)}(x), u_k^{(i)}(x), \gamma^{(i)}\right) \\
&= \mathcal{L}\left(V_k^{(i)}(x), u_{k-1}^{(i)}(x), \gamma^{(i)}\right) \\
&\quad + \left(u_k^{(i)}(x) - u_{k-1}^{(i)}(x)\right)^\top R\left(u_k^{(i)}(x) - u_{k-1}^{(i)}(x)\right) \geq 0.
\end{aligned} \tag{20}$$

Therefore, $V_{k+1}^{(i)}(x) = V_k^{(i)}(x)$ is at least a feasible solution of the policy evaluation at the $(k+1)$-th iteration. ■

Corollary 1 (Recursive Stability and Robustness): If the initial control policy makes the Hamiltonian constraint (17b) feasible, the closed-loop system with the control policy (18) at every iteration step is asymptotically stable and has an L_2-gain $\leq \gamma^{(i)}$.

Note that, at every inner-loop iteration, substituting the improved control policy (18) into the Hamiltonian constraint (17b) yields the converted inequality (14). Thus, Corollary 1 can be directly obtained from Theorem 1. The aforementioned analysis establishes the stability and disturbance attenuation performance of the controller during the implementation of the PI process in the inner loop of Algorithm 1. Because the controller generated by each iteration of the inner loop has desired performance, the setting of termination condition has become an open problem.

In order to apply convergence guidance to the value function, the two inequalities (17c) and (17d) are imposed during the policy evaluation step to ensure that the value function is monotonically non-increasing and semi-positive definite, respectively. To formulate an optimization problem in policy evaluation, the integral of value function in interested state space $\Omega \subseteq \mathbb{R}^{2n}$ is selected as the optimization objective (17a), where Ω is a compact set [17].

(Outer-loop) Attenuation Level Optimization: After policy improvement step, the gap hidden in the Hamiltonian constraint (20) implies that it is possible to find a smaller attenuation level $\gamma^{(i)} \leq \gamma^{(i-1)}$ to ensure that

$$\mathcal{L}\left(V_k^{(i-1)}(x), u_k^{(i-1)}(x), \gamma^{(i)}\right) \geq 0.$$

Therefore, the obtained controller $u_k^{(i-1)}(x)$ has the potential to achieve a smaller L_2-gain for the closed-loop system. A smaller attenuation level can be found by solving the sum of squares program (16), which has a non-empty feasible set that contains at least $\gamma^{(i-1)}$. By wrapping the optimization of attenuation level in the outer loop of the PI framework, a numerical method for approximating the solution to the H_∞ optimal control problem is obtained.

Remark 2: In the implementation of the algorithm, the value function is parameterized as

$$V(x) = \sum_{i=1}^{m} c_i \varphi_i(x), \tag{21}$$

where $\{\varphi_i(x)\}_{i=1}^m$ is a set of basis functions, such as polynomials, and $\{c_i\}_{i=1}^m$ are the parameters to be optimized. The attenuation level optimization (16) and the policy evaluation step (17) are constructed as sum of squares programs [17], which can be conveniently solved via SOSTOOLS.

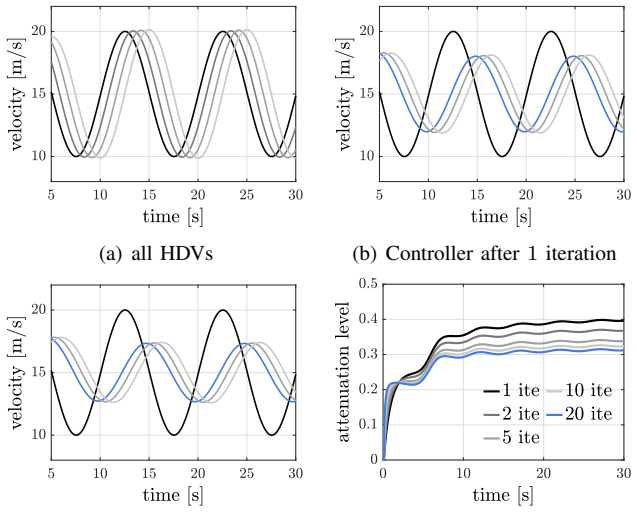
IV. Traffic Simulations

In this section, we present the nonlinear traffic simulations and analyze the performance of the developed model-based learning control policy. The nonlinear OVM model with a typical parameter setup [8] is employed for the HDVs.

For the parameter setup in the controller, the weight coefficients are set as $\theta_s = 0.03$, $\theta_v = 0.15$, $\theta_u = 1$. To conveniently solve sum of squares programs via SOSTOOLS toolbox, the piecewise function (2) is approximated by a quintic polynomial. In Algorithm 1, a quartic polynomial consisting of 144 terms is employed to approximate the value function by (21). The set Ω is in the range of $|s_i| \leq 4$ and $|v_i| \leq 5$. The initial controller is chosen as $u^{(0)} = 0.5s_1 - 1.0v_1$.

In the simulations, a sinusoidal disturbance signal $w(t) = 5\sin(20t/\pi)$ m/s is imposed on the head vehicle. We first consider a small-scale mixed traffic system with three following vehicles, *i.e.*, $n = 3$. As shown in Fig. 2(a), when all the vehicles are HDVs, the velocity oscillations persist during its propagation. For comparison, the velocity perturbations are apparently mitigated by the proposed controller even after one single iteration (see Fig. 2(b)). This result validates the effectiveness of the inner-loop PI paradigm in Algorithm 1: the proposed controller can stabilize the mixed traffic system at each iteration step with attenuation performance guarantees. Further, with more iterations conducted, the attenuation level can be gradually and continuously improved (see Fig. 2(c) for the performance after 20 iterations and Fig. 2(d) for the attenuation level during the simulations after different iterations). These results validate the performance of the outer-loop attenuation level optimization in Algorithm 1.

In addition, we also consider a moderate-scale mixed traffic system with $n = 15$, $m = 5$ to demonstrate the control performance (see Fig. 3). In this case, a quadratic polynomial function consisting of 900 terms is employed to approximate the value function, and after 50 outer-loop iterations, a centralized controller for the 5 CAVs is obtained. It can be clearly observed in Fig. 3 that our method enables



(a) all HDVs (b) Controller after 1 iteration

(c) Controller after 20 iterations (d) Attenuation comparison

Fig. 2. Small-scale simulation results with $n = 3$, $m = 1$. The black, gray and blue profiles represent the velocity of the head vehicle, the HDVs and the CAV, respectively. (a) Simulation results when all the vehicles are HDVs. (b)(c) Simulation results under the learned control policies after 1 or 20 iterations, respectively. (d) The attenuation level γ during the simulations under the controller after different iteration numbers.

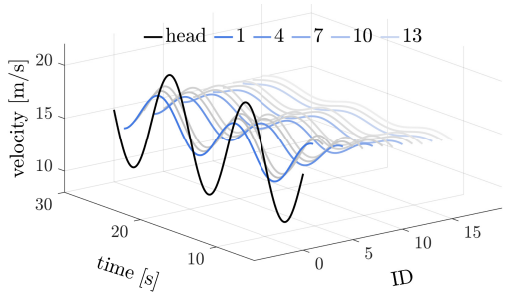


Fig. 3. Moderate-scale simulation results with $n = 15, m = 5$. The meaning of different profiles is consistent with that in Fig. 2.

the CAVs to cooperatively dampen traffic perturbations and smooth traffic flow.

V. Conclusion

This work investigates the optimal robust control problem of nonlinear mixed traffic systems. In order to reduce the influence of external disturbances from the head vehicle on the entire traffic flow, a zero-sum game is first formulated to optimize the worst-case performance. The converted Hamilton-Jacobi inequality is employed to derive robust controllers and reserve space for the optimization of disturbance attenuation performance. A model-based learning algorithm is then presented, combining inner-loop policy iterations and outer-loop attenuation level optimization. Simulation studies verify the effectiveness of the obtained control policy for the CAVs to mitigate traffic waves. Considering possible traffic model mismatches, one future direction is to design similar policy iteration algorithms to address the corresponding robust performance problem. Another interesting topic is to extend the presented method to a model-free learning version, which does not require any priori knowledge of nonlinear mixed traffic dynamics.

References

[1] Y. Sugiyama, M. Fukui, M. Kikuchi *et al.*, "Traffic jams without bottlenecks–experimental evidence for the physical mechanism of the formation of a jam," *New J. Phys.*, vol. 10, no. 3, p. 033001, 2008.

[2] K. Li, S. E. Li, F. Gao *et al.*, "Robust distributed consensus control of uncertain multiagents interacted by eigenvalue-bounded topologies," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3790–3798, 2020.

[3] R. E. Stern, S. Cui, M. L. Delle Monache *et al.*, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transp. Res. Part C Emerging Technol.*, vol. 89, pp. 205–221, 2018.

[4] 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 Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7445–7459, 2020.

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

[6] T. G. Molnár, D. Upadhyay, M. Hopka, M. Van Nieuwstadt, and G. Orosz, "Open and closed loop traffic control by connected automated vehicles," in *CDC*. IEEE, 2020, pp. 239–244.

[7] S. Cui, B. Seibold, R. Stern, and D. B. Work, "Stabilizing traffic flow via a single autonomous vehicle: Possibilities and limitations," in *IV*. IEEE, 2017, pp. 1336–1341.

[8] I. G. Jin and G. Orosz, "Optimal control of connected vehicle systems with communication delay and driver reaction time," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 2056–2070, 2017.

[9] J. Wang, Y. Zheng, C. Chen, Q. Xu, and K. Li, "Leading cruise control in mixed traffic flow: System modeling, controllability, and string stability," *IEEE Trans. Intell. Transp. Syst.*, 2022.

[10] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *Phys. Rev. E*, vol. 51, no. 2, p. 1035, 1995.

[11] C. Wu, A. R. Kreidieh, K. Parvate *et al.*, "Flow: A modular learning framework for mixed autonomy traffic," *IEEE Trans. Rob.*, vol. 38, no. 2, pp. 1270–1286, 2021.

[12] H. Shi, Y. Zhou, K. Wu, X. Wang, Y. Lin, and B. Ran, "Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment," *Transp. Res. Part C Emerging Technol.*, vol. 133, p. 103421, 2021.

[13] J. Wang, Y. Zheng, K. Li, and Q. Xu, "DeeP-LCC: Data-enabled predictive leading cruise control in mixed traffic flow," *IEEE Trans. Control Syst. Technol.*, 2023.

[14] J. Wang, Y. Lian, Y. Jiang *et al.*, "Distributed data-driven predictive control for cooperatively smoothing mixed traffic flow," *Transp. Res. Part C Emerging Technol.*, vol. 155, p. 104274, 2023.

[15] S. Wang, M. Shang, M. W. Levin, and R. Stern, "A general approach to smoothing nonlinear mixed traffic via control of autonomous vehicles," *Transp. Res. Part C Emerging Technol.*, vol. 146, p. 103967, 2023.

[16] M. Abu-Khalaf, F. L. Lewis, and J. Huang, "Policy iterations on the Hamilton–Jacobi–Isaacs equation for H_∞ state feedback control with input saturation," *IEEE Trans. Autom. Control*, vol. 51, no. 12, pp. 1989–1995, 2006.

[17] Y. Zhu, D. Zhao, X. Yang, and Q. Zhang, "Policy iteration for H_∞ optimal control of polynomial nonlinear systems via sum of squares programming," *IEEE Trans. Cyber.*, vol. 48, no. 2, pp. 500–509, 2018.

[18] L. Kong, W. He, C. Yang, and C. Sun, "Robust neurooptimal control for a robot via adaptive dynamic programming," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 6, pp. 2584–2594, 2020.

[19] Y. Zhu, D. Zhao, and Z. Zhong, "Adaptive optimal control of heterogeneous CACC system with uncertain dynamics," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 4, pp. 1772–1779, 2018.

[20] S. E. Li, *Reinforcement learning for sequential decision and optimal control*. Springer, 2022.

[21] 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 Trans. Cyber.*, vol. 52, no. 6, pp. 5267–5277, 2020.

[22] T. Başar and P. Bernhard, *H_∞ optimal control and related minimax design problems: A dynamic game approach*. Springer Science & Business Media, 2008.

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

[24] A. J. Van Der Schaft, "L_2-gain analysis of nonlinear systems and nonlinear state feedback H_∞ control," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 770–784, 1992.