2021 61st IEEE Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

A Cooperative Optimal Control Framework for Connected and Automated Vehicles in Mixed Traffic Using Social Value Orientation

Viet-Anh Le, *IEEE Student Member*, Andreas A. Malikopoulos, *IEEE Senior Member*

***Abstract*— In this paper, we develop a socially cooperative optimal control framework to address the motion planning problem for connected and automated vehicles (CAVs) in mixed traffic using social value orientation (SVO) and a potential game approach. In the proposed framework, we formulate the interaction between a CAV and a human-driven vehicle (HDV) as a simultaneous game where each vehicle minimizes a weighted sum of its egoistic objective and a cooperative objective. The SVO angles are used to quantify preferences of the vehicles toward the egoistic and cooperative objectives. Using the potential game approach, we propose a single objective function for the optimal control problem whose weighting factors are chosen based on the SVOs of the vehicles. We prove that a Nash equilibrium can be obtained by minimizing the proposed objective function. To estimate the SVO angle of the HDV, we develop a moving horizon estimation algorithm based on maximum entropy inverse reinforcement learning. The effectiveness of the proposed approach is demonstrated by numerical simulations of a vehicle merging scenario.**

I. Introduction

Coordination of connected and automated vehicles (CAVs) has received significant attention over the last years (see [1]–[3] for surveys). In recent work [4]–[6], we addressed optimal coordination of CAVs at different traffic scenarios and last-mile delivery applications, and showed potential benefits in reducing traffic congestion, travel time, and energy consumption. These approaches have considered 100% CAV penetration rate, which, however, might not be expected by 2060 [7]. Therefore, it is highly necessary to investigate motion planning and control strategies for CAVs that perform efficiently in mixed traffic environments.

A number of research articles have recently focused on developing different control methods for CAVs in mixed traffic scenarios, e.g., optimal control [8], model predictive control [9], game-theoretic control [10], or learning-based control [11]. Other research efforts [12]–[17] have utilized *social value orientation* (SVO), a concept from psychology that quantifies the level of an agent's selfishness or altruism, in control development to predict how human drivers will interact and cooperate with others. The first attempt taking social factors of human drivers into consideration via SVO was made in [12]. The authors modeled interactions between vehicles as a best-response game in which each agent maximizes its individual utility given optimal actions of other agents, and utilized SVO to better predict human driving intention. A centralized coordination algorithm was developed by Buckman *et al*. [13] to perform socially-compliant navigation at an intersection given the social preferences of the vehicles. Ozkan and Ma [14] investigated the impacts of a socially compatible control design in a car-following scenario through four different SVO levels for an automated vehicle. Larsson *et al*. [15] developed a pro-social model predictive control algorithm in which SVO was used to derive weighting strategies. Toghi *et al*. [16] proposed a cooperative sympathetic reward structure using SVO for multi-agent reinforcement learning in autonomous driving. Crosato *et al*. [17] considered a scenario with pedestrian crossing and utilized SVO to design a reinforcement learning reward function for an automated vehicle to avoid collision with the pedestrian.

Inspired by using SVO to understand the driving preferences of human drivers, in this paper, we develop a *socially cooperative optimal control framework* for a CAV while interacting with a HDV. Contrary to other efforts that have used the Stackelberg game with a leader-follower structure [12], [14], [16], which may require much computation for finding an equilibrium, our approach models the interaction between the CAV and the HDV as a *simultaneous game*. In the imposed game, the agents simultaneously take control actions to minimize their objective functions that are given by a weighted sum of their individual (egoistic) objective and a shared (cooperative) one. The SVO angles are used to quantify how the agent weights its individual objective against the shared objective. Based on the idea of *potential games* [18], we then propose an objective function for the control framework which is used to obtain a Nash equilibrium of the imposed game. The framework is implemented in a *receding horizon control* (or model predictive control) manner for robustness against stochastic human driving behavior. To estimate the SVO angle of the HDV, we employ *moving horizon estimation* which uses real-time data from both vehicles combined with the *maximum entropy inverse reinforcement learning* (IRL) technique. Based on the estimated SVO of the HDV, the SVO of the CAV is chosen to compensate for the level of altruism in CAV–HDV coordination. Finally, the proposed framework is illustrated and validated by a numerical example of a vehicle merging scenario.

The remainder of this paper is organized as follows. In Section II, we present the proposed socially cooperative optimal control framework for a CAV while interacting with a HDV, and introduce an illustrative example at a merging. In Section III, we provide the method for estimating the SVO angle of the HDV, and in Section IV, we demonstrate the performance of the proposed framework by numerical sim-

The authors are with the Department of Mechanical Engineering, University of Delaware, Newark, DE 19716 USA. E-mail: vietale@udel.edu, andreas@udel.edu.

978-1-6654-6761-2/22/$31.00 ©2022 IEEE

ulations. Finally, we draw concluding remarks and discuss potential directions for future research in Section V.

II. Control Framework for Connected and Automated Vehicles in Mixed Traffic

In this section, we first present a socially cooperative optimal control framework for a CAV while interacting with a HDV in mixed traffic, and illustrate it by a vehicle merging example.

A. Problem Formulation and Control Framework

We consider an interactive driving scenario that includes a CAV and a HDV indexed by 1 and 2, respectively. To facilitate connectivity, we assume that a coordinator is available to collect real-time trajectories of HDV–2 and transmit them to CAV–1 while also storing necessary information, e.g., physical parameters of the traffic scenario. We also consider that there is no error or delay during the communication between the vehicles and the coordinator.

The dynamics of the vehicles are described by discrete-time models. Let $\mathbf{x}_{i,k}$ and $\mathbf{u}_{i,k}$, $i = 1, 2$, be the vectors of states and control actions, respectively, at time $k \in \mathbb{N}$. Then the model of vehicle i is given by

$$\mathbf{x}_{i,k+1} = \mathbf{f}_i(\mathbf{x}_{i,k}, \mathbf{u}_{i,k}). \tag{1}$$

The goal is to develop a cooperative optimal control framework for motion planning of CAV–1 which considers the driving intention of HDV–2. In previous research efforts [12], [14], [16], CAV–HDV interaction was modeled as a non-cooperative Stackelberg game, in which a leader makes a decision, then a follower makes its optimal decision with respect to the leader's decision. Generally, computing an Stackelberg equilibrium in CAV–HDV interactions can be computationally expensive for real-time optimization [14]. In addition, the Stackelberg game might not ideally reflect CAV–HDV interaction since determining who should be the leader and the follower is not explicit in many traffic scenarios [12].

To overcome these issues, we consider a non-cooperative simultaneous game and derive a single optimal control problem that yields the Nash equilibrium of the game. In the imposed game, CAV–1 and HDV–2 take control actions at the same time to minimize their objective functions which include an individual (egoistic) term and a shared (cooperative) term. The egoistic term denotes the effort of each vehicle to achieve its own driving goal represented by functions of its states and actions, whereas the cooperative term is defined as the effort to achieve a common target involving the states and actions of both CAV–1 and HDV–2. Let $l_1\big(\mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}\big)$ and $l_2\big(\mathbf{x}_{2,k+1}, \mathbf{u}_{2,k}\big)$ be the egoistic terms of the objective functions of CAV–1 and HDV–2, respectively, at time k. Let $l_{12}\big(\mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}, \mathbf{x}_{2,k+1}, \mathbf{u}_{2,k}\big)$ be the cooperative term of their objective function at time k. Note that l_1, l_2, and l_{12} are composed appropriately through some weights to avoid one dominating the other.

The preferences of the vehicles toward the egoistic and cooperative terms can be described by social value orientation, a commonly used concept in psychology that has been recently employed in autonomous driving research [12]–[17]. Let ϕ_1 and ϕ_2 be the SVO angles for CAV–1 and HDV–2, respectively. The SVO angle is usually constrained between 0 and $\frac{\pi}{2}$. If the angle is equal to 0, it means that the vehicle is purely egoistic, i.e., it makes decisions that only benefit its own objective. In contrast, if the angle is equal to $\frac{\pi}{2}$ it implies that this is a purely altruistic vehicle, i.e., it optimizes the cooperative objective without concerning its own objective. Since in reality, most drivers have at least a minimal level of egoism and altruism, we consider that $0 < \phi_1, \phi_2 < \frac{\pi}{2}$.

The objective of each vehicle in the imposed game can be formed as a weighted sum of the egoistic and cooperative terms in which the weights are determined by trigonometric functions of the SVO angles. In particular, given the SVO angles ϕ_1 and ϕ_2 the objective function of CAV–1 and HDV–2 in the game is

$$\begin{aligned}\bar{l}_1(\mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}, \mathbf{x}_{2,k+1}, \mathbf{u}_{2,k}) = l_1(\mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}) \cos\phi_1 \\ + l_{12}(\mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}, \mathbf{x}_{2,k+1}, \mathbf{u}_{2,k}) \sin\phi_1,\end{aligned} \tag{2}$$

and

$$\begin{aligned}\bar{l}_2(\mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}, \mathbf{x}_{2,k+1}, \mathbf{u}_{2,k}) = l_2(\mathbf{x}_{2,k+1}, \mathbf{u}_{2,k}) \cos\phi_2 \\ + l_{12}(\mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}, \mathbf{x}_{2,k+1}, \mathbf{u}_{2,k}) \sin\phi_2,\end{aligned} \tag{3}$$

respectively.

In what follows, to simplify the notation we drop the time index k whenever this does not cause any ambiguity.

To find a Nash equilibrium for the simultaneous game described above, we propose an objective function recasting the game as a potential game [18]. In potential games, a Nash equilibrium at which all players minimize their own objective functions can be obtained by alternatively minimizing a single function called the potential function. For the game between CAV–1 and HDV–2 where their objectives are given in (2) and (3), the potential objective function can be derived as follows

$$\begin{aligned}l(\mathbf{x}_1, \mathbf{u}_1, \mathbf{x}_2, \mathbf{u}_2) = l_1(\mathbf{x}_1, \mathbf{u}_1) \cos\phi_1 \sin\phi_2 \\ + l_2(\mathbf{x}_2, \mathbf{u}_2) \sin\phi_1 \cos\phi_2 \\ + l_{12}(\mathbf{x}_1, \mathbf{u}_1, \mathbf{x}_2, \mathbf{u}_2) \sin\phi_1 \sin\phi_2.\end{aligned} \tag{4}$$

Note that in (4) the weights for the objective functions l_1, l_2, and l_{12} are determined by a weighting strategy using trigonometric functions of the SVO angles ϕ_1 and ϕ_2. In the following theorem, we prove that a Nash equilibrium can be obtained by minimizing (4).

Theorem 1: Consider the simultaneous game between CAV–1 and HDV–2 whose objective functions are given by (2) and (3). Then the minimum of the proposed objective function (4) is a Nash equilibrium of the game.

Proof: Given the vehicle dynamics in (1), the functions l, l_1, l_2, and l_{12} can be expressed as the functions of control actions $\mathbf{u}_1$ and $\mathbf{u}_2$. Hence, we use $l(\mathbf{u}_1, \mathbf{u}_2)$, $l_1(\mathbf{u}_1)$, $l_2(\mathbf{u}_2)$, and $l_{12}(\mathbf{u}_1, \mathbf{u}_2)$ for brevity. Denote $\mathbf{u}_1^*$ and $\mathbf{u}_2^*$ as the control actions that minimizing the objective function (4), i.e.,

$$l(\mathbf{u}_1^*, \mathbf{u}_2^*) \leq l(\mathbf{u}_1, \mathbf{u}_2^*), \forall \mathbf{u}_1, \tag{5}$$

which, from (4), is equivalent to

$$\begin{aligned} &l_1(\mathbf{u}_1^*)\cos\phi_1\sin\phi_2 + l_2(\mathbf{u}_2^*)\sin\phi_1\cos\phi_2 \\ &\quad + l_{12}(\mathbf{u}_1^*,\mathbf{u}_2^*)\sin\phi_1\sin\phi_2 \\ &\le l_1(\mathbf{u}_1)\cos\phi_1\sin\phi_2 + l_2(\mathbf{u}_2^*)\sin\phi_1\cos\phi_2 \\ &\quad + l_{12}(\mathbf{u}_1,\mathbf{u}_2^*)\sin\phi_1\sin\phi_2,\ \forall \mathbf{u}_1. \end{aligned} \tag{6}$$

Since $\phi_2 > 0$, from (6) we obtain

$$\begin{aligned} &l_1(\mathbf{u}_1^*)\cos\phi_1 + l_{12}(\mathbf{u}_1^*,\mathbf{u}_2^*)\sin\phi_1 \\ &\le l_1(\mathbf{u}_1)\cos\phi_1 + l_{12}(\mathbf{u}_1,\mathbf{u}_2^*)\sin\phi_1,\ \forall \mathbf{u}_1, \end{aligned} \tag{7}$$

or

$$\mathbf{u}_1^* = \arg\min_{\mathbf{u}_1} l_1(\mathbf{u}_1)\cos\phi_1 + l_{12}(\mathbf{u}_1,\mathbf{u}_2^*)\sin\phi_1. \tag{8}$$

Similarly, we have

$$\mathbf{u}_2^* = \arg\min_{\mathbf{u}_2} l_2(\mathbf{u}_2)\cos\phi_2 + l_{12}(\mathbf{u}_1^*,\mathbf{u}_2)\sin\phi_2. \tag{9}$$

From (8) and (9), the result follows. ■

In our socially cooperative optimal control framework for motion planning of CAV–1, we formulate a finite-time optimal control problem over a control horizon of length $H \in \mathbb{N}\setminus\{0\}$. For simplicity, we consider constant SVO angles over the current control horizon at each time step. Let t be the current time step and $\mathcal{I}_t = \{t,\dots,t+H-1\}$ be the set of all time steps in the control horizon at time step t. The finite-horizon optimal control problem for CAV–1 is given by

$$\underset{\{\mathbf{u}_{1,k},\mathbf{u}_{2,k}\}_{k\in\mathcal{I}_t}}{\text{minimize}} \quad \sum_{k\in\mathcal{I}_t} l(\mathbf{x}_{1,k+1},\mathbf{u}_{1,k},\mathbf{x}_{2,k+1},\mathbf{u}_{2,k}) \tag{10a}$$

subject to:

$$(1),\ \ i=1,2, \tag{10b}$$

$$g_j(\mathbf{x}_{1,k+1},\mathbf{u}_{1,k},\mathbf{x}_{2,k+1},\mathbf{u}_{2,k}) \le 0,\ \forall j \in \mathcal{J}_{\text{ieq}}, \tag{10c}$$

$$h_j(\mathbf{x}_{1,k+1},\mathbf{u}_{1,k},\mathbf{x}_{2,k+1},\mathbf{u}_{2,k}) = 0,\ \forall j \in \mathcal{J}_{\text{eq}}, \tag{10d}$$

where (10b)–(10d) hold for all $k \in \mathcal{I}_t$. The constraints (10c) and (10d) are inequality and equality constraints for system states and inputs with $\mathcal{J}_{\text{ieq}}$ and $\mathcal{J}_{\text{eq}}$ are sets of inequality and equality constraint indices, respectively.

Remark 1: The optimization variables of (10) consist of not only variables of CAV–1 but also of HDV–2 so that the proposed control framework can predict HDV–2's states over the next control horizon given the control objective that best describe driving behavior of the human driver. The individual objective function of HDV–2 can be recovered from historical human driving data through offline inverse reinforcement learning [19], [20]. If historical data are not available, the objective function can be predefined to represent rational human objectives in specific driving scenarios.

The optimal control problem (10) is implemented in a receding horizon control manner at every time step. By estimating the SVO angle ϕ_2 online, e.g., using the method presented in Section III, we can inspect the level of altruism of the human driver then adapt the SVO angle of CAV–1 accordingly. For example, in this paper, the SVO angle ϕ_1 of CAV–1 is chosen as

$$\phi_1 = \frac{\pi}{2} - \phi_2. \tag{11}$$

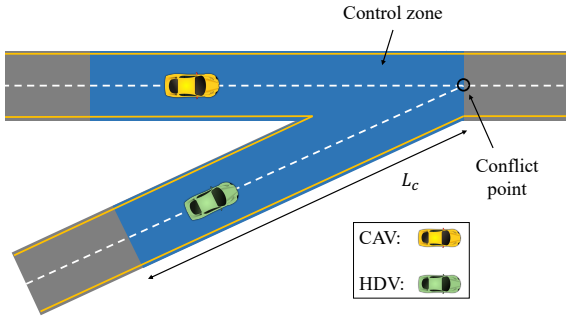


Fig. 1: A merging scenario with a CAV and a HDV.

The justification for (11) is that if HDV–2 is more egoistic then CAV–1 needs to be more altruistic, and vice versa.

B. Illustrative Example

To illustrate the presented control framework, we consider an example of a vehicle merging scenario illustrated in Fig. 1. The area that a lateral collision can occur is called a *conflict point*. We consider a *control zone* that is constituted by areas of length $L_c \in \mathbb{R}^+$ upstream of the conflict point in each road (Fig. 1). Inside the control zone, the vehicles can communicate with a coordinator. Note that CAV–1 is controlled by the proposed cooperative control framework only within the control zone, whereas outside the control zone, CAV–1 can use a speed profile given by a car-following model.

We consider that the dynamics of the vehicles in this example can be described by the following double-integrator longitudinal dynamics

$$\begin{aligned} p_{i,k+1} &= p_{i,k} + \Delta T v_{i,k} + \frac{1}{2}\Delta T^2 a_{i,k}, \\ v_{i,k+1} &= v_{i,k} + \Delta T a_{i,k}, \end{aligned} \tag{12}$$

where $\Delta T \in \mathbb{R}^+$ is the sampling time, $p_{i,k}$ is the longitudinal position of the vehicle with respect to the conflict point at time k, and $v_{i,k}$ and $a_{i,k}$ are the speed and acceleration of the vehicle i at time k, respectively. The state and control input of vehicle i are denoted by $\mathbf{x}_{i,k} = [p_{i,k}, v_{i,k}]^\top$ and $u_{i,k} = a_{i,k}$, respectively.

In this example, the egoistic term for each vehicle includes minimizing both the control input and the deviation from the maximum allowed speed to cross the merging zone, i.e.,

$$l_1(\mathbf{x}_{1,k+1}, u_{1,k}) = w_1 a_{1,k}^2 + w_2(v_{1,k+1} - v_{\max})^2, \tag{13}$$

and

$$l_2(\mathbf{x}_{2,k+1}, u_{2,k}) = w_3 a_{2,k}^2 + w_4(v_{2,k+1} - v_{\max})^2, \tag{14}$$

where w_1, w_2, w_3, and w_4 are positive normalized weights. In this example, the cooperative term takes the form of a penalty function corresponding to a collision avoidance constraint [21] as follows

$$l_{12}(\mathbf{x}_{1,k+1}, u_{1,k}, \mathbf{x}_{2,k+1}, u_{2,k}) = \frac{w_5}{p_{1,k+1}^2 + p_{2,k+1}^2 - r^2}, \tag{15}$$

where w_5 is a positive normalized weight and $r \in \mathbb{R}^+$ is a safety threshold. Intuitively, in this merging scenario, egoistic

human drivers accelerate to cross the merging zone quickly, while altruistic drivers slow down and let the other vehicle cross the merging zone prior to them.

Next, we impose the following state and control constraints for CAV–1,

$$v_{\min} \leq v_{1,k+1} \leq v_{\max}, \quad u_{\min} \leq a_{1,k} \leq u_{\max}, \tag{16}$$

for all $k \in \mathcal{I}_t$, where $u_{\min}$, $u_{\max}$ are the minimum deceleration and maximum acceleration, respectively, and $v_{\min}$, $v_{\max}$ are the minimum and maximum speed limits, respectively. Note that we do not consider state and input constraints for HDV–2 since those constraints can be violated by the behavior of human drivers.

The receding horizon control problem for CAV–1 in this example is thus formulated as follows

$$\underset{\{u_{1,k},u_{2,k}\}_{k\in\mathcal{I}_t}}{\text{minimize}} \quad \sum_{k\in\mathcal{I}_t} l(\mathbf{x}_{1,k+1}, u_{1,k}, \mathbf{x}_{2,k+1}, u_{2,k}) \tag{17a}$$

subject to:

$$(12),\ \forall k \in \mathcal{I}_t,\ i = 1, 2, \tag{17b}$$

$$(16),\ \forall k \in \mathcal{I}_t. \tag{17c}$$

III. Inverse Reinforcement Learning-Based Estimation for Social Value Orientation

In this section, we present a moving horizon estimation method for the SVO angle ϕ_2 of HDV–2 using the maximum entropy inverse reinforcement learning.

A. Inverse Reinforcement Learning

Inverse reinforcement learning (IRL) is a machine learning technique developed to learn the underlying objective or reward of an agent by observing its behavior [22]. Using IRL, we can estimate the SVO angle of HDV–2 that best fits predicted trajectories to actual trajectories. Recall that in Section II-A the control input $\mathbf{u}^*_{2,k}$ of HDV–2 at each time step k is the solution of the following problem

$$\mathbf{u}^*_{2,k} = \underset{\mathbf{u}_{2,k}}{\arg\min}\ \bar{l}_2 = l_2 \cos\phi_2 + l_{12} \sin\phi_2, \tag{18}$$

subject to the imposed state, control, and safety constraints.

We apply the feature-based IRL approach [19] in (18), where we let $\mathbf{f} = [l_2, l_{12}]^\top$ and $\boldsymbol{\theta}(\phi_2) = [\cos\phi_2, \sin\phi_2]^\top$ be the vector of the features and the vector of weights, respectively. Let $\mathcal{R}$ be the set of sample trajectory segments used in IRL and $\mathbf{r}_j \in \mathcal{R}$ be the j-th element in $\mathcal{R}$. The goal is to find the best possible value for ϕ_2 so that expected feature values can match observed feature values, i.e., $\mathbb{E}_p[\mathbf{f}] = \tilde{\mathbf{f}}$, where $\tilde{\mathbf{f}}$ is the vector of average observed feature values, and $\mathbb{E}_p[\mathbf{f}]$ denotes the expected feature values given a probability distribution over trajectories p. In general, there are many such probability distributions. In this paper, we choose the maximum entropy IRL [22] that utilizes an exponential family distribution and maximizes the entropy of the distribution, yielding the following optimization problem

$$\begin{aligned} &\underset{\phi_2}{\text{maximize}} \sum_{\mathbf{r}_j\in\mathcal{R}} \log\, p\big(\mathbf{r}_j \,|\, \phi_2\big) \\ &\text{subject to:} \quad 0 < \phi_2 < \frac{\pi}{2}, \end{aligned} \tag{19}$$

where

$$p\big(\mathbf{r}_j \,|\, \phi_2\big) = \frac{\exp\big(-\boldsymbol{\theta}^\top(\phi_2)\mathbf{f}(\mathbf{r}_j)\big)}{Z\big(\boldsymbol{\theta}(\phi_2)\big)}, \tag{20}$$

and $Z\big(\boldsymbol{\theta}(\phi_2)\big)$ is the partition function [22]. The probability distribution in (20) implies that the agent exponentially chooses the trajectory with a lower objective. The constraint in (22) is relaxed by parameterizing ϕ_2 with the sigmoid function, i.e.,

$$\phi_2 = \frac{\pi}{2}\sigma(\psi) \tag{21}$$

where $\sigma(\psi) = \frac{1}{1+\exp(-\psi)}$ and $\psi \in \mathbb{R}$, leading to the following unconstrained optimization problem

$$\underset{\psi}{\text{maximize}} \quad \mathcal{L}(\psi) = \sum_{\mathbf{r}_j\in\mathcal{R}} \log\, p\big(\mathbf{r}_j \,|\, \psi\big) \tag{22}$$

It is not possible to solve the optimization problem (22) analytically, but the gradient $\nabla\mathcal{L}_\psi$ of the objective function in (20) with respect to ψ can be computed using the chain rule as follows

$$\nabla\mathcal{L}_\psi = \nabla\mathcal{L}_{\boldsymbol{\theta}}^\top \begin{bmatrix} -\sin\phi_2 \\ \cos\phi_2 \end{bmatrix} \frac{\pi}{2}\sigma(\psi)\big(1-\sigma(\psi)\big), \tag{23}$$

where $\nabla\mathcal{L}_{\boldsymbol{\theta}}$ is the gradient of the probability distribution (20) with respect to the vector of weights $\boldsymbol{\theta}$. It can be shown that this gradient is the difference between the expected and the empirical feature values [22]

$$\nabla\mathcal{L}_{\boldsymbol{\theta}} = \tilde{\mathbf{f}} - \mathbb{E}_p[\mathbf{f}]. \tag{24}$$

The average observed feature values $\tilde{\mathbf{f}}$ in (24) can be computed from an average of feature values for all training samples. Meanwhile, it is generally impossible to exactly compute $\mathbb{E}_p[\mathbf{f}]$. In [23], an approximation of the expected feature values was proposed to compute the feature values of the most likely trajectories, instead of computing expectations by sampling, as follows

$$\mathbb{E}_p[\mathbf{f}] \approx \mathbf{f}\big(\arg\max_{\mathbf{r}}\ \log\, p(\mathbf{r} \,|\, \psi)\big). \tag{25}$$

Using (25), the gradient of the objective function in (22) with respect to ψ can be computed and used to update the estimation of ϕ_2. More details on maximum entropy IRL can be found in [22].

B. Moving Horizon Estimation

We employ the IRL approach presented above in a moving horizon estimation manner to estimate the SVO angle ϕ_2. Let $L \in \mathbb{N} \setminus \{0\}$ be the length of the estimation horizon and t be the current time step. To simplify the notation, let $\mathbf{r}_j = (\mathbf{x}_{1,t-j}, \mathbf{x}_{2,t-j}, \mathbf{x}_{1,t-j+1}, \mathbf{x}_{1,t-j+1}, \mathbf{u}_{1,t-j}, \mathbf{u}_{2,t-j})$, for $j = 1, \ldots, L$, be the tuples representing trajectory segments collected over the estimation horizon. In other words, at each time step, we utilize the L most recent sample trajectories to update ϕ_2. If the number of existing trajectory segments is less than the length of estimation horizon, CAV–1 utilizes all existing trajectory segments for the estimation.

Given L sample trajectories over the estimation horizon, the IRL procedure for estimating ϕ_2 can be detailed as follows. At each time step, we initialize ϕ_2 with the value

Algorithm 1 IRL-based MHE for SVO

Require: $L \in \mathbb{N} \setminus \{0\}$, $\eta \in \mathbb{R}^+$

Initialize ψ or re-use the previous estimate
Collect sample trajectories $\mathbf{r}_j$, $j = 1, \dots, L$ over the estimation horizon
for $j = 1, \dots, L$ **do**
    Compute $\mathbf{f}(\mathbf{r}_j)$ for the sample trajectory $\mathbf{r}_j$
Compute the empirical feature vector $\tilde{\mathbf{f}}$ by (26)
for $j = 1, \dots, L$ **do**
    Find the optimized trajectory $\mathbf{r}_j^{\phi_2}$ with respect to ϕ_2.
    Compute $\mathbf{f}(\mathbf{r}_j^{\phi_2})$ for the optimized trajectory $\mathbf{r}_j^{\phi_2}$
Compute approximated expected feature vector $\tilde{\mathbb{E}}_p[\mathbf{f}]$ by (27)
Update ψ by (28) and ϕ_2 by (21)
return ϕ_2

computed at the previous time step, or with an arbitrary value between 0 and $\frac{\pi}{2}$ if at the first time step. Given an initialization of ϕ_2, we evaluate the features for all training samples and compute the empirical feature vector averaged over all samples by

$$\tilde{\mathbf{f}} = \frac{1}{L} \sum_{j=1}^{L} \mathbf{f}(\mathbf{r}_j). \tag{26}$$

Next, for j-th sample trajectory, we fix ϕ_2, the trajectory $\{\mathbf{x}_{1,k}, \mathbf{x}_{1,k+1}, \mathbf{u}_{1,k}\}$ of CAV–1, and the initial condition $\mathbf{x}_{2,k}$, then find the optimized control actions of HDV–2 $\mathbf{u}_{2,k}$ that minimize $\boldsymbol{\theta}^\top(\phi_2)\mathbf{f}(\mathbf{r}_j)$. We denote the system trajectories resulted from the optimized HDV–2's actions as $\{\mathbf{r}_1^{\phi_2}, \dots, \mathbf{r}_L^{\phi_2}\}$. Next, we evaluate the features for all optimized trajectories and compute the approximated expected feature values $\tilde{\mathbb{E}}_p[\mathbf{f}]$ by

$$\tilde{\mathbb{E}}_p[\mathbf{f}] = \frac{1}{L} \sum_{j=1}^{L} \mathbf{f}(\mathbf{r}_j^{\phi_2}). \tag{27}$$

Using the empirical and the expected feature values in (26) and (27), we can compute the gradient of the objective function in (22) with respect to ϕ_2 by (23) and (24), which can be used to update ϕ_2 by gradient ascent method as follows

$$\psi \leftarrow \psi + \eta \nabla \mathcal{L}_\psi, \tag{28}$$

where $\eta \in \mathbb{R}^+$ is the learning rate.

The IRL-based MHE for the SVO angle ϕ_2 at each time step is also summarized in Algorithm 1.

Remark 2: In Algorithm 1, at each time step, the estimation of ϕ_2 is updated once to limit the execution time of the algorithm in real-time applications. However, the algorithm can be repeated multiple times or until convergence.

IV. Simulations

In this section, we show simulation results for the vehicle merging example presented in Section II-B to demonstrate the performance of the proposed framework.

TABLE I: Parameters of the simulations

Parameters	Value	Parameters	Value
w_1	1	w_2	5
w_3	1	w_4	5
w_5	10^7	ΔT	0.1 s
$v_{\min}$	0 m/s	$v_{\max}$	30 m/s
$u_{\min}$	$-10\,\mathrm{m/s^2}$	$u_{\max}$	$5\,\mathrm{m/s^2}$
H	20	L	20
η	1.0		

A. Simulation Setup

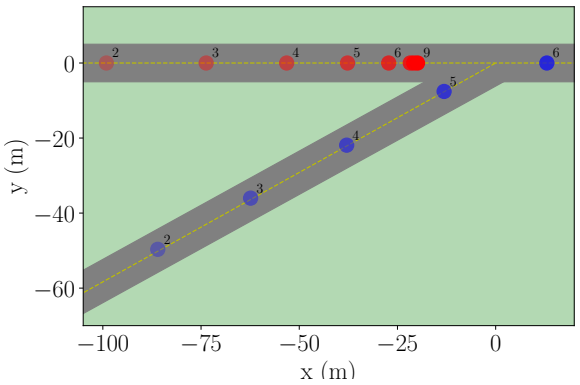
We simulate the behavior of HDV–2 by imitating driving behavior of a realistic human driver from Next Generation Simulation (NGSIM) dataset [24]. Using the IRL technique, we learn an offline machine learning model to capture human driving behavior at a highway on-ramp merging. More details on imitating human driving behavior via IRL can be found in [19], [20]. The learned human driving model is used to generate actions of HDV–2 in the control zone of length 120 m, while outside the control zone the speed of HDV–2 is assumed to be constant. By changing the weights of the obtained objective functions in the IRL human driving model, we can replicate behavior of more egoistic or altruistic human drivers to validate the methods with different human driving styles. The weights of the objective functions and other parameters of the RHC and MHE are given in Table I. We used Python for the implementation and CasADi [25] along with the built-in IPOPT solver for formulating and solving the optimization problems, respectively.

B. Results and Discussions

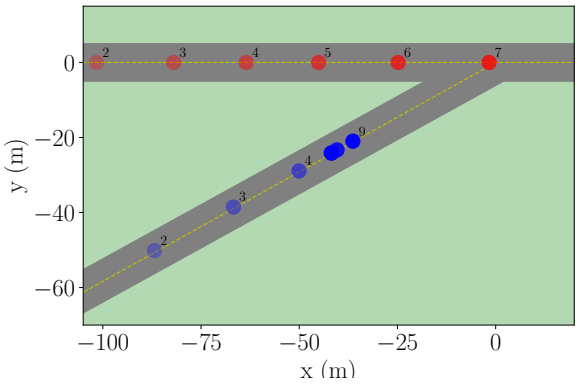
We evaluate the performance of the proposed method in two simulation scenarios where CAV–1 correspondingly deals with egoistic and altruistic drivers. Figure 2a shows a snapshot of the merging scenario with positions of the vehicles at every 1 second for the simulation with the egoistic driver. As can be seen from the figure, the egoistic HDV (blue dot) accelerates while approaching the merging zone, thus CAV–1 (red dot) slows down to let HDV–2 cross the merging zone first. The results for the second simulation, in which the human driver is more altruistic, are shown in Figure 2b. In this case, HDV–2 decelerates and then stops when moving close to the conflict point to give way to CAV–1. By noticing that the human driver is highly cooperative, CAV–1 then speeds up to pass the merging zone prior to HDV–2. Using the proposed socially cooperative optimal control framework, CAV–1 behaves differently depending on human driving preferences. More simulation results can be found at the supplemental site, https://sites.google.com/view/ud-ids-lab/socially-cooperative-control.

V. Conclusions

In this paper, we presented a socially cooperative optimal control framework for a CAV to efficiently and safely interact with a HDV in mixed traffic. Using the SVO angles of the vehicles, we synthesized an objective function that allows the CAV to compensate appropriately given the level of cooperation of the human driver. The SVO of the HDV was



(a) Simulation with an egoistic human driver



(b) Simulation with an altruistic human driver

Fig. 2: Snapshots of the merging scenario with vehicle positions at every 1 second in two simulations. The CAV and the HDV are denoted by red and blue dots, respectively.

estimated online using inverse reinforcement learning-based moving horizon estimation. The performance of the proposed method was validated by a vehicle merging example. Future work should focus on extending this framework to consider the interactions between multiple CAVs and HDVs. Another direction for future research should investigate the practical effectiveness of the proposed approach using experiments in a scaled robotic environment [26].

REFERENCES

[1] L. Zhao and A. A. Malikopoulos, "Enhanced mobility with connectivity and automation: A review of shared autonomous vehicle systems," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 1, pp. 87–102, 2022.

[2] J. Guanetti, Y. Kim, and F. Borrelli, "Control of connected and automated vehicles: State of the art and future challenges," *Annual reviews in control*, vol. 45, pp. 18–40, 2018.

[3] T. Ersal, I. Kolmanovsky, N. Masoud, N. Ozay, J. Scruggs, R. Vasudevan, and G. Orosz, "Connected and automated road vehicles: state of the art and future challenges," *Vehicle system dynamics*, vol. 58, no. 5, pp. 672–704, 2020.

[4] A. I. Mahbub, A. A. Malikopoulos, and L. Zhao, "Decentralized optimal coordination of connected and automated vehicles for multiple traffic scenarios," *Automatica*, vol. 117, no. 108958, 2020.

[5] B. Remer and A. A. Malikopoulos, "The multi-objective dynamic traveling salesman problem: Last mile delivery with unmanned aerial vehicles assistance," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 5304–5309.

[6] A. M. I. Mahbub and A. A. Malikopoulos, "Conditions to Provable System-Wide Optimal Coordination of Connected and Automated Vehicles," *Automatica*, vol. 131, no. 109751, 2021.

[7] A. Alessandrini, A. Campagna, P. Delle Site, F. Filippi, and L. Persia, "Automated vehicles and the rethinking of mobility and cities," *Transportation Research Procedia*, vol. 5, pp. 145–160, 2015.

[8] I. G. Jin and G. Orosz, "Connected cruise control among human-driven vehicles: Experiment-based parameter estimation and optimal control design," *Transportation research part C: emerging technologies*, vol. 95, pp. 445–459, 2018.

[9] S. Feng, Z. Song, Z. Li, Y. Zhang, and L. Li, "Robust platoon control in mixed traffic flow based on tube model predictive control," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 4, pp. 711–722, 2021.

[10] R. Chandra and D. Manocha, "Gameplan: Game-theoretic multi-agent planning with human drivers at intersections, roundabouts, and merging," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2676–2683, 2022.

[11] S. Bae, D. Saxena, A. Nakhaei, C. Choi, K. Fujimura, and S. Moura, "Cooperation-aware lane change maneuver in dense traffic based on model predictive control with recurrent neural network," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 1209–1216.

[12] W. Schwarting, A. Pierson, J. Alonso-Mora, S. Karaman, and D. Rus, "Social behavior for autonomous vehicles," *Proceedings of the National Academy of Sciences*, vol. 116, no. 50, pp. 24 972–24 978, 2019.

[13] N. Buckman, A. Pierson, W. Schwarting, S. Karaman, and D. Rus, "Sharing is caring: Socially-compliant autonomous intersection negotiation," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 6136–6143.

[14] M. F. Ozkan and Y. Ma, "Socially compatible control design of automated vehicle in mixed traffic," *IEEE Control Systems Letters*, vol. 6, pp. 1730–1735, 2021.

[15] J. Larsson, M. F. Keskin, B. Peng, B. Kulcsár, and H. Wymeersch, "Pro-social control of connected automated vehicles in mixed-autonomy multi-lane highway traffic," *Communications in Transportation Research*, vol. 1, p. 100019, 2021.

[16] B. Toghi, R. Valiente, D. Sadigh, R. Pedarsani, and Y. P. Fallah, "Cooperative autonomous vehicles that sympathize with human drivers," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 4517–4524.

[17] L. Crosato, C. Wei, E. S. Ho, and H. P. Shum, "Human-centric autonomous driving in an av-pedestrian interactive environment using svo," in *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*. IEEE, 2021, pp. 1–6.

[18] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[19] M. Kuderer, S. Gulati, and W. Burgard, "Learning driving styles for autonomous vehicles from demonstration," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 2641–2646.

[20] Z. Wu, L. Sun, W. Zhan, C. Yang, and M. Tomizuka, "Efficient sampling-based maximum entropy inverse reinforcement learning with application to autonomous driving," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5355–5362, 2020.

[21] W. Cao, M. Mukai, T. Kawabe, H. Nishira, and N. Fujiki, "Cooperative vehicle path generation during merging using model predictive control with real-time optimization," *Control Engineering Practice*, vol. 34, pp. 98–105, 2015.

[22] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey, *et al.*, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[23] M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard, "Feature-based prediction of trajectories for socially compliant navigation." in *Robotics: science and systems*, 2012.

[24] U.S. Department of Transportation Federal Highway Administration. (2016) Next generation simulation (ngsim) vehicle trajectories and supporting data. [Online]. Available: https://data.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Trajector/8ect-6jqj

[25] J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "Casadi: a software framework for nonlinear optimization and optimal control," *Mathematical Programming Computation*, vol. 11, no. 1, pp. 1–36, 2019.

[26] B. Chalaki, L. E. Beaver, A. M. I. Mahbub, H. Bang, and A. A. Malikopoulos, "A research and educational robotic testbed for real-time control of emerging mobility systems: From theory to scaled experiments," *IEEE Control Systems Magazine*, 2022 (in press).