Robust game-theoretic decision making for autonomous vehicles

Gokul S. Sankar, Kyoungseok Han and Ilya Kolmanovsky

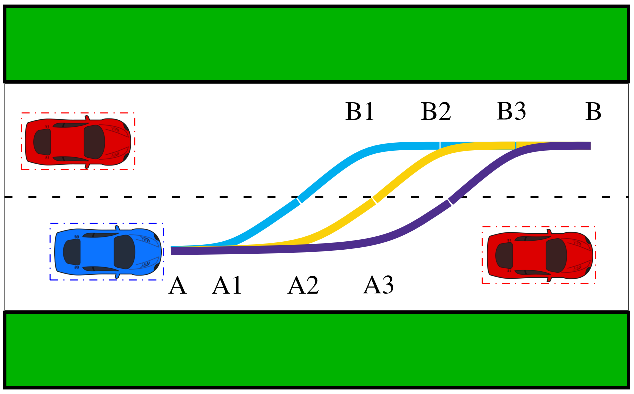

Fig. 1. Possible lane changing sequences $\{(A, A1), (A1, B1), (B1, B)\}$, $\{(A, A2), (A2, B2), (B2, B)\}$ and $\{(A, A3), (A3, B3), (B3, B)\}$ that can be chosen by the autonomous vehicle (blue) based on the motion of other non-autonomous vehicles (red).

***Abstract*—**

I. Introduction

A. Notation

II. Vehicle dynamic model

The vehicle dynamics are represented by the following discrete model

$$x(t+1) = x(t) + v(t)\cos(\psi(t) + \beta(t))\,\Delta t + w_x(k) \tag{1a}$$

$$y(t+1) = y(t) + v(t)\sin(\psi(t) + \beta(t))\,\Delta t + w_y(k) \tag{1b}$$

$$\psi(t+1) = \psi(t) + \frac{v(t)}{l_r}\sin(\beta(t))\,\Delta t \tag{1c}$$

$$v(t+1) = v(t) + a(t)\,\Delta t \tag{1d}$$

$$\beta(t) = \arctan\left(\frac{l_r}{l_r + l_f}\tan(\delta_f(t))\right), \tag{1e}$$

where t denotes the discrete time instant; the pair $(x(t), y(t))$ represent the global position of the center of mass of the vehicle; the vehicle's speed is denoted by $v(t)$; $\beta(t)$ is the angle of $v(t)$ with respect to the longitudinal axis of the vehicle; $\psi(t)$ denotes the vehicle's yaw angle (the angle between the vehicle's heading direction and the global x-direction); $a(t)$ denotes the vehicle's acceleration at time t; Δt denotes the time step size; $\delta_f(t)$ represents the front steering angle; and l_f and l_r are the distance of the center of the mass of the vehicle to the front and rear axles, respectively; $w_x(k)$ and $w_x(k)$ denote the uncertainty in the position of the center of mass, respectively. It is assumed the uncertainties originate from a closed and compact disturbance set, $\mathcal{W} := \left\{w = (w_x, w_y) \,|\, \zeta w \leq \theta, \, \zeta \in \mathbb{R}^{a\times 2}, \, \theta \in \mathbb{R}^b\right\}$, with $b \in 2\mathbb{Z}^+$. The disturbance set is assumed to contain the origin. Furthermore, it is assumed that the rear wheels cannot be steered. The input to the model (1), represented by $\gamma(t) = (a(t), \delta_f(t))$, is the acceleration and front steering angle pair.

Gokul S. Sankar, Kyoungseok Han and Ilya Kolmanovsky are with the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI, USA. Emails: ggowrisa@umich.edu (G. Sankar), kyoungsh@umich.edu (K. Han) and ilya@umich.edu (I. Kolmanovsky)

III. Game-theoretic decision making

At each time instant, each vehicle selects an input pair from the finite action set, $\mathbf{\Gamma} = \{(0, 0), (0, \delta_{f,\max}), (0, -\delta_{f,\max}), (a_{\max}, 0), (-a_{\max}, 0), (-2 \times a_{\max}, 0)\}$, where $a_{\max}$ and $\delta_{f,\max}$ are the maximum acceleration and front steer angle, respectively. The inputs pairs in $\mathbf{\Gamma}$ correspond to the actions, {maintain, turn left, turn right, accelerate, decelerate, brake}, respectively. The input pair to be applied at every time step is decided based on optimizing a reward function.

A. Action choice

The decision making process of the vehicle in choosing the optimal input pair follows a receding horizon strategy. A sequence of actions, $\boldsymbol{\gamma_t} = \{\gamma_t, \gamma_{t+1}, \cdots, \gamma_{t+N-1}\}$, is chosen that maximizes a cumulative reward given by

$$\mathcal{R}(\boldsymbol{\gamma_t}) = \sum_{j=0}^{N-1} \lambda^{j-1} R_{t+j}(\gamma_{t+j}), \tag{2}$$

where $R_{t+j}(\gamma_{t+j})$ is the stage reward at a prediction step j determined at time step t for an input, $\gamma_{t+j} \in \mathbf{\Gamma}$; $\lambda \in [0, 1]$ is the discount factor. By the receding horizon strategy, the input applied to (1), $\gamma(t)$, is the first element of $\boldsymbol{\gamma_t}^* = \left\{\gamma_t^*, \gamma_{t+1}^*, \cdots, \gamma_{t+N-1}^*\right\}$ is applied at each time instant t, i.e., $\gamma(t) = \gamma_t^*$. The stage reward at a prediction step j, $R_{t+j}(\gamma_{t+j})$, is defined as

$$R_{t+j}(\gamma_{t+j}) = R_{t+j}(\gamma_{t+j}|s_{t+j}) = \boldsymbol{\alpha}^T \boldsymbol{\phi_{t+j}} \tag{3}$$

where s_{t+j} is the traffic state at prediction step j; $\boldsymbol{\phi_{t+j}} = \{\phi_{1,\,t+j},\, \phi_{2,\,t+j},\, \cdots,\, \phi_{m,\,t+j}\}$ is the feature vector at step j and the weights for these features are in $\boldsymbol{\alpha} = \{\alpha_1,\, \alpha_2,\, \cdots,\, \alpha_m\}$, in which $\alpha_i > 0, \forall\, i \in \mathbb{Z}_{[0:m]}$. For the lane changing scenario in Fig. 1, the features considered are described below.

Rectangular outer approximation of the geometric contour of each vehicle is considered as shown by the dash-dotted boxes in Fig. 1. This outer approximation is referred as the collision avoidance zone (c-zone). The features, $\phi_{1,\,t}$, $\phi_{2,\,t}$ and $\phi_{3,\,t}$, are indicator functions based on the c-zone of the vehicles that respectively characterize:

- Collision status - The intersection of the c-zone of the ego vehicle with that of any other vehicle indicates a collision or a danger of collision. If an overlap is detected then $\phi_{1,\,t}$ is assigned a value -1; and 0, otherwise.
- On-road status - The intersection of the c-zone of the ego vehicle with that of green regions shown in Fig. 1 indicates that the ego vehicle is outside the road boundaries. The feature $\phi_{2,\,t} = -1$ if an overlap is detected; $\phi_{2,\,t} = 0$, otherwise.
- Safe zone violation status - A safe zone (s-zone) of a vehicle is a rectangular area that subsumes the c-zone of the vehicles with a safety margin. The safety margin is chosen based on the minimum distance to be maintained from the surrounding vehicles. If an overlap of the s-zone of the ego vehicle with that of another vehicle is detected then $\phi_{3,\,t}$ is assigned a value -1; and 0, otherwise.

The other features considered in this work characterize:

- Distance to objective - In order to encourage the ego vehicle to change lane and reach a reference point in the new lane, $\left(x^{\text{ref}},\, y^{\text{ref}}\right)$, the feature $\phi_{4,\,t}$ is defined as
$$\phi_{4,\,t} = -\left(\left|x_t - x^{\text{ref}}\right| + \left|y_t - y^{\text{ref}}\right|\right). \tag{4}$$
- Distance to lane center - The feature, $\phi_{5,\,t}$, defined as
$$\phi_{5,\,t} = -\left|y_t - y^{lc}\right|, \tag{5}$$
where y^{lc} is the y-coordinate of the center of the current lane, is included to encourage the ego vehicle to be at the middle of the current lane.
- Velocity error - The deviation of the velocity of the ego vehicle from a reference velocity, v^{ref}, is described by the feature $\phi_{6,\,t}$ as
$$\phi_{6,\,t} = -\left|v_t - v^{\text{ref}}\right|, \tag{6}$$
where the reference velocity is typically chosen as the legislated speed limit.

B. Level-k framework

The state of the traffic, s_{t+j}, at prediction steps $j = 0,\, 1,\, \cdots,\, N-1$ have to be known to compute the cumulative reward in (2).

, where

$$\boldsymbol{\gamma_t}^* = \arg\max_{\boldsymbol{\gamma_t}\in\boldsymbol{\Gamma}} \mathcal{R}\left(\boldsymbol{\gamma_t}\right), \tag{7}$$

We use a game theoretic model for this prediction.

Numerous experimental results from psychology, cognitive science, and economics have suggested a hierarchical structure in human reasoning in games, see [13], [14], [15], [16]. The study of this reasoning hierarchy and its applications in game theoretic settings are addressed by the "level-k game theory." In [5], [6], [17], the level-k game theory is exploited to model vehicle interactions in highway traffic. The model has been compared to human traffic data in [6]. In this paper, we also exploit level-k game theory to model vehicle interactions, in particular, at intersections. Recently, level-k modeling of human agents was also considered in aerospace and energy applications [18], [19], [20], where human-to- human and human-to-automation interactions play a central role.

The model is premised on the idea that strategic agents (drivers/vehicles at an intersection, in our setting) have different reasoning levels. In particular, the level k indicates an agent's reasoning depth. The reasoning hierarchy starts from level-0. A level-0 agent makes instinctive decisions to pursue its goal without considering the interactions between itself and the others. On the contrary, a level-1 agent takes into account such interactions in its decision making process, in particular, by assuming that all the other agents in the game are level-0. Specifically, a level-1 agent assumes that all the other agents are level-0 so they make instinctive decisions; the level-1 agent predicts their actions as well as the evolutions of the game resulting from their actions based on this assumption; the level-1 agent then makes its own decision as the best response to such evolutions to pursue its own goal. Similarly, a level-k agent assumes that all the other agents are level-(k-1), makes predictions based on this assumption, and responds accordingly.

In this paper, a level-0 driver/vehicle treats the other vehicles at the intersection as stationary obstacles, and selects its action sequence, 0, accordingly. In this setting, a level- 0 driver/vehicle may represent an aggressive driver in real traffic, who usually assumes that other drivers will yield the right of way.

After the level-0 is defined, the action selection procedure for level-k, k 1, in the case of 2-agent interactions has the following form,

Based on the definitions of the features, the rewards of a vehicle not only depend on its own states and actions, but also depend on the states and actions of its opponent vehicle (e.g., 1(t) and 4(t)). Such an interdependence reflects the interactive nature of vehicle decision making in a multi- vehicle traffic scenario. The receding-horizon optimal control problem is thus formulated as: At each time step t, to select

represents the state of the traffic, which contains both the ego vehicle's states and the opponent vehicle's states; ego(t + j) is the ego vehicle's action at t + j over the horizon and is to be optimized, and opp.(t + j) is the opponent vehicle's action at t + j over the horizon and is to be predicted. We note that for the two interacting vehicles, either is the "ego vehicle" from its own perspective, and is also the "opponent vehicle" from the other's perspective, that is, (4) can be used to describe the decision making of either of the two vehicles.

IV. Simulation results

V. Conclusions