Adaptive Robust Game-Theoretic Decision Making for Autonomous Vehicles

Gokul S. Sankar, Kyoungseok Han and Ilya Kolmanovsky

***Abstract*—This paper proposes a robust decision-making strategy for autonomous vehicles based on the game theory. Autonomous vehicles should be controlled by sophisticated negotiation skills with the other road participants, e.g., human-driven vehicles, pedestrians. In this paper, by modeling the interactions between the vehicles, autonomous vehicle can make reasonable decision that is not too conservative or aggressive. The behavior of the vehicles is modeled by cognitive hierarchy theory, e.g., level-*k* framework, which demonstrates its effectiveness when describing human thinking process. By inferring other vehicle's thoughts, autonomous vehicle can avoid the biased decision, i.e., conservative or aggressive. The effectiveness of the proposed approach is verified with the lane change scenario at the highway.**

I. Introduction

Despite many recent advances in connected and automated vehicles (CAVs) technology, the full automation systems that can provide the similar or better driving proficiency compared to the human drivers are still inherently flawed to be deployed in the market [1]. One of the most significant challenges is to plan the motion of the automated vehicles at the mixed-traffic conditions where the automated car coexists with the human-driven vehicles [2]. In particular, describing the human decision-making process is the most difficult since the humans do not always exhibit the optimized behavior due to limited rationality [3]. In order to ensure the safety, a conservative driving policy of the automated vehicles where the all possible driving situations are considered has been suggested [4], [5], but it may cause the disharmony with the human drivers. That is, too conservative behavior of the automated vehicles sometimes causes adverse effect on traffics, e.g., road congestion, car accidents, due to the uncertainties of the human drivers.

If the human-driven vehicle's response according to traffics, i.e., behavior of the surrounding vehicles, can be predicted, far less conservative motion planning of the automated vehicle can be designed [6]. As mentioned, however, modeling of the interaction between the vehicles may not very accurate. Moreover, since the communication time with the surrounding vehicles in reality is not sufficient to build the deterministic human driver models.

To overcome such shortcomings, we exploit the "level-*k* game theory" framework where the human thought processes in strategic games, i.e., driving at the mixed-traffics, are modeled in hierarchical structure [7]. The game theoretic approach has already proved its effectiveness when describing the interactions between the vehicles [8], [9], [10]. Although these approaches effectively describe rational decision-making for the human-driven vehicles, the model uncertainties that affect to the vehicle position are not considered. Therefore, it is assumed that the center of gravity of the vehicle follows the deterministic kinematic model with the constant safety constraint that prevents the physical safety violation between the vehicles, and depending on the size of the safety constraint, the level of conservatism in the autonomous vehicle's motion is determined.

In this paper, we adaptively adjust the safety constraint sizes of the interactive human-driven vehicles according to the confidence level to establish a balanced motion planning between the aggressive and conservative motion planning. The interactive vehicle's aggressiveness level is estimated and it is proportional to the size of safety constraint of the human-driven vehicle. Based on this, the autonomous vehicle's future trajectory is planned, which maximizes the cumulative reward. In [11], the caution level of the human-driven vehicle is used to describe the aggressiveness of the HDV, but only constant caution level is assumed, which eventually determine the level of conservatism of AV. To resolve this, we incorporate the level-*k* game theory with the caution level so that adaptive caution level is available, which enables adjustable motion planning of AV depending on the interactive vehicle's aggressiveness.

The contributions of this paper are summarized as follows. Firstly, we modeled the human-driven vehicle's decision-making processes at high way using game theory, representing the a degree of the aggressiveness. Secondly, the uncertainties that is imposed in vehicle position is represented by the identified drivers aggressiveness and these bounds are utilized to make the robust decision-making strategy for the autonomous vehicles. Lastly, we verify the benefit of the proposed approach through simulations when comparing with the results that do not consider the position uncertainty.

The rest of the paper is organized as follows. In Section II, the vehicle dynamic model with the bounded uncertainty is introduced. We next present the game-theoretic decision making algorithm and motion planning procedure in Section III, the effectiveness of the proposed approach is described in Section IV, and the paper is concluded in Section V.

[12]

[13]

[14]

[15]

[16]

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)

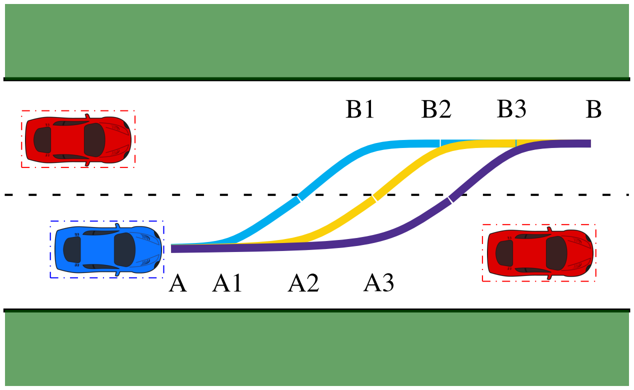

Fig. 1. Possible lane changing sequences that can be chosen by the autonomous vehicle (blue) based on the motion of other non-autonomous vehicles (red) are $\{(A, A1), (A1, B1), (B1, B)\}$, $\{(A, A2), (A2, B2), (B2, B)\}$ and $\{(A, A3), (A3, B3), (B3, B)\}$.

A. Notation

The symbol $\mathbb{Z}_{[a,\, b]}$ denotes a set of consecutive integers from a to b and $2\mathbb{Z}^+$ denotes set of positive even integers. For a vector x, $x > 0$ denotes element-wise inequality. The operator $\oplus$ denotes the Minkowski addition, defined for the sets $\mathcal{A}$ and $\mathcal{B}$ as

$$\mathcal{A} \oplus \mathcal{B} := \{a + b | a \in \mathcal{A} \forall b \in \mathcal{B}\}. \tag{1}$$

II. Probelm statement

As illustrated in Fig. 1, depending on the motions of interactive vehicles (red), the autonomous vehicle can plan the different motions. If the interactive vehicles behave aggressively, the AV should choose the conservative motion planning, and vice versa. In addition, to avoid the too conservative or aggressive driving policy, the balanced motion planning is preferred in the actual application.

The problem we treat is to model the intent of the interactive vehicles using game theoretical approach, then AV exploits the estimated interactive vehicle's driving intention when making the decision in real-time. The confidence level of the estimated intention is represented by dashed-dotted boxes in Fig. 1 and their sizes are adjusted adaptively according to the confidence levels.

III. Vehicle dynamic model

The vehicle dynamics are represented by the following discrete model [17]

$$x(t+1) = x(t) + v(t)\cos(\psi(t) + \beta(t))\,\Delta t + w_x(t) \tag{2a}$$

$$y(t+1) = y(t) + v(t)\sin(\psi(t) + \beta(t))\,\Delta t + w_y(t) \tag{2b}$$

$$\psi(t+1) = \psi(t) + \frac{v(t)}{l_r}\sin(\beta(t))\,\Delta t \tag{2c}$$

$$v(t+1) = v(t) + a(t)\,\Delta t \tag{2d}$$

$$\beta(t) = \arctan\left(\frac{l_r}{l_r + l_f}\tan(\delta_f(t))\right), \tag{2e}$$

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; $w_x(t)$ and $w_x(t)$ 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 defined as

$$\mathcal{W} := \left\{w = (w_x,\, w_y) \,|\, \zeta w \leq \theta,\, \zeta \in \mathbb{R}^{a\times 2},\, \theta \in \mathbb{R}^b\right\}, \tag{3}$$

where $b \in 2\mathbb{Z}^+$. The disturbance set is assumed to contain the origin. Furthermore, it is assumed that the rear wheels cannot be steered. Therefore, the control input to the model (2), represented by $\gamma(t) = (a(t), \delta_f(t))$, is the acceleration and front steering angle pair.

IV. Robust game-theoretic decision making

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

A. Reward function

The decision making process of the vehicle in choosing the optimal input pair follows a receding horizon strategy. The nominal model used within the prediction horizon is given by

$$x_{t+j+1} = x_{t+j} + v_{t+j}\cos(\psi_{t+j} + \beta_{t+j})\,\Delta t \tag{4a}$$

$$y_{t+j+1} = y_{t+j} + v_{t+j}\sin(\psi_{t+j} + \beta_{t+j})\,\Delta t \tag{4b}$$

$$\psi_{t+j+1} = \psi_{t+j} + \frac{v_{t+j}}{l_r}\sin(\beta_{t+j})\,\Delta t \tag{4c}$$

$$v_{t+j+1} = v_{t+j} + a_{t+j}\Delta t \tag{4d}$$

$$\beta_{t+j} = \arctan\left(\frac{l_r}{l_r + l_f}\tan(\delta_{f,\,t+j})\right), \tag{4e}$$

where $j \in \mathbb{Z}_{[0,\,N-1]}$ represents the prediction step, and $\gamma_{t+j} = (a_{t+j}, \delta_{f,\,t+j})$ denotes the input pair applied to (4) at a prediction step j. 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

$$\mathcal{R}(\boldsymbol{\gamma_t}) = \sum_{j=0}^{N-1} \lambda^j R_{t+j}, \tag{5}$$

where R_{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 (2), $\gamma(t)$, 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. The stage reward at a prediction step j, R_{t+j}, is defined as

$$R_{t+j} = \boldsymbol{\alpha}^T \boldsymbol{\phi_{t+j}}, \tag{6}$$

where $\boldsymbol{\phi_{t+j}} = \{\phi_{1,\,t+j}, \phi_{2,\,t+j}, \cdots, \phi_{m,\,t+j}\}$ 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). The features, $\phi_{1,\,t}$, $\phi_{2,\,t}$ and $\phi_{3,\,t}$, are indicator functions based on the collision avoidance zone of the vehicles that respectively characterize:

- Collision status - The intersection of the collision avoidance 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 collision avoidance 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) of a vehicle is a rectangular area that subsumes the collision avoidance 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 safe 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

$$\phi_{4,\,t} = -\left(\left|x_t - x^{\text{ref}}\right| + \left|y_t - y^{\text{ref}}\right|\right). \tag{7}$$

- Distance to lane center - The feature, $\phi_{5,\,t}$, defined as

$$\phi_{5,\,t} = -\left|y_t - y^{lc}\right|, \tag{8}$$

where y^{lc} is the y-coordinate of the center of the current lane that 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

$$\phi_{6,\,t} = -\left|v_t - v^{\text{ref}}\right|, \tag{9}$$

where the reference velocity is typically chosen as the legislated speed limit.

B. Level-k framework

In a multi-agent traffic scenario, the interactive nature of the decision making process is taken into account by the features, $\phi_{1,\,t}$ and $\phi_{3,\,t}$, of the stage reward in (6) that depend on the states of other vehicles. To compute the cumulative reward in (5), for a given sequence of actions of the l^{th} autonomous vehicle, $\boldsymbol{\gamma_t[l]} = \{\gamma_t[l], \gamma_{t+1}[l], \cdots, \gamma_{t+N-1}[l]\}$, it is required to predict the actions of other agents, $\boldsymbol{\gamma_t[i]} = \{\gamma_t[i], \gamma_{t+1}[i], \cdots, \gamma_{t+N-1}[i]\}$, $\forall i \in \mathcal{O}$, where $\mathcal{O} = \left\{i | i \in \mathbb{Z}_{[1,\,n]},\, i \neq l\right\}$ with n representing the number of agents, and the corresponding state of the traffic, s_{t+j}, at prediction steps $j = 0, 1, \cdots, N-1$, where $s_t = [x_t[1], y_t[1], v_t[1], \theta_t[1], \cdots, x_t[n], y_t[n], v_t[n], \theta_t[n]]^T$. In this paper, level-k game theory [18], [19] is utilized to model the vehicle-to-vehicle interactions and thus predict the actions of the other agents over the horizon.

In level-k game theory, it is assumed that the decisions taken by the strategic agents are based on the predictions of the actions of the other agents and the agents can have have different reasoning levels. The reasoning depth of an agent is indicated by $k \in \{0, 1, \cdots\}$. The hierarchy begins with agents at level-0, where the agents make instinctive decisions to achieve the objective without accounting for the interactions between other agents. On the other hand, the agents at level-k $\forall k > 0$, consider the interactions by assuming that the other agents are at level-$(k-1)$ and take decisions accordingly. For instance, a l^{th} level-1 agent assumes other agents are at level-0 and predicts their action sequences, $\boldsymbol{\gamma_t^{(0)}[i]} = \left\{\gamma_t^{(0)}[i], \gamma_{t+1}^{(0)}[i], \cdots, \gamma_{t+N-1}^{(0)}[i]\right\}$ $\forall i \in \mathcal{O}$, to compute its own action sequence, $\boldsymbol{\gamma_t^{(1)}[l]} = \left\{\gamma_t^{(1)}[l], \gamma_{t+1}^{(1)}[l], \cdots, \gamma_{t+N-1}^{(1)}[l]\right\}$.

The level-k game theory was adapted to model the vehicle-to-vehicle interactions at an unsignalized four-way intersection in [9]. It is assumed that level-0 vehicles consider the other vehicles in the traffic scenario as stationary obstacles. Therefore, these level-0 drivers implicitly assume the others will yield the right of way, and can be regarded 'aggressive'. And level-1 drivers consider other drivers to be aggressive and take 'cautious' actions. In [9], the drivers are categorized into level-0, 1 and 2. Since, the behavior of level-2 driver will be similar to that of the level-0 drivers, in this paper, only level-0 and 1 drivers are considered. The stage reward value obtained for l^{th} level-k agent at a prediction step j for an action $\gamma_{t+j}^{(k)}[l]$, depend on the current traffic state, s_0, the ego agent's actions, $\left\{\gamma_t^{(k)}[l], \gamma_{t+1}^{(k)}[l], \cdots, \gamma_{t+j-1}^{(k)}[l]\right\}$, and the actions of other agents, $\left\{\gamma_t^{(k-1)}[i], \gamma_{t+1}^{(k-1)}[i], \cdots, \gamma_{t+j-1}^{(k-1)}[i]\right\}$ $\forall i \in \mathcal{O}$, is given as

$$\begin{aligned} R_{t+j}^{(k)}[l] = R_{t+j}\Big(\gamma_{t+j}^{(k)}[l]\,\Big|\, & s_0, \gamma_t^{(k)}[l], \gamma_{t+1}^{(k)}[l], \cdots, \gamma_{t+j-1}^{(k)}[l], \\ & \gamma_t^{(k-1)}[i], \gamma_{t+1}^{(k-1)}[i], \cdots, \gamma_{t+j-1}^{(k-1)}[i]\Big), \end{aligned} \tag{10}$$

and its cumulative reward is

$$\mathcal{R}^{(k)}\left(\boldsymbol{\gamma_t^{(k)}[l]}\right) = \sum_{j=0}^{N-1} \lambda^j R_{t+j}^{(k)}[l]. \tag{11}$$

C. Multi-model strategy

Human drivers, initially, do not have perfect knowledge about other drivers. However, they gain better understandings of other driver's characteristics through interactions, and therefore, resolve conflicts effectively. Similarly, the autonomous vehicles in a multi-agent traffic scenario, hold an initial belief about the driver model (level-0 and 1) of the other vehicles as a probability distribution over both models. Subsequently, based on the actual action applied by the other agents, the estimate of the probability distribution is updated at every step.

From the perspective of an l^{th} autonomous agent, the probability that the i^{th} other agent can be modeled as level-k is represented by $P_{K_i^l=k}$. The probability of the model k is increased when it matches the actual action by

$$k^* = \arg \min_{k \in \{0,\,1\}} \left\| \gamma[i](t) - \gamma_t^{(k)}[i] \right\| \tag{12a}$$

$$\tilde{P}_{K_i^l=k^*}(t) = P_{K_i^l=k^*}(t-1) + \Delta P \tag{12b}$$

$$P_{K_i^l=k}(t) = \frac{\tilde{P}_{K_i^l=k}(t)}{\sum_{\tilde{k}=0}^{1} \tilde{P}_{K_i^l=\tilde{k}}(t)}, \forall k \in \{0,\,1\}, \tag{12c}$$

where $\gamma[i](t)$ and $\gamma_t^{(k)}[i]$ represent the actual and predicted action taken by i^{th} agent assuming level-k model, respectively; $\Delta P > 0$ is a constant that denotes the rate of increment of the probability; $\left\| \gamma[i](t) - \gamma_t^{(k)}[i] \right\| = \left| a(t) - a_t^{(k)} \right| + \left| \delta_f(t) - \delta_{f,\,t}^{(k)} \right|$. When the input pair of the actual action is equal to that of the predicted action, the probability distribution remains unchanged.

In order to incorporate the multi-model strategy in the decision making process and select the optimal action according to the model of other agents, the expected cumulative reward for the l^{th} agent, using (10) and (11), is given by

$$\mathcal{R}_P(\boldsymbol{\gamma_t}[l]) = \sum_{k=0}^{1} P_{K_i^l=k}(t)\, \mathcal{R}^{(k)}\left(\boldsymbol{\gamma_t^{(k)}}[l]\right), \forall i \in \mathcal{O}. \tag{13}$$

D. Adaptive robust decision making

The mismatch between the actual position of the center of mass of the vehicle which is used to determine the rectangular outer approximation of the vehicle (see Section IV-A) and its predictions obtained using the dynamic model in (4) might lead to collision. In the multi-agent traffic scenario under consideration, there are two sources of modeling errors: (i) the uncertainties, $w^{(m)} = \left(w_x^{(m)},\, w_y^{(m)}\right) \in \mathcal{W}_m$, resulting due to the use of a simplified model in (4); and (ii) the uncertainty, $w^{(d)} = \left(w_x^{(d)},\, w_y^{(d)}\right) \in \mathcal{W}_d$, arising due to unknown driver model. Hence, the disturbance set defined in (3) is

$$\mathcal{W} = \mathcal{W}_m \oplus \mathcal{W}_d. \tag{14}$$

Robust approaches can be used to account for these uncertainties while computing the control actions. Since a discrete set of input actions is considered in this work, feedback min-max strategy [12] is utilized in this work for considering the uncertainties originating from the disturbance set $\mathcal{W}$. Since the autonomous agents update the driver model of the other agents in the multi-agent traffic scenario at each step according to (12), an adaptive scheme is proposed to incorporate the confidence on the driver models and leverage the fact that level-1 drivers are cautious. It is achieved by considering the following adaptive disturbance set in the robust decision making process of the autonomous agent l,

$$\bar{\mathcal{W}}_i^l(t) = \mathcal{W}_m \oplus P_{K_i^l=0}(t)\mathcal{W}_d, \forall i \in \mathcal{O} \tag{15}$$

where at time t, $P_{K_i^l=0}(t)$ is the probability that the i^{th} agent is a level-0 driver from the perspective of the l^{th} autonomous agent, and $\bar{\mathcal{W}}_i^l(t)$ denotes the disturbance set. It is assumed that, initially, all the agents are level-0 drivers, i.e., $P_{K_i^l=0}(0) = 1, \forall i \in \mathcal{O}$. Essentially, this assumption allows the autonomous agent to be cautious with another interacting agent when there is no/less information about that agent. If an i^{th} agent is level-0, as time evolves, $P_{K_i^l=0}(t)$ will continue to be equal to one, and hence, the autonomous agent take conservative actions (or behave like level-1 driver). On the other hand, when the i^{th} agent is a level-1 driver, $P_{K_i^l=0}(t)$ will decrease, resulting in a reduced disturbance set size, thereby, allowing the autonomous agent to take less conservative actions and adapt to the behavior of the interacting agents while capable of handling the uncertainties arising due to the use of a simple prediction model.

When interacting with an agent $i \in \mathcal{O}$, the objective the autonomous agent l is to maximize the expected cumulative reward (11), while accounting for the effect of the worst-case uncertainty from the possible disturbance realizations from the adaptive disturbance set, $\bar{\mathcal{W}}_i^l(t)$. The optimal control sequence, $\boldsymbol{\gamma_t^*}[\boldsymbol{l}] = \left\{\gamma_t^*[l],\, \gamma_{t+1}^*[l],\, \cdots,\, \gamma_{t+N-1}^*[l]\right\}$, is obtained by solving the following optimization problem

$$\boldsymbol{\gamma_t^*}[\boldsymbol{l}] = \arg \max_{\boldsymbol{\gamma_t}[\boldsymbol{l}] \in \boldsymbol{\Gamma}} \min_{w_{t+j}^p \in \bar{\mathcal{W}}_i^l(t)} \mathcal{R}_P(\boldsymbol{\gamma_t}[\boldsymbol{l}]) \tag{16a}$$

$$\text{s. t.} \forall j \in \mathbb{Z}_{0,\,N-1},\ \text{(4c), (4d), (4e)}, \forall i \in \mathcal{O}, \forall p \in \mathcal{P}$$

$$\tilde{x}_{t+j+1} = \tilde{x}_{t+j} + v_{t+j}\cos(\psi_{t+j} + \beta_{t+j})\,\Delta t + w_{x,\,t+j}^p \tag{16b}$$

$$\tilde{y}_{t+j+1} = \tilde{y}_{t+j} + v_{t+j}\sin(\psi_{t+j} + \beta_{t+j})\,\Delta t + w_{x,\,t+j}^p, \tag{16c}$$

where $w_{t+j}^p = \left(w_{x,\,t+j}^p,\, w_{y,\,t+j}^p\right)$ denote a possible realization of the uncertainty in the global position of the center of mass in x and y directions, respectively; and $\mathcal{P}$ represents the set of indexes of the realizations. The autonomous agent then applies the first element $\gamma_t^*[l]$ of the optimal control action sequence, i.e., $\gamma(t) = \gamma_t^*[l]$.

V. Simulation results

The proposed adaptive robust approach is validated for the lane changing maneuver on a three lane highway section. Consider the multi-agent traffic scenario shown in the subfigure $[1,\,1]$ (first element represents row and second element represents column) in Fig. 2. The objective of the autonomous agent (blue) is to change from lane II to lane III, while the human agents keep their respective lane. All the human drivers are assumed to be level-1, i.e., they exhibit cautious behavior, however, it is unknown to the AV. The sampling time is set to $0.5\,\mathrm{s}$ and two step prediction horizon is considered. The

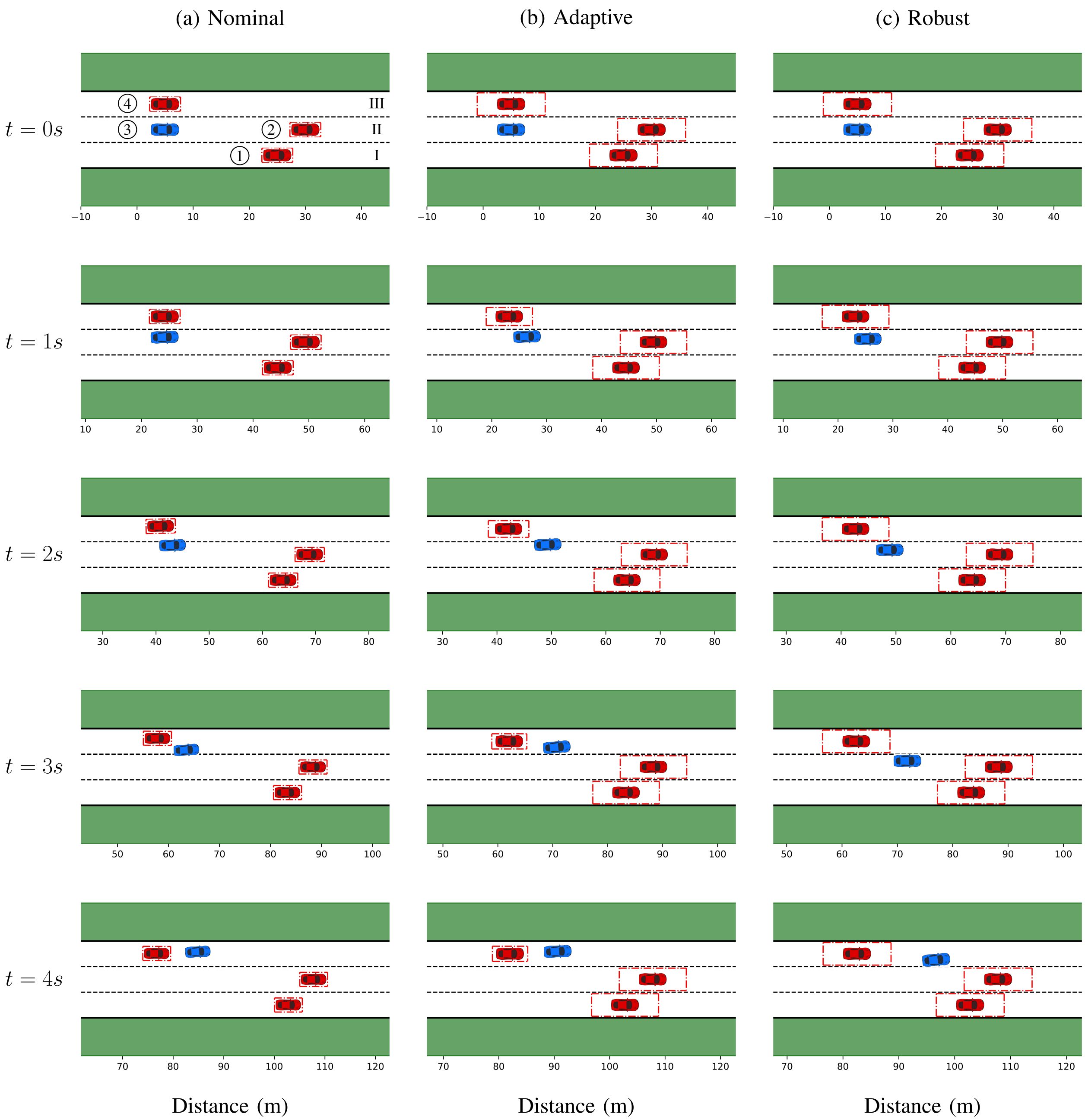

Fig. 2. A four second simulation sequence with a one second time interval (see along the column) showing the lane changing maneuver performed by the autonomous vehicle under (a) nominal; (b) adaptive; and (c) robust decision making strategies in a multi-traffic scenario. The circled numbers indicate the vehicle ids and the Roman numerals denote the lane number in the subfigure $[1,\ 1]$. The dashed lines indicate the set $s \oplus \bar{\mathcal{W}}_i^l(t)$, $\forall i \in \mathcal{O}$ from the perspective of the autonomous agent.

proposed methodology is compared to the nominal strategy, that assumes $\bar{\mathcal{W}}_i^l(t) = 0$ while deciding the control action using (16); and the robust strategy, that considers $\bar{\mathcal{W}}_i^l(t) = \mathcal{W}$ as defined in (14) during the decision making process.

The traffic simulation under the nominal decision making strategy is shown in the first column of subfigures in Fig. 2. Since the disturbances are not considered in this case, it can be noted the AV chooses to steer left as soon as the simulation begins which provides the maximum reward. This move is considered to be aggressive. However, the human vehicle 4, being cautious, reacts by steering left at time $t = 2\,\mathrm{s}$ and returning to the lane center at $t = 4\,\mathrm{s}$ once the AV has passed. The AV completes the lane change between $60\,\mathrm{m}$ and $70\,\mathrm{m}$. On the other hand,

We next illustrate the results when adopting the conservative lane change strategy as shown in Fig. 2(c). Unlike the previous case, the constant large uncertainty sizes are assumed to all interactive vehicles, meaning that the AV believes that the other vehicles will show the aggressive behaviors, which is not true. As a result, very conservative driving policy is designed as confirmed in Fig. 2(c). Compared to the previous case, the AV changes the lane late ($90m$ - $100m$) because incorrect beliefs of the aggressiveness of the interactive vehicles are utilized. Obviously, the AV can avoid the collision and safety

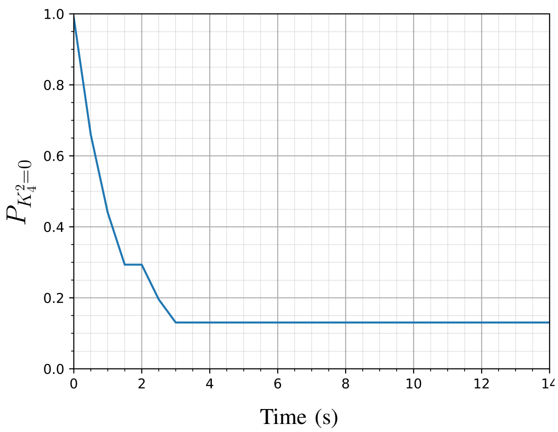

Fig. 3. Probability that vehicle 4 can be modeled as level-0 from the perspective of the autonomous vehicle 2.

is ensured with this conservative driving policy. However, compared to the human-decision making, it cannot be said to be reasonable motions.

As compared to two previous result, we show the effectiveness of the proposed approach through Fig. 2(b). We can observe the more reasonable behavior of the AV. Starting from the belief of large uncertainty bound of the left HV, the AV adjusts its belief as the historical actions of the left HV are collected based on (12). As a result, the lane change happens around $70m$ that is between the those of the previous results. The real-time level estimation is illustrated in Fig. 3. Initially, it is believe that the HV at the left is modeled as level-0 but it true level is identified within $3s$. However, the initial beliefs to the two front HVs do not vary because their expected actions between level-0 and level-1 modeling are same for $4s$, so the AV is very cautious with the front HVs.

VI. Conclusions

We proposed the robust decision-making strategy for autonomous vehicles when they share the road with the other road participants.We modeled the vehicle interactions using a level-k framework, then the decision of the autonomous vehicle was made based on the identified other vehicle's level in real-time. To avoid the conservative or aggressive behavior of autonomous vehicle, the accuracy of the level estimation was utilized to represent the uncertainty bounds of the interactive vehicles. Based on updated level estimation accuracy, the uncertainty bounds could be adjusted, and autonomous vehicle adaptively plans its motion for a given prediction horizon. Through the simulation verification, we could observe the reasonable behavior of autonomous vehicle, compared to the behaviors of conventional decision-making approaches. However, extending to handle the large number of vehicles has not been researched in this study, and computational challenge is anticipated. For the future research, we address computational tractability when the more number of interactive vehicles is given.

Acknowledgement

We would like to express our appreciation to Anouck Girard and Nan Li who provided expertise that greatly assisted the research and it should be noted that the code publicly available (https://github.com/gokulsivasankar/RobustDecisionMaking) is based on his previous study [9].

References

[1] R. Okuda, Y. Kajiwara, and K. Terashima, "A survey of technical trend of adas and autonomous driving," in *Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test*, pp. 1–4, IEEE, 2014.

[2] D. A. Lazar, R. Pedarsani, K. Chandrasekher, and D. Sadigh, "Maximizing road capacity using cars that influence people," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 1801–1808, IEEE, 2018.

[3] T. L. Griffiths, F. Lieder, and N. D. Goodman, "Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic," *Topics in cognitive science*, vol. 7, no. 2, pp. 217–229, 2015.

[4] L. Claussmann, A. Carvalho, and G. Schildbach, "A path planner for autonomous driving on highways using a human mimicry approach with binary decision diagrams," in *2015 European Control Conference (ECC)*, pp. 2976–2982, IEEE, 2015.

[5] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic decision-making under uncertainty for autonomous driving using continuous pomdps," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 392–399, IEEE, 2014.

[6] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions.," in *Robotics: Science and Systems*, vol. 2, Ann Arbor, MI, USA, 2016.

[7] D. O. Stahl, "Evolution of smartn players," *Games and Economic Behavior*, vol. 5, no. 4, pp. 604–617, 1993.

[8] N. Li, D. W. Oyler, M. Zhang, Y. Yildiz, I. Kolmanovsky, and A. R. Girard, "Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems," *IEEE Transactions on control systems technology*, vol. 26, no. 5, pp. 1782–1797, 2017.

[9] N. Li, I. Kolmanovsky, A. Girard, and Y. Yildiz, "Game theoretic modeling of vehicle interactions at unsignalized intersections and application to autonomous vehicle control," in *2018 Annual American Control Conference (ACC)*, pp. 3215–3220, IEEE, 2018.

[10] R. Tian, S. Li, N. Li, I. Kolmanovsky, A. Girard, and Y. Yildiz, "Adaptive game-theoretic decision making for autonomous vehicle control at roundabouts," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 321–326, IEEE, 2018.

[11] G. I. Jin, S. Bastian, M. M. Richard, and A. Matthias, "Risk-aware motion planning for automated vehicle among human-driven cars," in *2019 American Control Conference (ACC)*, pp. –, IEEE, 2019.

[12] P. O. Scokaert and D. Mayne, "Min-max feedback model predictive control for constrained linear systems," *IEEE Transactions on Automatic control*, vol. 43, no. 8, pp. 1136–1142, 1998.

[13] D. Q. Mayne, M. M. Seron, and S. Raković, "Robust model predictive control of constrained linear systems with bounded disturbances," *Automatica*, vol. 41, no. 2, pp. 219–224, 2005.

[14] A. Richards and J. P. How, "Model predictive control of vehicle maneuvers with guaranteed completion time and robust feasibility," in *American Control Conference, 2003. Proceedings of the 2003*, vol. 5, pp. 4034–4040 vol.5, June 2003.

[15] G. S. Sankar, R. C. Shekhar, C. Manzie, T. Sano, and H. Nakada, "Fast calibration of a robust model predictive controller for diesel engine airpath," *IEEE Transactions on Control Systems Technology*, pp. 1–15, 2019.

[16] G. S. Sankar, R. C. Shekhar, C. Manzie, T. Sano, and H. Nakada, "Model predictive controller with average emissions constraints for diesel airpath," *Control Engineering Practice*, vol. 90, pp. 182 – 189, 2019.

[17] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli, "Kinematic and dynamic vehicle models for autonomous driving control design," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1094–1099, IEEE, 2015.

[18] M. A. Costa-Gomes and V. P. Crawford, "Cognition and behavior in two-person guessing games: An experimental study," *American Economic Review*, vol. 96, no. 5, pp. 1737–1768, 2006.

[19] M. A. Costa-Gomes, N. Iriberri, and V. P. Crawford, "Comparing models of strategic thinking in van huyck, battalio, and beil's coordination games," *Journal of the European Economic Association*, vol. 7, no. 2/3, pp. 365–376, 2009.