Interaction-Aware Game-Theoretic Motion Planning for Automated Vehicles using Bi-level Optimization

Christoph Burger[1], Johannes Fischer[1], Frank Bieder[2], Ömer Şahin Taş[2] and Christoph Stiller[1]

***Abstract*— Planning an interactive and cooperative behavior in the vicinity of multiple decision-makers is a challenging task. The game-theoretic perspective provides a suitable framework to describe such interactive scenarios. In this paper, we introduce a motion planning algorithm to generate interaction-aware behavior for highly interactive scenarios. Our algorithm is based upon a reformulation of a bi-level optimization problem which frames interactions among two decision makers as a Stackelberg game. In contrast to existing works in this field, we solve the prediction and planning problem simultaneously, which enables the generation of efficient behaviors even in highly interactive situations. The main novelty of our algorithm evolves around its ability to consider general nonlinear constraints. Further, we present mechanisms to introduce courtesy and cooperation into behavior planning which prevents overly aggressive driving, as issue that has been identified in existing interaction-aware planning approaches. Finally, we evaluate our approach in the context of automated driving. Our evaluation first investigates the algorithm's ability to purposefully influence and exploit the response of surrounding vehicles. We then illustrate how the approach can be used for cooperative and courteous planning.**

I. INTRODUCTION

While much progress has been made in motion planning for automated vehicles (AVs), generating an efficient driving behavior in interactive scenarios is still an open field of research. Many planning approaches follow a *predict-then-plan* scheme, where the future motion of surrounding agents is predicted first, and in a subsequent step, the motion of the AV is planned considering these predictions as fixed. Due to this separation of prediction and planning, interactions with other agents are neglected. The influence the AV's actions might have on other agents can not be considered during planning which can result in suboptimal, overly defensive driving. Some approaches are already able to overcome this limitation by solving the prediction and planning task simultaneously. Most techniques can be categorized into the following three distinct classes: Multi-agent planning, forward simulation methods, and game-theoretic planning.

In *multi-agent planning*, the underlying assumption is that all agents in the environment are part of the same team, i.e., working towards a common goal [1]–[4]. Coupled behaviors are generated by minimizing a joint cost function, further assuming that every agent will follow the joint plan. In such multi-agent planning approaches, different driver types or asymmetries in the traffic scene can be incorporated by using different weights for the costs of individual agents in the joint cost function [5]. In real traffic, however, the assumption of a common cost function might not be valid, since some drivers are only interested in optimizing their own costs. In [6] and [7], the behavior of other agents is tracked to infer if the agents optimize for the same objective or something entirely different.

A different way of generating interactive behavior is to determine the reaction of other agents by forward simulations of the current traffic scene. To utilize these methods, transition models are required which describe how other agents respond to changes in the traffic scene due to actions taken by the ego agent. We refer to such techniques as

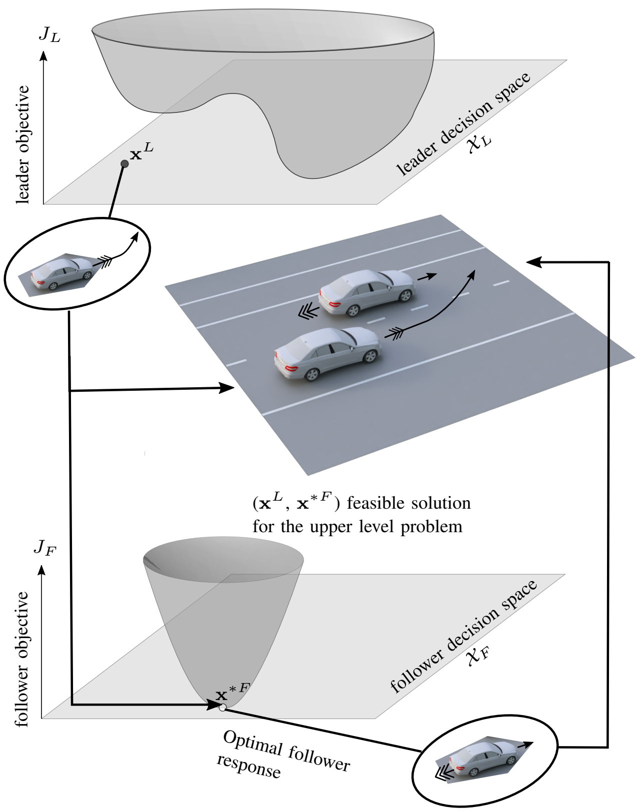


Fig. 1: Illustrated is the structure of a Stackelberg game modeled as a bi-level optimization problem. Here, the follower optimizes its objective function as a response to the given actions of the leader.

[1]Christoph Burger, Johannes Fischer and Christoph Stiller are with the Institute of Measurement and Control Systems, Karlsruhe Institute of Technology (KIT), 76131, Karlsruhe, GERMANY. {christoph.burger,johannes.fischer, stiller}@kit.edu

[2]Frank Bieder and Ömer Şahin Taş are with FZI Research Center for Information Technology, 76131, Karlsruhe, GERMANY. {bieder,tas}@fzi.de

forward simulation methods. Most sampling-based planning methods that consider interactions can be associated with this category. One approach is to generate a set of candidate motion profiles that consider the disturbance on other agents [8], another approach is to model motion planning as a sequential decision-making problem, *e.g.* as a POMDP model, and cover interactions in environment state updates [9]. The reactions of other agents in the environment are typically modeled with specific driver models such as the Intelligent Driver Model (IDM) [10] or the Minimum Overall Braking Induced by Lane change (MOBIL) model [11]. In contrast to optimization-based methods, the influence the ego agent exerts on others is not explicitly given, but must be determined by trying out several actions and subsequent forward simulations.

Game-theoretic approaches have proven to be effective for capturing the interactions between agents with different objectives and have been successfully used for lane changes, merge scenarios, intersection crossing, traversing roundabouts, and overtaking [12]–[20]. Apart from these driving applications, these game-theoretic planning approaches have been applied for agile maneuvering of multiple ground vehicles in close proximity [21], and automated car racing [22]–[25], where they are shown to be superior to baseline MPC planners.

In game-theoretic formulations, there is no optimal solution in the traditional sense, but depending on the game's structure, different solutions are possible, also referred to as *equilibria*. In literature, it is distinguished between *Nash* and *Stackelberg* equilibria. A Nash equilibrium describes a set of strategies where no individual agent can benefit from unilaterally changing its strategy, given that all other agents will stick to their strategy. This type of equilibrium is e.g. used in [15], [16], [21], [22], [26], [27]. In contrast to a Nash equilibrium, a Stackelberg equilibrium involves turn taking and, therefore considers an asymmetry in the decision-making process. It is typically applied to two player games, with one agent being the *leader* and one being the *follower*. The leader selects its strategy first and then the follower optimizes its strategy considering the leader's decision, see Fig. 1. Motion planning approaches that solve for Stackelberg equilibria are presented in [12], [13], [17], [23], [28]–[30].

In this work, we model interaction-aware motion planning as a Stackelberg game. We assume a turn-taking structure in interactive scenarios, where the AV initiates an interaction and the human driver (HD) reacts to the actions of the AV. This leads to a bi-level optimization problem, where the HD's optimization is an optimization inside the AV's optimization. To efficiently solve the bi-level problem, it is reformulated to a single-level representation using the KKT conditions of the HD's optimization problem.

Compared to previous works, our algorithm i) is able to consider general nonlinear constraints, thereby ensuring feasibility of the solution, ii) is able to consider the impact that own actions have on the driving behavior and comfort of other agents, thereby providing a mechanism to introduce cooperation and courtesy into motion planning which can then be used to cope with the overly aggressive behavior produced by previous interaction-aware planning approaches.

The remainder of this paper is structured as follows: In Section II we present the game-theoretic problem formulation for interaction-aware planning in the considered AV-HD system. Next, in Section III we present how this game can be approximated by a bi-level optimization problem. In Section IV, we transform the bi-level optimization problem into a single-level representation to efficiently use derivative-based optimization methods. After presenting the details on implementation, results from simulation studies are discussed in Section V. Finally, in Section VI, we summarize our findings and conclude our work.

II. Problem Statement

In this work, we develop a model to directly capture interactions between automated vehicles and human drivers based on a game theoretic formulation. We consider a simplified system with one AV, representing the leader L, and one HD, representing the follower F. Following the formalism of [13], the state of the system at time t is given by the leader's and follower's state $\mathbf{x}_t^L, \mathbf{x}_t^F \in \mathcal{X}$. The evolutions of the leader's and the follower's state are described by their trajectories $\xi_L, \xi_F : [0,T] \rightarrow \mathcal{X}$. Further, each agent has its individual objective function denoted by J_L and J_F. The objective is minimized subject to the initial state of the vehicle $\xi(0) = \mathbf{x}_{\text{init}}$ and $\xi(t)$ which is only allowed to pass through the set of feasible states $\mathcal{X}_{\text{feasible}}(t) \subseteq \mathcal{X}$. $\mathcal{X}_{\text{feasible}}(t)$ encodes for instance collision avoidance and the system dynamics. Additionally, dynamic constraints can be enforced by $F(\xi(t), \dot{\xi}(t), \ddot{\xi}(t), \dots) = 0$.

In contrast to traditional multi-agent systems, the follower is assumed to optimize its own trajectory in response to the trajectory of the leader. To do so, the follower makes a prediction of the leader's future motion $\tilde{\xi}_L$ and plans its future motion by minimizing its objective function J_F given this prediction. Therefore, the follower's optimal trajectory can be stated as:

$$\xi_F^*(\tilde{\xi}_L) = \arg\min_{\xi_F} J_F(\xi_F, \tilde{\xi}_L) \tag{1}$$

In our work, we assume that for short time horizons, a HD can predict the trajectory of an AV sufficiently well, such that the prediction $\tilde{\xi}_L$ can be assumed to be the actual trajectory ξ_L of the AV. Hence, the optimal trajectory of the HD as a function of the AV's actual trajectory ξ_L is given as:

$$\xi_F^*(\xi_L) = \arg\min_{\xi_F} J_F(\xi_F, \xi_L) \tag{2}$$

Equation (2) gives the AV the ability to reason about the HD's response and allows to indirectly control the HD's future trajectory.

This link with the follower's actions allows the leader in turn to optimize its own behavior using:

$$\xi_L^* = \arg\min_{\xi_L} J_L(\xi_L, \xi_F^*(\xi_L)) \tag{3}$$

The derived model can be described as a Stackelberg game, where the leader decides on its behavior first and the follower optimizes its behavior given the decision of the leader.

III. Bi-Level Formulation

If the best response of the follower to the leader's actions is known in closed form, Eq. (3) can be solved as a standard optimal control problem (OCP). However, in our formulation, the follower's response is itself given as an OCP, resulting in a nested optimization. To efficiently solve the OCP, we use model-predictive control (MPC) with a multiple shooting method and discretize the time horizon T into N intervals. With a slight abuse of notation we subsume the state and input sequences of leader and follower as $\mathbf{x} := (\mathbf{x}_1, \dots, \mathbf{x}_N)$ and $\mathbf{u} := (\mathbf{u}_0, \dots, \mathbf{u}_{N-1})$. In the following OCP formulations, the equality constraints $\mathbf{h}$ can be used to represent constraints imposed by the system dynamics model while the inequality constraints $\mathbf{g}$ collect bound constraints, collision constraints, and dynamic constraints.

A. OCP of the Follower

Given state sequence $\mathbf{x}^L$ of the leader, the follower's OCP can be formulated as:

$$\begin{aligned} \min_{\mathbf{x}^F, \mathbf{u}^F} \quad & J_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) && \text{(4a)} \\ \text{s.t.} \quad & \mathbf{h}_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) = 0, && \text{(4b)} \\ & \mathbf{g}_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) \leq 0 && \text{(4c)} \end{aligned}$$

B. OCP of the Leader

Following Eq. (3) the leader's OCP is given by:

$$\begin{aligned} \min_{\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L, \mathbf{u}^F} \quad & J_L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L) && \text{(5a)} \\ \text{s.t.} \quad & \mathbf{h}_L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L) = 0, && \text{(5b)} \\ & \mathbf{g}_L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L) \leq 0, && \text{(5c)} \\ & (\mathbf{x}^F, \mathbf{u}^F) \text{ solves Eq. (4) for } \mathbf{x}^L && \text{(5d)} \end{aligned}$$

making use of the follower's optimal response.

IV. Single Level Reformulation

If the follower OCP is convex, the Karush Kuhn Tucker (KKT) conditions are necessary and sufficient for optimality. In general this is not the case but can be achieved by linearizing the constraints and approximating the objective with a second order Taylor expansion.

Henceforth, the bi-level optimization problem, Eq. (5), can be reformulated as a single level problem by replacing the follower's optimization problem, Eq. (4), with its KKT conditions, resulting in

$$\begin{aligned} \min_{\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L, \mathbf{u}^F, \boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & J_L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L) && \text{(6a)} \\ \text{s.t.} \quad & \mathbf{h}_L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L) = 0, && \text{(6b)} \\ & \mathbf{g}_L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^L) \leq 0, && \text{(6c)} \\ & \nabla_{(\mathbf{x}^F, \mathbf{u}^F)} L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0, && \text{(6d)} \\ & \mathbf{h}_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) = 0, && \text{(6e)} \\ & \mathbf{g}_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) \leq 0, && \text{(6f)} \\ & \boldsymbol{\mu} \geq 0, && \text{(6g)} \\ & \boldsymbol{\mu} \perp \mathbf{g}_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) && \text{(6h)} \end{aligned}$$

with the Lagrangian

$$\begin{aligned} L(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F, \boldsymbol{\lambda}, \boldsymbol{\mu}) = & J_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) \\ & + \boldsymbol{\lambda}^T \mathbf{h}_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F) \\ & + \boldsymbol{\mu}^T \mathbf{g}_F(\mathbf{x}^L, \mathbf{x}^F, \mathbf{u}^F). \end{aligned}$$

A. Solving the Complementarity Constraints

The leader OCP forms an instance of a mathematical program with complementarity constraints (MPCC). Due to the complementarity constraints, MPCCs are non-smooth and non-convex, which makes them particularly challenging to solve: At every feasible point, ordinary constraint qualifiers (CQ) such as LICQ or Mangasarian-Fromovitz CQ are violated [31]. To solve the MPCC, we reformulate it using relaxation methods [32]. Here, the complementarity constraints are relaxed as follows:

$$-\epsilon \leq \boldsymbol{\mu}^T \mathbf{g}_F \tag{7}$$

V. Results and Evaluation

In our experiments, we first investigate the ability of the leader to deliberately influence and exploit the follower's response.

Since in real driving applications, the goal of the AV is to drive efficiently and comfortably rather than to influence the state of other vehicles, we also illustrate how the approach can be used for cooperative and courteous planning.

To this end, we showcase the efficacy of the approach in four selected minimal examples. The leader's objective function is augmented with $J_{\text{influence}}$ to set the example specific incentives. For the conducted experiments we assume a relatively good estimation of the human's objective function is given.

A. Trajectory Optimization for AVs

1) Vehicle Model: The vehicle state $\mathbf{x} = (x, y, \psi, v)$ of the leader and the follower is described by the lateral and longitudinal position (x, y) of the vehicle's center of gravity, the orientation ψ, and the absolute velocity v. Together with the input $\mathbf{u} = (\delta, a)$ consisting of steering angle δ and acceleration a, the dynamics of the vehicles are given by the kinematic single-track model

$$\dot{\mathbf{x}} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\psi} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} v \cos(\psi + \beta) \\ v \sin(\psi + \beta) \\ \frac{v}{l} \tan(\delta) \cos(\beta) \\ a \end{pmatrix} \tag{8}$$

Parameter	Value
N	30
T	6 s
Q	$\mathrm{diag}(0, 1, 0, 100)$
R_u	$\mathrm{diag}(1, 1)$
$R_{\dot{u}}$	$\mathrm{diag}(10000, 1000)$
$v_{\min}, v_{\max}$	$0\,\frac{\mathrm{m}}{\mathrm{s}}$, $30\,\frac{\mathrm{m}}{\mathrm{s}}$
$\delta_{\max}$	30°
$a_{\min}, a_{\max}$	$-8\,\frac{\mathrm{m}}{\mathrm{s}^2}$, $3\,\frac{\mathrm{m}}{\mathrm{s}^2}$
$j_{\min}, j_{\max}$	$-10\,\frac{\mathrm{m}}{\mathrm{s}^3}$, $6\,\frac{\mathrm{m}}{\mathrm{s}^3}$
l	4 m
l_r	2 m

TABLE I: MPC parameters

with slip angle $\beta = \arctan\left(\frac{l_r}{l}\tan(\delta)\right)$. Here, l is the wheelbase and l_r is the difference between the center of gravity and the rear axle. A discrete dynamics model is obtained using a fourth order Runge-Kutta method.

To enforce realistic dynamic limits, bound constraints on the states and input signals are introduced. The values used are given in Table I. To ensure the validity of the single-track model, we additionally limit the lateral acceleration by $|v_k\dot{\psi}_k| \leq 4\,\frac{\mathrm{m}}{\mathrm{s}^2}$ [33].

2) Collision Avoidance Constraints: We approximate the collision constraints by overapproximating each vehicle with two circles (x_i, y_i, r_i). The circles are then pairwise checked for overlap using:

$$\left(\frac{x_i - x_j}{2r}\right)^2 + \left(\frac{y_i - y_j}{2r}\right)^2 \geq 1 \tag{9}$$

3) Objective Function: We use the following cost function to penalize deviations from a desired reference state or trajectory $\mathbf{x}_{\text{ref}}$:

$$J_{\text{base}}(\mathbf{x}, \mathbf{u}) = \sum_{k=1}^{N}(\mathbf{x}_k - \mathbf{x}_{\text{ref}})^T Q(\mathbf{x}_k - \mathbf{x}_{\text{ref}}) \tag{10}$$

$$+ \sum_{k=0}^{N-1} \mathbf{u}_k{}^T R_u \mathbf{u}_k \tag{11}$$

$$+ \sum_{k=1}^{N-1}(\mathbf{u}_k - \mathbf{u}_{k-1})^T R_{\dot{u}}(\mathbf{u}_k - \mathbf{u}_{k-1}) \tag{12}$$

$$+ (\mathbf{u}_0 - \hat{\mathbf{u}})^T R_{\dot{u}}(\mathbf{u}_0 - \hat{\mathbf{u}}) \tag{13}$$

where $\hat{\mathbf{u}}$ is the input from the previous step.

The follower uses this cost function directly with the weights given in Table I. The leader's objective function is also based on J_{base} but has an additional scenario specific cost term $J_{\text{influence}}$.

B. Base Scenario

We evaluate our approach in a multi-lane road with two vehicles, the leader (blue) and the follower (gray), see Fig. 2. In the considered scenario, the goal of the leader is to perform a lane change. Unless otherwise stated, the initial and reference states used for the leader and follower are listed in Table II.

Parameter	Value
$\mathbf{x}_0^L$	$[12.0\,\mathrm{m}, 3.0\,\mathrm{m}, 0^\circ, 10.0\,\frac{\mathrm{m}}{\mathrm{s}}]^T$
$\mathbf{x}_0^F$	$[2.0\,\mathrm{m}, 5.0\,\mathrm{m}, 0^\circ, 10.0\,\frac{\mathrm{m}}{\mathrm{s}}]^T$
$\mathbf{x}_{\text{ref}}^L = \mathbf{x}_{\text{ref}}^F$	$[\text{free}, 5.0\,\mathrm{m}, 0^\circ, 10.0\,\frac{\mathrm{m}}{\mathrm{s}}]^T$

TABLE II: Initial and reference states.

C. Influence the state of the human

To investigate the leader's ability to influence the follower's state, the objective function of the leader is chosen as a weighted sum of the leader's individual costs J_{base} and the scenario specific $J_{\text{influence}}$:

$$J = w_L J_{\text{base}} + w_{\text{influence}} J_{\text{influence}} \tag{14}$$

where w_L and $w_{\text{influence}}$ are the corresponding weights.

1) Slow down the human: To incentivise the leader to slow down the follower, deviations of the follower's velocity along the road to a certain reference velocity v_{ref}^F are penalized in the leader's objective:

$$J_{\text{influence}} = \sum_{k=1}^{N}(v_k^F - v_{\text{ref}}^F)^2 \tag{15}$$

The results for $v_{\text{ref}}^F = 7.5\,\frac{\mathrm{m}}{\mathrm{s}}$ are illustrated in Fig. 2. As can be seen, the leader changes to the left lane to get in front of the follower. The leader then starts to brake which forces the follower to also slow down, see Fig. 2b.

2) Push the human to the adjacent lane: To demonstrate the ability to influence the follower in the lateral direction, the follower's deviation to a certain lateral position is penalized:

$$J_{\text{influence}} = \sum_{k=1}^{N}(y_k^F - y_{\text{ref}}^F)^2 \tag{16}$$

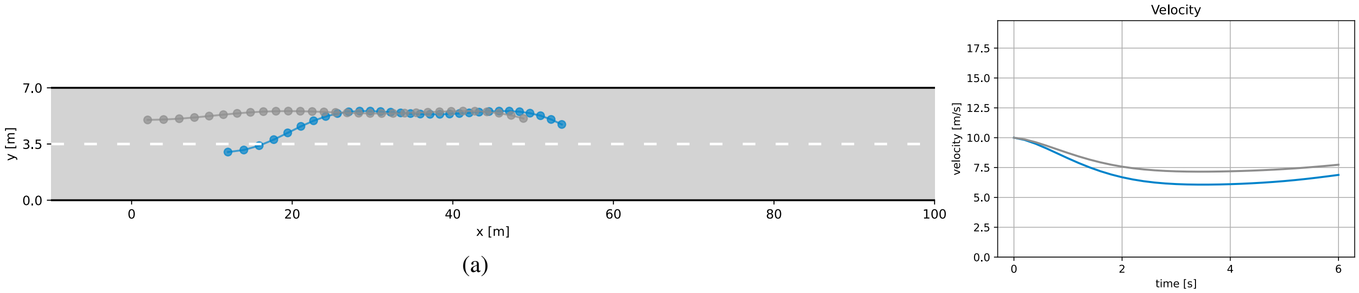
In our experiment, the goal of the leader is to push the human to the adjacent left lane, which is encoded by setting $y_{\text{ref}}^F = 8.5\,\mathrm{m}$. The results are shown in Fig. 3a. The leader brakes harshly and steers to the left to push the follower to the left lane as soon as possible. The velocity profile is shown in Fig. 3b.

D. Exploiting interaction

In the following experiment, the leader solely optimizes its own costs J_{base}. To better show the effect, the desired velocity of the follower is increased to $v_{\text{ref}}^F = 15.0\,\frac{\mathrm{m}}{\mathrm{s}}$

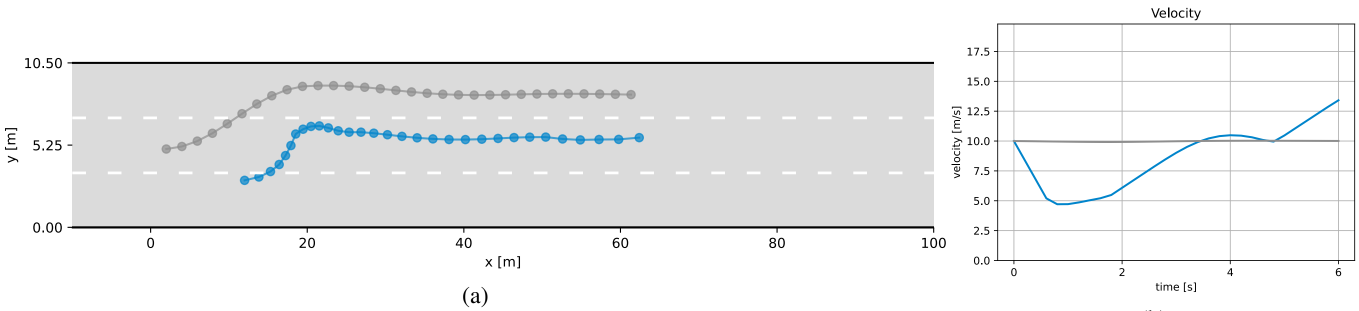
The resulting trajectories are shown in Fig. 4a. Since the leader expects the follower to react to him, he changes lanes right away. To avoid a collision, the follower has to slow down and is forced to stay behind the slower driving leader, see Fig. 4b.

By exploiting the follower's response, the leader plans a more aggressive driving behavior to further optimize its costs.



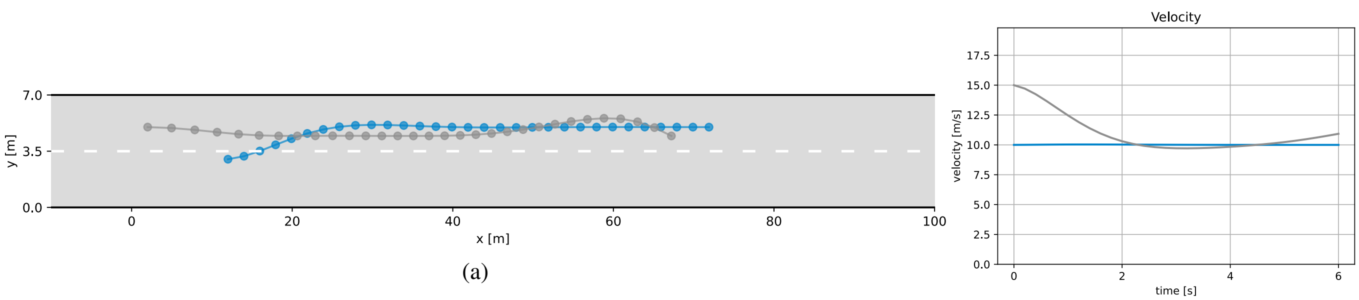
(a) (b)

Fig. 2: By changing lanes, the leader drives in front of the follower to slow him down.



(a) (b)

Fig. 3: To push the follower to the leftmost lane, the leader brakes harshly and steers to the left.



(a) (b)

Fig. 4: When only considering it's own costs the leader performs an aggressive lane change.

E. Introducing Cooperative Behavior

The bi-level formulation gives the leader the ability to anticipate the reactions of the follower. To prevent the leader from exploiting this ability, as it was shown in the previous experiment, we introduce a cooperative cost function. This cost function includes the costs of the follower and the leader:

$$J_{\text{cooperative}} = \alpha J^F_{\text{base}} + (1-\alpha) J^L_{\text{base}} \tag{17}$$

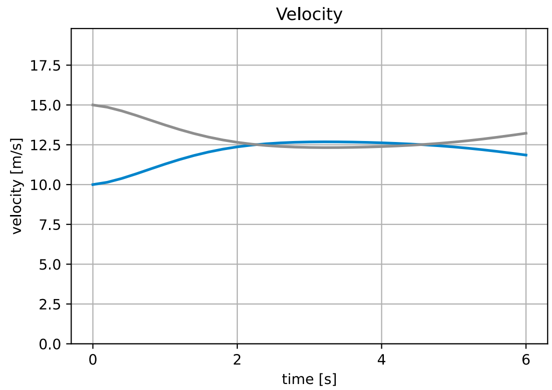
In this formulation the variable α determines to which extent the leader's and the follower's costs are considered. It therefore provides a way to design a driving behavior that is between overly conservative and overly aggressive. The impact the parameter α has on the generated behavior is investigated in the following.

The effect is best illustrated by the different velocity profiles. The case with $\alpha = 0$ represents the egoistic case which was presented in the previous experiment, see Fig. 4b. The different velocities for $\alpha = 0.5$ and $\alpha = 0.99$ are illustrated in Fig. 5. Compared to the $\alpha = 0$ case, with $\alpha = 0.5$ the leader accelerates and drives faster than its desired velocity. With $\alpha = 0.99$ the leader basically only considers the costs of the follower and tries to intervene with the optimal plans of the follower as little as possible. As shown in Fig. 5b, this value of α can lead to an overly conservative driving behavior, similar to common predict-then-react planners.

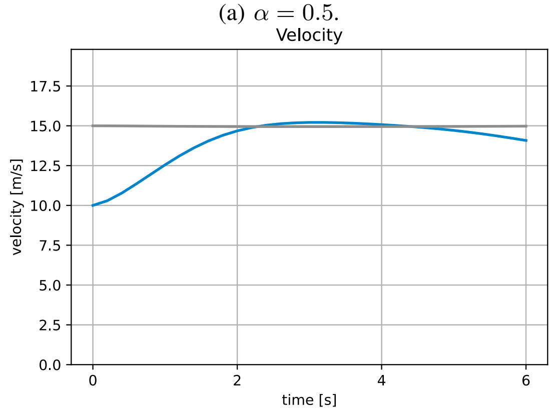
F. Courtesy constraints

While considering courtesy is not necessary in predict-then-plan algorithms, it becomes important in interaction aware method. An alternative to the cooperative formulation presented in Section V-E is introducing *courtesy* constraints.

With these constraints the impact the automated vehicle imposes on others can be limited. E.g., the automated vehicle is allowed to maximally cause a deceleration of a_{min} to

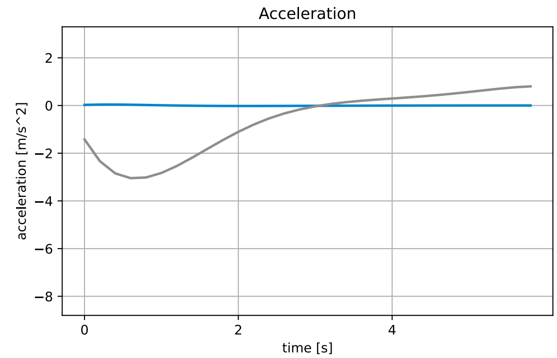


(a) $\alpha = 0.5$.

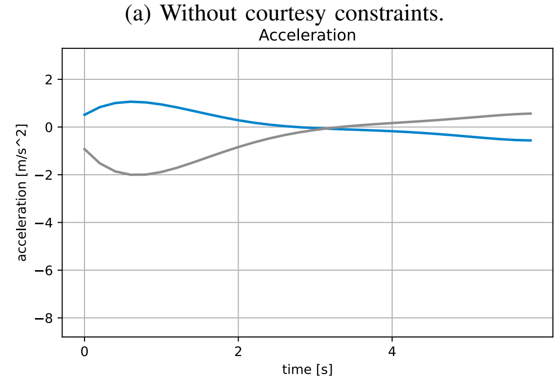


(b) $\alpha = 0.99$.

Fig. 5: With the parameter α the leader's level of cooperativity can be set.



(a) Without courtesy constraints.



(b) Courtesy constraints with $a_{\min} = 2$.

Fig. 6: With the added courtesy constraints the maximal deceleration of the follower can be reduced.

surrounding vehicles.

To enforce this, the following constraints are added to the OCP of the leader Eq. (6a):

$$\mathbf{g}_{\text{courtesy},k} = a_k^F + a_{\min} \geq 0 \tag{18}$$

where a_k^F is the acceleration of the follower in step k.

The effect for $a_{\min} = 2$ is illustrated in Fig. 6. In Fig. 6a the acceleration of the leader and the follower without courtesy constraints are shown. The leader does not adapt its velocity during the lane change and the follower has to brake harshly. In contrast, in Fig. 6b the same scenario is shown with courtesy constraints. To limit the induced deceleration, the leader accelerates, which successfully reduces the deceleration of the follower to at most $a_{\min}$.

G. Discussion

The presented bi-level optimization formulation enables the automated vehicle to anticipate how surrounding vehicles will react to its future motion. This gives the AV the possibility to indirectly influence the state of the human, as shown in Section V-C. Care must be taken to not generate an overly aggressive driving behavior, as shown in experiment Section V-D.

To avoid this, the objective of the leader is extended to also consider the costs of the follower in Section V-E. The parameter α can be used to tune the resulting behavior between being egoistic and courteous.

Finally, in Section V-F we present a strategy to introduce courtesy constraints to the planning algorithm. These constraints allow realizing different behaviors. For instance, the vehicle can be made to act egoistically, provided that a maximally acceptable deceleration imposed on other vehicles is not exceeded.

Even though the focus of the evaluation was to investigate the AV's ability to indirectly control the state of the human, the intended use case is to combine cooperative planning with courtesy constraints.

VI. Conclusions and Future Work

This work presents a bi-level optimization scheme for interactive and courteous driving behavior. Based on the assumption that human drivers can reasonably well predict and react to other vehicles on short time horizons, we formulate a game-theoretic approach to motion planning. The human's best response to a trajectory of the automated vehicle can be determined by solving an optimal control problem (OCP). Optimizing the autonomous vehicle's trajectory subject to the human's optimal response results in a bi-level problem formulation. To efficiently solve the bi-level problem, it is reformulated as a single-level problem using the KKT conditions of the human's OCP.

Our evaluation first showcases different ways to influence or exploit the human driver. We then continue with implementations of cooperative driving behavior. This can be achieved by considering the human's cost function as part of the objective or by imposing constraints on the acceleration induced on the human. Our approach allows interactive motion planning while satisfying constraints on the surrounding vehicles' responses.

Our method is able to overcome the limitations of traditional predict-then-plan approaches enabling the efficient planning in highly interactive situations.

Promising directions for future research are to apply the method to more realistic traffic situations. Additionally, the reward function parameters can be learned from real driving data using constrained inverse reinforcement learning [34] to produce more realistic behavior.

Acknowledgements

The authors thank the German Research Foundation (DFG) for being funded within the German collaborative research center "SPP 1835 – Cooperative Interacting Automobiles" (CoInCar).

References

[1] G. R. de Campos, P. Falcone, and J. Sjöberg, "Autonomous cooperative driving: A velocity-based negotiation approach for intersection crossing", in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, Oct. 2013, pp. 1456–1461.

[2] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard, "Socially compliant mobile robot navigation via inverse reinforcement learning", *The International Journal of Robotics Research*, vol. 35, no. 11, pp. 1289–1307, Sep. 2016, ISSN: 0278-3649, 1741-3176.

[3] D. Lenz, T. Kessler, and A. Knoll, "Tactical cooperative planning for autonomous highway driving using Monte-Carlo Tree Search", in *2016 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2016, pp. 447–453.

[4] S. Bansal, A. Cosgun, A. Nakhaei, and K. Fujimura. (Aug. 7, 2018). Collaborative Planning for Mixed-Autonomy Lane Merging. arXiv: `1808.02550 [cs]`, [Online]. Available: `http://arxiv.org/abs/1808.02550` (visited on 09/02/2021).

[5] C. Burger and M. Lauer, "Cooperative Multiple Vehicle Trajectory Planning using MIQP", in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 602–607.

[6] J. Schulz, K. Hirsenkorn, J. Löchner, M. Werling, and D. Burschka, "Estimation of collective maneuvers through cooperative multi-agent planning", in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2017, pp. 624–631.

[7] C. Burger, T. Schneider, and M. Lauer, "Interaction aware cooperative trajectory planning for lane change maneuvers in dense traffic", in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, Sep. 2020, pp. 1–8.

[8] N. Evestedt, E. Ward, J. Folkesson, and D. Axehill, "Interaction aware trajectory planning for merge scenarios in congested traffic situations", in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2016, pp. 465–472.

[9] C. Hubmann, J. Schulz, G. Xu, D. Althoff, and C. Stiller, "A Belief State Planner for Interactive Merge Maneuvers in Congested Traffic", in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 1617–1624.

[10] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations", *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, Aug. 1, 2000, ISSN: 1063-651X, 1095-3787.

[11] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model mobil for car-following models", *Transp. Res. Rec. J. of the Transp. Res. Board*, no. 1999, pp. 86–94, 2007.

[12] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan, "Information gathering actions over human internal state", in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2016, pp. 66–73.

[13] 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 XII*, Robotics: Science and Systems Foundation, 2016, ISBN: 978-0-9923747-2-3.

[14] D. Sadigh, N. Landolfi, S. S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for cars that coordinate with people: Leveraging effects on human actions for planning and active information gathering over human internal state", *Autonomous Robots*, vol. 42, no. 7, pp. 1405–1426, Oct. 1, 2018, ISSN: 1573-7527.

[15] D. Fridovich-Keil, E. Ratner, L. Peters, A. D. Dragan, and C. J. Tomlin. (Mar. 18, 2020). Efficient Iterative Linear-Quadratic Approximations for Nonlinear Multi-Player General-Sum Differential Games. arXiv: `1909.04694 [cs, eess]`, [Online]. Available: `http://arxiv.org/abs/1909.04694` (visited on 09/10/2021).

[16] A. Dreves and M. Gerdts, "A generalized Nash equilibrium approach for optimal control problems of autonomous cars", *Optimal Control Applications and Methods*, vol. 39, no. 1, pp. 326–342, 2018, ISSN: 1099-1514.

[17] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan. (Oct. 12, 2018). Hierarchical Game-Theoretic Planning for Autonomous Vehicles. arXiv: `1810.05766 [cs, math]`, [Online]. Available: `http://arxiv.org/abs/1810.05766` (visited on 07/22/2020).

[18] M. Wang, N. Mehr, A. Gaidon, and M. Schwager, "Game-Theoretic Planning for Risk-Aware Interactive Agents", in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA: IEEE, Oct. 24, 2020, pp. 6998–7005, ISBN: 978-1-72816-212-6.

[19] S. Lecleac'h, M. Schwager, and Z. Manchester, "LUCIDGames: Online Unscented Inverse Dynamic Games for Adaptive Trajectory Prediction and Planning", *IEEE Robotics and Automation Letters*, pp. 1–1, 2021, ISSN: 2377-3766, 2377-3774.

[20] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game-Theoretic Planning for Self-Driving Cars in Multivehicle Competitive Scenarios", *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1313–1325, Aug. 2021, ISSN: 1941-0468.

[21] G. Williams, B. Goldfain, P. Drews, J. M. Rehg, and E. A. Theodorou, "Best Response Model Predictive Control for Agile Interactions Between Autonomous Ground Vehicles", in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018, pp. 2403–2410.

[22] M. Wang, Z. Wang, J. Talbot, J. Christian Gerdes, and M. Schwager, "Game Theoretic Planning for Self-Driving Cars in Competitive Scenarios", in *Robotics: Science and Systems XV*, Robotics: Science and Systems Foundation, Jun. 22, 2019, ISBN: 978-0-9923747-5-4.

[23] A. Liniger and J. Lygeros, "A Noncooperative Game Approach to Autonomous Racing", *IEEE Transactions on Control Systems Technology*, vol. 28, no. 3, pp. 884–897, May 2020, ISSN: 1558-0865.

[24] G. Notomista, M. Wang, M. Schwager, and M. Egerstedt, "Enhancing Game-Theoretic Autonomous Car Racing Using Control Barrier Functions", in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France: IEEE, May 2020, pp. 5393–5399, ISBN: 978-1-72817-395-5.

[25] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game-Theoretic Planning for Self-Driving Cars in Multivehicle Competitive Scenarios", *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1313–1325, Aug. 2021, ISSN: 1552-3098, 1941-0468.

[26] S. L. Cleac'h, M. Schwager, and Z. Manchester. (May 30, 2021). ALGAMES: A Fast Augmented Lagrangian Solver for Constrained Dynamic Games. arXiv: `2104.08452 [cs]`, [Online]. Available: `http://arxiv.org/abs/2104.08452` (visited on 09/03/2021).

[27] A. Britzelmeier, A. Dreves, and M. Gerdts, "Numerical solution of potential games arising in the control of cooperative automatic vehicles", in *2019 Proceedings of the Conference on Control and Its Applications (CT)*, ser. Proceedings, Society for Industrial and Applied Mathematics, Jan. 1, 2019, pp. 38–45.

[28] J. Yoo and R. Langari, "Stackelberg Game Based Model of Highway Driving", Oct. 17, 2012.

[29] J. H. Yoo and R. Langari, "A Stackelberg Game Theoretic Driver Model for Merging", presented at the ASME 2013 Dynamic Systems and Control Conference, American Society of Mechanical Engineers Digital Collection, Mar. 6, 2014.

[30] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan. (Aug. 15, 2018). Courteous Autonomous Cars. arXiv: `1808.02633 [cs]`, [Online]. Available: `http://arxiv.org/abs/1808.02633` (visited on 09/01/2021).

[31] H. Chen, H. Kremling, and F. Allgöwer, "Nonlinear predictive control of a benchmark cstr", *Proceedings of the 3rd European Control Conference, Rome-Italy.*, pp. 3247–3252, Jan. 1995.

[32] T. Hoheisel, C. Kanzow, and A. Schwartz, "Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints", *Mathematical Programming*, vol. 137, no. 1, pp. 257–288, Feb. 1, 2013, ISSN: 1436-4646.

[33] P. Polack, F. Altché, B. Novel, and A. de La Fortelle, "The kinematic bicycle model: A consistent model for planning feasible trajectories for autonomous vehicles?", Jun. 2017, pp. 812–818.

[34] J. Fischer, C. Eyberg, M. Werling, and M. Lauer, "Sampling-based Inverse Reinforcement Learning Algorithms with Safety Constraints", in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2021, pp. 791–798.