Game-Theoretic Planning for Risk-Aware Interactive Agents

Mingyu Wang[1], Negar Mehr[1], Adrien Gaidon[2], and Mac Schwager[1]

***Abstract*— Modeling the stochastic behavior of interacting agents is key for safe motion planning. In this paper, we study the interaction of risk-aware agents in a game-theoretical framework. Under the entropic risk measure, we derive an iterative algorithm for approximating the intractable feedback Nash equilibria of a risk-sensitive dynamic game. We use an iteratively linearized approximation of the system dynamics and a quadratic approximation of the cost function in solving a backward recursion for finding feedback Nash equilibria. In this respect, the algorithm shares a similar structure with DDP and iLQR methods. We conduct experiments in a set of challenging scenarios such as roundabouts. Compared to ignoring the game interaction or the risk sensitivity, we show that our risk-sensitive game-theoretic framework leads to more time-efficient, intuitive, and safe behaviors when facing underlying risks and uncertainty.**

I. INTRODUCTION

To act safely in a dynamic and uncertain environment, robots must (i) consider the risk associated with their actions, and (ii) account for feedback interactions with other risk-sensitive agents. Existing work typically ignores one or both of these two elements. To address this problem, we present iLEQGames, an algorithm for computing approximate feedback Nash equilibria for interacting risk-aware agents.

Uncertainties are intrinsic to robots; robots may be subject to disturbances, modeling ambiguity, and uncertain localization. To take into account such uncertainties, several stochastic formulations of trajectory planning methods have been proposed [1]–[5], where an expected performance metric is optimized subject to a set of potentially probabilistic constraints. However, these methods are not sufficient when one needs to account for risks that are associated with an uncertain environment.

In contrast, risk-sensitivity based planning methods have proven to be a practical and safe method of capturing uncertainties [6]–[10]. These methods take advantage of a notion of risk measure to avoid potentially unsafe rare events. In addition, recent results in inverse reinforcement learning and human modeling indicate that humans are not risk-neutral, they tend to be risk-aware in their decision making [11]–[13], reinforcing the applicability and relevancy of risk measures. The common risk measures utilized in risk-sensitive planning are entropic risk and Conditional Value at Risk (CVaR) [14], [15]. Entropic risk has been widely used in optimal control due to its simplicity and tractability [16], while recently, CVaR has been incorporated in trajectory optimization due to its interpretability [8]. In risk-aware trajectory optimization, an agent has an inherent risk tolerance, which determines how conservatively the agent acts, i.e. how sensitive the agent is to the underlying risks. For instance, the more risk-sensitive an agent is, the further the average minimum distance between the agent and an obstacle may get.

Going beyond obstacle avoidance, in interactive scenarios, such as autonomous driving, robots need to interact with other intelligent agents such as human drivers or other robotic cars. Such settings are fundamentally game-theoretic [17], [18]. In the absence of uncertainties, it was shown in [19] and [20]–[22] that by treating the interaction as a game, robots can reason about the impact of their actions as well as the intentions of other agents. Inspired by these results, when dealing with uncertainties, assuming that all agents capture a notion of risk in their planning, we argue that to generate a more realistic and time efficient behavior for autonomous agents, it becomes crucial to model the *interaction* of risk-sensitive agents. We show that during interactions, the extent to which agents exhibit risky maneuvers is not solely determined by their inherent risk tolerance, it depends on how risk-sensitive the other agents are too.

In this paper, we model the interaction of risk-aware agents in a game-theoretical framework. Through different case studies where uncertainties are involved, we demonstrate that by being aware of the underlying risk during interactions, our algorithm leads to safer behaviors at a higher distance from other agents. Moreover, the proposed approach is not overly conservative either. By anticipating the feedback game-theoretic interactions, our algorithm can leverage other agent's risk-awareness and plan a time efficient trajectory.

We model the interaction of risk-aware agents via the equilibrium of a dynamic game between agents, where every agent minimizes an entropic risk measure of their underlying cost function. In particular, we study the feedback Nash equilibrium of a dynamic game between such risk-aware agents. However, finding the exact Nash equilibria of our game, an instance of a general-sum dynamic game with nonlinear cost functions, is in general intractable [23]. By drawing on results from [24] and [25], we derive an iterative algorithm for approximating the feedback Nash equilibria of our risk-sensitive dynamic game. At every iteration, we use a linearized approximation of the system dynamics and a quadratic approximation of the cost function in solving a backward recursion for finding feedback Nash equilibria.

We demonstrate the consequences of our framework in a set of simulation studies. We compare the performance of our framework with two baselines: 1) disregarding the interaction while planning for risk-sensitive agents, and 2)

[1]Mingyu Wang, Negar Mehr and Mac Schwager are with Stanford University, Stanford, CA 94305, USA. {mingyuw, nmehr, schwager}@stanford.edu

[2]Adrien Gaidon is with Toyota Research Institute, Los Altos, CA, 94022, USA. adrien.gaidon@tri.global

disregarding the risk-sensitivity while planning for interactive agents. We showcase that the behaviors that emerge out of risk-sensitive interactive planning results in higher distance between agents when facing higher uncertainties. And it is more time efficient compared with non-game planning.

The organization of this paper is as follows. In Section II, we review the preliminaries and prior results. We describe the problem statement in Section III, and discuss our solution to the problem in Section IV. In Section V, we present the consequences of our framework via our case studies. Finally, we conclude the paper in Section VI.

II. Preliminaries

A. Risk-Sensitive Optimal Control

The Linear Exponential Quadratic Gaussian (LEQG) problem is the most common and well-studied form of risk-senstive stochastic optimal control problems [16], [26]. Unlike the Linear Quadratic Gaussian (LQG) formulation which is risk-neutral, LEQG not only considers the expected cost but also the higher order moments of uncertainty by using an entropic objective function. Consider the system with linear dynamics:

$$x_{t+1} = A_t x_t + B_t u_t + w_t, \tag{1}$$

where $x_t \in X$ is the system state, $u_t \in U$ is the control input, and $w_t \sim \mathcal{N}(0, W_t)$ is the system noise. The agent incurs the following cost during a finite horizon T:

$$\Psi = \sum_{t=0}^{t-1} \frac{1}{2}\left(x_t^T Q_t x_t + u_t^T R_t u_t\right) + \frac{1}{2} x_T^T Q_T x_T,$$

where $Q_t \succeq 0$, and $R_t \succ 0$. The exponential risk measure is

$$J = \frac{1}{\theta} \log \mathbb{E}\left[e^{(\theta \Psi)}\right] = R_\theta(\Psi),$$

where θ is the risk-sensitivity parameter. An LEQG problem is to find the optimal control policy that minimizes J, with system dynamics (1). To understand how J captures risk, we use Taylor series expansion of the risk sensitive cost function and obtain

$$J = R_\theta(\Psi) = \mathbb{E}[\Psi] + \frac{\theta}{2}\text{var}[\Psi] + o(\theta).$$

Hence, $R_\theta(\cdot)$ is a linear combination of the expectation and higher order moments of the original random variable. With $\theta = 0$, $J = \mathbb{E}[\Psi]$, and the problem reverts to the LQG problem. When $\theta > 0$, in addition to the expected cost, the variance and higher order moments are penalized. Thus, the resulting policy leads to risk-averse behaviors. Conversely, risk-seeking behaviors can be achieved with negative θ where higher variance and other moments are preferred.

The LEQG problem is known to have a closed form solution using dynamic programming where a modified Riccati equation is used to obtain the closed form solution in a recursive fashion [27]. In the LEQG setting, the optimal feedback control policy is linear in the state, and further, the value function is quadratic, similar to LQG. However, the difference is that the feedback policy is dependent on the distribution of the noise which violates the certainty equivalence principle [26].

B. Dynamic Games

In this section, we introduce discrete-time infinite dynamic games, which is important when an agent interacts with another intelligent agent. Consider a two-player discrete-time system with dynamics modeled as

$$x_{t+1} = f_t(x_t, u_t^1, u_t^2, w_t), \tag{2}$$

where x_t is the state of the system at time step t, u_t^1 and u_t^2 are the control inputs of player 1 and 2 at time t, respectively. Moreover, $w_t \sim \mathcal{N}(0, W_t)$ is the system noise which is assumed to be normally distributed with covariance matrix W_t. We assume that agent i has a finite-horizon cost function of the form

$$\Psi^i = \sum_{t=0}^{T} \left(g_{x,t}^i(x_t) + g_{u,t}^i(u_t^1, u_t^2)\right), \tag{3}$$

where $g_{x,t}^i$ is state cost and $g_{u,t}^i$ is control cost. And agent i's objective is to minimize the expected cost

$$J^i = \mathbb{E}[\Psi^i].$$

Preferably, we want to compute feedback policies for the agents. Let the strategy sets be $\{\Gamma^i, i = 1, 2\}$, where the t-th block of Γ^i is the strategy space for player i at time step t. We assume a feedback information pattern for all players, i.e. at each time step, all players have access to the perfect current state information x_t. Under this information structure, any permissive strategy at time t is a mapping from $X \to U$. Thus, for any strategy tuple $\{\gamma^i \in \Gamma^i, i = 1, 2\}$, where $\gamma^i = \{\gamma_t^i \in \Gamma_t^i\}_{t=0,T-1}$, the control input for agent i is understood to be $u_t^i = \gamma_t^i(x_t)$. We now define the feedback Nash equilibria of risk-sensitive dynamic games.

Definition 2.1: A pair of feedback policies $\{\gamma^{1*}, \gamma^{2*}\}$ constitutes a Nash equilibrium solution if and only if the following inequalities are satisfied for all $\{\gamma^i \in \Gamma^i, i = 1, 2\}$

$$J^{1*} \triangleq J^1(\gamma^{1*}; \gamma^{2*}) \leq J^1(\gamma^1; \gamma^{2*}),$$
$$J^{2*} \triangleq J^2(\gamma^{1*}; \gamma^{2*}) \leq J^2(\gamma^{1*}; \gamma^2).$$

In general, a dynamic game could admit multiple Nash equilibria, The uniqueness of the Nash equilibria is problem dependent. Moreover, it is generally computationally intractable [23]. However, under certain assumptions on the structure of the problem, such as linear (affine) dynamics and quadratic cost functions, Nash equilibria can be found via a Riccati-type set of equations as discussed in [24, Chap. 6].

III. Problem Statement

As we discussed in the preliminaries, when facing uncertainties, optimizing the expected cost is often not intuitive and may incur high risk. In this work, we consider two agents who are both risk-sensitive. Note that for simplicity, we focus on the two-player setting; however, the problem formulation and the proposed solution can be extended to more than two players. The system dynamics follow the definition in (2). And the cost function is given as in (3). Assuming that every agent is risk-sensitive, the corresponding risk-sensitive

cost function of agent i with risk-sensitivity parameter θ^i is defined as

$$J^i(x_0;\theta^i) = R_{\theta^i}(\Psi^i) = \frac{1}{\theta^i}\log\mathbb{E}\left[e^{(\theta^i\Psi^i)}\right]. \tag{5}$$

With the risk sensitive utility, the strategy space and Nash equilibria strategies are defined accordingly. In the context of risk-sensitive games, for linear exponential quadratic games, the solution to feedback Nash equilbria was introduced in [25]. In this paper, for our risk-sensitive dynamic game, we leverage the result of [25] and use an iterative approach to extend the solution to non-linear systems dynamics and cost functions. The proposed approach iteratively approximates the original nonlinear problem with linear dynamics and quadratic costs, and aims to approximate a Nash equilibrium in the sense of the approximated system.

IV. Iterative LEQ Game

In this section, we first describe the solution to an extended set of LEQ games with linear dynamics and affine-quadratic cost functions. Then, we present our proposed iterative LEQ game solution for general risk–sensitive games.

A. LEQ game

The solution of risk sensitive discrete-time dynamic games was first introduced in [25] for systems with linear dynamics and quadratic costs. First, we extend this result to consider a more general case with affine-quadratic cost functions. For linear dynamics, the system dynamics (2) are

$$x_{t+1} = A_t x_t + B_t^1 u_t^1 + B_t^2 u_t^2 + w_t, \tag{6}$$

where A_t, B_t^1, and B_t^2 are matrices of appropriate dimensions. We assume that the cost function for agent i is given in affine-quadratic form as:

$$\begin{aligned}\Psi^i = \sum_{t=0}^{T-1}\left[\frac{1}{2}x_t^T Q_t^i x_t + l_t^{iT} x_t + \frac{1}{2}\sum_j u_t^{jT} R_t^{ij} u_t^j\right] + \\ \frac{1}{2}x_T^T Q_T^i x_T + l_T^{iT} x_T,\end{aligned} \tag{7}$$

where $Q_t^i \succeq 0, R_t^{ij} \succ 0$ and l_t^i are matrices of appropriate dimensions. We assume that every agent optimizes the risk-sensitive cost (5) with Ψ^i being defined via (7).

Lemma 4.1: For a two-agent risk-sensitive game, let θ^1 and θ^2 be the risk-sensitivity parameters for the two agents, respectively. For every agent $i = 1, 2$, let P_t^i and α_t^i be matrices of appropriate dimensions that satisfy the following sets of linear matrix equations:

$$\begin{aligned}\left[R_t^{ii} + B_t^{iT}\tilde{Z}_{t+1}^i B_t^i\right]P_t^i + B_t^{iT}\tilde{Z}_{t+1}^i\sum_{j\neq i}B_t^j P_t^j = \\ B_t^{iT}\tilde{Z}_{t+1}^i A_t,\end{aligned} \tag{8a}$$

$$\begin{aligned}\left[R_t^{ii} + B_t^{iT}\tilde{Z}_{t+1}^i B_t^i\right]\alpha_t^i + B_t^{iT}\tilde{Z}_{t+1}^i\sum_{j\neq i}B_t^j \alpha_t^j = \\ B_t^{iT}W_t^{-1}(W_t^{-1} - \theta^i Z_{t+1}^i)^{-1}\zeta_{t+1}^i,\end{aligned} \tag{8b}$$

where $\tilde{Z}_t^i, Z_t^i$, and ζ_t^i are recursively obtained from the following

$$\tilde{Z}_{t+1}^i = Z_{t+1}^i + \theta^i Z_{t+1}^i(W_t^{-1} - \theta^i Z_{t+1}^i)^{-1}Z_{t+1}^i, \tag{9}$$

$$Z_t^i = Q_t^i + \sum_j P_t^{jT} R_t^{1j} P_t^j + F_t^T\tilde{Z}_{t+1}^i F_t, \tag{10}$$

$$\begin{aligned}\zeta_t^{iT} = l_t^{iT} + \sum_j \alpha_t^{jT} R_t^{ij} P_t^j + \beta_t^T\tilde{Z}_{t+1}^i F_t + \\ \zeta_{t+1}^{iT}(W_t^{-1} - \theta^i Z_{t+1}^i)^{-1}W_t^{-1}F_t,\end{aligned} \tag{11}$$

where

$$F_t = A_t - \sum_{j=1,2} B_t^j P_t^j,\ \beta_t = -\sum_{j=1,2} B_t^j \alpha_t^j. \tag{12}$$

The terminal conditions for Equations (10) and (11) are

$$Z_T^i = Q_T^i,\ \zeta_T^i = l_T^i. \tag{13}$$

Note it is required that

$$W^{-1} - \theta^i Z_t^i \succ 0, \forall t \in T, i = 1, 2 \tag{14}$$

to avoid "neurotic breakdown". Because if this condition is not satisfied, the risk-sensitive cost (5) becomes infinity. See Appendix I for a more detailed discussion. The following Corollary is an extension of Corollary 1 in [25]

Corollary 4.1: A two-person linear exponential quadratic game defined by system dynamics (6), cost function (7) and risk sensitive cost function (5) admits a unique feedback Nash equilibrium solution if, and only if, (8) admits unique solution sets $\{P_t^{i*}, \alpha_t^{i*}, t \in T, i = 1, 2\}$. Furthermore, the equilibrium strategies are given by

$$\gamma_t^{i*}(x_t) = -P_t^{i*}x_t - \alpha_t^{i*}. \tag{15}$$

For completeness and clarity, we have provided the outline of our proof in the Appendix I.

Note that with the risk-sensitivity parameters $\theta^1 = \theta^2 = 0$, the above equations revert to the case of linear quadratic games [24, Chap. 6], where both players are considered to be risk-neutral. Note that with $\theta^1 = \theta^2 = 0$, the Gaussian covariance matrix W_t does not appear in Lemma 4.1 anymore. Hence risk-neutral control policy is indifferent to the level of noise.

B. Iterative LEQ problem

To handle general nonlinear dynamics and cost functions, we propose an iterative algorithm that proceeds as follows. We start with a nominal strategy sequence $\{P_t^i, \alpha_t^i\}$ for every time step $t \in T$, and every agent $i = 1, 2$. If a nominal policy is not available, a trivial initialization could be initializing all matrices to zeros. Then, at every iteration, a nominal state trajectory and nominal action trajectories $\eta = \{\bar{x}, \bar{u}^1, \bar{u}^2\}$ are obtained from forward simulation of the system dynamics using the nominal strategy. Let $\delta x_t = x_t - \bar{x}_t$, $\delta u_t^i = u_t^i - \bar{u}_t^i$, we can then acquire a linear approximation of the dynamics (2) as

$$\delta x_{t+1} \approx A_t\delta x_t + \sum_{i=1,2} B_t^i \delta u_t^i, \tag{16}$$

where $A_t = D_x f_t(\cdot)$ and $B_t^i = D_{u_t^i} f_t(\cdot)$ are the Jacobians of the original nonlinear dynamics function with respect to x_t, and u_t^i, respectively. Furthermore, the cost function (3) is approximated using quadratic functions:

$$g_{x,t}^i(\bar{x}_t + \delta x_t) \approx g_{x,t}^i(\bar{x}_t) + \frac{1}{2}\delta x_t^T Q_t^i \delta x_t + l_t^{iT} \delta x_t, \quad (17)$$

where $Q_t^i = D_{x_t x_t} g_{x,t}^i(\cdot)$ and $l_t^i = D_{x_t} g_{x,t}^i(\cdot)$ are the Hessian and the gradient of the cost function $g_{x,t}^i(\cdot)$ with respect to x_t. Note that our formulation only considers nonlinear costs on state variables, and $g_{u,t}^i(\cdot)$ is quadratic. For a more general case, where the cost function is also nonlinear on control inputs, a similar approximation could be used to derive the quadratic terms and linear terms in u_t^i.

All the approximations A_t, B_t^i, Q_t^i, l_t^i are evaluated at η. For the linearized system dynamics and quadratized cost function, we obtain a new LEQ game problem with new variable sequences $\delta x, \delta u^1$, and δu^2. These approximations result in a new game that can be solved using Lemma. 4.1. Once the approximated game is solved, we obtain a new sequence of control inputs.

$$\{\bar{u}_t^i + \delta u_t^*, t = 0, \cdots, T-1\}, \quad (18)$$

where u_t^* is the solution for previous iteration. A new $\bar{x}_t$ is attained from the forward simulation of the original system dynamics (2) using the newly obtained control inputs. We repeat the above process until convergence, i.e. the deviation of the new state trajectory from the state trajectory in the previous iteration lies within tolerance. In practice, we limit the maximum number of iterations for real-time implementations of our algorithm. The outline of our algorithm is summarized in Algorithm 1.

Algorithm 1 Iterative Linear Exponential Quadratic Game

Inputs
system dynamics (2), risk sensitive utility (5)
risk sensitive parameters θ^1, θ^2
Initialization
initialize the control policy using $P_t^i = \alpha_t^i = 0, \forall i$
forward simulation and obtain $(\bar{x}_0, \bar{x}_1, \bar{x}_2, \cdots, \bar{x}_T)$, $(\bar{u}_0^1, \cdots, \bar{u}_{T-1}^1)$, and $(\bar{u}_0^2, \cdots, \bar{u}_{T-1}^2)$
while not converged **do**
    linear approximation of (2)
    quadratic approximation of (5)
    solve the backward recursion with (8-13)
    forward simulation and obtain the new trajectories
end while
return policy P_t^i, α_t^i

Remark 1: Applying u_t^i directly from (18) may lead to non-convergence since the resulting trajectory could deviate from the original non-linear systems which we approximated around η_t. As in other iterative optimal control problems [28], [29], we augment our algorithm with a line search. At each iteration, rather than (18), we apply the following control input

$$\bar{u}_t^i - P_t^i \delta x_t - \epsilon \alpha_t^i, \quad (19)$$

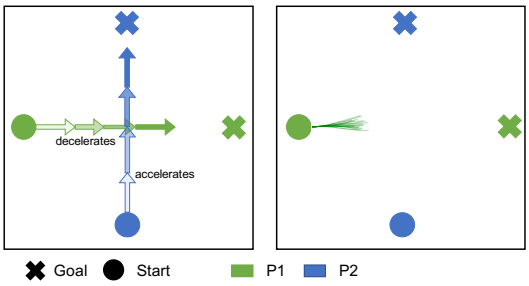


Fig. 1: Illustration of the cross interaction scenario. **Left:** One possible realization of the interaction at the intersection. **Right:** A visualization of the uncertainty of P1.

where ϵ is the step size for improving our control strategy. Initially, we set $\epsilon = 1$. If the new trajectory deviates too far from the nominal trajectory, we reject the trajectory and decrease ϵ by half.

Remark 2: As mentioned in Lemma 4.1, there is a critical value θ^{crit} in the case of a positive risk-sensitivity parameter. The intuition is that, if a player is too risk-averse, then, due to the rapid growth of the exponential function, the agent's risk-sensitive cost (5) under uncertainty may approach infinity with any possible choice of policy. In our work, we implement backtracking on positive risk-sensitivity parameters. An initial risk-sensitivity parameter is used in the backward computation. At every stage, (14) is asserted. If the criteria is violated, then we decrease θ by half.

V. Case Studies

In this section, we demonstrate the performance of our proposed iterative algorithm for solving risk-sensitive dynamic games in different systems and simulation environments. We first show that our algorithm generates intuitive, interpretable and safe trajectories. Then, to highlight the importance of capturing the interactions via games, we compare our framework with a baseline where the interaction of risk-sensitive agents is disregarded, and every risk-aware agent treats the other agent as a nonreactive obstacle. Moreover, we compare the performance of our algorithm with that of interaction between risk-neutral agents to illustrate the impact of risk-awareness. We consider the following scenarios in autonomous driving: a cross intersection, an onramp merging maneuver, and entering a roundabout in CARLA simulator [30]. Throughout this section, we use P1 and P2 to denote the two agents involved in the scenario of focus.

A. Cross intersection

Consider an intersection scenario similar to Fig. 1. Two risk-sensitive players start from their starting positions and try to advance towards their goal positions at a constant speed. The configuration is selected such that without collision avoidance, the two players will cross each other at the center point. The dynamics of the two agents is modeled as two extended unicycles

$$x = [p_1^x, p_1^y, v_1, \theta_1, p_2^x, p_2^y, v_2, \theta_2], \; u^i = [a_i, \dot{\theta}_i], \; i = 1, 2.$$

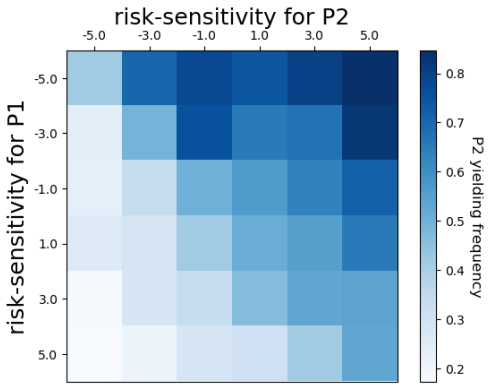


Fig. 2: Heatmap of the percentage of simulations where player 1 crossed the interaction before player 2.

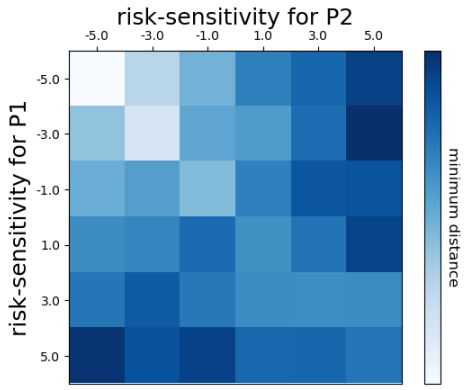


Fig. 3: Heatmap of minimum distance between two players. The minimum distance increases as relative risk-sensitivity increases.

The state vector x includes the position of the two agents p^x, p^y, their speeds v_1, v_2, and their headings θ_1, θ_2. For every agent i, the control inputs are acceleration a_i, and yaw rate $\dot{\theta}_i$. We use $g^i(x,u)$ to denote the cost function of each player and separate it into tracking cost, control cost, and collision cost

$$g^i(x,u) = g^i_{\text{track}}(x) + g^i_{\text{ctrl}}(u) + g^i_{\text{coll}}(x),$$

with

$$\begin{aligned} g^i_{\text{track}}(x) &= (x_t - x_t^{ref})^T W_t (x_t - x_t^{ref}), \\ g^i_{\text{ctrl}}(u) &= \sum_j u_t^{jT} W_u^{ij} u_t^j, \\ g^i_{\text{coll}}(x) &= W_c (a_c \cdot d + 1)^{-c}, \end{aligned}$$

where W_t, W_u^{ij}, W_c are the weight matrices in each cost respectively, d is the distance between two players, and a_c, c are the collision cost parameters that penalize for unsafe distances. We let the risk tolerance of the two players θ_1 and θ_2 vary from -5.0 (risk-seeking) to $+5.0$ (risk-averse). For each pair of risk tolerances (θ_1, θ_2), we conducted 150 simulations, in which an additive random noise on velocity was added to both players dynamics during the simulations.

Fig. 2 shows the percentage of simulations where P2 yielded to P1 to pass first. We can observe from the heatmap that as the risk-sensitivity value of one player increases, the percentage of it yielding to the other player increases. Moreover, when the risk-sensitivity values of both player are the same (diagonal grids in the heatmap), the yielding frequency is everywhere close to 50% in this symmetric setup. We also notice that, for the parameters pairs which have the same difference (corresponds to the lines of grids parallel to the diagonal), the yielding frequencies are similar. Intuitively, this indicates that it is the *relative* risk sensitivity of agents rather than the absolute risk sensitivity of agents which influences the qualitative behavior during the interaction. In Fig. 3, we show the average of minimum distance between the two players. As the relative risk-sensitivity increases, the min distance also increases.

Comparison with baselines

Noise	0.03	0.05	0.07	0.09	0.11
RS game	1.03± 0.05	1.05± 0.05	1.06± 0.04	1.07± 0.04	1.09± 0.03
RN game	1.01± 0.05	1.01± 0.05	1.00± 0.05	1.01± 0.05	1.01± 0.05

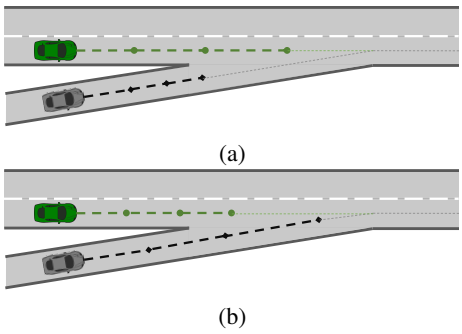
TABLE I: Intersection scenario: Minimum distance between the two agents with different Gaussian uncertainty covariance.

1) Risk-Neutral Games (RN games): We first demonstrate the importance of capturing risk when planning in the presence of uncertainties. As opposed to risk-aware agents, risk-neutral agents only optimize for the expected cost value, and they are insensitive to the level of noise. We implemented iterative linear quadratic Gaussian games as a risk-neutral baseline by setting $\theta_1 = \theta_2 = 0$. Table I illustrates the minimum distance between the two agents as we change the level of uncertainties. For risk-sensitive agents (RS game), we can see that the trajectories adapt to the noise level. Moreover, the minimum distance increases as the uncertainty increases. However, for the risk-neutral case, the control input is indifferent to noise.

2) Risk-sensitivity in isolation: Another important aspect of our method is the game-theoretic reasoning. By enabling the autonomous agents to reason about their influence on other agents, we can avoid overly-conservative behaviors and achieve higher efficiency. We compare in Table. II the time that it takes for P2 to pass the intersection using our approach with that of a non game-theoretic setting, where P1 is regarded as a non-reactive dynamic obstacle by P2 and vice versa. We can see that compared to this baseline, our proposed approach consistently spends less time to pass the intersection. The main reason for better efficiency of our method is the fact that in the game-theoretic setting, every agent leverages its knowledge of the other agent's risk tolerance. Due to space limits, we only show the result for a fixed risk-sensitivity for P1 (risk neutral in Table. II). The same conclusion can be drawn with other parameters.

P2 θ_2	-5	-3	-1	1	3	5
game	3.21±0.26	3.27±0.29	3.31±0.32	3.40±0.33	3.44±0.33	3.46±0.32
no-game	3.87±0.52	3.97±0.49	3.90±0.50	3.99±0.52	3.94±0.54	3.92±0.52

TABLE II: Intersection scenario: Time to cross intersection for P2 with different risk-sensitivity.



(a)

(b)

Fig. 4: Demonstration of the interactions between two risk-aware agents during a merge maneuver. The two cars start from a configuration where which one should yield is ambiguous. The relative risk-sensitivity determines the interaction outcome.

B. Merging

Consider a highway merging scenario with two cars. One car starts from the onramp and wants to merge into the highway with the presence of another car in the main lane as shown in Fig. 4. In the cases where the two cars are close in longitudinal direction and have similar speed, successfully executing the merging maneuver is a very challenging task. The challenge comes from the ambiguity that arises from the uncertainty of the other car's future trajectory and the order of merging. Two possible interactions are illustrated in Fig. 4a and Fig. 4b. We use this scenario to demonstrate the effect of relative risk-sensitivity during such challenging interactions.

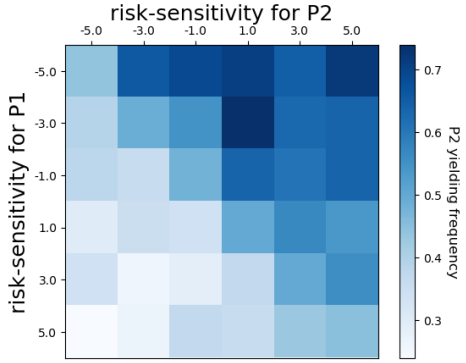
We assume both cars follow the center line in their current driving lanes and only consider the control of vehicle's speed to finish the merging maneuver. In other words, we assume a steering controller will be executed separately for each car to remain in their lanes. The state of our system includes 2D position and the speed of the two cars.

$$x = [p_1^x, p_1^y, v_1, p_2^x, p_2^y, v_2]$$

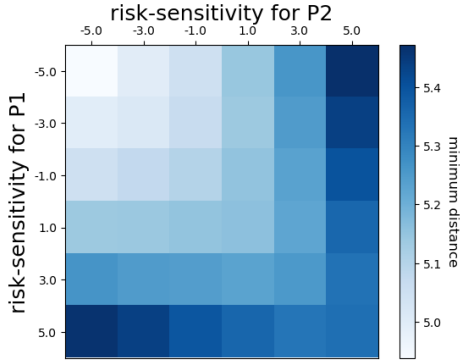
The control input of each player is acceleration $u^1 = a_1$, $u^2 = a_2$. We use the same form of cost functions as in Sec. V-A. The initial position and speed of the two players are set so that they will collide if no control is applied. Our results in minimum distance and yielding behavior are reported in Fig. 5. As the risk tolerances change, we can observe that the patterns of yield/pass behaviors and minimum distance in the merging scenario are very similar to the intersection scenario.

Comparison with baselines

Similar to the intersection scenario, two comparisons are conducted. The comparison with risk-neutral games is shown in Table III while the comparison with non interactive risk-sensitive agents is shown in Table IV.



(a) Percentage of player 2 yielding to player 1.



(b) Minimum distance along the trajectory.

Fig. 5: Simulation results from merging scenario risk-sensitive players.

Noise	0.03	0.05	0.07	0.09	0.11
RS game	5.13±0.03	5.25±0.05	5.44±0.04	5.67±0.04	5.70±0.03
RN game	5.02±0.09	5.04±0.06	5.00±0.11	5.03±0.06	5.02±0.08

TABLE III: Merge scenario: Minimum distance between the two agents with different Gaussian uncertainty covariance.

C. Entering Roundabout with High Fidelity Simulator

We also evaluate our algorithm in the CARLA simulator [30], a high fidelity open-source simulator for autonomous driving. Fig. 6 shows the roundabout scenario we use. P1 starts in the roundabout and P2's initial position is in an entering lane. Our algorithm plans risk-sensitive trajectories for the two vehicles at 3Hz. The closed-loop trajectory is passed to a low-level controller which computes the throttle value to achieve the desired acceleration. We set the risk sensitivity parameter of each agent to be -10 (risk-seeking), 0, and 10 (risk-averse). For each parameter pair, we conducted 10 simulations and initial position for P2 was randomized. The qualitative behavioral is shown in Fig. 7. For the roundabout simulation, we report our comparison with the no-game baseline. The time to finish entering roundabout for P2 is shown in Table. V. It can be observed that in this case, the risk-aware game-theoretic planner achieved much more

P2 θ_2	-5	-3	-1	1	3	5
game	3.83±0.18	3.84±0.19	3.84±0.20	3.95±0.25	3.94±0.30	3.98±0.36
no-game	3.91±0.21	3.90±0.23	3.93±0.23	4.08±0.39	4.09±0.39	4.20±0.36

TABLE IV: Merge scenario: Time to finish merging maneuver for P2 with different risk-sensitivity.

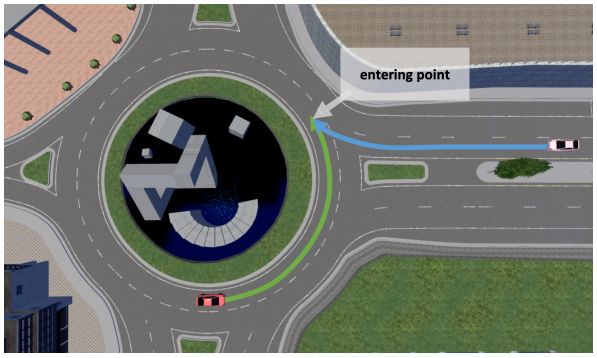


Fig. 6: A snapshot of CARLA roundabout environment. Red car (bottom) is P1 and light pink car (right) is P2.

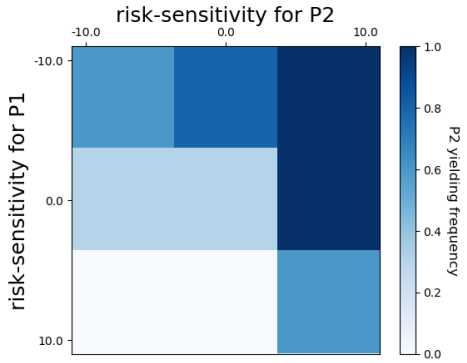


Fig. 7: Frequency of P2 yielding to P1 when entering the roundabout.

efficient trajectories compared to the baseline.

VI. Discussion

In this work, we presented a game-theoretic planning approach for risk-aware agents. The formulation of risk-sensitive dynamic games provides insights to the interaction between players. Compared to risk-neutral games, our framework introduces new parameters that lead to safer and more intuitive behaviors. The game aspect is also captured to address the mutual influence between agents. Compared with non-game risk-sensitive control, our algorithm achieves better efficiency without sacrificing safety. The proposed iterative linear exponential quadratic method is used to solve nonlinear dynamic systems with nonlinear costs in real-time. The performance is demonstrated in several case studies, including a cross-intersection, an onramp merging maneuver and a roundabout entering in driving simulator.

Limitations and future work: While the introduced framework is insightful for understanding the interaction between agents and has the potential to increase interpretability of autonomous agents' motion, we have not addressed the problem of interaction with human players. To achieve this goal, a human model is needed. For example, [31], [32] uses inverse reinforcement learning to learn a reward function. [33], [34] approaches the problem as a preference-based learning problem. More recently, [35] estimates model parameters online with a filtering algorithm. We would like to pursue this direction for future work.

P2 θ_2	-10.0	0.0	10.0
game.	6.86±1.55	7.09±1.59	9.00±0.33
no-game	9.65±1.34	14.07±3.56	16.37±0.41

TABLE V: Roundabout scenario: Time to finish entering roundabout for P2 with different risk-sensitivity values.

Acknowledgment

This work was supported in part by ONR grant N00014-18-1-2830. Toyota Research Institute ("TRI") provided funds to assist the authors with their research but this article solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity.

Appendix I
Risk Sensitive Game with Linear Dynamics and Affine-Quadratic Costs

First, we give the following equation, which would become useful when deriving the Riccati equations for LEQ game. The results could be derived using basic calculus.

Let

$$J = R_\theta(\frac{1}{2}(z^T P z + s^T z)) = \frac{1}{\theta}\log \mathbb{E}\exp(\frac{\theta}{2}(z^T P z + s^T z)),$$

where $z \sim \mathcal{N}(\bar{z}, Z)$. Then, if $Z^{-1} > \theta P$

$$J = -\frac{1}{2\theta}\log\det(I - \theta P Z) - \frac{1}{2\theta}c, \tag{21}$$

where

$$c = \bar{z}^T Z^{-1}\bar{z} - (\frac{\theta}{2}s + Z^{-1}\bar{z})^T(Z^{-1} - \theta P)^{-1}(\frac{\theta}{2}s + Z^{-1}\bar{z}).$$

$J = \infty$ if $z^{-1} \not> \theta P$. This is called "neurotic breakdown" when θ is too large.

Now, we consider a two-person discrete-time infinite dynamic game. The dynamics equation of the system is given as in (6) and the risk sensitive objective function in (5). We use dynamic programming and induction to derive the solution. Suppose at time step t, the optimal cost-to-go for player i is

$$V^i_{t+1}(x_{t+1}) = \frac{1}{2}x^T_{t+1}Z^i_{t+1}x_{t+1} + \zeta^{iT}_{t+1}x_{t+1} + n^i_{t+1} \tag{22}$$

Then, going backwards, at time t,

$$\begin{aligned} J^i_t(x_t) =& \frac{1}{2}x^T_t Q^i_t x_t + l^{iT}_t x_t + \sum_j \frac{1}{2}u^{jT}_t R^{ij}_t u^j_t + \\ & R_{\theta^i}\left(V^i_{t+1}(x_{t+1})\right) \end{aligned} \tag{23}$$

where $x_{t+1} \sim \mathcal{N}(A_t x_t + B^1_t u^1_t + B^2_t u^2_t, W_t)$. Plug in (21, 22) and rearrange the above equation as a function of u^i_t. To

minimize J_t^i, the first-order and second-order condition for u_t^{i*} are

$$\frac{\partial}{\partial u_t^i} J_t^i(x_t)\Big|_{u_t^{i*}} = B_t^{iT} \tilde{Z}_{t+1}^i A_t x_t + B_t^{iT} \tilde{Z}_{t+1}^i \sum_{j \neq i} B_t^j u_t^j + B_t^{iT} W_t^{-1} (W_t^{-1} - \theta^i Z_{t+1}^i)^{-1} \zeta_{t+1}^i = 0, \tag{24}$$

$$\frac{\partial^2}{\partial u_t^{i2}} J_t^i(x_t)\Big|_{u_t^{i*}} = R_t^{11} + B_t^{1T} \tilde{Z}_{t+1}^1 B_t^1 > 0 \tag{25}$$

Assume the form of the optimal feedback policy is given as $u_t^i = -P_t^i x_t - \alpha_t^i$. Then, by substitute the expression into (24-25), and requiring them to be satisfied for all x_t, we obtain P_t^i, α_t^i as the solution using equations in Lemma 4.1. Plugging in the optimal control policy, we arrive at a quadratic form value function at time t. Thus, we finish the induction.

REFERENCES

[1] J. A. Primbs and C. H. Sung, "Stochastic receding horizon control of constrained linear systems with state and control multiplicative noise," *IEEE transactions on Automatic Control*, vol. 54, no. 2, pp. 221–230, 2009.

[2] L. Blackmore, M. Ono, and B. C. Williams, "Chance-constrained optimal path planning with obstacles," *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1080–1094, 2011.

[3] M. Cannon, B. Kouvaritakis, S. V. Rakovic, and Q. Cheng, "Stochastic tubes in model predictive control with probabilistic constraints," *IEEE Transactions on Automatic Control*, vol. 56, no. 1, pp. 194–200, 2010.

[4] D. Bernardini and A. Bemporad, "Stabilizing model predictive control of stochastic constrained linear systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1468–1480, 2011.

[5] J. Fleming, M. Cannon, and B. Kouvaritakis, "Stochastic tube mpc for lpv systems with probabilistic set inclusion conditions," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 4783–4788.

[6] T. Osogami, "Robustness and risk-sensitivity in markov decision processes," in *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012, pp. 233–241.

[7] Y.-L. Chow and M. Pavone, "A framework for time-consistent, risk-averse model predictive control: Theory and algorithms," in *2014 American Control Conference*. IEEE, 2014, pp. 4204–4211.

[8] Y. Chow, A. Tamar, S. Mannor, and M. Pavone, "Risk-sensitive and robust decision-making: a cvar optimization approach," in *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, pp. 1522–1530.

[9] S. Samuelson and I. Yang, "Safety-aware optimal control of stochastic systems using conditional value-at-risk," in *2018 American Control Conference*. IEEE, 2018, pp. 6285–6290.

[10] M. P. Chapman, J. Lacotte, A. Tamar, D. Lee, K. M. Smith, V. Cheng, J. F. Fisac, S. Jha, M. Pavone, and C. J. Tomlin, "A risk-sensitive finite-time reachability approach for safety of stochastic dynamic systems," in *2019 American Control Conference*. IEEE, 2019, pp. 2958–2963.

[11] A. Majumdar, S. Singh, A. Mandlekar, and M. Pavone, "Risk-sensitive inverse reinforcement learning via coherent risk models," in *Proceedings of Robotics: Science and Systems*, Cambridge, Massachusetts, July 2017.

[12] L. J. Ratliff and E. Mazumdar, "Inverse risk-sensitive reinforcement learning," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 1256 – 1263, 2019.

[13] M. Kwon, E. Biyik, A. Talati, K. Bhasin, D. P. Losey, and D. Sadigh, "When humans aren't optimal: Robots that collaborate with risk-aware humans," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, 2020, p. 43–52.

[14] H. Föllmer and T. Knispel, "Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations," *Stochastics and Dynamics*, vol. 11, no. 02n03, pp. 333–351, 2011.

[15] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of banking & finance*, vol. 26, no. 7, pp. 1443–1471, 2002.

[16] P. Whittle and P. R. Whittle, *Risk-sensitive optimal control*. Wiley New York, 1990, vol. 20.

[17] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan, "Hierarchical game-theoretic planning for autonomous vehicles," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 9590–9596.

[18] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 797–803.

[19] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions," in *Proceedings of Robotics: Science and Systems*, AnnArbor, Michigan, June 2016.

[20] R. Spica, D. Falanga, E. Cristofalo, E. Montijano, D. Scaramuzza, and M. Schwager, "A real-time game theoretic planner for autonomous two-player drone racing," in *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018.

[21] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game theoretic planning for self-driving cars in competitive scenarios," in *Proceedings of Robotics: Science and Systems*, FreiburgimBreisgau, Germany, June 2019.

[22] D. Fridovich-Keil, E. Ratner, A. D. Dragan, and C. J. Tomlin, "Efficient iterative linear-quadratic approximations for nonlinear multi-player general-sum differential games," in *International Conference on Robotics and Automation (ICRA)*, 2020.

[23] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a nash equilibrium," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 195–259, 2009.

[24] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.

[25] M. B. Klompstra, "Nash equilibria in risk-sensitive dynamic games," *IEEE Transactions on Automatic Control*, vol. 45, no. 7, pp. 1397–1401, 2000.

[26] W. H. Fleming and W. M. McEneaney, "Risk sensitive optimal control and differential games," in *Stochastic theory and adaptive control*. Springer, 1992, pp. 185–197.

[27] P. Whittle, "Risk-sensitive linear/quadratic/gaussian control," *Advances in Applied Probability*, vol. 13, no. 4, pp. 764–777, 1981.

[28] S. Yakowitz, "Algorithms and computational techniques in differential dynamic programming," *Control and Dynamical Systems: Advances in Theory and Applications*, vol. 31, pp. 75–91, 2012.

[29] J. Van Den Berg, S. Patil, and R. Alterovitz, "Motion planning under uncertainty using iterative local optimization in belief space," *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1263–1278, 2012.

[30] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 78. PMLR, 13–15 Nov 2017, pp. 1–16.

[31] D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, "Cooperative inverse reinforcement learning," in *Advances in neural information processing systems*, 2016, pp. 3909–3917.

[32] D. S. Brown, Y. Cui, and S. Niekum, "Risk-aware active inverse reinforcement learning," in *Proceedings of The 2nd Conference on Robot Learning*, vol. 87. PMLR, 29–31 Oct 2018, pp. 362–372.

[33] A. Jain, S. Sharma, T. Joachims, and A. Saxena, "Learning preferences for manipulation tasks from online coactive feedback," *The International Journal of Robotics Research*, vol. 34, no. 10, pp. 1296–1313, 2015.

[34] M. Palan, G. Shevchuk, N. C. Landolfi, and D. Sadigh, "Learning reward functions by integrating human demonstrations and preferences," in *Proceedings of Robotics: Science and Systems*, FreiburgimBreisgau, Germany, June 2019.

[35] R. Bhattacharyya, R. Senanayake, K. Brown, and M. J. Kochenderfer, "Online parameter estimation for human driver behavior prediction," in *2020 American Control Conference*, 2020.