

Chance-Constrained Iterative Linear-Quadratic Stochastic Games

Hai Zhong , Student Member, IEEE, Yutaka Shimizu , and Jianyu Chen 

Abstract—Dynamic game arises as a powerful paradigm for multi-robot planning, for which safety constraint satisfaction is crucial. Constrained stochastic games are of particular interest, as real-world robots need to operate and satisfy constraints under uncertainty. Existing methods for solving stochastic games handle chance constraints using exponential penalties with hand-tuned weights. However, finding a suitable penalty weight is nontrivial and requires trial and error. In this letter, we propose the chance-constrained iterative linear-quadratic stochastic games (CCILQGames) algorithm. CCILQGames solves chance-constrained stochastic games using the augmented Lagrangian method. We evaluate our algorithm in three autonomous driving scenarios, including merge, intersection, and roundabout. Experimental results and Monte Carlo tests show that CCILQGames can generate safe and interactive strategies in stochastic environments.

Index Terms—Multi-robot systems, motion and path planning, optimization and optimal control.

I. INTRODUCTION

WITH the recent advances in multi-robot systems, dynamic game has emerged as a new paradigm for multi-robot planning [1], [2]. Dynamic game naturally captures the interactive nature of multi-agent planning, as the ego agent's strategy accounts for the fact that other agents' plans could change based on the ego agent's action. A few works also consider safety constraints in the dynamic game settings [3], [4], which is another key aspect of multi-robot planning (e.g., avoiding collisions with other robots or obstacles).

Most of the existing work on dynamic games focuses on deterministic settings. However, robots operating in the real world must reason about uncertainties, resulting in stochastic games. For example, robots need to reason about their unmodeled or disturbed system dynamics, noisy sensor perceptions, and other

Manuscript received 8 July 2022; accepted 22 November 2022. Date of publication 8 December 2022; date of current version 14 December 2022. This letter was recommended for publication by Associate Editor D. Aksaray and Editor M. A. Hsieh upon evaluation of the reviewers' comments. This work was supported by the Ministry of Science and Technology of the People's Republic of China, the 2030 Innovation Megaprojects "Program on New Generation Artificial Intelligence" under Grant 2021AAA0150000. (Corresponding author: Jianyu Chen.)

Hai Zhong is with the Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China (e-mail: zhongahimanchester@gmail.com).

Yutaka Shimizu is with the Tier IV, Inc., Shinagawa-ku, Tokyo 140-0001, Japan (e-mail: purewater0901@gmail.com).

Jianyu Chen is with the Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China, and also with the Shanghai Qizhi Institute, Shanghai 200232, China (e-mail: jianyuchen@tsinghua.edu.cn).

This letter has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2022.3227867>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2022.3227867

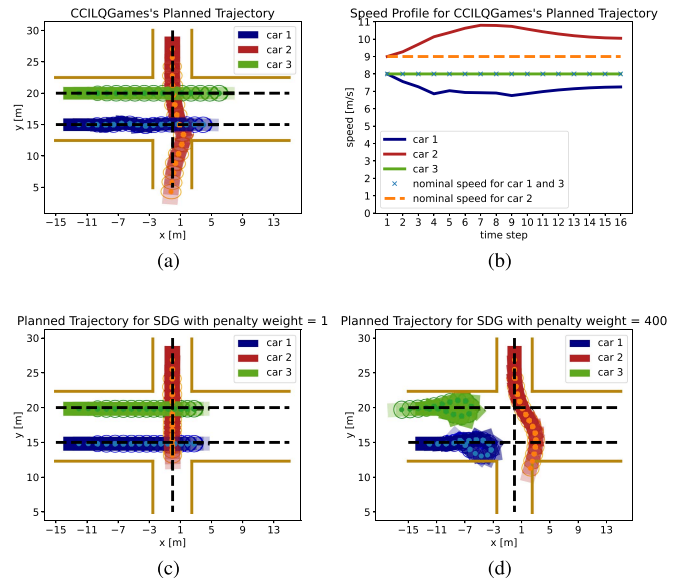


Fig. 1. SDG with a large penalty weight could lead to conservative strategies (as shown in (d)), while a small penalty weight could cause unsafe trajectories (as shown in (c)). CCILQGames could ensure performance while satisfying constraints without manually tuning penalty weights (as shown in (a)).

agents' motion and intention uncertainties. Also, the complexity of satisfying safety constraints in stochastic games is further evolved, as the agent needs to guarantee safety under uncertainty. Schwarting et al. [5] recently proposed the stochastic dynamic game (SDG) algorithm to solve stochastic games, considering process and observation uncertainties. SDG simply adds exponential penalties for chance constraint violations. Yet, this approach is far from being satisfactory: choosing the penalty weight is nontrivial. Inappropriate penalty weights will result in undesired behaviors: a small penalty leads to dangerous collisions (Fig. 1(c)), while a large penalty leads to conservative suboptimal behaviors (Fig. 1(d)). Tuning penalty terms requires trial and error and is generally failure-prone, which prohibits the potential for real-time applications. To the best of our knowledge, no existing work systematically handles safety constraints for stochastic dynamic games (with both process and observation uncertainties).

This letter proposes the chance-constrained iterative linear-quadratic stochastic games (CCILQGames) algorithm to overcome the hurdle above. CCILQGames can incorporate chance constraints in stochastic dynamic games, which allows specification of the probability threshold for successfully satisfying constraints. We use the augmented Lagrangian method to

construct an outer and inner loop framework, with the benefits of automatically tuning the penalty weights. The chance constraint violation is calculated to update penalty and Lagrange multiplier terms in the outer loop. We solve the corresponding unconstrained stochastic game in the inner loop by extending the iterative linear-quadratic games (ILQGames) algorithm [2], [6] to the stochastic game setting. We evaluate the proposed method in lane merging, three-player intersection, and round-about scenarios. Experimental results and Monte Carlo tests show that our approach generates interactive and safe strategies under uncertainty.

II. RELATED WORK

A. Game-Theoretical Planning

A family of algorithms for solving differential games hinges on transforming a differential game into one or a series of single-agent optimal control problems. Kavuncu et al. [7] pursue this idea by utilizing potential differential games, for which Nash equilibria could be obtained by solving an associated optimal control problem. Iterative best response scheme is another family of algorithms that takes advantage of tools for solving optimal control problems [8], [9]. This approach solves Nash equilibria by solving single-agent optimal control problems in a round-robin fashion. The usefulness of this type of algorithm is demonstrated in racing games [8], [9]. A pitfall for this approach is that there are no formal guarantees for convergence to Nash equilibria. Another thrust of game-theoretic solvers focuses on solving for open-loop Nash equilibrium. The key to this type of approach is to formulate the necessary conditions of open-loop Nash equilibria by concatenating the Karush–Kuhn–Tucker conditions for each player. Di et al. [10] solved the corresponding nonlinear program based on Newton’s method. In [3], augmented Lagrangian is combined with Newton’s method to handle state and control constraints. In [4], the constrained nonlinear program is addressed by projected gradient or Douglas-Rachford splitting method. As this type of approach solves for open-loop Nash equilibrium (i.e., controls are functions of time instead of states), special treatments such as model predictive control are required to help to capture the reactive game-theoretic nature [3]. Another limitation of the aforementioned approaches is that they only consider deterministic games and can not handle uncertainties in dynamics and observations.

An increasingly popular branch of game-theoretic solvers aims to compute local feedback Nash equilibrium [2], [5], [6], [10], [11]. The iterative linear-quadratic game method [2] exploits the analytical solutions of linear-quadratic games and solves the game as a sequence of approximated linear-quadratic games. Schwarting et al. [5] propose a belief space variant of differential dynamic programming algorithm for stochastic games, which could handle dynamics and observation uncertainties. However, their work handles chance constraints using exponential penalties, which requires adjusting penalty weights by trial and error. Our work builds upon the iterative linear-quadratic game method and extends to stochastic games with dynamics and observation uncertainty. In contrast to [5], our work directly handles chance constraints using the augmented Lagrangian method. Our approach does not require manually

adjusting penalty weights and achieves a good balance of safety and performance.

B. Chance-Constrained Planning

Chance constraint formulation provides a unified framework for quantifying probabilistic constraint violations, which allows the specification of desired probabilities for constraint violations. However, calculating chance constraint violations is intractable in general, which makes chance-constrained planning challenging. A branch of works exploits the structure of Gaussian belief distributions [12], [13], whether it be a true belief distribution or an approximation. We follow this line of work, exploiting the structure of linear-Gaussian systems via linearization. We further incorporate this technique into an augmented Lagrangian framework. A different line of research exploits moment-based approaches to handle non-Gaussian uncertainties [14]. Han et al. [14] proposed a moment-based method that can handle general chance-constrained motion planning problems. The key to handling general chance-constrained planning problems is to transform chance constraints into a set of deterministic constraints on the moments of the probability distributions of the states, along with an exact moment-based uncertainty propagation method to relate the moments of the uncertain states to control inputs.

III. PROBLEM FORMULATION

We consider the following nonlinear stochastic N -player dynamics:

$$x_{k+1} = f(x_k, u_k^{1:N}, w_k), \quad w_k \sim \mathcal{N}(0, \Sigma_{w_k}), \quad (1)$$

where x_k and $x_{k+1} \in \mathbb{R}^n$ are the states at time step k and $k + 1$. $u_k^i \in \mathbb{R}^{m_i}$, $i \in \{1, \dots, N\}$ is the control input of player i at time step k , and $u_k^{1:N} = \{u_k^1, \dots, u_k^N\}$ is the collection of all players’ controls at time step k . $w_k \in \mathbb{R}^d$ represents the Gaussian process noise at time step k . Σ_{w_k} is the corresponding covariance for w_k . Although the dynamics model considers Gaussian noises, it is still general as the Gaussian noises could go through arbitrary nonlinear transformations through f .

Following [5], we assume that each agent shares the same observation model and state measurement. At the core of CCILQGames is solving a series of linear-quadratic stochastic games. Linear-quadratic stochastic games with certain structures are computationally tractable [15], [16], thanks to the separation principle. In this work, assumption 1 is imposed to make the linear-quadratic stochastic games tractable. We state this assumption formally below:

Assumption 1: All agents share the same joint measurement model and the same state (i.e., x_k) measurement.

With assumption 1, we consider the following measurement model:

$$y_{k+1} = h(x_{k+1}, v_{k+1}), \quad v_{k+1} \sim \mathcal{N}(0, \Sigma_{v_{k+1}}), \quad (2)$$

where $y_{k+1} \in \mathbb{R}^z$ and $v_{k+1} \in \mathbb{R}^s$ are the sensor measurement and the Gaussian measurement noise at time step $k + 1$. $\Sigma_{v_{k+1}}$ is the covariance of the measurement noise.

System dynamics (1) and measurement model (2) jointly define a partially observable Markov decision process, in which

agents maintain a probability distribution of the state x_k . We use belief b_k to represent the probability distribution of state x_k conditioned on control inputs and measurements:

$$b_k = Pr(x_k | y_0, \dots, y_k, u_0^{1:N}, \dots, u_k^{1:N}). \quad (3)$$

Furthermore, we consider chance constraints on state:

$$Pr(g_k(x_k) \leq 0) \geq p, \quad (4)$$

where $g_k(x_k) \leq 0$ is the nonlinear constraint on state x_k at time step k . p is the chance constraint threshold. We assume $g_k(x_k)$ is a scalar.

Each agent has a time-varying feedback control policy $\gamma_i(k, y_1, \dots, y_k) \in \Gamma_i$, which maps the current time step and observations to control input (i.e., $u_k^i = \gamma_i(k, y_1, \dots, y_k)$). Γ_i is the strategy space for player i . Each agent has a cost function J_i defined as the expectation of states and measurements:

$$J_i(\gamma_1, \dots, \gamma_N) = \mathbb{E}_{X_0, \dots, X_L, Y_0, \dots, Y_L} \left[c_L^i(x_L) + \sum_{k=0}^{L-1} c_k^i(x_k, u_k^{1:N}) \right], \quad (5)$$

where L is the planning horizon, and $X_0, \dots, X_L, Y_0, \dots, Y_L$ are random variables for states and measurements. Now we are ready to define the generalized Nash equilibrium, which is the Nash equilibrium under chance constraints:

Definition 1: A set of control policies $\{\gamma_i^*, i \in \{1, \dots, N\}\}$, is a generalized Nash equilibrium if the following inequality holds for all players while satisfying chance constraints (4):

$$J_i(\gamma_1^*, \dots, \gamma_i^*, \dots, \gamma_N^*) \leq J_i(\gamma_1^*, \dots, \gamma_i, \dots, \gamma_N^*), \quad \forall \gamma_i \in \Gamma_i, i \in \{1, \dots, N\}. \quad (6)$$

As we describe in detail in Section IV-C, our work builds upon ILQGames. Since the computation of a Nash equilibrium is generally intractable, ILQGames attempts to find local Nash equilibria (i.e., the inequality in (6) holds within a neighborhood around each player's local Nash equilibrium strategy) [17, Def. 1]. ILQGames solves a series of linear-quadratic game approximations to the original game. More precisely, ILQGames obtains the feedback Nash equilibrium [1, Def. 6.2] solution to the approximate linear-quadratic game. The feedback Nash equilibrium solution is strongly time-consistent [1, Ch. 6, Th. 6.6], which is appealing since the feedback Nash equilibrium strategy remains optimal at a future time step even if there are deviations from the feedback strategy in the past time steps. Given that ILQGames is at the core of our algorithm, CCILQGames also aims to find local Nash equilibrium in practice.

With the above components, we can finally derive the target problem for this work (i.e., finding the generalized Nash equilibrium), which is summarized as follows:

Problem 1: Solve for a generalized Nash equilibrium with the following cost functions, dynamics model, measurement model, and chance constraints:

$$J_i(\gamma_1, \dots, \gamma_N) = \mathbb{E} \left[c_L^i(x_L) + \sum_{k=0}^{L-1} c_k^i(x_k, u_k^{1:N}) \right], \\ x_{k+1} = f(x_k, u_k^{1:N}, w_k), \quad w_k \sim \mathcal{N}(0, \Sigma_{w_k}),$$

$$y_{k+1} = h(x_{k+1}, v_{k+1}), \quad v_{k+1} \sim \mathcal{N}(0, \Sigma_{v_{k+1}}), \\ Pr(g_k^{i,m}(x_k) \leq 0) \geq p_{i,m}, \quad m \in \{1, \dots, M_k^i\}, \\ i \in \{1, \dots, N\},$$

where m , $p_{i,m}$, and M_k^i are the index for the m th chance constraint, chance constraint threshold and the total number of constraints at time step k for agent i .

IV. CHANCE-CONSTRAINED ITERATIVE LINEAR-QUADRATIC GAME

A. System Linearization and Belief Dynamics

Given the current belief b_k and next time step's measurement y_{k+1} , the belief dynamics could be described using the Bayesian filter [18]. Following [13], [19], we linearize the dynamical system around a nominal trajectory and then apply the Kalman filter to approximately track the belief dynamics. Kalman filter approximates the true belief as a Gaussian distribution, such that $b_k = (\hat{x}_k, \Sigma_{x_k})$, where \hat{x}_k is the mean and Σ_{x_k} is the covariance.

Given a nominal trajectory $\bar{b} = \{\bar{b}_k = (\bar{x}_k, \bar{\Sigma}_{x_k}), k \in \{0, \dots, L\}\}$, $\bar{u}_0^{1:N}, \dots, \bar{u}_{L-1}^{1:N}$, satisfying $\bar{x}_{k+1} = f(\bar{x}_k, \bar{u}_k^{1:N}, 0)$, we linearize the dynamics (1) and measurement models (2):

$$x_{k+1} = \bar{x}_{k+1} + A_k(x_k - \bar{x}_k) + \sum_{j=1}^N B_j(u_k^j - \bar{u}_k^j) + W_k w_k, \\ y_{k+1} = h(\bar{x}_{k+1}, 0) + H_{k+1}(x_{k+1} - \bar{x}_{k+1}) + V_{k+1} v_{k+1}, \quad (7)$$

where $A_k, B_j, W_k, H_{k+1}, V_{k+1}$ are:

$$A_k = \frac{\partial f(\bar{x}_k, \bar{u}_k^{1:N}, 0)}{\partial x_k}, \quad B_j = \frac{\partial f(\bar{x}_k, \bar{u}_k^{1:N}, 0)}{\partial u_k^j}, \\ W_k = \frac{\partial f(\bar{x}_k, \bar{u}_k^{1:N}, 0)}{\partial w_k}, \quad H_{k+1} = \frac{\partial h(\bar{x}_{k+1}, 0)}{\partial x_{k+1}}, \\ V_{k+1} = \frac{\partial h(\bar{x}_{k+1}, 0)}{\partial v_{k+1}}. \quad (8)$$

Now we can use Kalman filter to propagate next time step's belief $b_{k+1} = (\hat{x}_{k+1}, \Sigma_{x_{k+1}})$ with the linearized dynamics, observation model, and current belief $b_k = (\hat{x}_k, \Sigma_{x_k})$:

$$\hat{x}_{k+1}^p = \bar{x}_{k+1} + A_k(\hat{x}_k - \bar{x}_k) + \sum_{j=1}^N B_j(u_k^j - \bar{u}_k^j), \\ \Sigma_{x_{k+1}}^p = A_k \Sigma_{x_k} A_k^T + W_k \Sigma_{w_k} W_k^T, \\ \hat{x}_{k+1} = \bar{x}_{k+1} + A_k(\hat{x}_k - \bar{x}_k) + \sum_{j=1}^N B_j(u_k^j - \bar{u}_k^j) \\ + K_{k+1}(y_{k+1} - (h(\bar{x}_{k+1}, 0) + H_{k+1}(\hat{x}_{k+1}^p - \bar{x}_{k+1}))), \\ \Sigma_{x_{k+1}} = (I - K_{k+1} H_{k+1}) \Sigma_{x_{k+1}}^p, \\ K_{k+1} = \Sigma_{x_{k+1}}^p H_{k+1}^T (H_{k+1} \Sigma_{x_{k+1}}^p H_{k+1}^T + V_{k+1} \Sigma_{v_{k+1}} V_{k+1}^T)^{-1}, \quad (9)$$

where \hat{x}_{k+1}^p is the state's prior estimation, $\Sigma_{x_{k+1}}^p$ is the covariance's prior estimation, and K_{k+1} is the Kalman gain. The innovation term $y_{k+1} - (h(\bar{x}_{k+1}, 0) + H_{k+1}(\hat{x}_k - \bar{x}_k))$ is random, which renders the mean propagation stochastic. Although the mean propagation is stochastic, the covariance propagation is deterministic, dependent neither on measurements nor controls.

B. Outer Loop: Augmented Lagrangian Scheme

1) *Augmented Lagrangian Formulation*: The augmented Lagrangian associated with player i is formulated by augmenting the cost function (5) with Lagrange multiplier terms and quadratic penalty terms:

$$\begin{aligned} & l_i(X_0, \dots, X_L, Y_0, \dots, Y_L, u_0^{1:N}, \dots, u_{L-1}^{1:N}) \\ &= \mathbb{E}_{X_0, \dots, X_L, Y_0, \dots, Y_L} \left[c_L^i(x_L) + \sum_{k=0}^{L-1} c_k^i(x_k, u_k^{1:N}) \right] \\ &+ \sum_{k=1}^L \sum_{m=1}^{M_k} \lambda_{i,m} [p_{i,m} - \Pr(g_k^{i,m}(x_k) \leq 0)] \\ &+ \sum_{k=1}^L \sum_{m=1}^{M_k} \frac{I_{i,m}}{2} [p_{i,m} - \Pr(g_k^{i,m}(x_k) \leq 0)]^2, \end{aligned} \quad (10)$$

where $\lambda_{i,m} \in \mathbb{R}$ is the Lagrange multiplier. $I_{i,m}$ is defined as:

$$I_{i,m} = \begin{cases} 0, & p_{i,m} - \Pr(g_k^{i,m}(x_k) \leq 0) < 0 \wedge \lambda_{i,m} = 0, \\ \mu_{i,m}, & \text{otherwise.} \end{cases} \quad (11)$$

where $\mu_{i,m} \in \mathbb{R}$ is the quadratic penalty term.

2) *Augmented Lagrangian Update*: Given a trajectory, the Lagrange multipliers and penalty terms are updated as follows [20]:

$$\begin{aligned} \lambda_{i,m} &= \max(0, \lambda_{i,m} + \mu_{i,m}(p_{i,m} - \Pr(g_k^{i,m}(x_k) \leq 0))), \\ \mu_{i,m} &= \phi \mu_{i,m}, \end{aligned} \quad (12)$$

where $\phi > 1$ is the increasing schedule.

3) *Evaluating Chance Constraints Violations*: Evaluating chance constraint violations is necessary to update the Lagrange multipliers. However, directly calculating chance constraint violations could be intractable in general. Hence, we follow the approach of [12], [13] to linearize chance constraints.

Given a trajectory with belief $\bar{b}_k = (\bar{x}_k, \bar{\Sigma}_{x_k})$ and control $\bar{u}_k^{1:N}$, we get a linearization of $g_k^{i,m}(x_k)$ around \bar{x}_k :

$$G_k^{i,m} x_k + q_k^{i,m}, \quad (13)$$

where

$$G_k^{i,m} = \frac{\partial g_k^{i,m}(\bar{x}_k)}{\partial x_k}, \quad q_k^{i,m} = g_k^{i,m}(\bar{x}_k) - G_k^{i,m} \bar{x}_k. \quad (14)$$

Now we can consider the following linearized chance constraints:

$$\Pr(G_k^{i,m} x_k + q_k^{i,m} \leq 0) \geq p_{i,m}. \quad (15)$$

To step further, we can decompose the Gaussian random variable x_k into its mean \bar{x}_k and $\bar{e}_k = x_k - \bar{x}_k \sim \mathcal{N}(0, \bar{\Sigma}_{x_k})$:

$$\Pr(G_k^{i,m}(\bar{x}_k + \bar{e}_k) + q_k^{i,m} \leq 0) \geq p_{i,m}. \quad (16)$$

By rearranging terms, chance constraint (15) can be transformed to a deterministic linear constraint on mean \bar{x}_k [12]:

$$G_k^{i,m} \bar{x}_k \leq -q_k^{i,m} - \bar{\rho}_k^{i,m}. \quad (17)$$

$\bar{\rho}_k^{i,m}$ could be calculated using the quantile function for univariate Gaussian, as $G_k^{i,m} \bar{e}_k \sim \mathcal{N}(0, G_k^{i,m} \bar{\Sigma}_{x_k} (G_k^{i,m})^T)$:

$$\begin{aligned} \Pr(G_k^{i,m} e_k \leq \bar{\rho}_k^{i,m}) &= p_{i,m}, \\ \bar{\rho}_k^{i,m} &= \sqrt{2G_k^{i,m} \bar{\Sigma}_{x_k} (G_k^{i,m})^T} \operatorname{erf}^{-1}(2p_{i,m} - 1), \end{aligned} \quad (18)$$

where erf^{-1} is the inverse error function. Since mean \bar{x}_k and covariance $\bar{\Sigma}_{x_k}$ are known for the nominal trajectory, the violation of constraint (17) can be directly calculated.

C. Inner Loop: Iterative Linear Quadratic Game

The inner loop iteratively solves a linear-quadratic stochastic game. We present how to construct the approximate linear-quadratic game from the original game and solve it.

1) *Approximate Linear-Quadratic Game*: The inner loop of the proposed CCILQGames performs an iterative linear-quadratic stochastic game step, which involves iteratively solving a linear-quadratic stochastic game. We now present how to obtain an approximated linear-quadratic stochastic game with a given nominal trajectory $\bar{b} = \{b_k = (\bar{x}_k, \bar{\Sigma}_{x_k}), k \in \{0, \dots, L\}\}$, control strategies $\{\bar{\gamma}_i\}$, $i \in \{1, \dots, N\}$ and controls $\bar{u}_0^{1:N}, \dots, \bar{u}_{L-1}^{1:N}$.

Toward this end, we first obtain the linearized dynamics and observation model (7) around the nominal trajectory, as described in Section IV-A.

Moreover, a quadratic approximation to the running costs $c_k^i(x_k, u_k^{1:N})$ is necessary. Yet, before we are able to proceed, we need to put the Lagrange multiplier and penalty terms associated with chance constraints inside the expectation operator.

We proceed in the same manner as in Section IV-B.(3). By linearizing around the nominal trajectory, we obtain a linear approximation to chance constraints in the form of a deterministic linear constraint on the mean \hat{x}_k :

$$\begin{aligned} G_k^{i,m} \hat{x}_k &\leq -q_k^{i,m} - \rho_k^{i,m}, \\ \rho_k^{i,m} &= \sqrt{2G_k^{i,m} \Sigma_{x_k} (G_k^{i,m})^T} \operatorname{erf}^{-1}(2p_{i,m} - 1), \end{aligned} \quad (19)$$

where $q_k^{i,m}$ is defined in (14). Since the covariance propagation is deterministic as we mentioned in Section IV-A, we can precompute the covariance before solving the game.

The linear constraint (19) now replaces the original chance constraint. The Lagrange multiplier term can be directly put inside the expectation operator:

$$\begin{aligned} & \lambda_{i,m} (G_k^{i,m} \hat{x}_k + q_k^{i,m} + \rho_k^{i,m}) \\ &= \mathbb{E}_{X_k} [\lambda_{i,m} (G_k^{i,m} x_k + q_k^{i,m} + \rho_k^{i,m})]. \end{aligned} \quad (20)$$

For any probability distribution of X_k , $I_{i,m}$ is determined since the mean \hat{x}_k is determined. Now, the penalty term can be transformed as follows:

$$\begin{aligned}
& \frac{I_{i,m}}{2} (G_k^{i,m} \hat{x}_k + q_k^{i,m} + \rho_k^{i,m})^2 \\
&= \frac{I_{i,m}}{2} \left[\hat{x}_k^T G_k^{i,m T} G_k^{i,m} \hat{x}_k + 2(q_k^{i,m} + \rho_k^{i,m})^T G_k^{i,m} \hat{x}_k \right. \\
&\quad \left. + (q_k^{i,m} + \rho_k^{i,m})^T (q_k^{i,m} + \rho_k^{i,m}) \right] \\
&= \frac{I_{i,m}}{2} \mathbb{E}_{X_k} \left[x_k^T G_k^{i,m T} G_k^{i,m} x_k + 2(q_k^{i,m} + \rho_k^{i,m})^T G_k^{i,m} x_k \right. \\
&\quad \left. + (q_k^{i,m} + \rho_k^{i,m})^T (q_k^{i,m} + \rho_k^{i,m}) \right] \\
&\quad - \frac{I_{i,m}}{2} \text{trace}(G_k^{i,m T} G_k^{i,m} \Sigma_{x_k}), \tag{21}
\end{aligned}$$

Since $G_k^{i,m}$ and Σ_{x_k} are constants after linearization around the nominal trajectory, $\text{trace}(G_k^{i,m T} G_k^{i,m} \Sigma_{x_k})$ is also constant. But the penalty $I_{i,m}$ indeed depends on the control, as the control affects the probability distribution of X_k . So when constructing the linear-quadratic game, the trace term $\text{trace}(G_k^{i,m T} G_k^{i,m} \Sigma_{x_k})$ should be taken into account. However, calculating the derivative and hessian of it requires calculating second and third order derivatives of the dynamics, which could add heavy computational costs. So we omit the trace term in each agent's augmented Lagrangian as an approximation in this work.

Combining (20) and (21), the Lagrange multiplier and penalty terms are transformed into the expectation operator. We denote the modified running cost as $\tilde{c}_k^i(x_k, u_k^{1:N})$. Finally, we are equipped to obtain a quadratic approximation for the running costs:

$$\begin{aligned}
\tilde{c}_k^i(x_k, u_k^{1:N}) &\approx \tilde{c}_k^i(\bar{x}_k, \bar{u}_k^{1:N}) + \frac{1}{2} \delta x_k^T (Q_k^i \delta x_k + 2l_k^i) \\
&\quad + \frac{1}{2} \sum_{j=1}^N \delta u_k^{j T} \left(R_k^{ij} \delta u_k^j + 2r_k^{ij} \right), \tag{22}
\end{aligned}$$

where $\delta x_k = x_k - \bar{x}_k$, $\delta u_k^j = u_k^j - \bar{u}_k^j$, Q_k^i, R_k^{ij} are the Hessians with respect to x_k and u_k^j , and l_k^i, r_k^{ij} are the gradients with respect to x_k and u_k^j . We take the value of $\frac{I_{i,m}}{2}$ as the same as the nominal trajectory and treat it as a constant as an approximation when constructing the quadratic cost approximation. Same as [2], we omit the mix partial derivatives.

2) *Separation Principle for Linear-Quadratic Stochastic Game*: With (7) and (22), we have constructed a linear-quadratic stochastic game. Next, we present the separation principle for linear-quadratic stochastic games, which would be exploited to solve the linear-quadratic stochastic game.

Theorem 1 (Separation principle for linear-quadratic stochastic games): With assumption 1, We consider stochastic games with linear dynamics, and quadratic costs for each player i :

$$x_{k+1} = A_k x_k + \sum_{j=1}^N B_j u_k^j + W_k w_k,$$

$$y_{k+1} = H_{k+1} x_{k+1} + V_{k+1} v_{k+1},$$

$$\begin{aligned}
J_i &= \mathbb{E}_{X_0, \dots, X_L, Y_0, \dots, Y_L} \left[\frac{1}{2} x_L^T (Q_L^i x_L + 2l_L^i) \right. \\
&\quad \left. + \sum_{k=0}^{L-1} \left(\frac{1}{2} x_k^T (Q_k^i x_k + 2l_k^i) \right) \right. \\
&\quad \left. + \frac{1}{2} \sum_{j=1}^N u_k^{j T} \left(R_k^{ij} u_k^j + 2r_k^{ij} \right) \right],
\end{aligned}$$

where w_k and v_{k+1} are zero mean Gaussian white noise. The initial Gaussian belief $(\hat{x}_k, \Sigma_{x_k})$ is known and independent of the additive noises.

Then the Nash equilibrium strategy of the above stochastic linear-quadratic game is the Nash equilibrium strategy of the deterministic linear-quadratic game applied to the estimated state \hat{x}_k provided by the Kalman filter.

The proof of Theorem 1 is in the appendix. Thanks to Theorem 1, we can obtain the Nash equilibrium strategy of the stochastic linear-quadratic game by solving the deterministic linear-quadratic game, which has an analytical solution [1].

D. Algorithm for Chance-Constrained Iterative Linear-Quadratic Stochastic Game

The proposed algorithm is summarized in algorithm 1. After obtaining the Nash equilibrium, a line searching procedure is necessary to find a suitable step size. We use the definition of step size as in (7) in [2]. We define the merit function to be $M = \frac{1}{2} \|\frac{\partial l^i}{\partial u^i}\|^2$ and apply a backtracking line search procedure to find a step size that satisfies the Armijo condition [20]. We follow [5], [21] to use a zero-noise realization to forward the dynamics. Then we use the extended Kalman filter [18] to propagate the belief to obtain the next iteration's nominal trajectory.

V. EXPERIMENTS

A. Experimental Setup

In this section, we evaluate the performance of CCILQGames in three autonomous driving scenarios: lane merging, three-player intersection, and roundabout. We also further conduct a Monte Carlo study to test the proposed algorithm in noisy environments. We compare our algorithm to SDG with different constraint penalty weights. Since SDG is not open-sourced, we implement SDG by ourselves. All experiments are conducted on a 2.2 GHz Intel Core i7 laptop. We run CCILQGames 50 times for each scenario to calculate the computation time's mean and standard deviation. Computation time for lane merging, roundabout, and three-player intersection are $0.1017 \pm 0.007.3$ s, 0.4230 ± 0.0212 s, and 0.0866 ± 0.0086 s respectively.

1) *Dynamics and Measurement Model*: We use the unicycle model as the vehicle dynamic model throughout the three scenarios. The state of the vehicle contains its position, heading angle, and velocity. The agent's control input is composed of its angular velocity and scalar acceleration.

Algorithm 1: Chance-Constrained Iterative Linear-Quadratic Stochastic Game.

Input: Initial belief $b_0 = (\hat{x}_0, \Sigma_{x_0})$, nominal control \bar{u}
Output: Belief trajectory $\bar{b} = \{\bar{b}_0, \bar{b}_1, \dots, \bar{b}_L\}$, \bar{u} and feedback control policy $\{\gamma_i\}, i \in \{1, \dots, N\}$

```

1  $\bar{b} \leftarrow$  Propagate belief with  $\bar{u}$ ;
2 while max chance constraint error > tolerance do
3   Linearize chance constraints around  $\bar{b}$  and evaluate chance constraint violations;
4   update  $\lambda, \mu$ ;
5   while not converge do
6     Linearize the system and chance constraints around the nominal trajectory  $\bar{b}, \bar{u}$ ;
7     Transform Lagrange multiplier and penalty terms, then add to the running costs;
8     Obtain a quadratic cost approximation;
9      $\{\gamma_i\} \leftarrow$  Solve the linear-quadratic stochastic game;
10    Backtracking line search for step size  $\alpha$ ;
11     $\bar{b}, \bar{u} \leftarrow$  Propagate belief with control policy  $\{\gamma_i\}$ ;
12  end
13 end

```

We use the following measurement model for each agent in all three scenarios:

$$y_{k+1}^i = x_{k+1}^i + v_{k+1}^i, \quad v_{k+1}^i \sim \mathcal{N}(0, \Sigma_{v_{k+1}^i}), \quad (23)$$

where x_{k+1}^i, y_{k+1}^i are the state and measurement of agent i at time step $k+1$. v_{k+1}^i is the associate noise. The joint measurement model is just the concatenation of all agents' measurement models.

2) *Costs and Chance Constraints:* We consider quadratic costs for distance from the lane center, deviation from the nominal speed, and control efforts, which are in the same form as in [2]. We consider chance constraints on the proximity and collisions with obstacles. For the proximity constraints, the minimum distance threshold between two agents is set to 3 m. The obstacle avoidance constraint is formulated using the convex feasible set [13], [22]. We augment the cost of the SDG simply with quadratic penalties for proximity constraint violations and linear penalties for obstacle avoidance constraint violation, all weighted by a penalty weight (1 or 400). We set the chance constraint threshold p to 0.95 in all experiments.

B. Lane Merging

As shown in Fig. 2, two vehicles need to merge into the same lane in the lane merging scenario. Also, vehicles need to avoid collisions between each other and obstacles on the lane boundary. The planning horizon is 3 seconds with 20 time steps.

Fig. 2(a) illustrates the planned trajectory using CCILQGames, in which the ellipses represent the covariance of the state. As demonstrated in the speed profile (Fig. 2(b)), the red vehicle speeds up to cut into the lane ahead of the blue vehicle. By contrast, the blue vehicle slows down and

allows the blue vehicle to merge. As shown in Fig. 2(c), a small penalty weight of 1 leads to unsafe strategies: two vehicles do not avoid collisions with each other. The Monte Carlo tests further validate a high probability of constraint violations. A large penalty weight of 400 leads the blue vehicle to slow down to let the red vehicle passes first. Notice that the penalty weight of 400 does not lead to a conservative strategy in this case. But the corresponding strategy is less safe than CCILQGames's strategy, as we would illustrate in the Monte Carlo Tests.

C. Three-Player Intersection

In the three-player intersection scenario, the planning horizon is 2.5 s with 16 time steps. Fig. 1(a) illustrates the trajectory planned by CCILQGames. The red vehicle accelerates to pass the intersection ahead of the green vehicle. On the other hand, the blue vehicle slows down to wait for the red vehicle to go across the intersection. Interestingly, the red vehicle tweaks to avoid collisions. SDG with a small penalty weight could lead to unsafe behaviors. All vehicles go straight across the intersection without avoiding collisions (Fig. 1(c)). However, a large penalty weight again leads to conservative strategies, as only the red vehicle could pass the intersection while the others turn back (Fig. 1(d)).

D. Roundabout

The planning horizon is 2.5 s with 16 time steps for the roundabout scenario. Fig. 2(e) illustrates the trajectory planned by CCILQGames. The red vehicle accelerates to escape the roundabout and cuts into the lane earlier than the blue vehicle. The blue vehicle's strategy is to slow down to let the red vehicle pass first. The green vehicle also slows down to leave space for the blue vehicle. Fig. 2(g) shows the trajectory for SDG with a penalty weight of 1. The red vehicle turns around after escaping the roundabout since the cost only encourages to stay close to the lane center instead of specifying a direction. The strategy leads to collisions, which are validated by the Monte Carlo Tests. By comparison, a large penalty weight again leads to conservative behaviors, as only the blue vehicle navigates successfully (Fig. 2(h)). The green vehicle slows down and could not get out of the Roundabout while the red vehicle turns back.

E. Monte Carlo Tests

We conduct a Monte Carlo study to test the performance of CCILQGames in stochastic environments. We run 100 trials for each scenario. The process noises are sampled from $\mathcal{N}(0, 0.1I)$, with the exception of heading angle's variance being 0.05. The observation noises are sampled from $\mathcal{N}(0, 0.6I)$, with the exception of angular velocity's variance being 0.1. The noises' covariances are the same for planning (the planned trajectories are shown in previous subsections) and Monte Carlo testing. Recall that our constraints are imposed on all the time steps in all three experiments. As shown in our problem formulation, we treat the same type of constraints at different time step as different constraints (e.g., the proximity constraint at time steps 1 and 2 are counted as two different constraints). Therefore, we calculate the statistics of constraint violations for each constraint

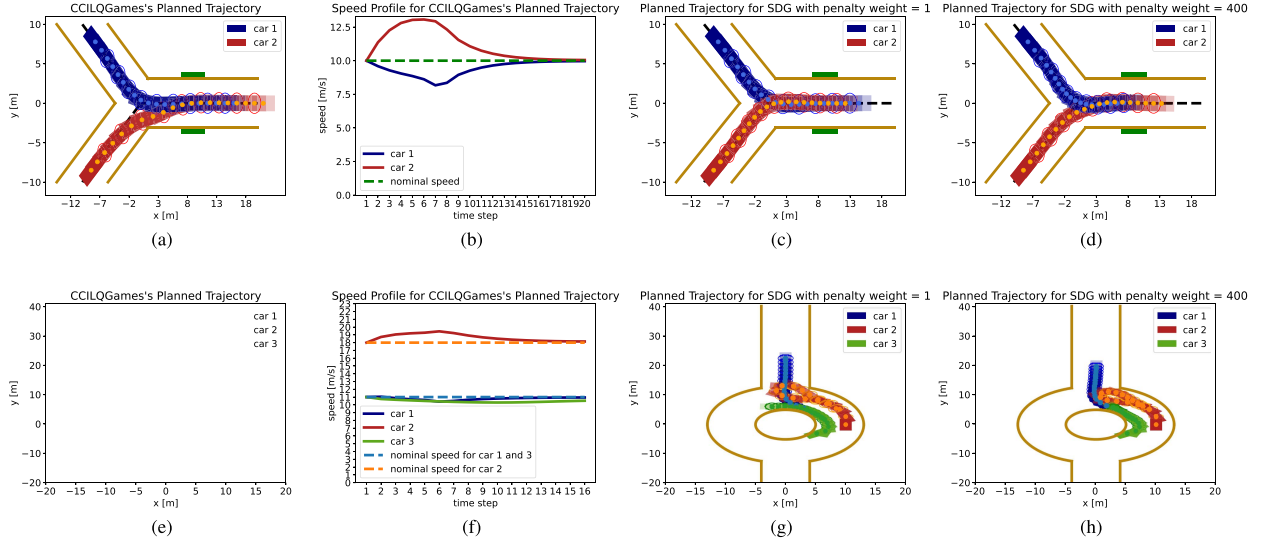


Fig. 2. The first row illustrates the results for the lane merging scenario, while the second row shows the results for the roundabout scenario.

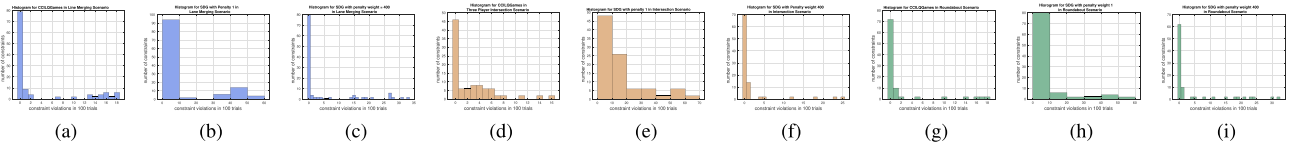


Fig. 3. Histograms of constraint violations for all the constraints in 100 Monte Carlo Tests, with sampled process and observation noises.

in 100 trials. Fig. 3 shows the histograms of constraint violations in 100 trials for CCILQGames and SDG in the three scenarios.

The maximum number of constraint violations among all the constraints for CCILQGames are 18,16,18 in lane merging, intersection and roundabout scenarios in 100 trials. By comparison, the maximum number of constraint violations among all the constraints for SDG with penalty weights 1(400) are 54 (33), 62 (25), 54 (22) in lane merging, intersection and roundabout scenarios in 100 trials.

F. Limitations

We would like to discuss several limitations of CCILQGames. First, since we handle the chance constraints via linear approximations, there are no strict chance constraint satisfaction guarantees. Additionally, the linearization techniques could only specify a chance constraint threshold for each individual constraint instead of a joint threshold for satisfying all the constraints. As a result, CCILQGames achieves rates of 48%, 51%, and 51% of constraint satisfaction for all the constraints in lane merging, intersection and roundabout scenarios in 100 trials. By comparison, SDG with the penalty weight 1 (400) achieves rates of 13% (35%), 6% (65%), 19% (32%) in lane merging, intersection and roundabout scenarios in 100 trials. We argue that the above limitations could be mitigated or resolved via a Model Predictive Control (MPC) formulation. For example, [23] uses the same technique to handle chance constraints in an MPC framework and achieves zero chance constraint violations in

experiments. Another possible way to mitigate this limitation is to use the Bonferroni correction [24] to assign a chance constraint threshold to each individual constraint. One more limitation of CCILQGames is that the Kalman filter could experience numerical instabilities in some cases (e.g., the covariance matrix becomes not symmetric positive definite).

VI. CONCLUSION

We have presented a novel algorithm for solving stochastic dynamic games under chance constraints. Our work extends the deterministic ILQGames to stochastic games with both observation and process noises while handling chance constraints using the framework of augmented Lagrangian. We showcased the proposed algorithm in the lane merging, three-player intersection, and roundabout scenarios. The experimental results proved the effectiveness of the proposed approach.

APPENDIX

A. Proof for Theorem 1

Proof: We can reformulate the cost function as follows:

$$J_i = \mathbb{E}_{X_0, \dots, X_L, Y_0, \dots, Y_L} \left[\frac{1}{2} x_L^T (Q_L^i x_L + 2l_L^i) \right. \\ \left. + \sum_{k=0}^{L-1} \left(\frac{1}{2} x_k^T (Q_k^i x_k + 2l_k^i) \right) \right]$$

$$\begin{aligned}
& \left. + \frac{1}{2} \sum_{j=1}^N u_k^{jT} (R_k^{ij} u_k + 2r_k^{ij}) \right) \Bigg], \\
= & \mathbb{E}_{Y_0, \dots, Y_L} \left[\frac{1}{2} \hat{x}_L^T (Q_L^i \hat{x}_L + 2l_L^i) \right. \\
& + \sum_{k=0}^{L-1} \left(\frac{1}{2} \hat{x}_k^T (Q_k^i \hat{x}_k + 2l_k^i) \right. \\
& \left. \left. + \frac{1}{2} \sum_{j=1}^N u_k^{jT} (R_k^{ij} u_k + 2r_k^{ij}) \right) \right] \\
& + \sum_{k=0}^L \frac{1}{2} \text{trace}(Q_k^i \Sigma_{x_k}), \quad (24)
\end{aligned}$$

where \hat{x}_k is the mean of the state at time step k . Notice that the trace terms $\text{trace}(Q_k^i \Sigma_{x_k})$ are constants and do not depend on control strategies. Due to assumption 1, the dynamics of \hat{x}_k are the same for different agents. In other words, given the same realizations of the observation noises, all the agents would get the same \hat{x}_k from their Kalman filters. Now since \hat{x}_k is fully observable (given by Kalman filter [25]), we have reduced the problem to a stochastic linear-quadratic game with closed-loop perfect state information pattern. By Corollary 6.4 in [1], the Nash equilibrium of the linear-quadratic stochastic game with exact state information coincides with the deterministic version of the linear-quadratic game. We add linear costs of states and controls into the cost function, which is slightly different than the original setting in [1]. But the corollary still holds, with minor modifications in the derivation, which is not included due to page limits. \square

REFERENCES

- [1] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA, USA: SIAM, 1998.
- [2] D. Fridovich-Keil, E. Ratner, L. Peters, A. D. Dragan, and C. J. Tomlin, "Efficient iterative linear-quadratic approximations for nonlinear multi-player general-sum differential games," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 1475–1481.
- [3] S. L. Cleach, M. Schwager, and Z. Manchester, "ALGAMES: A fast solver for constrained dynamic games," in *Proc. Robot.: Sci. Syst.*, Corvallis, Oregon, USA, 2020.
- [4] B. Di and A. Lamperski, "Local first-order algorithms for constrained nonlinear dynamic games," in *Proc. IEEE Amer. Control Conf.*, 2020, pp. 5358–5363.
- [5] W. Schwarting, A. Pierson, S. Karaman, and D. Rus, "Stochastic dynamic games in belief space," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 2157–2172, Dec. 2021.
- [6] D. Fridovich-Keil, V. Rubies-Royo, and C. J. Tomlin, "An iterative quadratic method for general-sum differential games with feedback linearizable dynamics," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 2216–2222.
- [7] T. Kavuncu, A. Yaraneri, and N. Mehr, "Potential iLQR: A potential-minimizing controller for planning multi-agent interactive trajectories," in *Proc. Robotics: Sci. Syst.*, 2021.
- [8] Z. Wang, R. Spica, and M. Schwager, "Game theoretic motion planning for multi-robot racing," in *Distributed Autonomous Robotic Systems*. Berlin, Germany: Springer, 2019, pp. 225–238.
- [9] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game-theoretic planning for self-driving cars in multivehicle competitive scenarios," *IEEE Trans. Robot.*, vol. 37, no. 4, pp. 1313–1325, Aug. 2021.
- [10] B. Di and A. Lamperski, "Newton's method, Bellman recursion and differential dynamic programming for unconstrained nonlinear dynamic games," *Dyn. Games Appl.*, vol. 12, pp. 394–442, 2021.
- [11] F. Laine, D. Fridovich-Keil, C.-Y. Chiu, and C. Tomlin, "The computation of approximate generalized feedback nash equilibria," 2021, *arXiv:2101.02900*.
- [12] A. Carvalho, Y. Gao, S. Lefevre, and F. Borrelli, "Stochastic predictive control of autonomous vehicles in uncertain environments," in *Proc. 12th Int. Symp. Adv. Veh. Control*, 2014, pp. 712–719.
- [13] J. Chen, Y. Shimizu, L. Sun, M. Tomizuka, and W. Zhan, "Constrained iterative LQG for real-time chance-constrained Gaussian belief space planning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 5801–5808.
- [14] W. Han, A. Jasour, and B. Williams, "Non-Gaussian risk bounded trajectory optimization for stochastic nonlinear systems in uncertain environments," in *Proc. Int. Conf. Robot. Automat.*, 2022, pp. 11044–11050.
- [15] I. Rhodes and D. Luenberger, "Stochastic differential games with constrained state estimators," *IEEE Trans. Autom. Control*, vol. 14, no. 5, pp. 476–481, Oct. 1969.
- [16] P. Kumar and J. V. Schuppen, "On Nash equilibrium solutions in stochastic dynamic games," *IEEE Trans. Autom. Control*, vol. 25, no. 6, pp. 1146–1149, Dec. 1980.
- [17] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "On the characterization of local nash equilibria in continuous games," *IEEE Trans. Autom. Control*, vol. 61, no. 8, pp. 2301–2307, Aug. 2016.
- [18] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. Cambridge, MA, USA: MIT Press, 2005.
- [19] W. Li and E. Todorov, "Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system," *Int. J. Control*, vol. 80, no. 9, pp. 1439–1453, 2007. [Online]. Available: <https://doi.org/10.1080/00207170701364913>
- [20] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer Series in Operations Research and Financial Engineering), 2nd ed. New York, NY, USA: Springer, 2006.
- [21] J. Van Den S. Berg Patil and R. Alterovitz, "Motion planning under uncertainty using iterative local optimization in belief space," *Int. J. Robot. Res.*, vol. 31, no. 11, pp. 1263–1278, 2012.
- [22] C. Liu, C.-Y. Lin, and M. Tomizuka, "The convex feasible set algorithm for real time optimization in motion planning," *SIAM J. Control Optim.*, vol. 56, no. 4, pp. 2712–2733, 2018.
- [23] H. Zhu and J. Alonso-Mora, "Chance-constrained collision avoidance for MAVs in dynamic environments," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 776–783, Apr. 2019.
- [24] T. J. VanderWeele and M. B. Mathur, "Some desirable properties of the Bonferroni correction: Is the Bonferroni correction really so bad?," *Amer. J. Epidemiol.*, vol. 188, no. 3, pp. 617–618, 2019.
- [25] B. D. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*. Chelmsford, MA, USA: Courier Corporation, 2007.