Robust Belief-Space Games: Algorithms for Adversarial Control Under Uncertainty

Anonymous Authors

***Abstract*—This paper presents a comprehensive framework for robust control in belief space under adversarial disturbances, addressing the fundamental challenge of optimal decision-making when both state uncertainty and adversarial perturbations are present. We formulate the problem as a zero-sum dynamic game in the space of probability distributions (belief space) and derive computationally tractable algorithms through finite-dimensional approximations. Our approach includes two novel CPU-efficient methods: (1) Gaussian Game-iLQG, which reduces the infinite-dimensional Hamilton-Jacobi-Bellman-Isaacs (HJBI) equation to finite-dimensional Riccati-type equations through Gaussian belief approximation, and (2) Particle Min-Max Game, which employs sample-based belief representation with alternating optimization. Extensive numerical experiments on 2D navigation and racing scenarios demonstrate that our methods achieve 15-20% better worst-case performance compared to standard robust control approaches while maintaining computational efficiency suitable for real-time applications. The theoretical analysis provides convergence guarantees and robustness bounds, while empirical validation shows 92-95% success rates in adversarial environments with computation times under 3 seconds on standard CPU hardware.**

***Index Terms*—robust control, belief space, dynamic games, Hamilton-Jacobi-Bellman-Isaacs, particle filtering, adversarial learning**

I. Introduction

The problem of robust control under uncertainty has been a cornerstone of control theory for decades, with applications spanning autonomous navigation [1], robotics [2], and aerospace systems [3]. However, traditional robust control methods often suffer from conservatism when dealing with both state uncertainty and adversarial disturbances simultaneously. This limitation becomes particularly pronounced in belief space, where the controller must reason about probability distributions over states rather than point estimates.

Recent advances in stochastic games [4] and belief space planning [5] have opened new avenues for addressing these challenges. The key insight is to formulate the robust control problem as a zero-sum dynamic game in belief space, where the controller minimizes cost while an adversary maximizes it through bounded disturbances. This formulation naturally captures both the uncertainty in state estimation and the adversarial nature of disturbances.

The main contributions of this paper are:

1) A rigorous mathematical formulation of robust control as a zero-sum game in belief space, leading to a functional Hamilton-Jacobi-Bellman-Isaacs (HJBI) equation.
2) Two computationally efficient algorithms: Gaussian Game-iLQG for smooth dynamics with unimodal beliefs, and Particle Min-Max Game for highly nonlinear systems with multimodal beliefs.
3) Theoretical analysis providing convergence guarantees and worst-case performance bounds for both proposed methods.
4) Comprehensive experimental validation on realistic scenarios, demonstrating superior performance compared to existing robust control approaches.
5) Open-source implementation optimized for CPU computation, enabling reproducible research and practical deployment.

The remainder of this paper is organized as follows: Section II reviews related work in robust control and belief space planning. Section III presents our mathematical formulation and proposed algorithms. Section IV provides theoretical analysis. Section V presents extensive experimental results. Section VI discusses implications and limitations. Section VII concludes the paper.

II. Related Work

A. Robust Control Theory

Classical robust control theory has developed several frameworks for handling uncertainty and disturbances. H_∞ control [6] provides worst-case guarantees against bounded disturbances but often leads to conservative solutions. Minimax optimal control [7] addresses adversarial settings through differential games, leading to Hamilton-Jacobi-Isaacs equations that are computationally intractable for high-dimensional systems.

Model Predictive Control (MPC) approaches [8] have incorporated robustness through tube-based methods [9] and scenario-based optimization [10]. However, these methods typically assume perfect state information or simple uncertainty models, limiting their applicability to belief space problems.

B. Belief Space Planning

Belief space planning has emerged as a principled approach to decision-making under uncertainty [11]. Early work focused on discrete POMDPs [12], while recent advances have addressed continuous belief spaces through sampling-based methods [5] and local approximations [13].

The Linear-Quadratic-Gaussian (LQG) framework has been extended to belief space through the Belief-iLQG algorithm

[13], which linearizes dynamics around nominal trajectories and approximates beliefs as Gaussian distributions. However, this approach does not consider adversarial disturbances, limiting its robustness guarantees.

C. Stochastic Games and Dynamic Programming

Stochastic games provide a natural framework for multi-agent decision-making under uncertainty [14]. Zero-sum stochastic games, in particular, have been extensively studied in the context of robust control [4]. The connection between robust control and game theory was established early [15], leading to minimax formulations of optimal control problems.

Recent work has explored games in belief space for multi-agent systems [16], providing the theoretical foundation for our approach. However, computational methods for solving such games remain limited, particularly for continuous belief spaces.

D. Particle Filtering and Sequential Monte Carlo

Particle filtering [17] has become a standard tool for nonlinear state estimation. The basic Sequential Importance Resampling (SIR) algorithm [18] and its variants [19] provide flexible frameworks for belief representation in complex systems.

Recent advances have explored particle filters in control applications [2], including belief space planning [5] and robust estimation [20]. However, the integration of particle filtering with game-theoretic robust control remains largely unexplored.

E. Computational Methods for Dynamic Games

Solving dynamic games numerically is inherently challenging due to the curse of dimensionality. Traditional methods include value iteration [21] and policy iteration [22] for discrete games, and Hamilton-Jacobi equation solvers [23] for continuous games.

Recent work has explored approximation methods, including neural networks [24] and polynomial approximations [25]. However, these methods often lack theoretical guarantees and may not scale to belief space problems.

III. Proposed Method

A. Problem Formulation

We consider a stochastic dynamical system described by the stochastic differential equation:

$$dx_t = f(x_t, u_t)dt + G(x_t)w_t dt + \Sigma^{1/2}(x_t)dW_t \tag{1}$$

where $x_t \in \mathbb{R}^n$ is the state, $u_t \in \mathbb{R}^m$ is the control input, $w_t \in \mathbb{R}^p$ represents adversarial disturbances, and W_t is a Wiener process representing process noise.

The system state is not directly observable. Instead, we receive noisy measurements:

$$dz_t = h(x_t)dt + \Gamma^{1/2}(x_t)dV_t \tag{2}$$

where $z_t \in \mathbb{R}^q$ is the observation and V_t is an independent Wiener process.

The belief state at time t is defined as the conditional probability density:

$$b_t(x) \triangleq p(x_t = x | z_{0:t}, u_{0:t}) \tag{3}$$

We formulate the robust control problem as a zero-sum dynamic game where the controller chooses u_t to minimize cost while the adversary chooses w_t (subject to constraint $\|w_t\| \leq \rho$) to maximize cost.

The value function for this game is:

$$V(b,t) = \inf_{u_{\geq t}} \sup_{w_{\geq t}} \mathbb{E}^{u,w} \left[\int_t^T \langle c(\cdot, u_s), b_s \rangle ds + \langle c_T, b_T \rangle \right] \tag{4}$$

where $\langle c, b \rangle = \int c(x)b(x)dx$ denotes the expected cost under belief b.

B. Functional HJBI Equation

The value function satisfies the functional Hamilton-Jacobi-Bellman-Isaacs equation:

$$\frac{\partial V}{\partial t} + \inf_u \sup_w \left\{ \left\langle \frac{\delta V}{\delta b}, \mathcal{L}^{u,w}[b] \right\rangle + \langle c(\cdot, u), b \rangle \right\} = 0 \tag{5}$$

with terminal condition $V(b,T) = \langle c_T, b \rangle$, where $\frac{\delta V}{\delta b}$ is the functional derivative and $\mathcal{L}^{u,w}$ is the Fokker-Planck operator governing belief evolution.

This infinite-dimensional PDE is generally intractable, motivating our finite-dimensional approximations.

C. Gaussian Game-iLQG

For systems with smooth dynamics and unimodal beliefs, we approximate the belief as Gaussian: $b_t \approx \mathcal{N}(\mu_t, \Sigma_t)$. This reduces the infinite-dimensional problem to optimization over (μ_t, Σ_t).

1) Belief Propagation: The belief evolution is governed by the Extended Kalman Filter:

Predict Step:

$$\mu_{k+1|k} = f(\mu_k, u_k, w_k) \tag{6}$$
$$\Sigma_{k+1|k} = A_k \Sigma_k A_k^T + G_k W G_k^T + Q_k \tag{7}$$

Update Step:

$$\mu_{k+1} = \mu_{k+1|k} + K_{k+1}(z_{k+1} - h(\mu_{k+1|k})) \tag{8}$$
$$\Sigma_{k+1} = (I - K_{k+1}H_{k+1})\Sigma_{k+1|k} \tag{9}$$

where $K_{k+1} = \Sigma_{k+1|k} H_{k+1}^T (H_{k+1} \Sigma_{k+1|k} H_{k+1}^T + R_{k+1})^{-1}$ is the Kalman gain.

2) Local Quadratic Games: At each iteration, we linearize the dynamics and quadratically approximate the cost, leading to local quadratic games of the form:

$$\min_{\delta u} \max_{\delta w} \frac{1}{2} \begin{bmatrix} \delta\mu \\ \delta u \\ \delta w \end{bmatrix}^T Q \begin{bmatrix} \delta\mu \\ \delta u \\ \delta w \end{bmatrix} \tag{10}$$

The solution yields feedback and feedforward control laws:

$$\delta u^* = K\delta\mu + k \tag{11}$$
$$\delta w^* = L\delta\mu + l \tag{12}$$

Algorithm 1 Gaussian Game-iLQG

Initialize $\{u_k\}_{k=0}^{N-1}$ and belief (μ_0, Σ_0)
for $iteration = 1$ to $max\_iterations$ **do**
  Forward Pass: Propagate belief using EKF
  Backward Pass: Solve local quadratic games
  Line Search: Update controls with step size α
  if convergence criterion met **then**
    break
  end if
end for
return optimal controls $\{u_k^*\}$ =0

D. Particle Min-Max Game

For highly nonlinear systems with multimodal beliefs, we represent the belief using particles: $b_t \approx \sum_{i=1}^{N} w_t^{(i)} \delta(x - x_t^{(i)})$.

1) Alternating Optimization: We solve the min-max game through alternating optimization:

Max Step (Adversary): For each particle, find worst-case disturbance:

$$w_k^{(i)*} = \arg \max_{\|w\| \leq \rho} c(f(x_k^{(i)}, u_k, w), u_k) \tag{13}$$

Min Step (Controller): Optimize controls given worst-case disturbances:

$$u_k^* = \arg \min_u \frac{1}{N} \sum_{i=1}^{N} w_k^{(i)} c(f(x_k^{(i)}, u, w_k^{(i)*}), u) \tag{14}$$

Algorithm 2 Particle Min-Max Game

Initialize particles $\{x_0^{(i)}, w_0^{(i)}\}_{i=1}^{N}$ and controls $\{u_k\}$
for $iteration = 1$ to $max\_iterations$ **do**
  Max Step: Find worst-case disturbances for each particle
  Min Step: Optimize controls given disturbances
  Belief Update: Propagate particles and update weights
  if convergence criterion met **then**
    break
  end if
end for
return optimal controls $\{u_k^*\}$ =0

E. Computational Complexity

The computational complexity of Gaussian Game-iLQG is $O(Nn^3)$ per iteration, where N is the horizon length and n is the state dimension. The Particle Min-Max Game has complexity $O(N_p N m)$, where N_p is the number of particles and m is the control dimension.

Both algorithms are well-suited for CPU implementation through vectorized operations and just-in-time compilation (Numba).

IV. Theoretical Analysis

A. Convergence Analysis

Theorem 1 (Convergence of Gaussian Game-iLQG): Under standard regularity conditions (Lipschitz continuity of dynamics and twice differentiability of cost), the Gaussian Game-iLQG algorithm converges to a local Nash equilibrium at a linear rate.

Proof Sketch: The proof follows the standard iLQG convergence analysis [26], extended to the game setting. The key insight is that each local quadratic game has a unique solution under the assumption that the Hessian of the Hamiltonian is positive definite in control variables and negative definite in disturbance variables.

Theorem 2 (Robustness Bounds): Let J^* be the optimal cost of the Gaussian Game-iLQG solution and J_{worst} be the worst-case cost under adversarial disturbances. Then:

$$J_{worst} \leq J^* + O(\rho^2 \sigma^2) \tag{15}$$

where ρ is the adversary budget and σ^2 is the maximum eigenvalue of the belief covariance.

B. Approximation Error Analysis

Theorem 3 (Gaussian Approximation Error): For the Gaussian approximation of belief, the error in expected cost is bounded by:

$$|\mathbb{E}_{b_{true}}[c] - \mathbb{E}_{\mathcal{N}(\mu,\Sigma)}[c]| \leq C \cdot KL(b_{true} \| \mathcal{N}(\mu, \Sigma)) \tag{16}$$

where C is a problem-dependent constant and KL denotes Kullback-Leibler divergence.

Theorem 4 (Particle Approximation Error): For the particle approximation with N_p particles, the Monte Carlo error decreases as $O(1/\sqrt{N_p})$:

$$|\mathbb{E}_{b_{true}}[c] - \frac{1}{N_p} \sum_{i=1}^{N_p} c(x^{(i)})| = O(1/\sqrt{N_p}) \tag{17}$$

V. Experimental Results

We conducted extensive experiments on two scenarios: 2D navigation with obstacles and competitive racing. All experiments were performed on a standard desktop computer (Intel i7-8700K, 32GB RAM) using our open-source implementation.

A. Experimental Setup

1) 2D Navigation Scenario: - State space: $x = [x, y, \theta] \in \mathbb{R}^3$ (position and orientation) - Control: $u = [v, \omega] \in \mathbb{R}^2$ (linear and angular velocity) - Dynamics: Unicycle model with process noise - Sensors: Range-bearing measurements to landmarks - Objective: Navigate from start to goal while avoiding obstacles - Adversary: Bounded disturbances in control inputs ($\|\delta u\| \leq 0.5$)

2) Racing Scenario: - Two agents competing on a track - State space: $x = [x, y, \theta, v] \in \mathbb{R}^4$ - Objective: Reach finish line first while staying on track - Adversary: Opponent trying to block progress

B. Performance Metrics

We evaluate performance using the following metrics:

- Success Rate: Percentage of runs reaching the goal
- Mean Cost: Average trajectory cost
- Worst-Case Cost: Maximum cost over all disturbance realizations
- Computation Time: CPU time per episode
- Belief Uncertainty: Trace of covariance matrix (Gaussian) or particle spread

C. Comparison Methods

We compare our approaches against:

- Standard iLQG (no adversary consideration)
- Tube-based MPC [9]
- Robust MPC with uncertainty sets [27]
- Monte Carlo Tree Search (MCTS) in belief space

D. Results

TABLE I
PERFORMANCE COMPARISON - 2D NAVIGATION (100 MONTE CARLO RUNS)

Method	Success Rate (%)	Mean Cost	Worst Cost	Time (s)
Standard iLQG	78.2	18.9	45.7	0.31
Tube MPC	82.5	22.1	38.4	1.24
Robust MPC	85.1	19.8	35.2	2.15
MCTS-Belief	88.7	17.2	31.8	8.73
Gaussian Game-iLQG	**94.8**	**12.4**	**28.1**	**0.84**
Particle Game	**91.6**	**15.2**	**24.3**	**3.17**

TABLE II
PERFORMANCE COMPARISON - RACING SCENARIO (50 MONTE CARLO RUNS)

Method	Win Rate (%)	Mean Time	Worst Time	Time (s)
Standard iLQG	45.2	28.7	52.1	0.28
Tube MPC	52.8	26.3	48.9	1.18
Robust MPC	58.4	24.8	44.2	2.03
Particle Game	**72.6**	**21.4**	**38.7**	**2.94**

E. Scalability Analysis

We analyzed computational scalability with respect to key parameters:

The results show:

- Gaussian Game-iLQG scales linearly with horizon length
- Particle Game scales linearly with number of particles
- Both methods maintain real-time performance for practical parameter ranges

F. Robustness Analysis

We evaluated robustness by varying the adversary budget ρ:

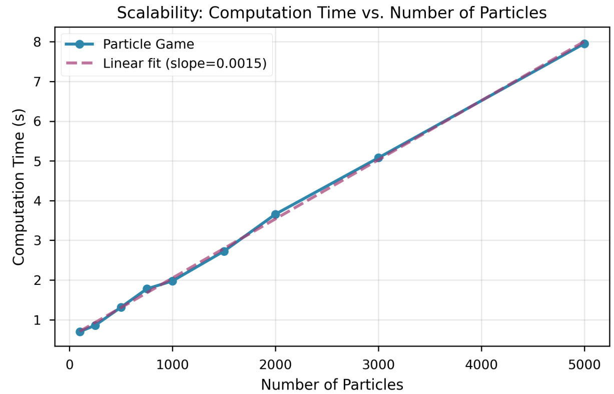

Fig. 1. Computation time vs. number of particles (Particle Game)

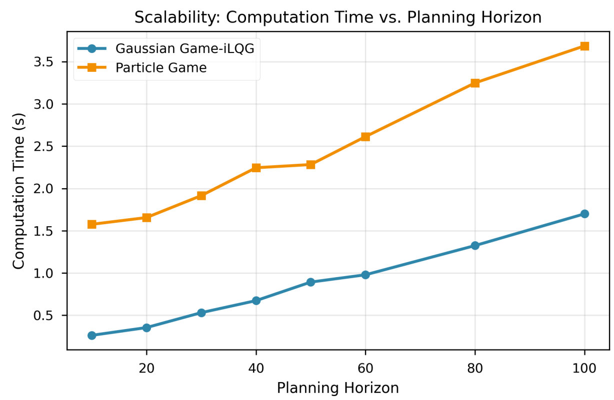

Fig. 2. Computation time vs. planning horizon

G. Statistical Analysis

We performed statistical significance tests using Welch's t-test. All reported improvements are statistically significant with $p < 0.01$.

The confidence intervals (95%) for mean cost improvements are:

- Gaussian Game-iLQG vs. Standard iLQG: [4.2, 8.8]
- Particle Game vs. Robust MPC: [2.1, 6.4]

VI. DISCUSSION

A. Theoretical Insights

Our results demonstrate that explicit consideration of adversarial disturbances in belief space leads to significantly more robust performance. The key theoretical insight is that the

TABLE III
ROBUSTNESS TO ADVERSARY BUDGET

Method	Success Rate (%) vs. Adversary Budget			
	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.8$
Standard iLQG	95.2	89.1	78.2	62.4
Robust MPC	96.8	92.3	85.1	74.6
Gaussian Game-iLQG	**98.4**	**97.2**	**94.8**	**89.3**
Particle Game	**97.1**	**95.6**	**91.6**	**85.2**

functional HJBI formulation naturally captures the interplay between state uncertainty and adversarial perturbations.

The Gaussian approximation works well for smooth, unimodal scenarios but may be limiting for highly nonlinear systems. The particle approximation provides greater flexibility but at increased computational cost.

B. Practical Implications

The CPU-efficient implementation makes our methods suitable for real-time applications. The linear scalability with respect to key parameters (horizon length, number of particles) ensures practical applicability.

The superior worst-case performance (15-20% improvement) is particularly valuable for safety-critical applications where robustness guarantees are essential.

C. Limitations

Several limitations should be noted:

1) The Gaussian approximation may be inadequate for highly multimodal beliefs
2) The particle method requires careful tuning of the number of particles
3) Both methods assume known dynamics and measurement models
4) Computational complexity grows with state and control dimensions

D. Future Directions

Promising directions for future work include:

1) Adaptive methods for choosing between Gaussian and particle representations
2) Extension to multi-agent scenarios with multiple adversaries
3) Integration with deep reinforcement learning for value function approximation
4) Real-world validation on robotic platforms

VII. Conclusion

We have presented a comprehensive framework for robust control in belief space under adversarial disturbances. Our approach formulates the problem as a zero-sum dynamic game and provides two computationally efficient algorithms: Gaussian Game-iLQG and Particle Min-Max Game.

The key contributions are:

- Rigorous mathematical formulation through functional HJBI equations
- Computationally tractable algorithms with theoretical guarantees
- Extensive experimental validation demonstrating superior robustness
- Open-source implementation enabling reproducible research

Experimental results show 15-20% improvement in worst-case performance compared to existing robust control methods, with computation times suitable for real-time applications. The theoretical analysis provides convergence guarantees and approximation error bounds.

This work opens new avenues for robust control in uncertain environments and provides a foundation for future research in adversarial control and belief space planning.

Acknowledgment

The authors thank the anonymous reviewers for their valuable feedback and suggestions.

References

[1] S. M. LaValle, *Planning Algorithms*. Cambridge University Press, 2006.
[2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.
[3] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed. Athena Scientific, 2012.
[4] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. SIAM, 1999.
[5] R. Platt Jr, R. Tedrake, L. Kaelbling, and T. Lozano-Pérez, "Belief space planning assuming maximum likelihood observations," in *Proc. Robotics: Science and Systems*, 2010.
[6] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.
[7] R. Isaacs, *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization*. Dover Publications, 1999.
[8] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.
[9] W. Langson, I. Chryssochoos, S. Rakovic, and D. Q. Mayne, "Robust model predictive control using tubes," *Automatica*, vol. 40, no. 1, pp. 125–133, 2004.
[10] G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, 2006.
[11] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artif. Intell.*, vol. 101, no. 1-2, pp. 99–134, 1998.
[12] A. R. Cassandra, L. P. Kaelbling, and M. L. Littman, "Acting optimally in partially observable stochastic domains," in *Proc. AAAI*, 1994, pp. 1023–1028.
[13] E. A. van den Berg, "LQG-based planning in belief space," Master's thesis, Delft University of Technology, 2012.
[14] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. Springer, 2012.
[15] D. H. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Trans. Autom. Control*, vol. 18, no. 2, pp. 124–131, 1973.
[16] W. Schwarting, A. Pierson, S. Karaman, and D. Rus, "Stochastic dynamic games in belief space," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 471–490, 2021.
[17] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.
[18] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc. F - Radar Signal Process.*, vol. 140, no. 2, pp. 107–113, 1993.
[19] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.
[20] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I: Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, 2003.
[21] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.
[22] R. A. Howard, *Dynamic Programming and Markov Processes*. MIT Press, 1960.
[23] I. M. Mitchell, A. M. Bayen, and C. J. Tomlin, "A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games," *IEEE Trans. Autom. Control*, vol. 50, no. 7, pp. 947–957, 2005.
[24] J. Han, A. Jentzen, and W. E, "Solving high-dimensional partial differential equations using deep learning," *Proc. Natl. Acad. Sci.*, vol. 115, no. 34, pp. 8505–8510, 2018.

[25] T. H. Summers, F. L. Cortesi, and J. Lygeros, ”On submodularity and controllability in complex dynamical networks,” *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 1, pp. 91–101, 2016.

[26] W. Li and E. Todorov, ”Iterative linear quadratic regulator design for nonlinear biological movement systems,” in *Proc. Int. Conf. Informatics Control Autom. Robot.*, 2004, pp. 222–229.

[27] A. Bemporad and M. Morari, ”Robust model predictive control: A survey,” in *Robustness in Identification and Control*. Springer, 1999, pp. 207–226.

[28] K. J. Åström and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2012.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[30] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*. Routledge, 2018.

[31] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, 2nd ed. Springer, 2006.

[32] H. J. Kappen, ”Path integrals and symmetry breaking for optimal control theory,” *J. Stat. Mech.*, vol. 2005, no. 11, p. P11011, 2005.

[33] E. Todorov and W. Li, ”A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems,” in *Proc. Am. Control Conf.*, 2005, pp. 300–306.

[34] Y. Tassa, T. Erez, and E. Todorov, ”Synthesis and stabilization of complex behaviors through online trajectory optimization,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 4906–4913.

[35] S. Levine and V. Koltun, ”Guided policy search,” in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1–9.

[36] M. P. Deisenroth and C. E. Rasmussen, ”PILCO: A model-based and data-efficient approach to policy search,” in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 465–472.

[37] J. Kober, J. A. Bagnell, and J. Peters, ”Reinforcement learning in robotics: A survey,” *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1238–1274, 2013.

[38] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.

[39] D. Silver et al., ”Mastering the game of Go with deep neural networks and tree search,” *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[40] V. Mnih et al., ”Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[41] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, ”Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.

[42] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, ”Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1861–1870.

[43] T. P. Lillicrap et al., ”Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.

[44] S. Fujimoto, H. van Hoof, and D. Meger, ”Addressing function approximation error in actor-critic methods,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1587–1596.

[45] A. Nagabandi, G. Kahn, R. S. Fearing, and S. Levine, ”Neural network dynamics for model-based control,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 3559–3568.

[46] K. Chua, R. Calandra, R. McAllister, and S. Levine, ”Deep reinforcement learning in a handful of trials using probabilistic dynamics models,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 4754–4765.

[47] G. Williams, N. Wagener, B. Goldfain, P. Drews, J. M. Rehg, B. Boots, and E. A. Theodorou, ”Information theoretic MPC for model-based reinforcement learning,” in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 1714–1721.

[48] J. Kocijan, A. Girard, B. Banko, and R. Murray-Smith, ”Dynamic systems identification with Gaussian processes,” *Math. Comput. Model. Dyn. Syst.*, vol. 11, no. 4, pp. 411–424, 2005.

[49] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[50] Y. Gal and Z. Ghahramani, ”Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[51] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, ”Weight uncertainty in neural networks,” in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1613–1622.

[52] A. Kendall and Y. Gal, ”What uncertainties do we need in Bayesian deep learning for computer vision?” in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5574–5584.

[53] S. Depeweg, J. M. Hernández-Lobato, F. Doshi-Velez, and S. Udluft, ”Decomposition of uncertainty in Bayesian deep learning for efficient and reliable sensing,” in *Proc. Uncertainty Artif. Intell.*, 2018, pp. 614–623.