Information-Theoretic Belief-Space Planning for Gaussian Systems with Nonlinear Observations

Haruki Nishimura

***Abstract*— We propose a novel belief-space planning algorithm based on model predictive path integral control (MPPI). We show that the existing MPPI framework can be applied to Gaussian belief systems with nonlinear observations if the underlying system dynamics is linear. We test our algorithm in simulation on a 2D robot navigation task. The proposed method empirically improves the performance more than three times over a state-of-the-art approximate dynamic programming approach.**

I. Introduction

Recent development of efficient batteries and devices for onboard computation has extended capabilities of robotic systems to operate in the wild. However, the robots still have difficulty in planning appropriate actions under uncertainty, which is crucial for robustness in accomplishing practical missions.

Planning under uncertainty is more challenging than its deterministic equivalent, and is inevitable since neither perception nor actuation of a robot can be fully deterministic in practice. The sources of uncertainty include unmodeled dynamics, stochastic disturbance, and imperfect sensing.

A principled approach to tackle this problem is planning in belief space. A belief space is a space of parameters that defines a probability distribution over the state space. The planner then chooses sequential control inputs based on the evolving belief states, with which the robot can appropriately take actions despite stochasticity and partial observability.

In this work we propose a novel information-theoretic control algorithm for belief-space planning. Unlike other approaches, our method is based on the model predictive path integral control (MPPI) [1]. The algorithm synthesizes control signals online by simulating belief state transitions under uncontrolled dynamics.

A. Related Work

Belief-space planning is known to be challenging for a few reasons. First, the belief state is continuous and can be high-dimensional even if the state space is small or discrete. Second, the dynamics that governs the belief state transitions is stochastic due to unknown future observations. In discrete time, some belief-space planning problems can be formulated as partially observable Markov decision processes (POMDPs). However POMDPs are shown to be PSPACE-complete even for finite horizon problems, and is uncomputable in infinite horizon cases [2]. Furthermore, the standard POMDP framework requires the reward function to be explicitly dependent on the state variable, while in belief-space planning the cost can be a function of the belief state.

For these reasons, tractable solution methods have made approximations to the original problem. One approach is to ignore stochasticity in the future observations by assuming that maximum likelihood observations will always occur [3], whereby the original stochastic problem is converted to a deterministic one. This approach has been popular since existing indirect or direct optimal control algorithms can be used, but the solutions are suboptimal since stochasticity is ignored. Moreover, methods based on nonlinear optimization techniques such as [4] yield locally optimal solutions even for the approximated problem.

Another class of techniques relies on solving approximate dynamic programming online. In [5], optimal manipulation of an object with unknown mass properties is addressed by formulating a Bayes-adaptive Markov decision process (BAMDP), which is then solved using a variant of Monte Carlo tree search (MCTS) [6] in the belief space. Approximate dynamic programming is advantageous as it takes into account the stochastic belief dynamics, but is computationally expensive since both the belief space and the action space need to be sufficiently explored.

In continuous-time stochastic optimal control theory, the path integral representation of the value function [7] is known to equal the information-theoretic free energy for control and noise affine dynamical systems under certain regularity conditions [8]. Williams et al. [1] exploit this property and present the model predictive path integral control (MPPI) algorithm. Instead of performing dynamic programming, the MPPI controller optimizes the sequence of control inputs by minimizing the KL-divergence between the induced distribution over the state trajectories and the optimal distribution as defined by the tight lower-bound of an information theoretic inequality. The algorithm is favorable as it does not require the sampling of different action sequences.

B. Contributions

In this paper we show that MPPI can be used for controlling nonlinear Gaussian belief systems. Specifically, we consider the systems with extended Kalman filter (EKF) and linear state transitions. The simple LQG controller cannot be used in this case due to nonlinearity of the measurement function.

This paper is submitted as a final project report for AA 203 Introduction to Optimal Control and Dynamic Optimization, class of 2017-18 Spring Quarter with Professor Pavone at Stanford University.

The author is with the Department of Aeronautics and Astronautics at Stanford University, Stanford, CA 94305, USA. Email: hnishimura@stanford.edu.

Although MPPI has been used for controlling stochastic systems including aggressive driving of an RC car [9], there are very few existing works that have applied MPPI to the belief-space planning. Recently Pan et al. [10] have proposed a belief-space control algorithm using MPPI under model uncertainty, in which the unknown parameters of the dynamics are learned online via sparse spectrum Gaussian processes. This is different from our work since there is no partial observability of the states or stochasticity in the belief transitions themselves in their problem, whereas we take into account both of them. To our knowledge this is the first paper that has applied MPPI to a belief-space planning problem of this kind.

The rest of the paper is organized as follows. In Section II we give conditions to the belief dynamics under which MPPI can be applied. In Section III we review MPPI and adapt it to the belief-space planning. Simulation results are given in Section IV, followed by conclusions and future work in Section V.

II. Extended Kalman Filter with Linear State Transitions

A. System Specifications

Suppose we have a linear discrete-time transition model for the unobserved state $x_t \in \mathbb{R}^n$ as given by

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

where $A_t \in \mathbb{R}^{n\times n}$ and $B_t \in \mathbb{R}^{n\times m}$ are possibly time-varying matrices, and $u_t \in U \subseteq \mathbb{R}^m$ is a control input. The system is corrupted by additive Gaussian white noise $v_t \sim \mathcal{N}(0, Q_t)$.

As opposed to the state transition model, the observation model can be nonlinearly dependent on the state as

$$y_t = g(x_t) + w_t, \tag{2}$$

where $y_t \in \mathbb{R}^l$ is the observation at time t and g is a differentiable function of x. Similarly to the state transition, the observation has an additive white Gaussian noise term $w_t \sim \mathcal{N}(0, R_t)$.

B. Belief Dynamics

In the belief-space planning, the probability distribution over the unobserved state x_t is represented by the belief state b_t. For Gaussian systems, the belief state is uniquely defined by the Cartesian product of the mean vector $\mu \in \mathbb{R}^n$ and the covariance matrix $\Sigma \in \mathbb{S}_+^n$ as

$$b_t = (\mu_t, \Sigma_t), \tag{3}$$

where $\mathbb{S}_+^n$ is the set of all symmetric positive definite n-by-n matrices.

The transition of the probability distribution follows the Bayes rule

$$p(x_{t+1}; b_{t+1}) \propto p(y_{t+1} \mid x_{t+1}) \int_x p(x_{t+1} \mid x, u_t) p(x; b_t) dx, \tag{4}$$

which induces the corresponding belief state transition model. Unfortunately the exact Bayesian update is intractable for systems with nonlinear observations, but we can locally linearize the observation model and approximate the posterior with a Gaussian distribution. The resulting EKF equations are given by

$$\mu_{t+1} = A_t \left(\mu_t + \Sigma_t C_t^{\mathrm{T}} H_t^{-1} (y_t - g(\mu_t))\right) + B_t u_t \tag{5}$$

$$\Sigma_{t+1} = A_t \left(\Sigma_t - \Sigma_t C_t^{\mathrm{T}} H_t^{-1} C_t \Sigma_t\right) A_t^{\mathrm{T}} + Q_t, \tag{6}$$

where $C_t = \frac{\partial}{\partial x} g(x)|_{\mu_t}$ and $H_t = C_t \Sigma_t C_t^{\mathrm{T}} + R_t$. Here we have used the EKF equations in the "update-then-predict" scheme.

In doing so, notice that the covariance transition in (6) becomes deterministic. Indeed, all the terms in (6) is either a constant or a deterministic function of $b_t = (\mu_t, \Sigma_t)$. On the other hand, the mean transition in (5) involves the stochastic observation term y_t. Thanks to the Gaussian approximation in EKF we can characterize the distribution over this observation as

$$p(y_t \mid \mu_t, \Sigma_t) = \mathcal{N}(g(\mu_t), H_t). \tag{7}$$

Therefore, the difference $y_t - g(\mu_t)$ is also approximated to be Gaussian with mean 0 and covariance H_t. Rewriting (5) yields

$$\mu_{t+1} = A_t \mu_t + B_t u_t + \eta_t, \tag{8}$$

where $\eta_t \sim \mathcal{N}(0, A_t \Sigma_t C_t^{\mathrm{T}} H_t^{-1} C_t \Sigma_t A_t^{\mathrm{T}})$. Let $S_t(\mu_t, \Sigma_t)$ denote this covariance matrix for η_t. S_t is a valid covariance matrix since it is symmetric and positive semidefinite.

In summary, we have shown that the belief dynamics can be compactly represented by

$$\mu_{t+1} = A_t \mu_t + B_t u_t + \eta_t \quad \eta_t \sim \mathcal{N}\left(0, S_t(\mu_t, \Sigma_t)\right) \tag{9}$$

$$\Sigma_{t+1} = h_t(\mu_t, \Sigma_t), \tag{10}$$

where h_t is the deterministic nonlinear function defined in (6). We have effectively separated the stochasticity in the belief dynamics and all the uncertainty in the transition is now in the mean vector only. Furthermore, the mean dynamics is affine in control and noise. This is not always the case since in general the nonlinear dynamics would make both μ and Σ evolve stochastically. In the next section we will see how this special property allows us to use the MPPI algorithm in the belief space.

III. Model Predictive Path Integral Control in Belief Space

A. Cost Model

Consider the following belief-state-dependent cost

$$J(b_{0:T}) = \phi(b_T) + \sum_{t=0}^{T-1} c(b_t), \tag{11}$$

with terminal cost ϕ and stage cost c. Notice that the control cost is not involved. Additive quadratic control cost terms will appear when we obtain the free energy equation below.

B. Information-Theoretic Inequality

Suppose that the covariance matrix $S_t(\mu_t, \Sigma_t)$ for the mean transition is full rank. Then it is positive definite and the conditional belief transition density is defined as

$$q_{u_t}(b_{t+1} \mid b_t) = Z \exp\Bigg(-\frac{1}{2} \left(\mu_{t+1} - (A_t\mu_t + B_t u_t)\right)^{\mathrm{T}} \times S_t(\mu_t, \Sigma_t)^{-1} \left(\mu_{t+1} - (A_t\mu_t + B_t u_t)\right) \Bigg), \tag{12}$$

where Z is a normalization constant. If S_t is not full rank, then the distribution is degenerate and the density function is ill-defined. This occurs when the dimension of the observation vector is lower than that of the state vector, for example. We can still define a density function by restricting the Lebesgue measure to a lower dimensional subspace, but the further analysis of this case is not considered in this paper and left for future work.

We are interested in the density ratio between the belief trajectory distributions induced by the controlled and the uncontrolled dynamics. Let $\mathbb{Q}_{u_{0:T-1}}$ be the controlled belief trajectory distribution and $\mathbb{P}$ be the uncontrolled one. The corresponding density functions are

$$q_{u_{0:T-1}}(b_{0:T}) = \prod_{t=0}^{T-1} q_{u_t}(b_{t+1} \mid b_t) p(b_0) \tag{13}$$

$$p(b_{0:T}) = \prod_{t=0}^{T-1} p(b_{t+1} \mid b_t) p(b_0), \tag{14}$$

respectively. $p(b_0)$ represents the prior distribution over b_0 and we used the Markov property for decoupling the joint distributions. The uncontrolled belief transition density $p(b_{t+1} \mid b_t)$ has the same form as (12), except that u_t is 0.

With the two distributions defined, we obtain the following information theoretic inequality:

$$-\lambda \log\left(\mathbb{E}_{\mathbb{P}}\left[\exp\left(-\frac{1}{\lambda} J(b_{0:T}) \right) \right] \right) \leq \mathbb{E}_{\mathbb{Q}_{u_{0:T-1}}}\left[J(b_{0:T}) + \frac{\lambda}{2} \sum_{t=0}^{T-1} u_t^{\mathrm{T}} S_t(\mu_t, \Sigma_t) u_t \right]. \tag{15}$$

The derivation is almost identical to the previous work [11] and is omitted for brevity. The key idea is to switch the distribution from $\mathbb{P}$ to $\mathbb{Q}$ in the left hand side of (15) and apply the Jensen's inequality. This operation is known as the Legendre transformation [12]. The left hand side of (15) is called the free energy. Notice that on the right hand side we have a standard cost of optimal control with the additive quadratic control cost, in which the coefficient matrix is given by S_t. $\lambda > 0$ is a user-defined parameter that determines the weight on the control effort.

In summary, we have obtained the lower bound on the cost of an optimal control problem. Furthermore, from the Jensen's inequality we know that the lower bound is tight for the density q^* given by

$$q^*(b_{0:T}) \propto \exp\left(-\frac{1}{\lambda} J(b_{0:T}) \right) p(b_{0:T}). \tag{16}$$

C. Control Optimization

In [9], [11] the authors have proposed a KL-divergence minimization approach to compute the control that achieves a controlled distribution as close as possible to $q^*(b_{0:T})$. Following this approach, we have the optimization problem

$$\begin{aligned} \underset{u_{0:T-1}}{\text{minimize}} \quad & \mathbb{D}_{\mathrm{KL}}\left(\mathbb{Q}^* \,||\, \mathbb{Q}_{u_{0:T-1}} \right) \\ \text{subject to} \quad & u_t \in U, \quad t = 0, \ldots, T-1. \end{aligned} \tag{17}$$

Using the definition of KL-divergence one can show that this optimization problem is equivalent to

$$\begin{aligned} \underset{u_{0:T-1}}{\text{maximize}} \quad & \int_{\Omega_b} q^*(b_{0:T}) \log\left(\frac{q_{u_{0:T-1}}(b_{0:T})}{p(b_{0:T})} \right) db_{0:T} \\ \text{subject to} \quad & u_t \in U, \quad t = 0, \ldots, T-1, \end{aligned} \tag{18}$$

where Ω_b is the space of all possible belief trajectories. We can further simplify the objective by substituting (13) and (14) into (18) and using (12). Then the $\log(\cdot)$ function becomes

$$\log\left(\frac{q_{u_{0:T-1}}(b_{0:T})}{p(b_{0:T})} \right) = \sum_{t=0}^{T-1} \Bigg\{ -\frac{1}{2} u_t^{\mathrm{T}} B_t^{\mathrm{T}} S_t(\mu_t, \Sigma_t)^{-1} B_t u_t + u_t^{\mathrm{T}} B_t^{\mathrm{T}} S_t(\mu_t, \Sigma_t)^{-1} (\mu_{t+1} - A_t\mu_t) \Bigg\}. \tag{19}$$

Therefore, the joint optimization with respect to the control trajectory $u_{0:T-1}$ can be decoupled into the optimization of each control signal u_t as

$$\begin{aligned} \underset{u_t}{\text{maximize}} \quad & -\frac{1}{2} u_t^{\mathrm{T}} B_t^{\mathrm{T}} X_t B_t u_t + u_t^{\mathrm{T}} B_t^{\mathrm{T}} z_t \\ \text{subject to} \quad & u_t \in U, \end{aligned} \tag{20}$$

where

$$X_t = \int_{\Omega_b} S_t(\mu_t, \Sigma_t)^{-1} q^*(b_{0:T}) db_{0:T} \tag{21}$$

$$z_t = \int_{\Omega_b} S_t(\mu_t, \Sigma_t)^{-1} (\mu_{t+1} - A_t\mu_t) q^*(b_{0:T}) db_{0:T}. \tag{22}$$

Note that X_t is positive definite since it is an integral of nonnegatively-weighted positive definite matrices. Thus, $B_t^{\mathrm{T}} X_t B_t$ is positive semidefinite and (20) becomes a convex optimization problem if the admissible control set U is given by a convex inequality constraint. In this case the solution can be obtained quite efficiently using an existing convex optimization solver.

Algorithm 1 MPPI for EKF Dynamics

INPUT: Current belief $b_0 = (\mu_0, \Sigma_0)$
OUTPUT: Control signal u_0

for i = 1:N **do**
  Sample uncontrolled belief trajectory $b^i_{0:T}$ ▷ (9),(10)
  Compute sample weight $w^i(b^i_{0:T})$ ▷ (23),(11)
end for
Compute normalized weights $\{\bar{w}^1, \dots, \bar{w}^N\}$
$X_0 \approx \sum_{i=1}^N \bar{w}^i S_0(\mu_0^i, \Sigma_0^i)^{-1}$
$z_0 \approx \sum_{i=1}^N \bar{w}^i S_0(\mu_0^i, \Sigma_0^i)^{-1}(\mu_1^i - A_0\mu_0^i)$
Solve (20) for u_0.
return u_0

D. Importance Sampling

We are left with computing the coefficient matrix $X_t = \mathbb{E}_{\mathbb{Q}^*}[S_t^{-1}]$ and the vector $z_t = \mathbb{E}_{\mathbb{Q}^*}[S_t^{-1}(\mu_{t+1} - A_t\mu_t)]$. In [11] an iterative importance sampling scheme is employed, where in each iteration the proposal distribution is the controlled distribution $\mathbb{Q}$ induced by the previously computed control sequence $u^i_{0:T-1}$, and the importance sampling weights are re-computed to find the updated sequence $u^{i+1}_{0:T-1}$. In our problem, however, this iterative update would require computing the time varying matrix $S_t = A_t\Sigma_t C_t^{\mathrm{T}} H_t^{-1} C_t \Sigma_t A_t^{\mathrm{T}}$ and inverting it many times for every time step, which can be computationally expensive for high-dimensional systems. Therefore we chose to use the uncontrolled distribution $\mathbb{P}$ as the proposal distribution. The importance sampling weight $w(b_{0:T})$ is given by

$$w(b_{0:T}) = \frac{q^*(b_{0:T})}{p(b_{0:T})} = \frac{1}{\eta}\exp\left(-\frac{1}{\lambda}J(b_{0:T})\right), \qquad (23)$$

where η is a normalization constant. In practice we do not need to compute η since the weights are normalized after the samples are drawn. The resulting sampling equations are

$$X_t = \mathbb{E}_{\mathbb{P}}[S_t(\mu_t, \Sigma_t)^{-1} w(b_{0:T})] \qquad (24)$$
$$z_t = \mathbb{E}_{\mathbb{P}}[S_t(\mu_t, \Sigma_t)^{-1}(\mu_{t+1} - A_t\mu_t) w(b_{0:T})]. \qquad (25)$$

The entire MPPI algorithm is summarized in Algorithm 1. The implementation is simple and highly parallelizable. The controller is applied in a receding horizon fashion to optimize the next control input over a finite horizon at each planning step.

IV. Simulation Results

In this section we apply the MPPI algorithm to a problem of 2D robot navigation under uncertainty. The task is to control a robot under stochastic disturbance so it successfully approaches a desired goal location. The robot employs noisy range observations of 7 waypoints for localization. The positions of the waypoints are known to the robot. A simple single integrator dynamics with $A_t = I_{2\times 2}$ and $B_t = I_{2\times 2}$ is assumed. The simulation environment is depicted in Figure 1.

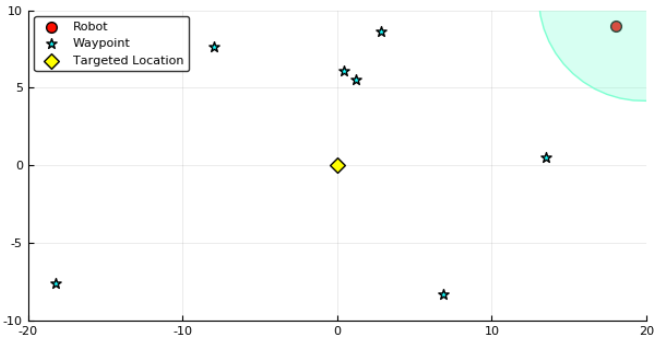

Fig. 1. Initial configuration of the 2D robot navigation problem. The robot depicted in red has to approach the yellow goal location using noisy range observations of the known waypoints only. The dynamics of the robot is subject to stochastic disturbance. The shaded blue area represents the 99% confidence ellipse.

To encode the desired behavior the following cost function

$$J(b_{0:T}) = \sum_{t=0}^{T} \frac{1}{2}\mathrm{tr}(C\Sigma_t) + \frac{1}{2}\mu_t^{\mathrm{T}} C \mu_t \qquad (26)$$

with $C = 10 \times I_{2\times 2}$ was used. The stage cost function $c(b_t) = \frac{1}{2}\mathrm{tr}(C\Sigma_t) + \frac{1}{2}\mu_t^{\mathrm{T}} C\mu_t$ corresponds to $\mathbb{E}_{b_t}[\frac{1}{2}x_t^{\mathrm{T}} C x_t]$ under the Gaussian belief assumption.

The MPPI algorithm was used to synthesize the control signals online with the horizon length of $T = 10$. The number of total steps in one simulation episode was 200. For the control constraint $u_t \in U$ we used the box constraint

$$(-0.1, -0.1)^{\mathrm{T}} \preceq u_t \preceq (0.1, 0.1)^{\mathrm{T}}. \qquad (27)$$

Finally the resulting convex program (20) was solved using the Convex.jl package in Julia [13].

We compared the performance of our approach against a state-of-the-art approximate dynamic programming algorithm called continuous upper confidence trees [14]. This algorithm is a variant of MCTS that uses double progressive widening (DPW) for gradually constructing the search tree in continuous state and action spaces. An advantage of this version of MCTS is that it can directly deal with continuous parameteric belief states as opposed to the standard MCTS method that requires discrete state space representations.

The comparison of the total cost can be found in Figure 2, where for each horizon length T we executed the two algorithms 10 times from the same initial configuration with different random seeds. The MCTS used the negative stage cost as the reward function. As can be seen, the proposed MPPI controller achieves significant performance improvement over MCTS. The results also suggest that MPPI is more sample-efficient than MCTS since the variance is small and the performance is not affected by the varying number of Monte Carlo samples.

V. Conclusions

In this paper we have proposed a novel belief-space planning algorithm for Gaussian belief dynamics with nonlinear observation models. Our method is based on the model predictive path integral control, which has been rarely used

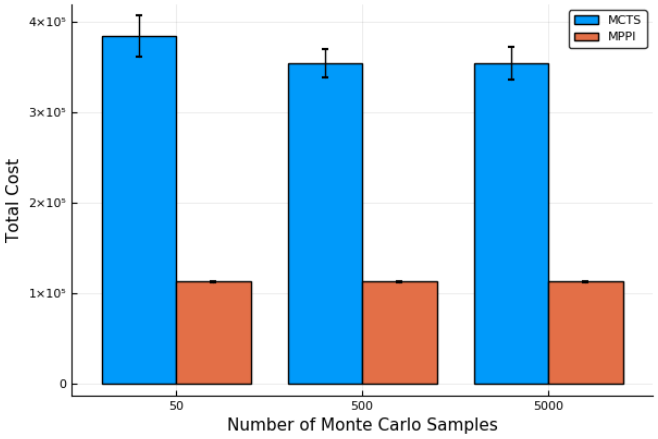

Fig. 2. Total cost of control with MPPI and MCTS algorithms. For each number of samples the results were averaged over 10 different simulation episodes. Overall proposed MPPI controller achieves cost values that are more than three times lower than MCTS, which is a significant performance improvement.

in belief-space planning. Starting from the EKF dynamics we have derived a receding horizon control law that is based on importance sampling under uncontrolled belief dynamics and convex optimization. In simulation we have confirmed that the proposed algorithm significantly outperforms a state-of-the-art approximate dynamic programming approach. In future work, we are interested in extending this method to the degenerate Gaussian density cases as well as applying to more complex belief-space planning problems of various kinds.

REFERENCES

[1] G. Williams, A. Aldrich, and E. A. Theodorou, "Model predictive path integral control: From theory to parallel computation," *Journal of Guidance Control Dynamics*, vol. 40, pp. 344–357, feb 2017.

[2] M. J. Kochenderfer, C. Amato, G. Chowdhary, J. P. How, H. J. D. Reynolds, J. R. Thornton, P. A. Torres-Carrasquillo, N. K. Üre, and J. Vian, *Decision Making Under Uncertainty: Theory and Application*, 1st ed. The MIT Press, 2015.

[3] R. Platt, R. Tedrake, L. Kaelbling, and T. Lozano-Perez, "Belief space planning assuming maximum likelihood observations," in *Robotics Science and Systems Conference (RSS)*, 2010.

[4] S. Patil, G. Kahn, M. Laskey, J. Schulman, K. Goldberg, and P. Abbeel, "Scaling up gaussian belief space planning through covariance-free trajectory optimization and automatic differentiation," in *WAFR*, ser. Springer Tracts in Advanced Robotics, vol. 107. Springer, 2014, pp. 515–533.

[5] P. Slade, P. Culbertson, Z. Sunberg, and M. Kochenderfer, "Simultaneous active parameter estimation and control using sampling-based bayesian reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, sep 2017, pp. 804–810.

[6] C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton, "A survey of monte carlo tree search methods," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 1–43, mar 2012.

[7] H. J. Kappen, "An introduction to stochastic control theory, path integrals and reinforcement learning," vol. 887, no. 149, 2007.

[8] E. A. Theodorou and E. Todorov, "Relative entropy and free energy dualities: Connections to path integral and kl control," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, dec 2012, pp. 1466–1473.

[9] G. Williams, P. Drews, B. Goldfain, J. M. Rehg, and E. A. Theodorou, "Aggressive driving with model predictive path integral control," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, may 2016, pp. 1433–1440.

[10] Y. Pan, K. Saigol, and E. A. Theodorou, "Belief space stochastic control under unknown dynamics," in *2017 American Control Conference (ACC)*, may 2017, pp. 3764–3770.

[11] 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 *2017 IEEE International Conference on Robotics and Automation (ICRA)*, may 2017, pp. 1714–1721.

[12] E. A. Theodorou, "Nonlinear stochastic control and information theoretic dualities: Connections, interdependencies and thermodynamic interpretations," *Entropy*, vol. 17, no. 5, pp. 3352–3375, 2015.

[13] M. Udell, K. Mohan, D. Zeng, J. Hong, S. Diamond, and S. Boyd, "Convex optimization in Julia," *SC14 Workshop on High Performance Technical Computing in Dynamic Languages*, 2014.

[14] A. Couëtoux, J.-B. Hoock, N. Sokolovska, O. Teytaud, and N. Bonnard, "Continuous upper confidence trees," in *Learning and Intelligent Optimization*, C. A. C. Coello, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 433–445.