CDC 2024

Discrete-Time Stochastic LQR via Path Integral Control and Its Sample Complexity Analysis

IEEE L-CSS 2024

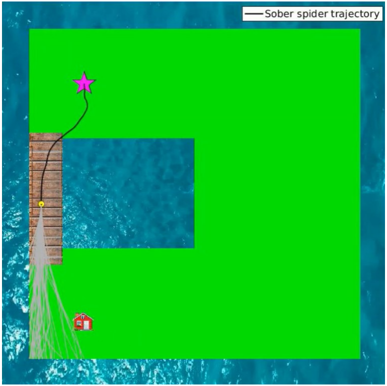
Apurva Patil (UT Austin)
Grani A. Hanasusanto (UIUC)
Takashi Tanaka (UT Austin)

What is Path Integral Control?

- A control algorithm inspired by the path integral formulation of quantum mechanics.
- It solves stochastic optimal control problems via real-time Monte Carlo simulations of open-loop systems.

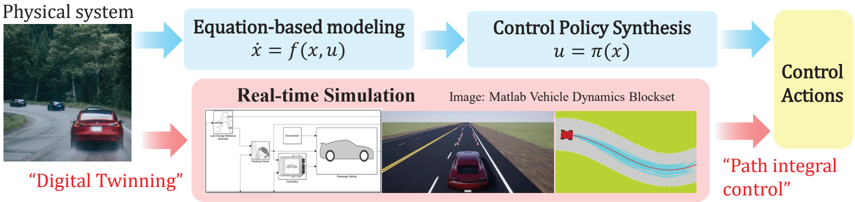
What is Path Integral Control?

- A control algorithm inspired by the path integral formulation of quantum mechanics.
- It solves stochastic optimal control problems via real-time Monte Carlo simulations of open-loop systems.



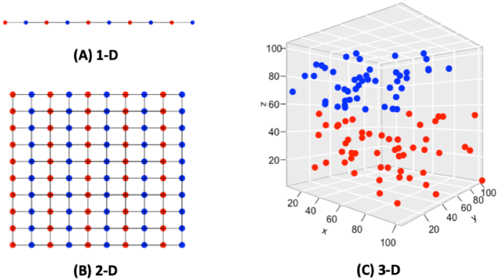
Why Path Integral Control?

- Applicable to nonlinear, stochastic optimal control problems.
- Simulator-driven: no analytical model required.



Why Path Integral Control?

- Less susceptible to the curse of dimensionality
- Monte Carlo simulations can be parallelized on GPUs which makes it effective for real-time control applications.



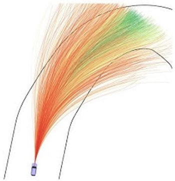


Figure: Path-integral approach

Figure: Grid-based approaches

Outline

Outline

Research Motivation and Prior Work

- The outcome of Monte Carlo simulation is probabilistic and suboptimal when the sample size is finite; hence applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.

[1] Yoon, Hyung-Jin, et al., "Sampling complexity of path integral methods for trajectory optimization," 2022 American Control Conference (ACC).

Research Motivation and Prior Work

- The outcome of Monte Carlo simulation is probabilistic and suboptimal when the sample size is finite; hence applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- To our knowledge not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][1].

[1] Yoon, Hyung-Jin, et al., "Sampling complexity of path integral methods for trajectory optimization," 2022 American Control Conference (ACC).

Research Motivation and Prior Work

- The outcome of Monte Carlo simulation is probabilistic and suboptimal when the sample size is finite; hence applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- To our knowledge not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][1].
- Contributions of [Yoon 2022]: The authors considered the continuous-time path integral control, and applied Chebyshev and Hoeffding inequalities to relate the instantaneous (pointwise-in-time) error bound in control input and the sample size of the Monte-Carlo simulations performed at that particular time instance.

[1] Yoon, Hyung-Jin, et al., "Sampling complexity of path integral methods for trajectory optimization," 2022 American Control Conference (ACC).

Research Motivation and Prior Work

- The outcome of Monte Carlo simulation is probabilistic and suboptimal when the sample size is finite; hence applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- To our knowledge not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][1].
- Contributions of [Yoon 2022]: The authors considered the continuous-time path integral control, and applied Chebyshev and Hoeffding inequalities to relate the instantaneous (pointwise-in-time) error bound in control input and the sample size of the Monte-Carlo simulations performed at that particular time instance.
- Limitations of [Yoon 2022]:

[1] Yoon, Hyung-Jin, et al., "Sampling complexity of path integral methods for trajectory optimization," 2022 American Control Conference (ACC).

Research Motivation and Prior Work

- The outcome of Monte Carlo simulation is probabilistic and suboptimal when the sample size is finite; hence applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- To our knowledge not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][1].
- Contributions of [Yoon 2022]: The authors considered the continuous-time path integral control, and applied Chebyshev and Hoeffding inequalities to relate the instantaneous (pointwise-in-time) error bound in control input and the sample size of the Monte-Carlo simulations performed at that particular time instance.
- Limitations of [Yoon 2022]:
  - The effect of time discretization is not addressed.

[1] Yoon, Hyung-Jin, et al., "Sampling complexity of path integral methods for trajectory optimization," 2022 American Control Conference (ACC).

Research Motivation and Prior Work

- The outcome of Monte Carlo simulation is probabilistic and suboptimal when the sample size is finite; hence applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- To our knowledge not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][1].
- Contributions of [Yoon 2022]: The authors considered the continuous-time path integral control, and applied Chebyshev and Hoeffding inequalities to relate the instantaneous (pointwise-in-time) error bound in control input and the sample size of the Monte-Carlo simulations performed at that particular time instance.
- Limitations of [Yoon 2022]:
  - The effect of time discretization is not addressed.
  - It is not clear how the pointwise-in-time bound can be translated into a more explicit, end-to-end (trajectory-level) error bound.

[1] Yoon, Hyung-Jin, et al., "Sampling complexity of path integral methods for trajectory optimization," 2022 American Control Conference (ACC).

Research Motivation and Prior Work

- The outcome of Monte Carlo simulation is probabilistic and suboptimal when the sample size is finite; hence applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- To our knowledge not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][1].
- Contributions of [Yoon 2022]: The authors considered the continuous-time path integral control, and applied Chebyshev and Hoeffding inequalities to relate the instantaneous (pointwise-in-time) error bound in control input and the sample size of the Monte-Carlo simulations performed at that particular time instance.
- Limitations of [Yoon 2022]:
  - The effect of time discretization is not addressed.
  - It is not clear how the pointwise-in-time bound can be translated into a more explicit, end-to-end (trajectory-level) error bound.
  - The work does not provide machinery to compute the required sample size to achieve an acceptable loss of control performance.

[1] Yoon, Hyung-Jin, et al., "Sampling complexity of path integral methods for trajectory optimization," 2022 American Control Conference (ACC).

Our Contributions

(1) Derivation of a path integral formulation for discrete-time stochastic Linear Quadratic Regulator (LQR) using Kullback-Leibler (KL) control problem

Our Contributions

(1) Derivation of a path integral formulation for discrete-time stochastic Linear Quadratic Regulator (LQR) using Kullback-Leibler (KL) control problem

(2) Derivation of an end-to-end (trajectory-level) bound on the error between the optimal control signal (computed by the classical Riccati solution) and the one obtained by the path integral method as a function of sample sizes

Our Contributions

(1) Derivation of a path integral formulation for discrete-time stochastic Linear Quadratic Regulator (LQR) using Kullback-Leibler (KL) control problem

(2) Derivation of an end-to-end (trajectory-level) bound on the error between the optimal control signal (computed by the classical Riccati solution) and the one obtained by the path integral method as a function of sample sizes
Our analysis reveals that the sample size required exhibits a logarithmic dependence on the dimension of the control input.

Our Contributions

(1) Derivation of a path integral formulation for discrete-time stochastic Linear Quadratic Regulator (LQR) using Kullback-Leibler (KL) control problem

(2) Derivation of an end-to-end (trajectory-level) bound on the error between the optimal control signal (computed by the classical Riccati solution) and the one obtained by the path integral method as a function of sample sizes
Our analysis reveals that the sample size required exhibits a logarithmic dependence on the dimension of the control input.

(3) Formulation of a chance-constrained optimization problem to quantify the worst-case performance of the path integral LQR control. This result, together with (2), allows us to relate the sample size and the worst-case control performance of the path integral method.

Our Contributions

(1) Derivation of a path integral formulation for discrete-time stochastic Linear Quadratic Regulator (LQR) using Kullback-Leibler (KL) control problem

(2) Derivation of an end-to-end (trajectory-level) bound on the error between the optimal control signal (computed by the classical Riccati solution) and the one obtained by the path integral method as a function of sample sizes
Our analysis reveals that the sample size required exhibits a logarithmic dependence on the dimension of the control input.

(3) Formulation of a chance-constrained optimization problem to quantify the worst-case performance of the path integral LQR control. This result, together with (2), allows us to relate the sample size and the worst-case control performance of the path integral method.

While the stochastic LQR problem can be efficiently solved by the backward Riccati recursion, our primary focus is to establish the foundation for a sample complexity analysis of the path integral method when the analytical expressions of optimal control law and the cost are available.

Outline

Discrete-Time KL Control Problem

- Deterministic state transition law: $x_{t+1} = F_t(x_t, u_t)$

$$P_{X_{t+1}|X_t,U_t}(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t,u_t)}(dx_{t+1})$$

Discrete-Time KL Control Problem

- Deterministic state transition law: $x_{t+1} = F_t(x_t, u_t)$

$$P_{X_{t+1}|X_t,U_t}(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t,u_t)}(dx_{t+1})$$

- Nominal policy (can be stochastic): $R_{U_t|X_t}$

Discrete-Time KL Control Problem

- Deterministic state transition law: $x_{t+1} = F_t(x_t, u_t)$

$$P_{X_{t+1}|X_t,U_t}(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t,u_t)}(dx_{t+1})$$

- Nominal policy (can be stochastic): $R_{U_t|X_t}$
- Control policy to be designed (can be stochastic): $Q_{U_t|X_t}$

Discrete-Time KL Control Problem

- Deterministic state transition law: $x_{t+1} = F_t(x_t, u_t)$

$$P_{X_{t+1}|X_t,U_t}(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t,u_t)}(dx_{t+1})$$

- Nominal policy (can be stochastic): $R_{U_t|X_t}$
- Control policy to be designed (can be stochastic): $Q_{U_t|X_t}$
- Probability distributions of the state-control trajectories:

$$Q_{X_{0:T},U_{0:T-1}} = \prod_{t=0}^{T-1} P_{X_{t+1}|X_t,U_t} Q_{U_t|X_t}$$

$$R_{X_{0:T},U_{0:T-1}} = \prod_{t=0}^{T-1} P_{X_{t+1}|X_t,U_t} R_{U_t|X_t}.$$

Discrete-Time KL Control Problem

- Deterministic state transition law: $x_{t+1} = F_t(x_t, u_t)$

$$P_{X_{t+1}|X_t,U_t}(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t,u_t)}(dx_{t+1})$$

- Nominal policy (can be stochastic): $R_{U_t|X_t}$
- Control policy to be designed (can be stochastic): $Q_{U_t|X_t}$
- Probability distributions of the state-control trajectories:

$$Q_{X_{0:T},U_{0:T-1}} = \prod_{t=0}^{T-1} P_{X_{t+1}|X_t,U_t} Q_{U_t|X_t}$$
$$R_{X_{0:T},U_{0:T-1}} = \prod_{t=0}^{T-1} P_{X_{t+1}|X_t,U_t} R_{U_t|X_t}.$$

- Path cost: $C_{t:T}(x_{t:T}, u_{t:T-1}) := \sum_{k=t}^{T-1} \underbrace{C_k(x_k, u_k)}_{\text{stage-wise cost}} + \underbrace{C_T(x_T)}_{\text{terminal cost}}$

Discrete-Time KL Control Problem

- Deterministic state transition law: $x_{t+1} = F_t(x_t, u_t)$

$$P_{X_{t+1}|X_t,U_t}(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t,u_t)}(dx_{t+1})$$

- Nominal policy (can be stochastic): $R_{U_t|X_t}$
- Control policy to be designed (can be stochastic): $Q_{U_t|X_t}$
- Probability distributions of the state-control trajectories:

$$Q_{X_{0:T},U_{0:T-1}} = \prod_{t=0}^{T-1} P_{X_{t+1}|X_t,U_t} Q_{U_t|X_t}$$
$$R_{X_{0:T},U_{0:T-1}} = \prod_{t=0}^{T-1} P_{X_{t+1}|X_t,U_t} R_{U_t|X_t}.$$

- Path cost: $C_{t:T}(x_{t:T}, u_{t:T-1}) := \sum_{k=t}^{T-1} \underbrace{C_k(x_k, u_k)}_{\text{stage-wise cost}} + \underbrace{C_T(x_T)}_{\text{terminal cost}}$
- KL control problem:

$$\min_{\{Q_{U_t|X_t}\}_{t=0}^{T-1}} \mathbb{E}_Q^{x_0} C_{0:T}(X_{0:T}, U_{0:T-1}) + \lambda D(Q_{X_{0:T},U_{0:T-1}} \| R_{X_{0:T},U_{0:T-1}}).$$

λ balances the trade-off between the path cost and KL divergence

KL Control Problem: Dynamic Programming

- Value function:

$$J_t(x_t) := \min_{\{Q_{U_k|X_k}\}_{k=t}^{T-1}} \mathbb{E}_Q^{x_t}\, C_{t:T}(X_{t:T}, U_{t:T-1}) + \lambda D(Q_{X_{t:T},U_{t:T-1}} \| R_{X_{t:T},U_{t:T-1}}).$$

KL Control Problem: Dynamic Programming

- Value function:

$$J_t(x_t) := \min_{\{Q_{U_k|X_k}\}_{k=t}^{T-1}} \mathbb{E}_Q^{x_t}\ C_{t:T}(X_{t:T}, U_{t:T-1}) + \lambda D(Q_{X_{t:T},U_{t:T-1}} \| R_{X_{t:T},U_{t:T-1}}).$$

- Theorem: For each t, the value function admits a representation

$$J_t(x_t) = -\lambda \log \mathbb{E}_R^{x_t} \exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right).$$

KL Control Problem: Dynamic Programming

- Value function:

$$J_t(x_t) := \min_{\{Q_{U_k|X_k}\}_{k=t}^{T-1}} \mathbb{E}_Q^{x_t}\, C_{t:T}(X_{t:T}, U_{t:T-1}) + \lambda D(Q_{X_{t:T},U_{t:T-1}} \| R_{X_{t:T},U_{t:T-1}}).$$

- Theorem: For each t, the value function admits a representation

$$J_t(x_t) = -\lambda \log \mathbb{E}_R^{x_t} \exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right).$$

The optimal policy $Q^*_{U_t|X_t}$ for the KL control problem is expressed as

$$Q^*_{U_t|X_t}(B|x_t) = \frac{\int_B \exp(-\rho_t(x_t, u_t)/\lambda) R_{U_t|X_t}(du_t|x_t)}{\int_{\mathcal{U}_t} \exp(-\rho_t(x_t, u_t)/\lambda) R_{U_t|X_t}(du_t|x_t)}$$

for each Borel set B, where $\rho_t(x_t, u_t) := C_t(x_t, u_t) + J_{t+1}(F(x_t, u_t))$.

KL Control Problem: Dynamic Programming

- Value function:

$$J_t(x_t) := \min_{\{Q_{U_k|X_k}\}_{k=t}^{T-1}} \mathbb{E}_Q^{x_t} \; C_{t:T}(X_{t:T}, U_{t:T-1}) + \lambda D(Q_{X_{t:T},U_{t:T-1}} \| R_{X_{t:T},U_{t:T-1}}).$$

- Theorem: For each t, the value function admits a representation

$$J_t(x_t) = -\lambda \log \mathbb{E}_R^{x_t} \exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right).$$

The optimal policy $Q^*_{U_t|X_t}$ for the KL control problem is expressed as

$$Q^*_{U_t|X_t}(B|x_t) = \frac{\int_B \exp(-\rho_t(x_t, u_t)/\lambda) R_{U_t|X_t}(du_t|x_t)}{\int_{\mathcal{U}_t} \exp(-\rho_t(x_t, u_t)/\lambda) R_{U_t|X_t}(du_t|x_t)}$$

for each Borel set B, where $\rho_t(x_t, u_t) := C_t(x_t, u_t) + J_{t+1}(F(x_t, u_t))$.

Proof: Use Bellman's optimality principle and the Legendre duality between the KL divergence and free energy.

KL Control Problem: Monte Carlo Simulations

- We proved: $J_t(x_t) = -\lambda \log \boxed{\mathbb{E}_R^{x_t}} \exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right)$.

KL Control Problem: Monte Carlo Simulations

- We proved: $J_t(x_t) = -\lambda \log \boxed{\mathbb{E}_R^{x_t}} \exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right)$.
- Let $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^n$ be an ensemble of n sample state-control trajectories under the reference policy $\{R_{U_k|X_k}\}_{k=t}^{T-1}$. Then

$$J_t(x_t) \approx -\lambda \log \left(\boxed{\frac{1}{n} \sum_{i=1}^{n}} r(i) \right)$$

where $r(i) := \exp\left(-\frac{1}{\lambda} C_{t:T}(x_{t:T}(i), u_{t:T-1}(i))\right)$ is the reward of sample path i.

KL Control Problem: Monte Carlo Simulations

- We proved: $J_t(x_t) = -\lambda \log \boxed{\mathbb{E}_R^{x_t}} \exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right)$.
- Let $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^n$ be an ensemble of n sample state-control trajectories under the reference policy $\{R_{U_k|X_k}\}_{k=t}^{T-1}$. Then

$$J_t(x_t) \approx -\lambda \log \left(\boxed{\frac{1}{n} \sum_{i=1}^{n}} r(i) \right)$$

where $r(i) := \exp\left(-\frac{1}{\lambda} C_{t:T}(x_{t:T}(i), u_{t:T-1}(i))\right)$ is the reward of sample path i.
- The expectation of the control input (we will use it later):

$$\begin{aligned} \mathbb{E}_{Q^*}(U_t|x_t) &= \int_{\mathcal{U}_t} u_t Q^*(du_t|x_t) \\ &= \frac{\mathbb{E}_R\left[U_t \exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right)\right]}{\mathbb{E}_R\left[\exp\left(-\frac{1}{\lambda} C_{t:T}(X_{t:T}, U_{t:T-1})\right)\right]} \\ &\approx \frac{\sum_{i=1}^n u_t(i) r(i)}{\sum_{i=1}^n r(i)}. \end{aligned}$$

Outline

Stochastic LQR: Classical Solution

- Compute the state feedback policy $u_t = k_t(x_t)$ that solves

$$\min_{\{k_t(\cdot)\}_{t=0}^{T-1}} \mathbb{E} \sum_{t=0}^{T-1} \left(\frac{1}{2} X_t^\top M_t X_t + \frac{1}{2} U_t^\top N_t U_t \right) + \mathbb{E} \left(\frac{1}{2} X_T^\top M_T X_T \right)$$

$$\text{s.t.} \quad X_{t+1} = A_t X_t + B_t U_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t), \quad X_0 = x_0.$$

where $\{M_t\}_{t=0}^{T}$ and $\{N_t\}_{t=0}^{T-1}$ are sequences of positive definite matrices.

Stochastic LQR: Classical Solution

- Compute the state feedback policy $u_t = k_t(x_t)$ that solves

$$\min_{\{k_t(\cdot)\}_{t=0}^{T-1}} \mathbb{E} \sum_{t=0}^{T-1} \left(\frac{1}{2} X_t^\top M_t X_t + \frac{1}{2} U_t^\top N_t U_t \right) + \mathbb{E} \left(\frac{1}{2} X_T^\top M_T X_T \right)$$
$$\text{s.t.} \quad X_{t+1} = A_t X_t + B_t U_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t), \quad X_0 = x_0.$$

  where $\{M_t\}_{t=0}^{T}$ and $\{N_t\}_{t=0}^{T-1}$ are sequences of positive definite matrices.

- Optimal policy by solving backward Riccati Recursion

$$u_t = k_t(x_t) = K_t x_t, \quad K_t = -(B_t^\top \Theta_{t+1} B_t + N_t)^{-1} B_t^\top \Theta_{t+1} A_t$$

Stochastic LQR: Classical Solution

- Compute the state feedback policy $u_t = k_t(x_t)$ that solves

$$\min_{\{k_t(\cdot)\}_{t=0}^{T-1}} \mathbb{E} \sum_{t=0}^{T-1} \left(\frac{1}{2} X_t^\top M_t X_t + \frac{1}{2} U_t^\top N_t U_t \right) + \mathbb{E} \left(\frac{1}{2} X_T^\top M_T X_T \right)$$
$$\text{s.t.} \quad X_{t+1} = A_t X_t + B_t U_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t), \quad X_0 = x_0.$$

  where $\{M_t\}_{t=0}^{T}$ and $\{N_t\}_{t=0}^{T-1}$ are sequences of positive definite matrices.

- Optimal policy by solving backward Riccati Recursion

$$u_t = k_t(x_t) = K_t x_t, \quad K_t = -(B_t^\top \Theta_{t+1} B_t + N_t)^{-1} B_t^\top \Theta_{t+1} A_t$$

  where $\{\Theta_t\}_{t=0}^{T}$ is a sequence of positive definite matrices computed by the backward Riccati recursion with $\Theta_T = M_T$:

$$\Theta_t = A_t^\top \Theta_{t+1} A_t + M_t - A_t^\top \Theta_{t+1} B_t (B_t^\top \Theta_{t+1} B_t + N_t)^{-1} B_t^\top \Theta_{t+1} A_t.$$

Stochastic LQR: Path Integral Solution

- Recover the classical LQR results using KL control framework
- Assumption[2]:For each t, there exists a positive definite matrix $\hat{\Omega}_t$ and $\lambda > 0$ such that $N_t = \lambda \hat{\Omega}_t^{-1}$ and $B_t \hat{\Omega}_t B_t^\top = \Omega_t$.

[2] This is a common assumption in the path integral control literature. See, e.g., [Kappen 2005] for its system theoretic interpretation.

Stochastic LQR: Path Integral Solution

- Recover the classical LQR results using KL control framework
- Assumption[2]:For each t, there exists a positive definite matrix $\hat{\Omega}_t$ and $\lambda > 0$ such that $N_t = \lambda \hat{\Omega}_t^{-1}$ and $B_t \hat{\Omega}_t B_t^\top = \Omega_t$.
- Consider a KL control problem with
  - state transition law: $F_t(x_t, u_t) = A_t x_t + B_t u_t$
  - Reference Policy: $R_{U_t|X_t}(u_t|x_t) = \mathcal{N}(0, \hat{\Omega}_t)$
  - cost function $C_t(x_t, u_t) = \frac{1}{2} x_t^\top M_t x_t$

[2] This is a common assumption in the path integral control literature. See, e.g., [Kappen 2005] for its system theoretic interpretation.

Stochastic LQR: Path Integral Solution

- Recover the classical LQR results using KL control framework
- Assumption[2]:For each t, there exists a positive definite matrix $\hat{\Omega}_t$ and $\lambda > 0$ such that $N_t = \lambda \hat{\Omega}_t^{-1}$ and $B_t \hat{\Omega}_t B_t^\top = \Omega_t$.
- Consider a KL control problem with
  - state transition law: $F_t(x_t, u_t) = A_t x_t + B_t u_t$
  - Reference Policy: $R_{U_t|X_t}(u_t|x_t) = \mathcal{N}(0, \hat{\Omega}_t)$
  - cost function $C_t(x_t, u_t) = \frac{1}{2} x_t^\top M_t x_t$
- Theorem: The optimal policy $Q^*_{U_t|X_t}$ for the above KL control problem:

$$Q^*_{U_t|X_t}(u_t|x_t) = \mathcal{N}(-\hat{H}_t^{-1} \hat{G}_t^\top x_t, \lambda \hat{H}_t^{-1})$$

where $\hat{G}_t = A_t^\top \hat{\Theta}_{t+1} B_t$, $\hat{H}_t = B_t^\top \hat{\Theta}_{t+1} B_t + \lambda \hat{\Omega}_t^{-1}$ and $\hat{\Theta}_t$ satisfies

$$\hat{\Theta}_t = A_t^\top \hat{\Theta}_{t+1} A_t + M_t - A_t^\top \hat{\Theta}_{t+1} B_t (B_t^\top \hat{\Theta}_{t+1} B_t + \lambda \hat{\Omega}_t^{-1})^{-1} B_t^\top \hat{\Theta}_{t+1} A_t$$

with $\hat{\Theta}_T = M_T$.

[2] This is a common assumption in the path integral control literature. See, e.g., [Kappen 2005] for its system theoretic interpretation.

Stochastic LQR: Path Integral Solution

- Recover the classical LQR results using KL control framework
- Assumption[2]:For each t, there exists a positive definite matrix $\hat{\Omega}_t$ and $\lambda > 0$ such that $N_t = \lambda \hat{\Omega}_t^{-1}$ and $B_t \hat{\Omega}_t B_t^\top = \Omega_t$.
- Consider a KL control problem with
  - state transition law: $F_t(x_t, u_t) = A_t x_t + B_t u_t$
  - Reference Policy: $R_{U_t|X_t}(u_t|x_t) = \mathcal{N}(0, \hat{\Omega}_t)$
  - cost function $C_t(x_t, u_t) = \frac{1}{2} x_t^\top M_t x_t$
- Theorem: The optimal policy $Q^*_{U_t|X_t}$ for the above KL control problem:

$$Q^*_{U_t|X_t}(u_t|x_t) = \mathcal{N}(-\hat{H}_t^{-1} \hat{G}_t^\top x_t, \lambda \hat{H}_t^{-1})$$

  where $\hat{G}_t = A_t^\top \hat{\Theta}_{t+1} B_t$, $\hat{H}_t = B_t^\top \hat{\Theta}_{t+1} B_t + \lambda \hat{\Omega}_t^{-1}$ and $\hat{\Theta}_t$ satisfies

$$\hat{\Theta}_t = A_t^\top \hat{\Theta}_{t+1} A_t + M_t - A_t^\top \hat{\Theta}_{t+1} B_t (B_t^\top \hat{\Theta}_{t+1} B_t + \lambda \hat{\Omega}_t^{-1})^{-1} B_t^\top \hat{\Theta}_{t+1} A_t$$

  with $\hat{\Theta}_T = M_T$.
- Under the above assumption, $\mathbb{E}_{Q^*}(u_t|x_t) = -\hat{H}_t^{-1} \hat{G}_t^\top x_t = K_t x_t$ coincides with the classical LQR solution.

[2] This is a common assumption in the path integral control literature. See, e.g., [Kappen 2005] for its system theoretic interpretation.

Stochastic LQR: Monte Carlo Simulations

- Recall that $\mathbb{E}_{Q^*}(u_t|x_t)$ of the KL control problem can be computed via Monte Carlo simulations.

Stochastic LQR: Monte Carlo Simulations

- Recall that $\mathbb{E}_{Q^*}(u_t|x_t)$ of the KL control problem can be computed via Monte Carlo simulations.
- This implies that LQR solution can be computed via Monte Carlo simulations:

Stochastic LQR: Monte Carlo Simulations

- Recall that $\mathbb{E}_{Q^*}(u_t|x_t)$ of the KL control problem can be computed via Monte Carlo simulations.
- This implies that LQR solution can be computed via Monte Carlo simulations:
  - At every time-step t, sample n_t trajectories $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{n_t}$ from the "uncontrolled" dynamics: $X_{t+1} = A_t X_t + W_t$, $W_t \sim \mathcal{N}(0, \Omega_t)$

Stochastic LQR: Monte Carlo Simulations

- Recall that $\mathbb{E}_{Q^*}(u_t|x_t)$ of the KL control problem can be computed via Monte Carlo simulations.
- This implies that LQR solution can be computed via Monte Carlo simulations:
  - At every time-step t, sample n_t trajectories $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{n_t}$ from the "uncontrolled" dynamics: $X_{t+1} = A_t X_t + W_t$, $W_t \sim \mathcal{N}(0, \Omega_t)$
  - Compute path cost of each sample path i:

$$r(i) = \exp\left(-\frac{1}{\lambda}\sum_{k=t}^{T} \frac{1}{2} x_k(i)^\top M_k x_k(i)\right)$$

Stochastic LQR: Monte Carlo Simulations

- Recall that $\mathbb{E}_{Q^*}(u_t|x_t)$ of the KL control problem can be computed via Monte Carlo simulations.
- This implies that LQR solution can be computed via Monte Carlo simulations:
  - At every time-step t, sample n_t trajectories $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{n_t}$ from the "uncontrolled" dynamics: $X_{t+1} = A_t X_t + W_t$, $W_t \sim \mathcal{N}(0, \Omega_t)$
  - Compute path cost of each sample path i:

$$r(i) = \exp\left(-\frac{1}{\lambda}\sum_{k=t}^{T} \frac{1}{2} x_k(i)^\top M_k x_k(i)\right)$$

  - Path integral LQR controller:

$$\hat{u}_t = \sum_{i=1}^{n_t} \frac{r(i)}{\sum_{i=1}^{n_t} r(i)} u_t(i).$$

Stochastic LQR: Monte Carlo Simulations

- Recall that $\mathbb{E}_{Q^*}(u_t|x_t)$ of the KL control problem can be computed via Monte Carlo simulations.
- This implies that LQR solution can be computed via Monte Carlo simulations:
  - At every time-step t, sample n_t trajectories $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{n_t}$ from the "uncontrolled" dynamics: $X_{t+1} = A_t X_t + W_t$, $W_t \sim \mathcal{N}(0, \Omega_t)$
  - Compute path cost of each sample path i:

$$r(i) = \exp\left(-\frac{1}{\lambda}\sum_{k=t}^{T} \frac{1}{2} x_k(i)^\top M_k x_k(i)\right)$$

  - Path integral LQR controller:

$$\hat{u}_t = \sum_{i=1}^{n_t} \frac{r(i)}{\sum_{i=1}^{n_t} r(i)} u_t(i).$$

- Does not require solving backward Riccati equation

Outline

Sample Complexity Analysis

- Define the empirical means of the numerator and the denominator as

$$\hat{E}_t^{ru} = \frac{\sum_{i=1}^{n_t} r(i)u_t(i)}{n_t} \text{ and } \hat{E}_t^{r} = \frac{\sum_{i=1}^{n_t} r(i)}{n_t}.$$

Sample Complexity Analysis

- Define the empirical means of the numerator and the denominator as

$$\hat{E}_t^{ru} = \frac{\sum_{i=1}^{n_t} r(i) u_t(i)}{n_t} \text{ and } \hat{E}_t^{r} = \frac{\sum_{i=1}^{n_t} r(i)}{n_t}.$$

- Theorem: Let $\{\epsilon_t\}_{t=0}^{T-1}$, $\{\alpha_t\}_{t=0}^{T-1}$ and $\{\beta_t\}_{t=0}^{T-1}$ be given sequences of positive numbers and $\epsilon := \sum_{t=0}^{T-1} \epsilon_t^2$, $\alpha := \sum_{t=0}^{T-1} \alpha_t$, $\beta := \sum_{t=0}^{T-1} \beta_t$. Suppose $\alpha + \beta < 1$.

Sample Complexity Analysis

- Define the empirical means of the numerator and the denominator as

$$\hat{E}_t^{ru} = \frac{\sum_{i=1}^{n_t} r(i) u_t(i)}{n_t} \text{ and } \hat{E}_t^{r} = \frac{\sum_{i=1}^{n_t} r(i)}{n_t}.$$

- Theorem: Let $\{\epsilon_t\}_{t=0}^{T-1}$, $\{\alpha_t\}_{t=0}^{T-1}$ and $\{\beta_t\}_{t=0}^{T-1}$ be given sequences of positive numbers and $\epsilon := \sum_{t=0}^{T-1} \epsilon_t^2$, $\alpha := \sum_{t=0}^{T-1} \alpha_t$, $\beta := \sum_{t=0}^{T-1} \beta_t$. Suppose $\alpha + \beta < 1$. If n_t satisfies

$$n_t \geq \frac{\left(\hat{E}_t^r \sqrt{2\|\hat{\Omega}_t\| \log \frac{2m}{\beta_t}} + \left(\epsilon_t \hat{E}_t^r + \|\hat{E}_t^{ru}\|_\infty\right)\sqrt{\frac{1}{2} \log \frac{2}{\alpha_t}}\right)^2}{\epsilon_t^2 (\hat{E}_t^r)^4}$$

and $\hat{E}_t^r > \sqrt{\frac{1}{2n_t} \log \frac{2}{\alpha_t}}$,

Sample Complexity Analysis

- Define the empirical means of the numerator and the denominator as

$$\hat{E}_t^{ru} = \frac{\sum_{i=1}^{n_t} r(i) u_t(i)}{n_t} \text{ and } \hat{E}_t^{r} = \frac{\sum_{i=1}^{n_t} r(i)}{n_t}.$$

- Theorem: Let $\{\epsilon_t\}_{t=0}^{T-1}$, $\{\alpha_t\}_{t=0}^{T-1}$ and $\{\beta_t\}_{t=0}^{T-1}$ be given sequences of positive numbers and $\epsilon := \sum_{t=0}^{T-1} \epsilon_t^2$, $\alpha := \sum_{t=0}^{T-1} \alpha_t$, $\beta := \sum_{t=0}^{T-1} \beta_t$. Suppose $\alpha + \beta < 1$. If n_t satisfies

$$n_t \geq \frac{\left(\hat{E}_t^r \sqrt{2\|\hat{\Omega}_t\| \log \frac{2m}{\beta_t}} + \left(\epsilon_t \hat{E}_t^r + \|\hat{E}_t^{ru}\|_\infty\right) \sqrt{\frac{1}{2} \log \frac{2}{\alpha_t}}\right)^2}{\epsilon_t^2 (\hat{E}_t^r)^4}$$

and $\hat{E}_t^r > \sqrt{\frac{1}{2n_t} \log \frac{2}{\alpha_t}}$, then $\|\hat{u} - u\|_\infty^2 := \sum_{t=0}^{T-1} \|\hat{u}_t - u_t\|_\infty^2 \leq \epsilon$ with probability greater than or equal to $1 - \alpha - \beta$.

Sample Complexity Analysis

- Define the empirical means of the numerator and the denominator as

$$\hat{E}_t^{ru} = \frac{\sum_{i=1}^{n_t} r(i) u_t(i)}{n_t} \text{ and } \hat{E}_t^{r} = \frac{\sum_{i=1}^{n_t} r(i)}{n_t}.$$

- Theorem: Let $\{\epsilon_t\}_{t=0}^{T-1}$, $\{\alpha_t\}_{t=0}^{T-1}$ and $\{\beta_t\}_{t=0}^{T-1}$ be given sequences of positive numbers and $\epsilon := \sum_{t=0}^{T-1} \epsilon_t^2$, $\alpha := \sum_{t=0}^{T-1} \alpha_t$, $\beta := \sum_{t=0}^{T-1} \beta_t$. Suppose $\alpha + \beta < 1$. If n_t satisfies

$$n_t \geq \frac{\left(\hat{E}_t^r \sqrt{2 \|\hat{\Omega}_t\| \log \frac{2m}{\beta_t}} + \left(\epsilon_t \hat{E}_t^r + \|\hat{E}_t^{ru}\|_\infty \right) \sqrt{\frac{1}{2} \log \frac{2}{\alpha_t}} \right)^2}{\epsilon_t^2 (\hat{E}_t^r)^4}$$

and $\hat{E}_t^r > \sqrt{\frac{1}{2n_t} \log \frac{2}{\alpha_t}}$, then $\|\hat{u} - u\|_\infty^2 := \sum_{t=0}^{T-1} \|\hat{u}_t - u_t\|_\infty^2 \leq \epsilon$ with probability greater than or equal to $1 - \alpha - \beta$.

- The required number of samples depends only logarithmically on the dimension of the control input m.

Outline

Impact of $\|\hat{u} - u\|$ on the control performance

- Let $\hat{U}_t$ be the path integral control input. The closed-loop dynamics is

$$X_{t+1} = A_t X_t + B_t \hat{U}_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t)$$

Impact of $\|\hat{u} - u\|$ on the control performance

- Let $\hat{U}_t$ be the path integral control input. The closed-loop dynamics is

$$X_{t+1} = A_t X_t + B_t \hat{U}_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t)$$

and the accrued LQR cost is

$$\mathcal{L} := \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\frac{1}{2}X_t^\top M_t X_t + \frac{1}{2}\hat{U}_t^\top N_t \hat{U}_t\right) + \frac{1}{2}X_T^\top M_T X_T\right].$$

Impact of $\|\hat{u} - u\|$ on the control performance

- Let $\hat{U}_t$ be the path integral control input. The closed-loop dynamics is

$$X_{t+1} = A_t X_t + B_t \hat{U}_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t)$$

and the accrued LQR cost is

$$\mathcal{L} := \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\frac{1}{2}X_t^\top M_t X_t + \frac{1}{2}\hat{U}_t^\top N_t \hat{U}_t\right) + \frac{1}{2}X_T^\top M_T X_T\right].$$

- Introducing $V_t := \hat{U}_t - U_t$, the dynamics can be rewritten as

$$X_{t+1} = \widetilde{A}_t X_t + B_t V_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t)$$

where $\widetilde{A}_t := A_t + B_t K_t$.

Impact of $\|\hat{u} - u\|$ on the control performance

- Let $\hat{U}_t$ be the path integral control input. The closed-loop dynamics is

$$X_{t+1} = A_t X_t + B_t \hat{U}_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t)$$

and the accrued LQR cost is

$$\mathcal{L} := \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\frac{1}{2} X_t^\top M_t X_t + \frac{1}{2} \hat{U}_t^\top N_t \hat{U}_t\right) + \frac{1}{2} X_T^\top M_T X_T\right].$$

- Introducing $V_t := \hat{U}_t - U_t$, the dynamics can be rewritten as

$$X_{t+1} = \widetilde{A}_t X_t + B_t V_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t)$$

where $\widetilde{A}_t := A_t + B_t K_t$. The LQR cost can be written in terms of V_t as

$$\mathcal{L} = \mathbb{E} \sum_{t=0}^{T-1} \left(\frac{1}{2} X_t^\top \widetilde{M}_t X_t + X_t^\top \widetilde{N}_t V_t + \frac{1}{2} V_t^\top N_t V_t\right) + \mathbb{E}\left[\frac{1}{2} X_T^\top M_T X_T\right]$$

where $\widetilde{M}_t := M_t + K_t^\top N_t K_t$ and $\widetilde{N}_t := K_t^\top N_t$.

Impact of $\|\hat{u} - u\|$ on the control performance

- Goal: formulate a problem to search for the state feedback policy $v_t = \pi_t(x_t)$ that maximizes $\mathcal{L}$ while satisfying $\sum_{t=0}^{T-1} \|v_t\|_\infty^2 \leq \epsilon$ with probability at least $1 - \alpha - \beta$.

Impact of $\|\hat{u} - u\|$ on the control performance

- Goal: formulate a problem to search for the state feedback policy $v_t = \pi_t(x_t)$ that maximizes $\mathcal{L}$ while satisfying $\sum_{t=0}^{T-1} \|v_t\|_\infty^2 \leq \epsilon$ with probability at least $1 - \alpha - \beta$.
- Chance-constrained LQR:

$$
\begin{aligned}
\max_{\{\pi_t(\cdot)\}_{t=0}^{T-1}} \quad & \mathbb{E} \sum_{t=0}^{T-1} \left(\frac{1}{2} X_t^\top \widetilde{M}_t X_t + X_t^\top \widetilde{N}_t V_t + \frac{1}{2} V_t^\top N_t V_t \right) + \mathbb{E} \left[\frac{1}{2} X_T^\top M_T X_T \right] \\
\text{s.t.} \quad & X_{t+1} = \widetilde{A}_t X_t + B_t V_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t) \\
& \Pr\left(\sum_{t=0}^{T-1} \|v_t\|_\infty^2 \leq \epsilon \right) \geq 1 - \alpha - \beta.
\end{aligned}
$$

Impact of $\|\hat{u} - u\|$ on the control performance

- Goal: formulate a problem to search for the state feedback policy $v_t = \pi_t(x_t)$ that maximizes $\mathcal{L}$ while satisfying $\sum_{t=0}^{T-1} \|v_t\|_\infty^2 \leq \epsilon$ with probability at least $1 - \alpha - \beta$.
- Chance-constrained LQR:

$$\begin{aligned}
\max_{\{\pi_t(\cdot)\}_{t=0}^{T-1}} \ & \mathbb{E} \sum_{t=0}^{T-1} \left(\frac{1}{2} X_t^\top \widetilde{M}_t X_t + X_t^\top \widetilde{N}_t V_t + \frac{1}{2} V_t^\top N_t V_t \right) + \mathbb{E}\left[\frac{1}{2} X_T^\top M_T X_T \right] \\
\text{s.t.} \quad & X_{t+1} = \widetilde{A}_t X_t + B_t V_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t) \\
& \Pr\left(\sum_{t=0}^{T-1} \|v_t\|_\infty^2 \leq \epsilon \right) \geq 1 - \alpha - \beta.
\end{aligned}$$

- Let f^* be the value of the above chance-constrained LQR. If n_t satisfies the sample complexity bound then $\boxed{\mathcal{L} \leq f^*}$.

Impact of $\|\hat{u} - u\|$ on the control performance

- Goal: formulate a problem to search for the state feedback policy $v_t = \pi_t(x_t)$ that maximizes $\mathcal{L}$ while satisfying $\sum_{t=0}^{T-1} \|v_t\|_\infty^2 \leq \epsilon$ with probability at least $1 - \alpha - \beta$.
- Chance-constrained LQR:

$$\max_{\{\pi_t(\cdot)\}_{t=0}^{T-1}} \mathbb{E} \sum_{t=0}^{T-1} \left(\frac{1}{2} X_t^\top \widetilde{M}_t X_t + X_t^\top \widetilde{N}_t V_t + \frac{1}{2} V_t^\top N_t V_t \right) + \mathbb{E}\left[\frac{1}{2} X_T^\top M_T X_T \right]$$

$$\text{s.t.} \quad X_{t+1} = \widetilde{A}_t X_t + B_t V_t + W_t, \quad W_t \sim \mathcal{N}(0, \Omega_t)$$

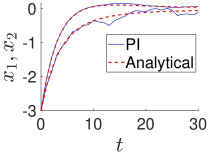
$$\Pr\left(\sum_{t=0}^{T-1} \|v_t\|_\infty^2 \leq \epsilon \right) \geq 1 - \alpha - \beta.$$

- Let f^* be the value of the above chance-constrained LQR. If n_t satisfies the sample complexity bound then $\boxed{\mathcal{L} \leq f^*}$.
- Finding a worst-case policy π_t that solves the above chance-constrained LQR is inherently challenging (left as a topic for future work).

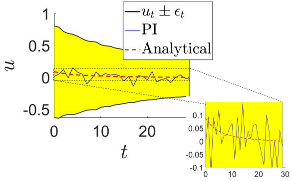
Outline

Example

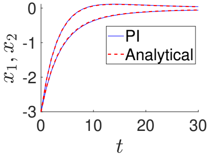
LQR problem: $A_t = \begin{bmatrix} 0.9 & -0.1 \\ -0.1 & 0.8 \end{bmatrix}$, $B_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\Omega_t = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$, $M_t = 0.1I$, $N_t = 10$. I represents an identity matrix of size 2×2.



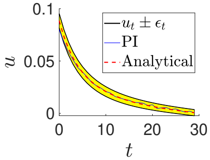
(a) State trajectory with $n = 10^3$



(b) Control input trajectory with $n = 10^3$



(c) State trajectory with $n = 10^7$



(d) Control input trajectory with $n = 10^7$

Outline

Summary

- We derived a path integral formulation for a discrete-time stochastic LQR problem.

Summary

- We derived a path integral formulation for a discrete-time stochastic LQR problem.
- An end-to-end bound on the error in the control signals was derived as a function of sample size.
  The required number of samples depends only logarithmically on the dimension of the control input.

Summary

- We derived a path integral formulation for a discrete-time stochastic LQR problem.
- An end-to-end bound on the error in the control signals was derived as a function of sample size.
  The required number of samples depends only logarithmically on the dimension of the control input.
- We formulated a chance-constrained optimization problem to quantify the worst-case control performance of the path integral LQR control.

Summary

- We derived a path integral formulation for a discrete-time stochastic LQR problem.
- An end-to-end bound on the error in the control signals was derived as a function of sample size.
  The required number of samples depends only logarithmically on the dimension of the control input.
- We formulated a chance-constrained optimization problem to quantify the worst-case control performance of the path integral LQR control.
- Future work:
  - Build upon this work to carry out sample complexity analysis of path integral for nonlinear continuous-time stochastic control problems.
  - Robustify the path integral control method by exploiting techniques from H^∞ control.