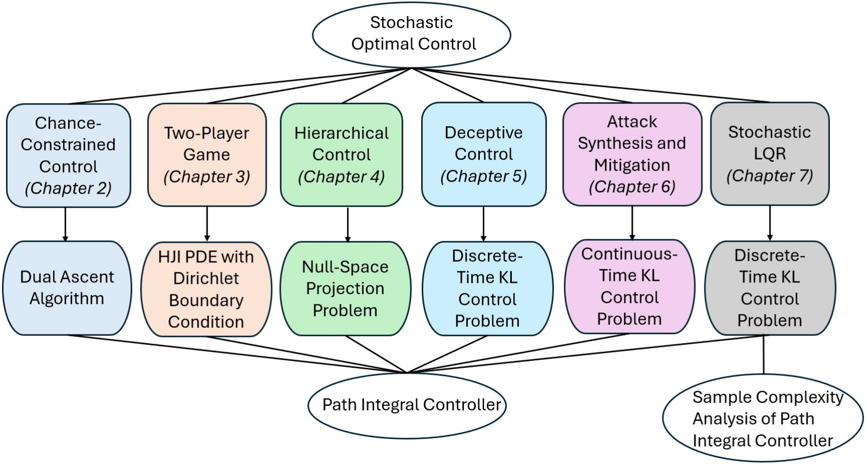
Ph.D. Defense

Advancing Frontiers of Path Integral Theory for Stochastic Optimal Control

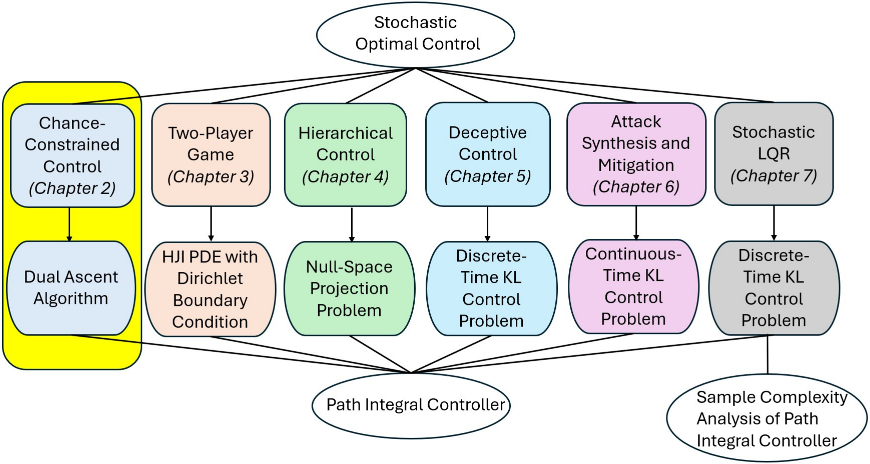
Apurva Patil

April 2, 2025

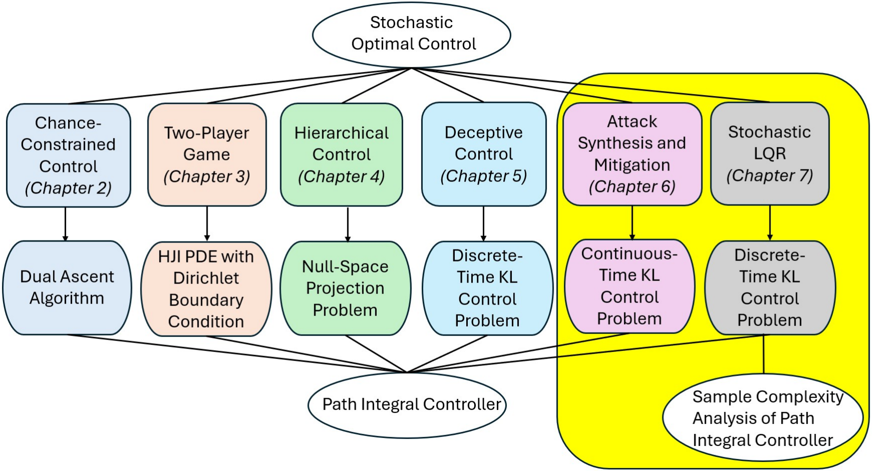
Outline of the Ph.D. Work



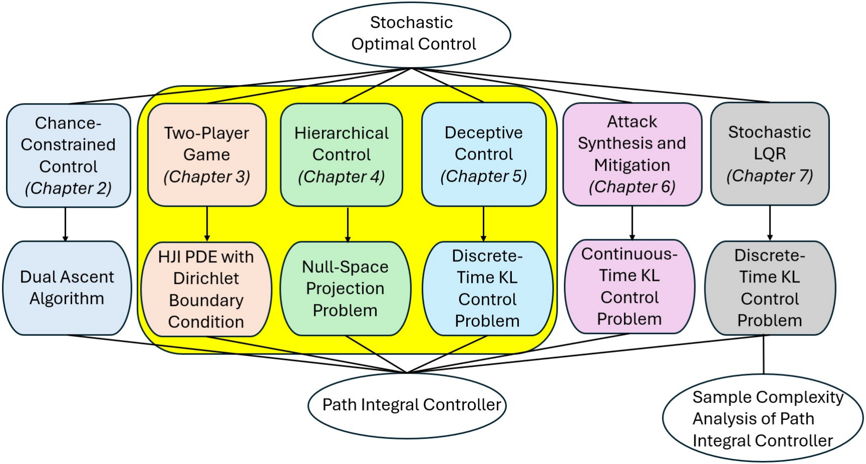
Proposal Talk: Recap



Today's Talk: Overview



Other Problems (if time permits)



Outline

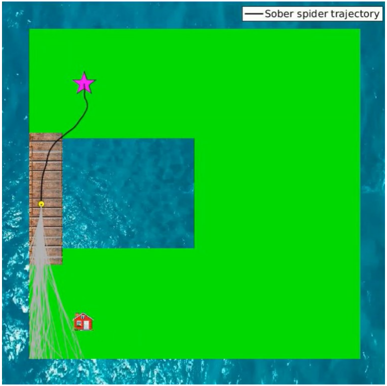
Outline

What is Path Integral Control?

- Path integral control solves stochastic optimal control problems. It computes optimal control input online (in real-time) via Monte-Carlo simulations.

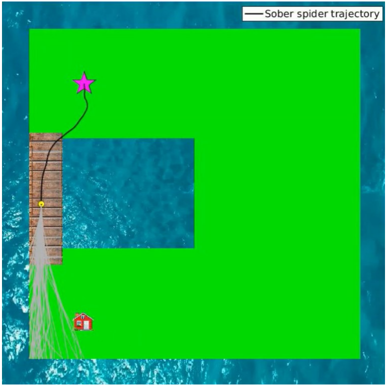
What is Path Integral Control?

- Path integral control solves stochastic optimal control problems. It computes optimal control input online (in real-time) via Monte-Carlo simulations.

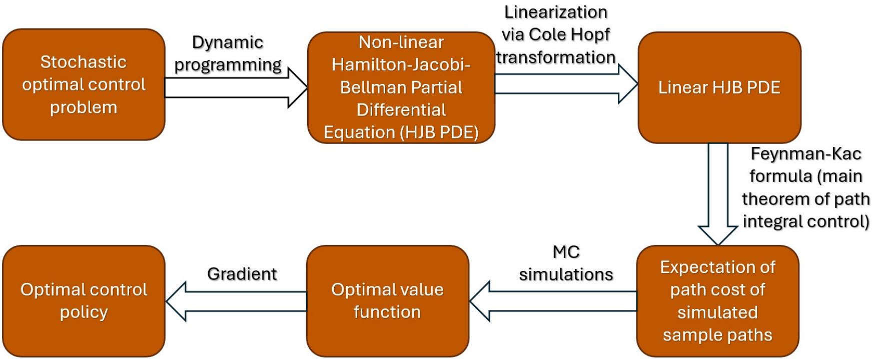


What is Path Integral Control?

- Path integral control solves stochastic optimal control problems. It computes optimal control input online (in real-time) via Monte-Carlo simulations.
- The optimal control input is computed via the empirical mean of the path cost ("path integral") of simulated sample paths.

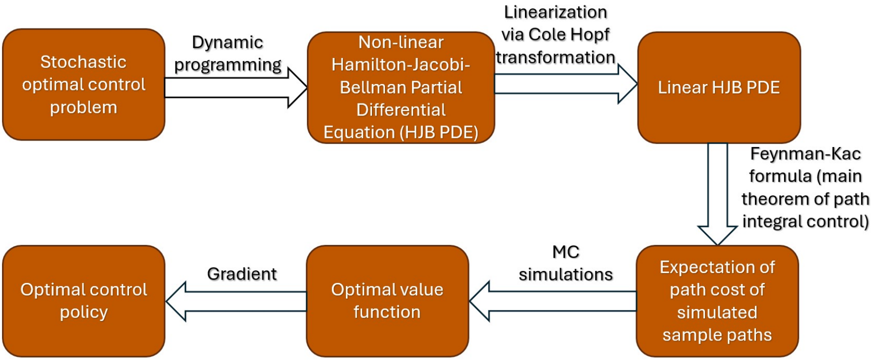


What is Path Integral Control? (Solution via HJB PDE)

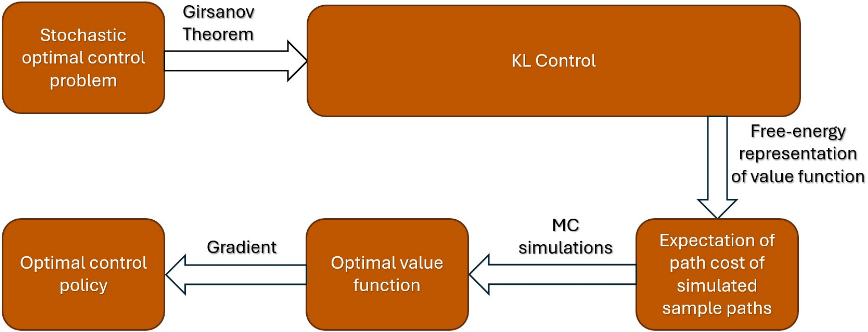


What is Path Integral Control? (Solution via HJB PDE)

Derivation by [Kappen 2005]

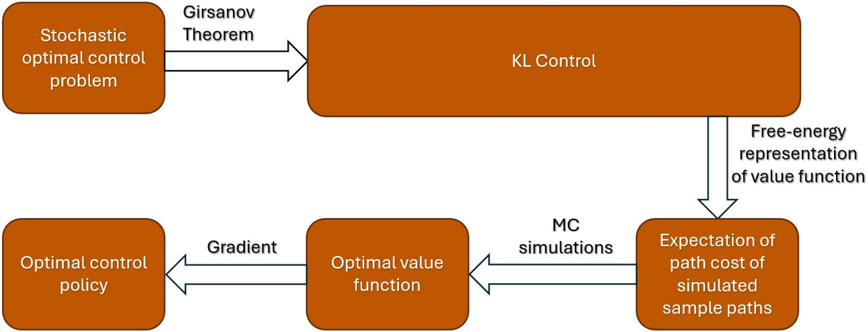


What is Path Integral Control? (Solution via KL Control)



What is Path Integral Control? (Solution via KL Control)

Derivation by [Theodorou and Todorov 2012]



Solution of KL Control using Path Integral Method

- P, Q: Probability distributions
- $C : \mathcal{X} \to \mathbb{R}$: cost function
- For $\lambda > 0$, the KL control problem:

$$\boxed{\min_{P} \mathbb{E}^{P}\left[C(x)\right] + \lambda \underbrace{D(P\|Q)}_{\text{KL Divergence}}}$$

KL Divergence: $D(P\|Q) := \int_{\mathcal{X}} \log \frac{dP}{dQ}(x) P(dx).$

Solution of KL Control using Path Integral Method

- P, Q: Probability distributions
- $C : \mathcal{X} \to \mathbb{R}$: cost function
- For $\lambda > 0$, the KL control problem:

$$\boxed{\min_P \mathbb{E}^P \left[C(x)\right] + \lambda \underbrace{D(P\|Q)}_{\text{KL Divergence}}}$$

$$\text{KL Divergence: } D(P\|Q) := \int_{\mathcal{X}} \log \frac{dP}{dQ}(x) P(dx).$$

Then according to the Legendre's duality,

$$\min_P \mathbb{E}^P \left[C(x)\right] + \lambda D(P\|Q) = \underbrace{-\lambda \log \boxed{\mathbb{E}^Q} \left[\exp\left\{-\frac{1}{\lambda} C(x)\right\}\right]}_{\text{Free energy}}$$

Solution of KL Control using Path Integral Method

- P, Q: Probability distributions
- $C : \mathcal{X} \to \mathbb{R}$: cost function
- For $\lambda > 0$, the KL control problem:

$$\boxed{\min_P \mathbb{E}^P \left[C(x)\right] + \lambda \underbrace{D(P\|Q)}_{\text{KL Divergence}}}$$

$$\text{KL Divergence: } D(P\|Q) := \int_{\mathcal{X}} \log \frac{dP}{dQ}(x) P(dx).$$

Then according to the Legendre's duality,

$$\min_P \mathbb{E}^P \left[C(x)\right] + \lambda D(P\|Q) = \underbrace{-\lambda \log \boxed{\mathbb{E}^Q} \left[\exp\left\{-\frac{1}{\lambda} C(x)\right\}\right]}_{\text{Free energy}}$$

$$\approx \underbrace{-\lambda \log \boxed{\frac{1}{N}\sum_{i=1}^{N}} \left[\exp\left\{-\frac{1}{\lambda} C^{(i)}(x)\right\}\right]}_{\text{Monte Carlo}}$$

Outline

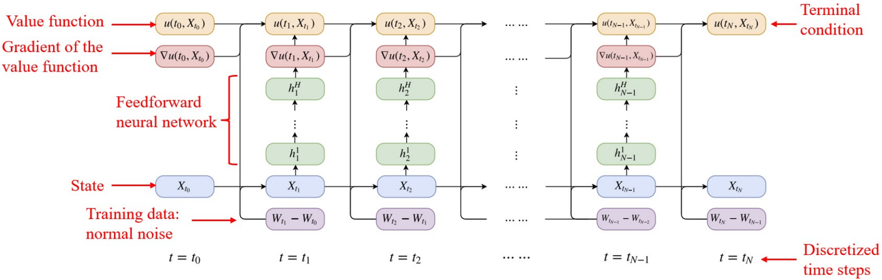
Why Path Integral Control?

Why Path Integral Control?

- Simulator-driven: no model required
- One shot, online method
- Works with non-linear systems and cost functions
- Can be applied to stochastic systems
- MC simulations can be parallelized on GPUs
- Less susceptible to curse of dimensionality
- No pretraining required

Why Path Integral Control?

Neural PDE solvers can solve high-dimensional HJB PDEs using deep neural networks (DNN).



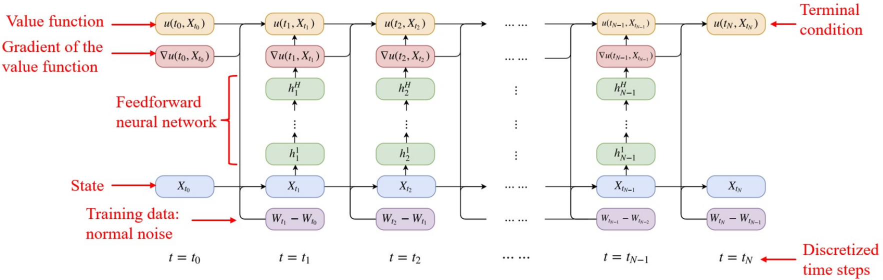
[1]

[1] Han et al., "Solving high-dimensional partial differential equations using deep learning", *Proceedings of the National Academy of Sciences* 115.34, 2018.

Why Path Integral Control?

Neural PDE solvers can solve high-dimensional HJB PDEs using deep neural networks (DNN).

- Extensive training required



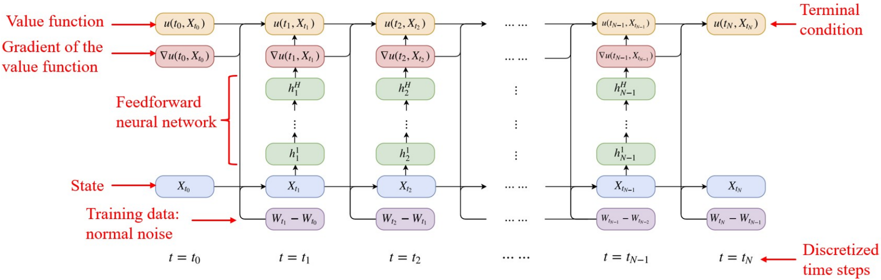
1

[1] Han et al., “Solving high-dimensional partial differential equations using deep learning", *Proceedings of the National Academy of Sciences* 115.34, 2018.

Why Path Integral Control?

Neural PDE solvers can solve high-dimensional HJB PDEs using deep neural networks (DNN).

- Extensive training required
- Careful DNN construction and hyperparameter tuning required



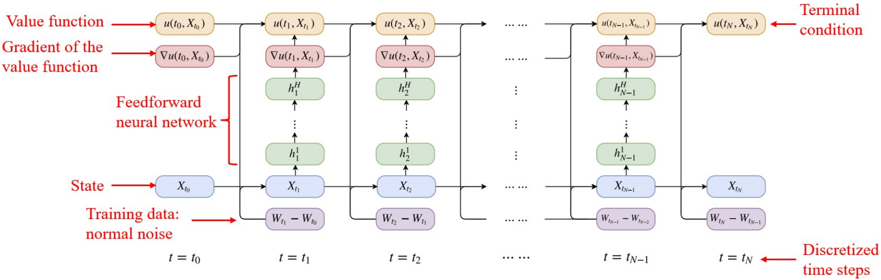
1

[1] Han et al., “Solving high-dimensional partial differential equations using deep learning", *Proceedings of the National Academy of Sciences* 115.34, 2018.

Why Path Integral Control?

Neural PDE solvers can solve high-dimensional HJB PDEs using deep neural networks (DNN).

- Extensive training required
- Careful DNN construction and hyperparameter tuning required
- Difficult to provide optimality guarantees



1

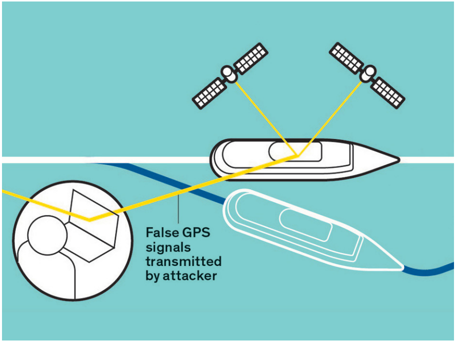
[1] Han et al., "Solving high-dimensional partial differential equations using deep learning", *Proceedings of the National Academy of Sciences* 115.34, 2018.

Outline

Stealthy Vehicle Misguidance and Its Mitigation

J. Bhatti and T. E. Humphreys, "Hostile control of ships via false GPS signals: Demonstration and detection," *NAVIGATION: Journal of the Institute of Navigation*, 2017.

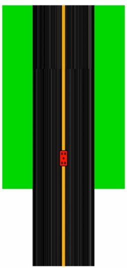
Stealthy Vehicle Misguidance and Its Mitigation

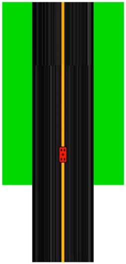


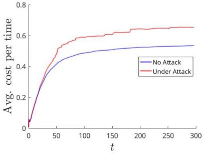
Inspired by the GPS spoofing demonstration, we formulate a stochastic zero-sum game to analyze the competition between

- Attacker, who tries to misguide the vehicle to an unsafe region covertly, and
- Controller, who tries to mitigate the impact of attack signals

Stealthy Attack on Cruise Control





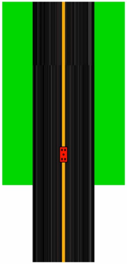


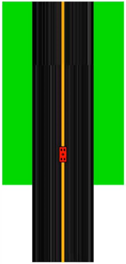
- The attacker injects a disturbance signal to degrade the control performance stealthily
- The legitimate controller tries to bring the vehicle state to a nominal level

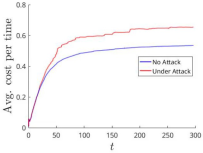
Stealthy Attack on Cruise Control

No attack

Under attack



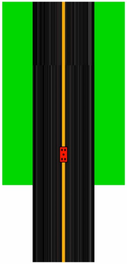




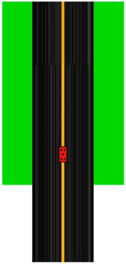
- The attacker injects a disturbance signal to degrade the control performance stealthily
- The legitimate controller tries to bring the vehicle state to a nominal level

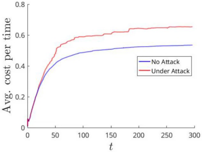
Stealthy Attack on Cruise Control

No attack



Under attack



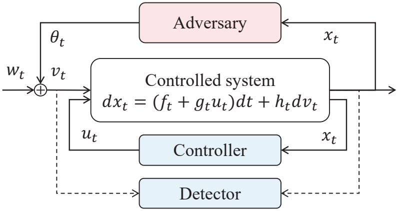


- The attacker injects a disturbance signal to degrade the control performance stealthily
- The legitimate controller tries to bring the vehicle state to a nominal level

- Q: How to synthesize the worst-case attack for nonlinear systems while remaining stealthy? (Attacker's problem)
- Q: How to mitigate the impact of stealthy attacks? (Controller's problem)

Outline

Problem Setup



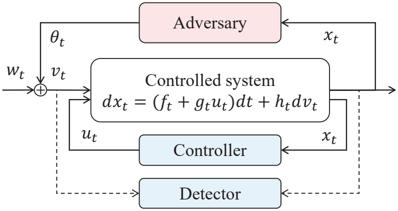
- Attack model:

$$dv_t = dw_t \quad \text{(No attack)}$$

$$dv_t = \theta_t dt + dw_t \quad \text{(Under attack)}$$

θ_t is a feedback policy.

Problem Setup



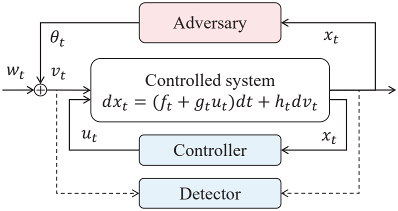
- Attack model:

$$dv_t = dw_t \quad \text{(No attack)}$$
$$dv_t = \theta_t dt + dw_t \quad \text{(Under attack)}$$

θ_t is a feedback policy.

- Controller can apply a legitimate control input u_t to combat with the noise w_t and the potential attack input θ_t

Problem Setup



- Attack model:

$$dv_t = dw_t \quad \text{(No attack)}$$
$$dv_t = \theta_t dt + dw_t \quad \text{(Under attack)}$$

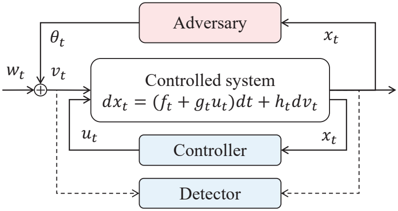
  θ_t is a feedback policy.

- Controller can apply a legitimate control input u_t to combat with the noise w_t and the potential attack input θ_t

$$\min_u \max_\theta \underbrace{\mathbb{E}^P \left[\int_0^T c_t(x_t, u_t) dt \right]}_{\text{e.g., penalty for going off-road}} - \lambda \times \text{(Attack Detectability)}$$

$\lambda > 0$ captures the trade-off between an attacker's desire to remain stealthy and its goal of degrading system performance.

Problem Setup



- Attack model:

  $$dv_t = dw_t \quad \text{(No attack)}$$
  $$dv_t = \theta_t dt + dw_t \quad \text{(Under attack)}$$

  θ_t is a feedback policy.

- Controller can apply a legitimate control input u_t to combat with the noise w_t and the potential attack input θ_t

$$\min_u \max_\theta \underbrace{\mathbb{E}^P \left[\int_0^T c_t(x_t, u_t) dt \right]}_{\text{e.g., penalty for going off-road}} - \lambda \times \text{(Attack Detectability)}$$

$\lambda > 0$ captures the trade-off between an attacker's desire to remain stealthy and its goal of degrading system performance.

Q: How to quantify the "attack detectability"?

KL Divergence: A Detectability Measure?

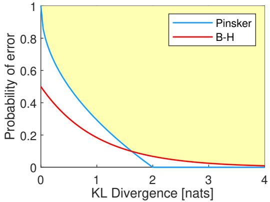
- Let P be the measure when the system is under attack, and Q be the measure when the system is under no attack

KL Divergence: A Detectability Measure?

- Let P be the measure when the system is under attack, and Q be the measure when the system is under no attack
- Type I error (false alarm): $Q(A)$
- Type II error (failure of detection): $P(A^c)$

KL Divergence: A Detectability Measure?

- Let P be the measure when the system is under attack, and Q be the measure when the system is under no attack
- Type I error (false alarm): $Q(A)$
- Type II error (failure of detection): $P(A^c)$
- KL divergence provides lower bounds to the total probability of errors

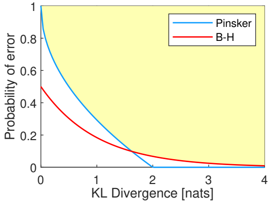


$$Q(A) + P(A^c) \geq 1 - \sqrt{\frac{1}{2} D(P \| Q)}$$ Pinsker's inequality

$$Q(A) + P(A^c) \geq \frac{1}{2} \exp(-D(P \| Q))$$ Bretagnolle-Huber inequality

KL Divergence: A Detectability Measure?

- Let P be the measure when the system is under attack, and Q be the measure when the system is under no attack
- Type I error (false alarm): $Q(A)$
- Type II error (failure of detection): $P(A^c)$
- KL divergence provides lower bounds to the total probability of errors



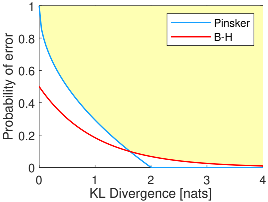
$$Q(A) + P(A^c) \geq 1 - \sqrt{\frac{1}{2} D(P\|Q)}$$ Pinsker's inequality

$$Q(A) + P(A^c) \geq \frac{1}{2} \exp(-D(P\|Q))$$ Bretagnolle-Huber inequality

Mini-max KL Control Problem: $$\min_u \max_\theta \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t) dt \right] - \lambda D(P\|Q)$$

KL Divergence: A Detectability Measure?

- Let P be the measure when the system is under attack, and Q be the measure when the system is under no attack
- Type I error (false alarm): $Q(A)$
- Type II error (failure of detection): $P(A^c)$
- KL divergence provides lower bounds to the total probability of errors



$$Q(A) + P(A^c) \geq 1 - \sqrt{\frac{1}{2} D(P \| Q)}$$ Pinsker's inequality

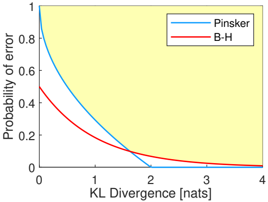
$$Q(A) + P(A^c) \geq \frac{1}{2} \exp(-D(P \| Q))$$ Bretagnolle-Huber inequality

Mini-max KL Control Problem: $$\min_u \max_\theta \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t) dt \right] - \lambda D(P \| Q)$$

The KL divergence captures the distance between the probability measures defined by the dynamics with and without attack signals

KL Divergence: A Detectability Measure?

- Let P be the measure when the system is under attack, and Q be the measure when the system is under no attack
- Type I error (false alarm): $Q(A)$
- Type II error (failure of detection): $P(A^c)$
- KL divergence provides lower bounds to the total probability of errors



$$Q(A) + P(A^c) \geq 1 - \sqrt{\frac{1}{2} D(P\|Q)}$$ Pinsker's inequality

$$Q(A) + P(A^c) \geq \frac{1}{2} \exp(-D(P\|Q))$$ Bretagnolle-Huber inequality

Mini-max KL Control Problem: $$\min_u \max_\theta \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t) dt \right] - \lambda D(P\|Q)$$

$$\text{s.t. } dx_t = (f_t + g_t u_t)\, dt + h_t\, (\theta_t dt + dw_t)$$

The KL divergence captures the distance between the probability measures defined by the dynamics with and without attack signals

Minimax KL Control, Risk-Sensitive Control and Two-Player Zero-Sum Stochastic Differential Game

Minimax KL control

$$\min_{u} \max_{\theta} \mathbb{E}^{P}\left[\int_{0}^{T} c(x_t, u_t)dt\right] - \lambda D(P||Q)$$

Minimax KL Control, Risk-Sensitive Control and Two-Player Zero-Sum Stochastic Differential Game

Minimax KL control

$$\min_u \max_\theta \mathbb{E}^P\left[\int_0^T c(x_t, u_t)dt\right] - \lambda D(P||Q)$$

Legendre Duality

Girsanov Theorem

Risk-sensitive control

$$\min_u \lambda \log \mathbb{E}^Q\left[\exp\left(\frac{1}{\lambda}\int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t\right\} dt\right)\right]$$

Two-player zero-sum stochastic differential game

$$\min_u \max_\theta \mathbb{E}^P \int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t - \frac{\lambda}{2}\|\theta_t\|^2\right\}$$

Minimax KL Control, Risk-Sensitive Control and Two-Player Zero-Sum Stochastic Differential Game

Minimax KL control

$$\min_u \max_\theta \mathbb{E}^P\left[\int_0^T c(x_t, u_t)dt\right] - \lambda D(P||Q)$$

Legendre Duality

Risk-sensitive control

$$\min_u \lambda \log \mathbb{E}^Q\left[\exp\left(\frac{1}{\lambda}\int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t\right\} dt\right)\right]$$

Girsanov Theorem

Two-player zero-sum stochastic differential game

$$\min_u \max_\theta \mathbb{E}^P \int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t - \frac{\lambda}{2}\|\theta_t\|^2\right\}$$

We will take the variational approach to solve these problems numerically!

Outline

Attacker's Problem

Suppose controller's policy u_t is fixed. Let's focus on the inner maximization problem

$$\min_u \boxed{\max_\theta \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t)dt\right] - \lambda D(P\|Q)}$$

Attacker's Problem

Suppose controller's policy u_t is fixed. Let's focus on the inner maximization problem

$$\min_u \boxed{\max_\theta \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t)dt\right] - \lambda D(P\|Q)}$$

Theorem (Legendre Duality)

The value function of the above problem
$V_t(x_t) := \max_\theta \mathbb{E}^P \left[\int_t^T c_s(x_s, u_s)ds\right] - \lambda D(P||Q)$ *exists, is unique and is given by*

$$V_t(x_t) = \underbrace{\lambda \log \mathbb{E}^Q \left[\exp\left\{\frac{1}{\lambda}\int_t^T c_s(x_s, u_s)ds\right\}\right]}_{\text{free energy}}$$

Furthermore, the optimal attack signal θ_t^* is given by

$$\theta_t^* dt = h_t^\top(x_t)\left(h_t(x_t)h_t(x_t)^\top\right)^{-1} \frac{\mathbb{E}^Q\left[exp\left\{\frac{1}{\lambda}\int_t^T c_s(x_s, u_s)ds\right\}h_t(x_t)dw_t\right]}{\mathbb{E}^Q\left[exp\left\{\frac{1}{\lambda}\int_t^T c_s(x_s, u_s)ds\right\}\right]}$$

Attacker's Problem

Suppose controller's policy u_t is fixed. Let's focus on the inner maximization problem

$$\min_u \boxed{\max_\theta \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t)dt\right] - \lambda D(P\|Q)}$$

Theorem (Legendre Duality)

The value function of the above problem
$V_t(x_t) := \max_\theta \mathbb{E}^P \left[\int_t^T c_s(x_s, u_s)ds\right] - \lambda D(P||Q)$ *exists, is unique and is given by*

$$V_t(x_t) = \underbrace{\lambda \log \mathbb{E}^Q \left[\exp\left\{\frac{1}{\lambda}\int_t^T c_s(x_s, u_s)ds\right\}\right]}_{\textit{free energy}}$$

Furthermore, the optimal attack signal θ_t^* is given by

$$\theta_t^* dt = h_t^\top(x_t)\left(h_t(x_t)h_t(x_t)^\top\right)^{-1} \frac{\mathbb{E}^Q\left[exp\left\{\frac{1}{\lambda}\int_t^T c_s(x_s, u_s)ds\right\} h_t(x_t)dw_t\right]}{\mathbb{E}^Q\left[exp\left\{\frac{1}{\lambda}\int_t^T c_s(x_s, u_s)ds\right\}\right]}$$

- Recall Q is the probability measure in which $dv_t = \theta_t dt + dw_t$ is a standard Browninan motion (i.e. $\theta_t = 0$)
- $\mathbb{E}^Q[\cdot]$ can be estimated from simulated trajectories of $dx_t = f_t dt + g_t u_t dt + h_t dw_t$

Attack Synthesis by Monte Carlo: Path Integral Control

- Direct computation of the value function by Monte Carlo (without solving backward HJB!)

$$\lambda \log \left[\boxed{\frac{1}{N} \sum_{i=1}^{N}} \exp \left\{ \frac{1}{\lambda} \int_t^T c_s(x_s^i, u_s^i) ds \right\} \right] \xrightarrow{a.s.} V_t(x_t)$$

Here, $\{x_s^i, u_s^i, t \leq s \leq T\}_{i=1}^N$ are randomly drawn sample paths by running $dx_t = f_t(x_t)dt + g_t(x_t)u_t dt + h_t(x_t)dw_t$

Attack Synthesis by Monte Carlo: Path Integral Control

- Direct computation of the value function by Monte Carlo (without solving backward HJB!)

$$\lambda \log \left[\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{ \frac{1}{\lambda} \int_t^T c_s(x_s^i, u_s^i) ds \right\} \right] \xrightarrow{a.s.} V_t(x_t)$$

Here, $\{x_s^i, u_s^i, t \le s \le T\}_{i=1}^N$ are randomly drawn sample paths by running $dx_t = f_t(x_t)dt + g_t(x_t)u_t dt + h_t(x_t)dw_t$

- Direct computation of the optimal control (worst and stealthiest attack)

$$h_t^\top(x_t)\left(h_t(x_t)h_t(x_t)^\top\right)^{-1} \frac{\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{ \frac{1}{\lambda} \int_t^T c_s(x_s^i, u_s^i) ds \right\} h_t(x_t)\epsilon}{\sqrt{\Delta t}\,\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{ \frac{1}{\lambda} \int_t^T c_s(x_s^i, u_s^i) ds \right\}} \xrightarrow{a.s.} \theta_t^*$$

where $\epsilon \sim \mathcal{N}(0,1)$

Outline

Minimax Game $\Rightarrow$ Risk Sensitive Control

Q: Can we solve the minimax game (equiv. risk-sensitive control problem) with Monte Carlo?

Minimax KL control

$$\min_u \boxed{\max_\theta \mathbb{E}^P\left[\int_0^T c(x_t, u_t)dt\right] - \lambda D(P||Q)}$$

Legendre Duality

Risk-sensitive control

$$\min_u \boxed{\lambda \log \mathbb{E}^Q\left[\exp\left(\frac{1}{\lambda}\int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t\right\} dt\right)\right]}$$

Minimax Game $\Rightarrow$ Risk Sensitive Control

Q: Can we solve the minimax game (equiv. risk-sensitive control problem) with Monte Carlo?

Minimax KL control

$$\min_u \boxed{\max_\theta \mathbb{E}^P\left[\int_0^T c(x_t, u_t)dt\right] - \lambda D(P||Q)}$$

Legendre Duality

Risk-sensitive control

$$\min_u \boxed{\lambda \log \mathbb{E}^Q\left[\exp\left(\frac{1}{\lambda}\int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t\right\} dt\right)\right]}$$

Assumption 1: The cost function c_t is quadratic in u_t:

$$c_t(x_t, u_t) = \ell_t(x_t) + \frac{1}{2}u_t^\top R_t(x_t)u_t \quad \text{where } R_t(x_t) \succeq 0 \text{ for all } t$$

Assumption 2: For all (x, t), there exists a constant $0 < \xi < \lambda$ satisfying:

$$\underbrace{h_t(x_t)h_t^\top(x_t)}_{\text{noise covariance}} = \xi g_t(x_t)\underbrace{R_t^{-1}(x_t)}_{\text{inverse of control cost}} g_t^\top(x_t).$$

Controller's Problem: Risk Sensitive Control

Define the value function

$$V_t(x_t) = \min_{u} \lambda \log \mathbb{E}^Q \left[\exp\left(\frac{1}{\lambda} \int_t^T \left\{ \ell_s(x_s) + \frac{1}{2} u_s^\top R_s u_s \right\} ds \right) \right]$$

Controller's Problem: Risk Sensitive Control

Define the value function

$$V_t(x_t) = \min_u \lambda \log \mathbb{E}^Q \left[\exp\left(\frac{1}{\lambda} \int_t^T \left\{ \ell_s(x_s) + \frac{1}{2} u_s^\top R_s u_s \right\} ds \right) \right]$$

Theorem

The solution of the above problem exists, is unique and is given by [a]

$$V_t(x_t) = -\gamma \log \mathbb{E}^Z \left[\exp \left\{ -\frac{1}{\gamma} \int_t^T \ell_s(x_s) ds \right\} \right] \quad \text{where } \gamma = \frac{\xi \lambda}{\lambda - \xi}$$

and Z is the probability measure defined by the "passive" dynamics $dx_t = f_t dt + h_t dw_t$. Furthermore, the optimal controller signal u_t^* is given by

$$u_t^* dt = \mathcal{H}_t(x_t) \frac{\mathbb{E}^Z \left[exp\left\{ -\frac{1}{\gamma} \int_t^T \ell_s(x_s) ds \right\} h_t(x_t) dw_t \right]}{\mathbb{E}^Z \left[exp\left\{ -\frac{1}{\gamma} \int_t^T \ell_s(x_s) ds \right\} \right]}$$

where

$$\mathcal{H}_t(x_t) = R_t^{-1} g_t^\top(x_t) \left(g_t(x_t) R_t^{-1} g_t^\top(x_t) - \frac{1}{\lambda} h_t(x_t) h_t(x_t)^\top \right)^{-1}$$

[a] Broek et al., "Risk sensitive path integral control", *arXiv preprint arXiv:1203.3523* 2012.

Policy Synthesis by Monte Carlo: Path Integral Control

- Direct computation of the value function by Monte Carlo (without solving backward HJB!)

$$-\gamma \log \left[\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{ -\frac{1}{\gamma}\int_t^T \ell_s(x_s^i)ds \right\} \right] \xrightarrow{a.s.} V_t(x_t)$$

Here, $\{x_s^i, t \leq s \leq T\}_{i=1}^N$ are randomly drawn sample paths by running $dx_t = f_t(x_t)dt + h_t(x_t)dw_t$

Policy Synthesis by Monte Carlo: Path Integral Control

- Direct computation of the value function by Monte Carlo (without solving backward HJB!)

$$-\gamma \log \left[\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{-\frac{1}{\gamma}\int_t^T \ell_s(x_s^i)ds\right\}\right] \xrightarrow{a.s.} V_t(x_t)$$

Here, $\{x_s^i, t \leq s \leq T\}_{i=1}^N$ are randomly drawn sample paths by running $dx_t = f_t(x_t)dt + h_t(x_t)dw_t$

- Direct computation of the optimal control (attack mitigating policy)

$$\mathcal{H}_t(x_t)\frac{\boxed{\frac{1}{N}\sum_{i=1}^{N}}\exp\left\{-\frac{1}{\gamma}\int_t^T \ell_s(x_s^i)ds\right\}h_t(x_t)\epsilon}{\sqrt{\Delta t}\boxed{\frac{1}{N}\sum_{i=1}^{N}}\exp\left\{-\frac{1}{\gamma}\int_t^T \ell_s(x_s^i)ds\right\}} \xrightarrow{a.s.} u_t^*$$

where $\epsilon \sim \mathcal{N}(0,1)$

Two-Player Zero-Sum Stochastic Differential Game

Girsanov Theorem:

$$D(P\|Q) = \mathbb{E}^P \log \frac{dP}{dQ} = \mathbb{E}^P \left[\int_0^T \theta_t^\top dw_t + \frac{1}{2} \int_0^T \|\theta_t\|^2 dt \right] = \frac{1}{2} \mathbb{E}^P \left[\int_0^T \|\theta_t\|^2 dt \right].$$

Minimax KL control

$$\min_u \max_\theta \mathbb{E}^P \left[\int_0^T c(x_t, u_t) dt \right] - \boxed{\lambda D(P\|Q)}$$

Legendre Duality

Girsanov Theorem

Risk-sensitive control

$$\min_u \lambda \log \mathbb{E}^Q \left[\exp\left(\frac{1}{\lambda} \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t \right\} dt \right) \right]$$

Two-player zero-sum stochastic differential game

$$\min_u \max_\theta \mathbb{E}^P \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t - \boxed{\frac{\lambda}{2} \|\theta_t\|^2} \right\}$$

Two-Player Zero-Sum Stochastic Differential Game

Girsanov Theorem:

$$D(P\|Q) = \mathbb{E}^P \log \frac{dP}{dQ} = \mathbb{E}^P \left[\int_0^T \theta_t^\top dw_t + \frac{1}{2} \int_0^T \|\theta_t\|^2 dt \right] = \frac{1}{2} \mathbb{E}^P \left[\int_0^T \|\theta_t\|^2 dt \right].$$

Minimax KL control

$$\min_u \max_\theta \mathbb{E}^P \left[\int_0^T c(x_t, u_t) dt \right] - \boxed{\lambda D(P||Q)}$$

Legendre Duality

Girsanov Theorem

Risk-sensitive control

$$\min_u \lambda \log \mathbb{E}^Q \left[\exp\left(\frac{1}{\lambda} \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t \right\} dt \right) \right]$$

Two-player zero-sum stochastic differential game

$$\min_u \max_\theta \mathbb{E}^P \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t - \boxed{\frac{\lambda}{2} \|\theta_t\|^2} \right\}$$

Assumption 3: For all (x, t), there exists a constant $\alpha > 0$ satisfying the following equation:

$$\underbrace{h_t(x_t) h_t^\top(x_t)}_{\text{noise covariance}} = \alpha \left(g_t(x_t) \underbrace{R_t^{-1}(x_t)}_{\text{inverse of control cost}} g_t^\top(x_t) - \frac{1}{\lambda} h_t(x_t) h_t^\top(x_t) \right).$$

Two-Player Zero-Sum Stochastic Differential Game

Define the value function

$$V_t(x_t) = \min_u \max_\theta \mathbb{E}^P \int_t^T \left(\ell_s(x_s) + \frac{1}{2} u_s^\top R_s u_s - \frac{\lambda}{2} \|\theta_s\|^2 \right) ds.$$

Two-Player Zero-Sum Stochastic Differential Game

Define the value function

$$V_t(x_t) = \min_u \max_\theta \mathbb{E}^P \int_t^T \left(\ell_s(x_s) + \frac{1}{2} u_s^\top R_s u_s - \frac{\lambda}{2} \|\theta_s\|^2 \right) ds.$$

Theorem

The solution of the above problem exists, is unique and is given by [a]

$$V_t(x_t) = -\alpha \log \mathbb{E}^Z \left[\exp \left\{ -\frac{1}{\alpha} \int_t^T \ell_s(x_s) ds \right\} \right]$$

and Z is the probability measure defined by the "passive" dynamics $dx_t = f_t dt + h_t dw_t$. Furthermore, the saddle-point policies are given by

$$u_t^* dt = \mathcal{H}_t^u \frac{\mathbb{E}^Z \left[exp\left\{ -\frac{1}{\alpha} \int_t^T \ell_s ds \right\} h_t dw_t \right]}{\mathbb{E}^Z \left[exp\left\{ -\frac{1}{\alpha} \int_t^T \ell_s ds \right\} \right]}, \quad \theta_t^* dt = \mathcal{H}_t^\theta \frac{\mathbb{E}^Z \left[exp\left\{ -\frac{1}{\alpha} \int_t^T \ell_s ds \right\} h_t dw_t \right]}{\mathbb{E}^Z \left[exp\left\{ -\frac{1}{\alpha} \int_t^T \ell_s ds \right\} \right]}$$

where

$$\mathcal{H}_t^u = R_t^{-1} g_t^\top \left(g_t R_t^{-1} g_t^\top - \frac{1}{\lambda} h_t h_t^\top \right)^{-1}, \quad \mathcal{H}_t^\theta = -\frac{1}{\lambda} h_t^\top \left(g_t R_t^{-1} g_t^\top - \frac{1}{\lambda} h_t h_t^\top \right)^{-1}$$

[a] Patil et al., "Risk-minimizing two-player zero-sum stochastic differential game via path integral control", *Conference on Decision and Control*, 2023.

Policy Synthesis by Monte Carlo: Path Integral Control

- Direct computation of the value function by Monte Carlo (without solving backward HJB!)

$$-\alpha \log \left[\boxed{\frac{1}{N} \sum_{i=1}^{N}} \exp \left\{ -\frac{1}{\alpha} \int_t^T \ell_s(x_s^i) ds \right\} \right] \xrightarrow{a.s.} V_t(x_t)$$

Here, $\{x_s^i, t \leq s \leq T\}_{i=1}^N$ are randomly drawn sample paths by running $dx_t = f_t(x_t)dt + h_t(x_t)dw_t$

Policy Synthesis by Monte Carlo: Path Integral Control

- Direct computation of the value function by Monte Carlo (without solving backward HJB!)

$$-\alpha \log \left[\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{ -\frac{1}{\alpha}\int_t^T \ell_s(x_s^i)ds \right\} \right] \xrightarrow{a.s.} V_t(x_t)$$

Here, $\{x_s^i, t \leq s \leq T\}_{i=1}^N$ are randomly drawn sample paths by running $dx_t = f_t(x_t)dt + h_t(x_t)dw_t$

- Direct computation of the optimal control (attack mitigating policy)

$$\mathcal{H}_t^u(x_t) \frac{\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{ -\frac{1}{\alpha}\int_t^T \ell_s(x_s^i)ds \right\} h_t(x_t)\epsilon}{\sqrt{\Delta t}\boxed{\frac{1}{N}\sum_{i=1}^{N}} \exp\left\{ -\frac{1}{\alpha}\int_t^T \ell_s(x_s^i)ds \right\}} \xrightarrow{a.s.} u_t^*$$

where $\epsilon \sim \mathcal{N}(0,1)$

Controller's Problem

Q: Are the path integral solutions of risk-sensitive control and two-player zero-sum stochastic differential game the same?

Minimax KL control

$$\min_u \max_\theta \mathbb{E}^P \left[\int_0^T c(x_t, u_t) dt \right] - \lambda D(P||Q)$$

Legendre Duality

Girsanov Theorem

Risk-sensitive control

$$\min_u \lambda \log \mathbb{E}^Q \left[\exp \left(\frac{1}{\lambda} \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t \right\} dt \right) \right]$$

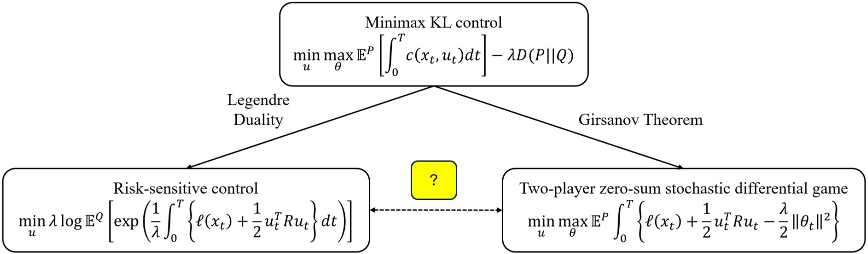
?

Two-player zero-sum stochastic differential game

$$\min_u \max_\theta \mathbb{E}^P \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t - \frac{\lambda}{2} \|\theta_t\|^2 \right\}$$

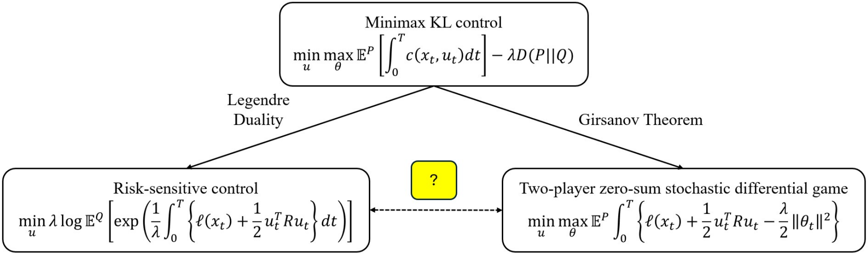
Controller's Problem

Q: Are the path integral solutions of risk-sensitive control and two-player zero-sum stochastic differential game the same?
We derived the path integral solutions under two seemingly different assumptions for each problem.



Controller's Problem

Q: Are the path integral solutions of risk-sensitive control and two-player zero-sum stochastic differential game the same?
We derived the path integral solutions under two seemingly different assumptions for each problem.



It turns out these two assumptions are essentially identical

Controller's Problem

Q: Are the path integral solutions of risk-sensitive control and two-player zero-sum stochastic differential game the same?
We derived the path integral solutions under two seemingly different assumptions for each problem.

Minimax KL control

$$\min_u \max_\theta \mathbb{E}^P \left[\int_0^T c(x_t, u_t) dt \right] - \lambda D(P||Q)$$

Legendre Duality

Girsanov Theorem

Risk-sensitive control

$$\min_u \lambda \log \mathbb{E}^Q \left[\exp \left(\frac{1}{\lambda} \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t \right\} dt \right) \right]$$

?

Two-player zero-sum stochastic differential game

$$\min_u \max_\theta \mathbb{E}^P \int_0^T \left\{ \ell(x_t) + \frac{1}{2} u_t^T R u_t - \frac{\lambda}{2} \|\theta_t\|^2 \right\}$$

It turns out these two assumptions are essentially identical $\Rightarrow$ The path integral solutions of risk-sensitive control and two-player zero-sum stochastic differential game are the same !!

Controller's Problem

Q: Are the path integral solutions of risk-sensitive control and two-player zero-sum stochastic differential game the same?
We derived the path integral solutions under two seemingly different assumptions for each problem.

Minimax KL control

$$\min_u \max_\theta \mathbb{E}^P\left[\int_0^T c(x_t, u_t)dt\right] - \lambda D(P||Q)$$

Legendre Duality

Girsanov Theorem

Risk-sensitive control

$$\min_u \lambda \log \mathbb{E}^Q\left[\exp\left(\frac{1}{\lambda}\int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t\right\} dt\right)\right]$$

1. Jacobson 1973.
2. Petersen, James, & Dupuis 2000.

Two-player zero-sum stochastic differential game

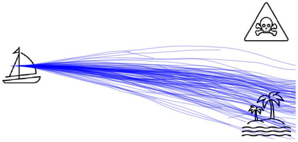
$$\min_u \max_\theta \mathbb{E}^P \int_0^T \left\{\ell(x_t) + \frac{1}{2}u_t^T R u_t - \frac{\lambda}{2}\|\theta_t\|^2\right\}$$

It turns out these two assumptions are essentially identical $\Rightarrow$ The path integral solutions of risk-sensitive control and two-player zero-sum stochastic differential game are the same !!

Outline

Simulation Results

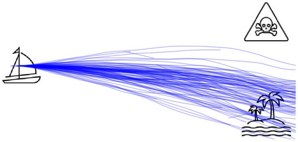
Simulation Results



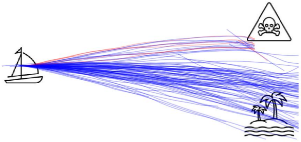
No attack, $P_{\text{crash}} \approx 0$

Simulation Results

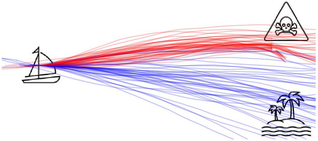
$$\max_{\theta} \mathbb{E}^{P}\left[\int_0^T c_t(x_t, u_t)dt\right] - \lambda D(P \| Q)$$



No attack, $P_{\mathrm{crash}} \approx 0$



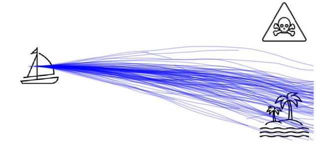
Stealthy attack, $\lambda = 1.5$, $P_{\mathrm{crash}} \approx 0.05$



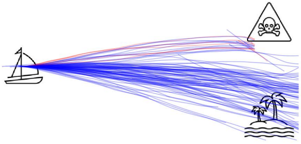
Stealthy attack, $\lambda = 0.05$, $P_{\mathrm{crash}} \approx 0.52$

Simulation Results

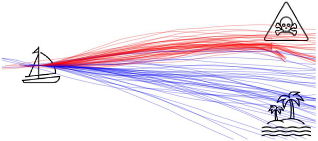
$$\min_{u} \max_{\theta} \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t) dt \right] - \lambda D(P \| Q)$$



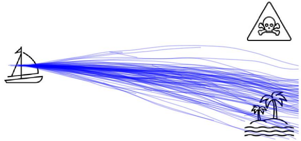
No attack, $P_{\text{crash}} \approx 0$



Stealthy attack, $\lambda = 1.5$, $P_{\text{crash}} \approx 0.05$



Stealthy attack, $\lambda = 0.05$, $P_{\text{crash}} \approx 0.52$



Attack mitigation, $\lambda = 0.05$, $P_{\text{crash}} \approx 0$

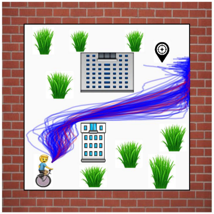
Simulation Results

No attack, $P_{\mathrm{crash}} \approx 0.01$

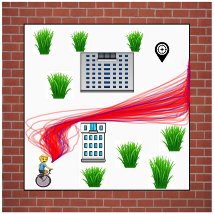
Simulation Results

$$\max_{\theta} \mathbb{E}^P \left[\int_0^T c_t(x_t, u_t) dt \right] - \lambda D(P \| Q)$$

No attack, $P_{\mathrm{crash}} \approx 0.01$



Stealthy attack, $\lambda = 2$, $P_{\mathrm{crash}} \approx 0.17$

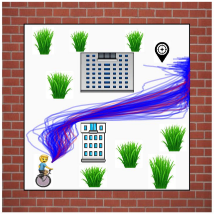


Stealthy attack, $\lambda = 0.8$, $P_{\mathrm{crash}} \approx 0.91$

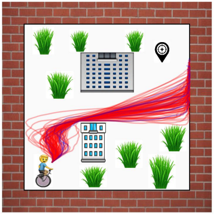
Simulation Results

$$\min_{u} \max_{\theta} \mathbb{E}^{P}\left[\int_{0}^{T} c_t(x_t, u_t)dt\right] - \lambda D(P\|Q)$$

No attack, $P_{\text{crash}} \approx 0.01$



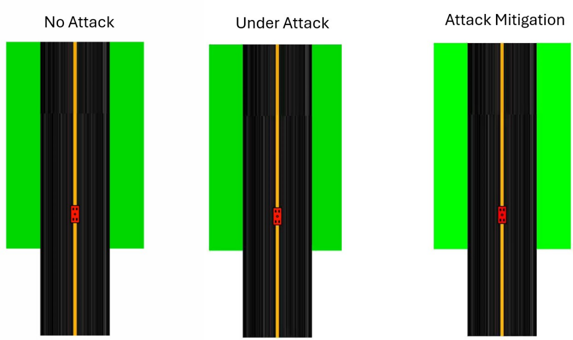
Stealthy attack, $\lambda = 2$, $P_{\text{crash}} \approx 0.17$

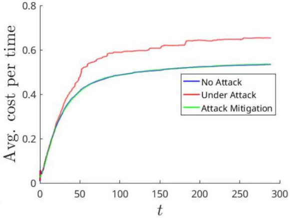


Stealthy attack, $\lambda = 0.8$, $P_{\text{crash}} \approx 0.91$

Attack mitigation, $\lambda = 0.8$, $P_{\text{crash}} \approx 0.02$

Simulation Results





Summary

- We developed a path-integral-based method to synthesize worst-case stealthy attacks in real time for nonlinear continuous-time systems

Summary

- We developed a path-integral-based method to synthesize worst-case stealthy attacks in real time for nonlinear continuous-time systems
- To mitigate the risk of stealthy attacks, we proposed a novel zero-sum game formulation

Summary

- We developed a path-integral-based method to synthesize worst-case stealthy attacks in real time for nonlinear continuous-time systems
- To mitigate the risk of stealthy attacks, we proposed a novel zero-sum game formulation
- We provided a path-integral-based approach for controller to synthesize the attack mitigating control inputs online

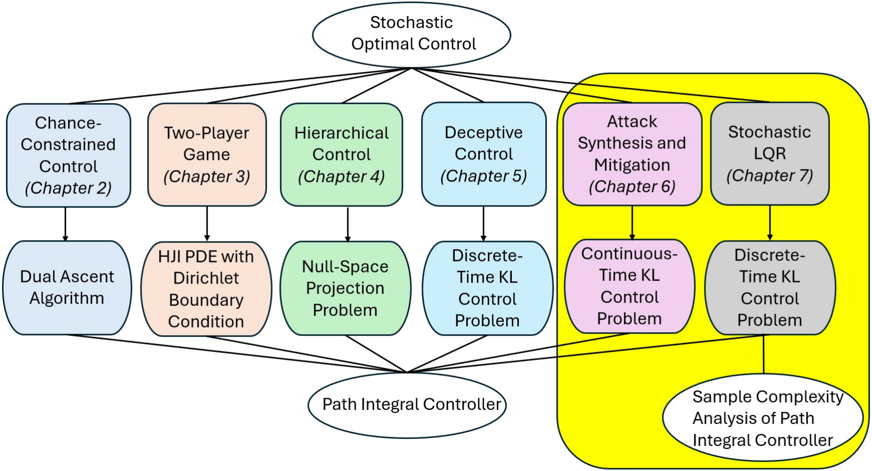
Summary

- We developed a path-integral-based method to synthesize worst-case stealthy attacks in real time for nonlinear continuous-time systems
- To mitigate the risk of stealthy attacks, we proposed a novel zero-sum game formulation
- We provided a path-integral-based approach for controller to synthesize the attack mitigating control inputs online
- We showed that the path integral solution to the risk-sensitive control coincides with that of two-player zero-sum stochastic differential game

Summary

- We developed a path-integral-based method to synthesize worst-case stealthy attacks in real time for nonlinear continuous-time systems
- To mitigate the risk of stealthy attacks, we proposed a novel zero-sum game formulation
- We provided a path-integral-based approach for controller to synthesize the attack mitigating control inputs online
- We showed that the path integral solution to the risk-sensitive control coincides with that of two-player zero-sum stochastic differential game
- Publications:
  - A. Patil*, M. Karabag*, U. Topcu, T. Tanaka, “Simulation-Driven Deceptive Control via Path Integral Approach," *2023 IEEE Conference on Decision and Control (CDC)*
  - A. Patil, K. Morgenstein, L. Sentis, T. Tanaka, “Stealthy Attack Synthesis and Its Mitigation for Nonlinear Cyber-Physical Systems: Path Integral Approach," *to be submitted*

Today's Talk: Overview



Outline

Research Motivation and Prior Work

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

[2] 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 $\Rightarrow$ applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- Not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][2].

[2] 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 $\Rightarrow$ applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- Not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][2].
- 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

[2] 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 $\Rightarrow$ applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- Not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][2].
- 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
- Limitations of [Yoon 2022]:

[2] 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 $\Rightarrow$ applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- Not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][2].
- 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
- Limitations of [Yoon 2022]:
  - The effect of time discretization is not addressed.

[2] 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 $\Rightarrow$ applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- Not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][2].
- 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
- 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 an end-to-end (trajectory-level) error bound.

[2] 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 $\Rightarrow$ applying path integral controller to safety-critical systems would require rigorous sample complexity analysis.
- Not enough work has been done on the sample complexity analysis of path integral control except the work by [Yoon 2022][2].
- 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
- 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 an end-to-end (trajectory-level) error bound.
  - The work does not compute the required sample size to achieve an acceptable loss of control performance.

[2] 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.

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

Stochastic LQR: Classical Solution

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

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

Stochastic LQR: Classical Solution

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

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

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

- 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

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

- 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

- 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: Path Integral Solution

- 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 the state-dependent 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: Path Integral Solution

- 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 the state-dependent 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: Path Integral Solution

- 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 the state-dependent path cost of each sample path i:

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

- 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

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

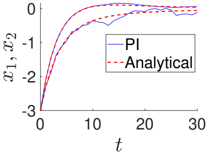
and $\hat{E}_t^r > \sqrt{\frac{1}{2n_t}\log\frac{2}{\alpha_t}}$, then $\|\hat{u} - u\|_\infty^2 := \sum_{t=0}^{T-1} \|\hat{u}_t - u_t\|_\infty^2 \leq \epsilon$ with probability greater than or equal to $1 - \alpha - \beta$.

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

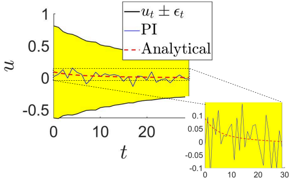
Outline

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.



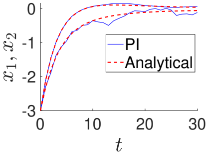
(a) State trajectory with $n = 10^3$



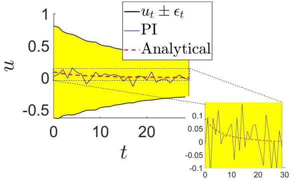
(b) Control input trajectory with $n = 10^3$

Example

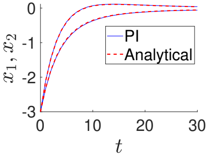
LQR problem: $A_t = \begin{bmatrix} 0.9 & -0.1 \\ -0.1 & 0.8 \end{bmatrix}$, $B_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\Omega_t = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$, $M_t = 0.1I$, $N_t = 10$. I represents an identity matrix of size 2×2.



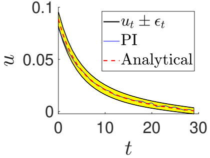
(a) State trajectory with $n = 10^3$



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

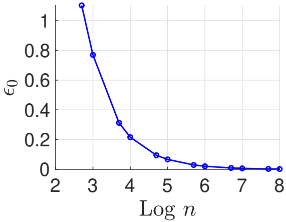


(c) State trajectory with $n = 10^7$

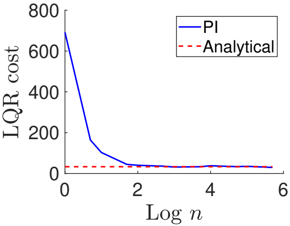


(d) Control input trajectory with $n = 10^7$

Example



(a) ϵ_0 vs sample size



(b) LQR costs vs sample size

Figure: ϵ_0 and LQR cost vs sample size

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.
- Future work:
  - Build upon this work to carry out sample complexity analysis of path integral for nonlinear continuous-time stochastic control problems.

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.
- Future work:
  - Build upon this work to carry out sample complexity analysis of path integral for nonlinear continuous-time stochastic control problems.
- Publication: A. Patil, G. Hanasusanto, T. Tanaka, “Discrete-Time Stochastic LQR via Path Integral Control and Its Sample Complexity Analysis," *IEEE Control Systems Letters (L-CSS)*

Publications

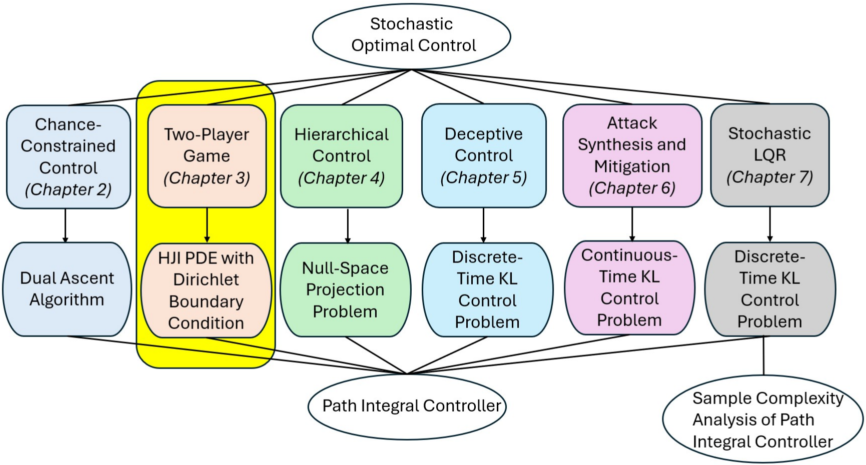
Journal Publications

- A. Patil, G. Hanasusanto, T. Tanaka, “Discrete-Time Stochastic LQR via Path Integral Control and Its Sample Complexity Analysis," *IEEE Control Systems Letters (L-CSS)*
- A. Patil, A. Duarte, F. Bisetti, T. Tanaka, “Strong Duality and Dual Ascent Approach to Continuous-Time Chance-Constrained Stochastic Optimal Control," *submitted to Transactions on Automatic Control*
- A. Patil, K. Morgenstein, L.Sentis, T. Tanaka, “Stealthy Attack Synthesis and Its Mitigation for Nonlinear Cyber-Physical Systems: Path Integral Approach," *to be submitted*
- M. Baglioni, A. Patil, L. Sentis, A. Jamshidnejad “Achieving Multi-UAV Best Viewpoint Coordination in Obstructed Environments," *to be submitted*

Conference Publications

- A. Patil, R. Funada, T. Tanaka, L. Sentis, “Task Hierarchical Control via Null-Space Projection and Path Integral Approach," *2025 American Control Conference (ACC)*
- A. Patil, G. Hanasusanto, T. Tanaka, “Discrete-Time Stochastic LQR via Path Integral Control and Its Sample Complexity Analysis," *2024 IEEE Conference on Decision and Control (CDC)*
- A. Patil*, M. Karabag*, U. Topcu, T. Tanaka, “Simulation-Driven Deceptive Control via Path Integral Approach," *2023 IEEE Conference on Decision and Control (CDC)*
- A. Patil, Y. Zhou, D. Fridovich-Keil, T. Tanaka, “Risk-Minimizing Two-Player Zero-Sum Stochastic Differential Game via Path Integral Control," *2023 IEEE Conference on Decision and Control (CDC)*
- A. Patil, T. Tanaka, “Upper and Lower Bounds for End-to-End Risks in Stochastic Robot Navigation," *2023 IFAC World Congress*
- A. Patil, A. Duarte, A. Smith, F. Bisetti, T. Tanaka, “Chance-Constrained Stochastic Optimal Control via Path Integral and Finite Difference Methods," *2022 IEEE Conference on Decision and Control (CDC)*
- A. Patil, T. Tanaka, “Upper Bounds for Continuous-Time End-to-End Risks in Stochastic Robot Navigation," *2022 European Control Conference (ECC)*
- C. Martin A. Patil, W. Li, T. Tanaka, D. Chen, “Model Predictive Path Integral Control for Roll-to-Roll Manufacturing," *to be submitted*

Two-Player Zero-Sum Game

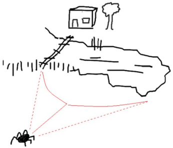


Zero-Sum Game Stochastic Differential Game (SDG)

- Control using an uncertain actuator:

$$dx(t) = f(x(t),t)dt + G(x(t),t)\Big(\underbrace{u(x(t),t)dt + v(x(t),t)dt + dw(t)}_{\text{Uncertain control input}}\Big)$$

- $v(x(t),t)$: Non-stochastic uncertainty: unmodeled bias, fatigue. It is reasonable to assume v is bounded but the control designer should assume the most pessimistic scenario.



- $w(t)$: Stochastic uncertainty
- Control designer wants to minimize $\mathbb{E}_{x_0,t_0}\left[\phi(x(t_f)) + \int_{t_0}^{t_f}\left(\frac{1}{2}u^\top R_u u + V\right)dt\right]$ under the presence of v and w.
- Zero-sum SDG

$$\min_u \max_v \mathbb{E}_{x_0,t_0}\left[\phi(x(t_f)) + \int_{t_0}^{t_f}\left(\frac{1}{2}u^\top R_u u - \frac{1}{2}v^\top R_v v + V\right)dt\right]$$
$$\text{s.t.}\quad dx = fdt + G_u udt + G_v vdt + \Sigma dw.$$

Zero-Sum Game Stochastic Differential Game (SDG)

- Our Contributions:
  - We convert the problem of two-player zero-sum SDG as a problem of solving a Hamilton-Jacobi-Isaacs (HJI) PDE with Dirichlet boundary condition.

Zero-Sum Game Stochastic Differential Game (SDG)

- Our Contributions:
  - We convert the problem of two-player zero-sum SDG as a problem of solving a Hamilton-Jacobi-Isaacs (HJI) PDE with Dirichlet boundary condition.
  - We develop a path-integral framework to solve the HJI PDE and establish the existence and uniqueness of the saddle point solution (optimal solution of the game).

Zero-Sum Game Stochastic Differential Game (SDG)

- Our Contributions:
  - We convert the problem of two-player zero-sum SDG as a problem of solving a Hamilton-Jacobi-Isaacs (HJI) PDE with Dirichlet boundary condition.
  - We develop a path-integral framework to solve the HJI PDE and establish the existence and uniqueness of the saddle point solution (optimal solution of the game).
  - We obtain explicit expressions for the saddle-point policies which can be numerically evaluated using Monte Carlo simulations.

Zero-Sum Game Stochastic Differential Game (SDG)

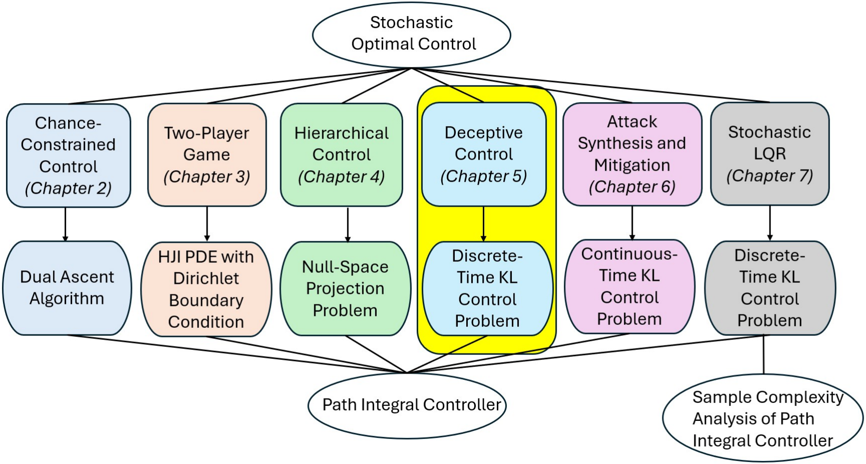
- Our Contributions:
  - We convert the problem of two-player zero-sum SDG as a problem of solving a Hamilton-Jacobi-Isaacs (HJI) PDE with Dirichlet boundary condition.
  - We develop a path-integral framework to solve the HJI PDE and establish the existence and uniqueness of the saddle point solution (optimal solution of the game).
  - We obtain explicit expressions for the saddle-point policies which can be numerically evaluated using Monte Carlo simulations.
  - Our approach allows the game to be solved online without the need for any offline training or precomputations.

Zero-Sum Game Stochastic Differential Game (SDG)

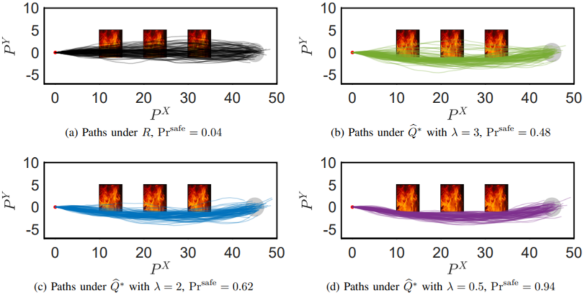
- Our Contributions:
  - We convert the problem of two-player zero-sum SDG as a problem of solving a Hamilton-Jacobi-Isaacs (HJI) PDE with Dirichlet boundary condition.
  - We develop a path-integral framework to solve the HJI PDE and establish the existence and uniqueness of the saddle point solution (optimal solution of the game).
  - We obtain explicit expressions for the saddle-point policies which can be numerically evaluated using Monte Carlo simulations.
  - Our approach allows the game to be solved online without the need for any offline training or precomputations.

  - Publication
    A. Patil, Y. Zhou, D. Fridovich-Keil, T. Tanaka, "Risk-Minimizing Two-Player Zero-Sum Stochastic Differential Game via Path Integral Control," *2023 IEEE Conference on Decision and Control (CDC)*

Deceptive Control



Optimal Deception by Path Integral Control



(a) Paths under R, $\Pr^{\text{safe}} = 0.04$

(b) Paths under $\widehat{Q}^*$ with $\lambda = 3$, $\Pr^{\text{safe}} = 0.48$

(c) Paths under $\widehat{Q}^*$ with $\lambda = 2$, $\Pr^{\text{safe}} = 0.62$

(d) Paths under $\widehat{Q}^*$ with $\lambda = 0.5$, $\Pr^{\text{safe}} = 0.94$

- Problem Setup
  - A supervisor wants an agent to reach the target as soon as possible (reference policy)
  - The agent, on the other hand, wishes to avoid the regions covered under fire (deviated policy)
  - How can the agent satisfy their own interest by deviating from the reference policy without being detected by the supervisor?

Our Contributions

- We formalize the synthesis of an optimal deceptive policy as a KL control problem. We introduce KL divergence as a stealthiness measure using motivations from hypothesis testing theory.

$$\min_{Q} \mathbb{E}_Q \sum_{t=0}^{T} C_t(X_t, U_t) + \lambda D(Q||R)$$

where R is the reference policy and Q is the deviated policy.

Our Contributions

- We formalize the synthesis of an optimal deceptive policy as a KL control problem. We introduce KL divergence as a stealthiness measure using motivations from hypothesis testing theory.

$$\min_{Q} \mathbb{E}_Q \sum_{t=0}^{T} C_t(X_t, U_t) + \lambda D(Q||R)$$

  where R is the reference policy and Q is the deviated policy.
- We solve the KL control problem using backward dynamic programming. Since dynamic programming suffers from the curse of dimensionality, we develop an algorithm based on path integral control to numerically compute the optimal deceptive actions online using Monte Carlo simulations without explicitly synthesizing the policy.

Our Contributions

- We formalize the synthesis of an optimal deceptive policy as a KL control problem. We introduce KL divergence as a stealthiness measure using motivations from hypothesis testing theory.

$$\min_{Q} \mathbb{E}_Q \sum_{t=0}^{T} C_t(X_t, U_t) + \lambda D(Q||R)$$

  where R is the reference policy and Q is the deviated policy.

- We solve the KL control problem using backward dynamic programming. Since dynamic programming suffers from the curse of dimensionality, we develop an algorithm based on path integral control to numerically compute the optimal deceptive actions online using Monte Carlo simulations without explicitly synthesizing the policy.
- We show that our proposed algorithm asymptotically converges to the optimal action distribution of the deceptive agent as the number of samples goes to infinity.

Our Contributions

- We formalize the synthesis of an optimal deceptive policy as a KL control problem. We introduce KL divergence as a stealthiness measure using motivations from hypothesis testing theory.

$$\min_{Q} \mathbb{E}_Q \sum_{t=0}^{T} C_t(X_t, U_t) + \lambda D(Q||R)$$

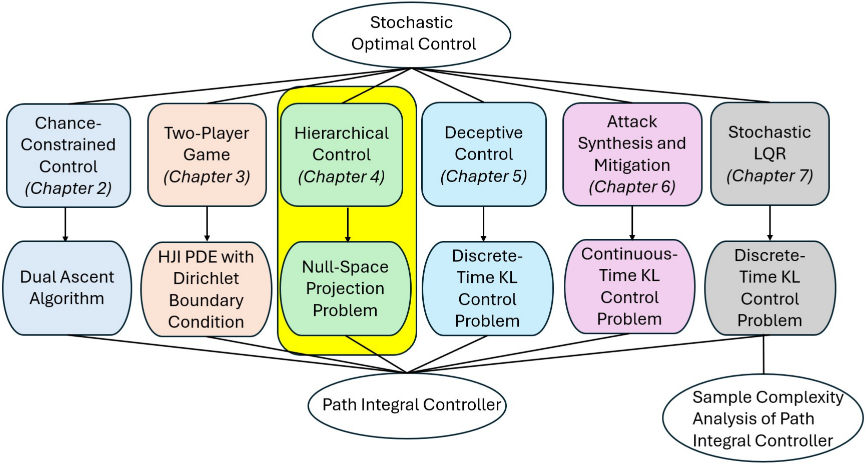
  where R is the reference policy and Q is the deviated policy.

- We solve the KL control problem using backward dynamic programming. Since dynamic programming suffers from the curse of dimensionality, we develop an algorithm based on path integral control to numerically compute the optimal deceptive actions online using Monte Carlo simulations without explicitly synthesizing the policy.

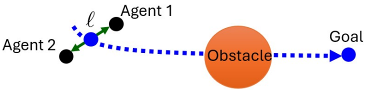
- We show that our proposed algorithm asymptotically converges to the optimal action distribution of the deceptive agent as the number of samples goes to infinity.

- Publication:
  A. Patil*, M. Karabag*, U. Topcu, T. Tanaka, “Simulation-Driven Deceptive Control via Path Integral Approach," *2023 IEEE Conference on Decision and Control (CDC)*

Hierarchical Control

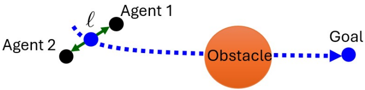


Conventional Task Hierarchical Control

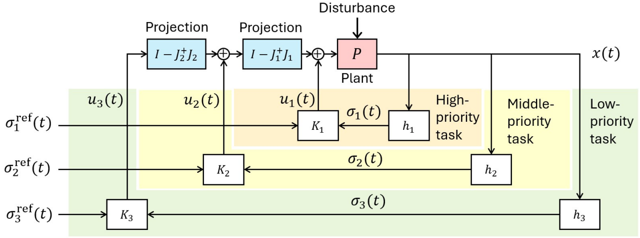


- Task 1: Avoid collisions with obstacles
- Task 2: Steer the platoon's centroid towards a goal position
- Task 3: Maintain specific distances between the agents

Conventional Task Hierarchical Control

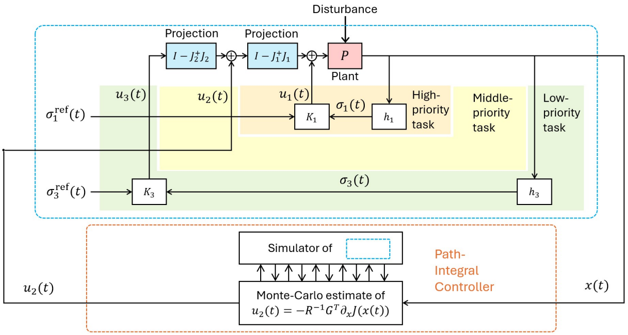


- Task 1: Avoid collisions with obstacles
- Task 2: Steer the platoon's centroid towards a goal position
- Task 3: Maintain specific distances between the agents



- Simple controllers (such as PID) are used for K_i to achieve reference tracking in task coordinate $\sigma_i(t)$
- Reference signals $\sigma_i^{\mathrm{ref}}(t)$ are often chosen manually.

Task Hierarchical Control via Path Integral Method



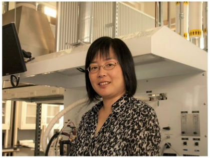
- Path integral controller seeks the optimal input for some of the tasks, while rudimentary controllers can be kept for other tasks.
- Manuscript:
  A. Patil, R. Funada, T. Tanaka, L. Sentis, “Task Hierarchical Control via Null-Space Projection and Path Integral Approach," *American Control Conference (ACC) 2025*

Acknowledgments

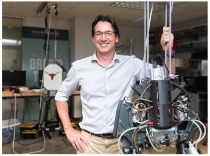

Takashi Tanaka

Acknowledgments

Takashi Tanaka



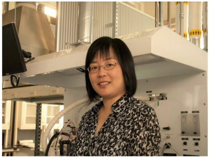
Maggie Chen



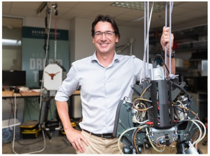
Luis Sentis

Acknowledgments

Takashi Tanaka



Maggie Chen

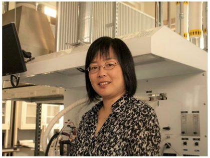


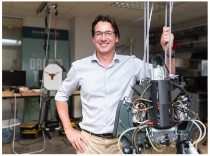
Luis Sentis

A big thanks to the rest of the PhD committee!

Acknowledgments

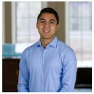

Takashi Tanaka


Maggie Chen


Luis Sentis

A big thanks to the rest of the PhD committee!


Grani Hanasusanto

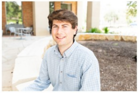

Alfredo Duarte

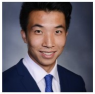

Fabrizio Bisetti


Mustafa Karabag


Ufuk Topcu


Riku Funada


David Fridovich-Keil


Yujing Zhou


Mirko Baglioni


Anahita Jamshidnejad


Aislinn Smith


Kyle Morgenstein


Christopher Martin


Wei Li

Acknowledgments

Questions?

apurvapatil@utexas.edu
Google Scholar | Website | LinkedIn | Github