IEEE Conference on Decision and Control 2023

Simulator-Driven Deceptive Control via Path Integral Approach

Apurva Patil* Mustafa O. Karabag*
Takashi Tanaka Ufuk Topcu

* indicates equal contribution.

Outline

- Background
- Existing Approaches
- Problem Formulation
- Synthesis of Optimal Deceptive Policies
- Simualtions
- Conclusion

Outline

Background

Existing Approaches

Problem Formulation

Synthesis of Optimal Deceptive Policies

Simualtions

Conclusion

Background

- A supervisor delegates an agent to perform a certain control task
- The agent is incentivized to deviate from the supervisor's policy to achieve its own goal
- Drone surveillance example

Background

- Synthesis of the optimal deceptive policies for an agent who attempts to hide its deviations from the supervisor's policy - KL control problem

Background

- Synthesis of the optimal deceptive policies for an agent who attempts to hide its deviations from the supervisor's policy - KL control problem
- Minimizing the KL divergence - minimizing the detection rate of the supervisor (log-likelihood ratio test, B-H inequality)

Background

- Synthesis of the optimal deceptive policies for an agent who attempts to hide its deviations from the supervisor's policy - KL control problem
- Minimizing the KL divergence - minimizing the detection rate of the supervisor (log-likelihood ratio test, B-H inequality)
- Nonlinear discrete-time continuous-state dynamics, arbitrary cost functions and reference policies

Background

- Synthesis of the optimal deceptive policies for an agent who attempts to hide its deviations from the supervisor's policy - KL control problem
- Minimizing the KL divergence - minimizing the detection rate of the supervisor (log-likelihood ratio test, B-H inequality)
- Nonlinear discrete-time continuous-state dynamics, arbitrary cost functions and reference policies
- Path integral control - simulator driven control synthesis framework

Background

- Synthesis of the optimal deceptive policies for an agent who attempts to hide its deviations from the supervisor's policy - KL control problem
- Minimizing the KL divergence - minimizing the detection rate of the supervisor (log-likelihood ratio test, B-H inequality)
- Nonlinear discrete-time continuous-state dynamics, arbitrary cost functions and reference policies
- Path integral control - simulator driven control synthesis framework
- The optimal deceptive policies can be numerically computed online using Monte Carlo sampling

Outline

Existing Approaches

- Deception framework [Shim et al. 2013, Wang et al. 2018]

Existing Approaches

- Deception framework [Shim et al. 2013, Wang et al. 2018]
- Deception problem in supervisory control for discrete-state systems [Karabag et al. 2021, Keroglou et al. 2018]

Existing Approaches

- Deception framework [Shim et al. 2013, Wang et al. 2018]
- Deception problem in supervisory control for discrete-state systems [Karabag et al. 2021, Keroglou et al. 2018]
- Detectability of an attacker that aims at maximizing the state estimation error of a controller using KL-divergence-based optimization problem [Bai et al. 2017, Kung et at. 2016]

Existing Approaches

- Deception framework [Shim et al. 2013, Wang et al. 2018]
- Deception problem in supervisory control for discrete-state systems [Karabag et al. 2021, Keroglou et al. 2018]
- Detectability of an attacker that aims at maximizing the state estimation error of a controller using KL-divergence-based optimization problem [Bai et al. 2017, Kung et at. 2016]
- KL control problem [Todorov 2007, Ito 2022]

Existing Approaches

- Deception framework [Shim et al. 2013, Wang et al. 2018]
- Deception problem in supervisory control for discrete-state systems [Karabag et al. 2021, Keroglou et al. 2018]
- Detectability of an attacker that aims at maximizing the state estimation error of a controller using KL-divergence-based optimization problem [Bai et al. 2017, Kung et at. 2016]
- KL control problem [Todorov 2007, Ito 2022]
- Path integral control [Kappen 2005, Theodorou 2010]: a sampling-based algorithm to solve nonlinear stochastic optimal control problems, less susceptible to the curse of dimensionality

Outline

Problem Formulation

- Agent's dynamics: $P(dx_{t+1}|x_t, u_t)$

Problem Formulation

- Agent's dynamics: $P(dx_{t+1}|x_t, u_t)$
- Supervisor's policy: $\{R_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$

Problem Formulation

- Agent's dynamics: $P(dx_{t+1}|x_t, u_t)$
- Supervisor's policy: $\{R_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$
- Path cost: $C_{0:T}(x_{0:T}, u_{0:T-1}) \coloneqq \sum_{t=0}^{T-1} C_t(x_t, u_t) + C_T(x_T)$

Problem Formulation

- Agent's dynamics: $P(dx_{t+1}|x_t, u_t)$
- Supervisor's policy: $\{R_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$
- Path cost: $C_{0:T}(x_{0:T}, u_{0:T-1}) := \sum_{t=0}^{T-1} C_t(x_t, u_t) + C_T(x_T)$
- Agent's policy: $\{Q_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$

Problem Formulation

- Agent's dynamics: $P(dx_{t+1}|x_t, u_t)$
- Supervisor's policy: $\{R_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$
- Path cost: $C_{0:T}(x_{0:T}, u_{0:T-1}) \coloneqq \sum_{t=0}^{T-1} C_t(x_t, u_t) + C_T(x_T)$
- Agent's policy: $\{Q_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$
- Distributions of the state-action paths under Q and R:

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

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

Problem Formulation

- Agent's dynamics: $P(dx_{t+1}|x_t, u_t)$
- Supervisor's policy: $\{R_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$
- Path cost: $C_{0:T}(x_{0:T}, u_{0:T-1}) \coloneqq \sum_{t=0}^{T-1} C_t(x_t, u_t) + C_T(x_T)$
- Agent's policy: $\{Q_{U_t|X_t}(\cdot|x_t)\}_{t=0}^{T-1}$
- Distributions of the state-action paths under Q and R:

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

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

- Log likelihood ratio (LLR):

$$\pi(x_{0:T}, u_{0:T-1}) = \log \frac{dQ_{X_{0:T}\times U_{0:T-1}}}{dR_{X_{0:T}\times U_{0:T-1}}}(x_{0:T}, u_{0:T-1})$$

Problem Formulation

- Expected LLR: $\Pi = \mathbb{E}_Q \left[\log \frac{dQ_{X_{0:T} \times U_{0:T-1}}}{dR_{X_{0:T} \times U_{0:T-1}}} (x_{0:T}, u_{0:T-1}) \right]$

Problem Formulation

- Expected LLR: $\Pi = \mathbb{E}_Q \left[\log \frac{dQ_{X_{0:T} \times U_{0:T-1}}}{dR_{X_{0:T} \times U_{0:T-1}}}(x_{0:T}, u_{0:T-1}) \right]$
- KL divergence:
  $\Pi = D(Q \| R)$

Problem Formulation

- Expected LLR: $\Pi = \mathbb{E}_Q \left[\log \frac{dQ_{X_{0:T} \times U_{0:T-1}}}{dR_{X_{0:T} \times U_{0:T-1}}} (x_{0:T}, u_{0:T-1}) \right]$
- KL divergence:
  $\Pi = D(Q \| R) = \mathbb{E}_Q \left[\sum_{t=0}^{T-1} D(Q_{U_t|X_t}(\cdot|X_t) \| R_{U_t|X_t}(\cdot|X_t)) \right]$

Problem Formulation

- Expected LLR: $\Pi = \mathbb{E}_Q\left[\log \frac{dQ_{X_{0:T}\times U_{0:T-1}}}{dR_{X_{0:T}\times U_{0:T-1}}}(x_{0:T}, u_{0:T-1})\right]$
- KL divergence:
  $\Pi = D(Q\|R) = \mathbb{E}_Q\left[\sum_{t=0}^{T-1} D(Q_{U_t|X_t}(\cdot|X_t)\|R_{U_t|X_t}(\cdot|X_t))\right]$
- B-H inequality: $\Pr(\mathcal{E}|R) + \Pr(\neg\mathcal{E}|Q) \geq \frac{1}{2}\exp(-D(Q||R))$

Problem Formulation

- Expected LLR: $\Pi = \mathbb{E}_Q\left[\log \frac{dQ_{X_{0:T}\times U_{0:T-1}}}{dR_{X_{0:T}\times U_{0:T-1}}}(x_{0:T}, u_{0:T-1})\right]$
- KL divergence:
  $\Pi = D(Q\|R) = \mathbb{E}_Q\left[\sum_{t=0}^{T-1} D(Q_{U_t|X_t}(\cdot|X_t)\|R_{U_t|X_t}(\cdot|X_t))\right]$
- B-H inequality: $\Pr(\mathcal{E}|R) + \Pr(\neg\mathcal{E}|Q) \geq \frac{1}{2}\exp(-D(Q||R))$
- Synthesis of optimal deceptive policy:

$$\min_{\{Q_{U_t|X_t}\}_{t=0}^{T-1}} \mathbb{E}_Q \sum_{t=0}^{T-1} \Big\{ C_t(X_t, U_t) + \lambda D(Q_{U_t|X_t}(\cdot|X_t)\|R_{U_t|X_t}(\cdot|X_t)) \Big\} + \mathbb{E}_Q C_T(X_T)$$

where λ is a positive weighting factor that balances the trade-off between the KL divergence and the path cost.

Outline

Synthesis of Optimal Policies: Backward DP

- Define for each $t \in \mathcal{T}$ and $x_t \in \mathcal{X}_t$, the value function:

$$J_t(x_t) := \min_{\{Q_{U_k|X_k}\}_{k=t}^{T-1}} \mathbb{E}_Q \sum_{k=t}^{T-1} \Big\{ C_k(X_k, U_k) + \lambda D(Q_{U_k|X_k}(\cdot|X_k) \| R_{U_k|X_k}(\cdot|X_k)) \Big\} + \mathbb{E}_Q C_T(X_T).$$

Synthesis of Optimal Policies: Backward DP

- Define for each $t \in \mathcal{T}$ and $x_t \in \mathcal{X}_t$, the value function:

$$J_t(x_t) := \min_{\{Q_{U_k|X_k}\}_{k=t}^{T-1}} \mathbb{E}_Q \sum_{k=t}^{T-1} \Big\{ C_k(X_k, U_k) + \lambda D(Q_{U_k|X_k}(\cdot|X_k) \| R_{U_k|X_k}(\cdot|X_k)) \Big\} + \mathbb{E}_Q C_T(X_T).$$

- Theorem 1: $J_t(x_t)$ satisfies the backward Bellman recursion with the terminal condition $J_T(x_T) = C_T(x_T)$:

$$J_t(x_t) = -\lambda \log \Big\{ \int_{\mathcal{U}_t} \exp\left(-\frac{C_t(x_t,u_t)}{\lambda}\right) \times \exp\left(-\frac{1}{\lambda}\int_{\mathcal{X}_{t+1}} J_{t+1}(x_{t+1}) P(dx_{t+1}|x_t,u_t)\right) R(du_t|x_t) \Big\} \ldots$$

Synthesis of Optimal Policies: Backward DP

...and the minimizer is given by

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

where $\rho_t(x_t, u_t) := C_t(x_t, u_t) + \int_{\mathcal{X}_{t+1}} J_{t+1}(x_{t+1}) P(dx_{t+1}|x_t, u_t)$ and B_{U_t} is a Borel set belonging to the $\sigma-$algebra $\mathcal{B}(\mathcal{U}_t)$.

Synthesis of Optimal Policies: Backward DP

...and the minimizer is given by

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

where $\rho_t(x_t, u_t) := C_t(x_t, u_t) + \int_{\mathcal{X}_{t+1}} J_{t+1}(x_{t+1}) P(dx_{t+1}|x_t, u_t)$ and B_{U_t} is a Borel set belonging to the $\sigma-$algebra $\mathcal{B}(\mathcal{U}_t)$.

- Recursive method to compute $J_t(x_t)$ and $Q^*_{U_t|X_t}$ backward in time.

Synthesis of Optimal Policies: Backward DP

...and the minimizer is given by

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

where $\rho_t(x_t, u_t) := C_t(x_t, u_t) + \int_{\mathcal{X}_{t+1}} J_{t+1}(x_{t+1}) P(dx_{t+1}|x_t, u_t)$ and B_{U_t} is a Borel set belonging to the $\sigma-$algebra $\mathcal{B}(\mathcal{U}_t)$.

- Recursive method to compute $J_t(x_t)$ and $Q^*_{U_t|X_t}$ backward in time.
- Suffers from the curse of dimensionality.

Synthesis of Optimal Policies: Path Integral Control

- Assumption: The state transition law is governed by a deterministic mapping $F_t : \mathcal{X}_t \times \mathcal{U}_t \to \mathcal{X}_{t+1}$ as $x_{t+1} = F_t(x_t, u_t)$.

Synthesis of Optimal Policies: Path Integral Control

- Assumption: The state transition law is governed by a deterministic mapping $F_t : \mathcal{X}_t \times \mathcal{U}_t \to \mathcal{X}_{t+1}$ as $x_{t+1} = F_t(x_t, u_t)$.
- Value function is recursively defined as

$$J_t(x_t) = -\lambda \log \left\{ \int_{\mathcal{U}_t} \exp\left(-\frac{C_t(x_t, u_t)}{\lambda}\right) \times \exp\left(-\frac{J_{t+1}\left(F_t(x_t, u_t)\right)}{\lambda}\right) R(du_t|x_t) \right\}.$$

Synthesis of Optimal Policies: Path Integral Control

- Exponentiated value function as $Z_t(x_t) := \exp\left(-\frac{1}{\lambda} J_t(x_t)\right)$

Synthesis of Optimal Policies: Path Integral Control

- Exponentiated value function as $Z_t(x_t) := \exp\left(-\frac{1}{\lambda} J_t(x_t)\right)$
- Linear recursion:

$$Z_t(x_t) = \int_{\mathcal{U}_t} \int_{\mathcal{X}_{t+1}} \exp\left(-\frac{C_t(x_t, u_t)}{\lambda}\right) Z_{t+1}(x_{t+1}) \times P(dx_{t+1} | x_t, u_t) R(du_t | x_t).$$

where $P(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t,u_t)}(dx_{t+1})$.

Synthesis of Optimal Policies: Path Integral Control

- Exponentiated value function as $Z_t(x_t) := \exp\left(-\frac{1}{\lambda} J_t(x_t)\right)$
- Linear recursion:

$$Z_t(x_t) = \int_{\mathcal{U}_t} \int_{\mathcal{X}_{t+1}} \exp\left(-\frac{C_t(x_t, u_t)}{\lambda}\right) Z_{t+1}(x_{t+1}) \times P(dx_{t+1}|x_t, u_t) R(du_t|x_t).$$

  where $P(dx_{t+1}|x_t, u_t) = \delta_{F_t(x_t, u_t)}(dx_{t+1})$.

- By recursive substitution:

$$Z_t(x_t) = \int_{\mathcal{U}_t} \int_{\mathcal{X}_{t+1}} \cdots \int_{\mathcal{U}_{T-1}} \int_{\mathcal{X}_T} \exp\left(-\frac{C_t(x_t, u_t)}{\lambda}\right) \times \cdots \times \exp\left(-\frac{C_T(x_T)}{\lambda}\right) R(dx_{t+1:T} \times du_{t:T-1}|x_t).$$

Synthesis of Optimal Policies: Path Integral Control

- Introducing the path cost function
$C_{t:T}(x_{t:T}, u_{t:T-1}) := \sum_{k=t}^{T-1} C_k(x_k, u_k) + C_T(x_T)$,

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

Synthesis of Optimal Policies: Path Integral Control

- Introducing the path cost function
  $C_{t:T}(x_{t:T}, u_{t:T-1}) := \sum_{k=t}^{T-1} C_k(x_k, u_k) + C_T(x_T)$,

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

- Numerical computation of $Z_t(x_t)$:
  Sample N independent paths $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^N$ under the distribution R. If $C_{t:T}(x_{t:T}(i), u_{t:T-1}(i))$ represents the path cost of the sample path i, then as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} \exp\left(-\frac{1}{\lambda} C_{t:T}(x_{t:T}(i), u_{t:T-1}(i))\right) \stackrel{a.s.}{\to} Z_t(x_t).$$

Synthesis of Optimal Policies: Path Integral Control

- It can be shown that

$$Q^*_{U_t|X_t}(B_{U_t}|x_t) = \frac{1}{Z_t(x_t)} \int_{\{\mathcal{X}_{t+1:T}, \mathcal{U}_{t:T-1} | u_t \in B_{U_t}\}} \exp\left(-\frac{C_{t:T}(x_{t:T}, u_{t:T-1})}{\lambda}\right) \times R(dx_{t+1:T} \times du_{t:T-1}|x_t).$$

Synthesis of Optimal Policies: Path Integral Control

- It can be shown that

$$Q^*_{U_t|X_t}(B_{U_t}|x_t) = \frac{1}{Z_t(x_t)} \int_{\{\mathcal{X}_{t+1:T}, \mathcal{U}_{t:T-1} | u_t \in B_{U_t}\}} \exp\left(-\frac{C_{t:T}(x_{t:T}, u_{t:T-1})}{\lambda}\right) \times R(dx_{t+1:T} \times du_{t:T-1}|x_t).$$

- Sampling u_t approximately from $Q^*_{U_t|X_t}(\cdot|x_t)$ by Monte Carlo simulations:

Synthesis of Optimal Policies: Path Integral Control

- It can be shown that

$$Q^*_{U_t|X_t}(B_{U_t}|x_t) = \frac{1}{Z_t(x_t)} \int_{\{\mathcal{X}_{t+1:T}, \mathcal{U}_{t:T-1} | u_t \in B_{U_t}\}} \exp\left(-\frac{C_{t:T}(x_{t:T}, u_{t:T-1})}{\lambda}\right) \times R(dx_{t+1:T} \times du_{t:T-1}|x_t).$$

- Sampling u_t approximately from $Q^*_{U_t|X_t}(\cdot|x_t)$ by Monte Carlo simulations:
  - Sample N independent paths $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^N$ under the distribution R.

Synthesis of Optimal Policies: Path Integral Control

- It can be shown that

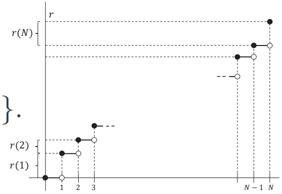
$$Q^*_{U_t|X_t}(B_{U_t}|x_t) = \frac{1}{Z_t(x_t)} \int_{\{\mathcal{X}_{t+1:T}, \mathcal{U}_{t:T-1}|u_t \in B_{U_t}\}} \exp\left(-\frac{C_{t:T}(x_{t:T}, u_{t:T-1})}{\lambda}\right) \times R(dx_{t+1:T} \times du_{t:T-1}|x_t).$$

- Sampling u_t approximately from $Q^*_{U_t|X_t}(\cdot|x_t)$ by Monte Carlo simulations:
  - Sample N independent paths $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^N$ under the distribution R.
  - Let $r_t(i) = C_{t:T}(x_{t:T}(i), u_{t:T-1}(i))$ represents the path cost of the sample path i and $r_t := \sum_{i=1}^N r_t(i)$.

Synthesis of Optimal Policies: Path Integral Control

- For each $t \in \mathcal{T}$, define

$$F_t(x) = \sum_{i=1}^{\lfloor x \rfloor} r_t(i), \quad F_t^{-1} : [0, r_t] \to \{1, 2, \dots, N\}.$$



$F_t(x)$

---

for $t \in \mathcal{T}$ **do**

Sample N paths $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{N}$ starting from x_t under the reference distribution R.

Compute $r_t(i)$ and $r_t := \sum_{i=1}^{N} r_t(i)$

Generate $d \sim \mathrm{unif}[0, r_t]$.

Select a sample ID by $j_t \leftarrow F_t^{-1}(d)$.

Select a control input as $u_t \leftarrow u_t(j_t)$.

Synthesis of Optimal Policies: Path Integral Control

- Theorem 2: Let $B_{U_t} \in \mathcal{B}(\mathcal{U}_t)$ be a Borel set. Suppose for a given collection of sample paths $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{N}$, u_t is computed by the above Algorithm and the probability of $u_t \in B_{U_t}$ is denoted by $\Pr\{u_t \in B_{U_t} | \{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{N}\}$. Then, as $N \to \infty$

$$\Pr\{u_t \in B_{U_t} | \{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{N}\} \overset{a.s.}{\to} Q^*_{U_t|X_t}(B_{U_t}|x_t).$$

Synthesis of Optimal Policies: Path Integral Control

- Theorem 2: Let $B_{U_t} \in \mathcal{B}(\mathcal{U}_t)$ be a Borel set. Suppose for a given collection of sample paths $\{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^N$, u_t is computed by the above Algorithm and the probability of $u_t \in B_{U_t}$ is denoted by $\Pr\{u_t \in B_{U_t} | \{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^N\}$. Then, as $N \to \infty$

$$\Pr\{u_t \in B_{U_t} | \{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^N\} \xrightarrow{a.s.} Q^*_{U_t|X_t}(B_{U_t}|x_t).$$

- Deceptive agent can numerically compute optimal actions via Monte Carlo simulations without explicitly synthesizing the policy.

Theorem 2: Sketch of Proof

- Let $\mathcal{I}_{B_{U_t}} = \{i \in \{1, 2, \dots, N\} | u_t(i) \in B_{U_t}\}$
- $r_{B_{U_t}} = \sum_{i \in \mathcal{I}_{B_{U_t}}} r_t(i)$.
- By construction of the Algorithm

$$\Pr\{u_t \in B_{U_t} | \{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{N}\} = \frac{r_{B_{U_t}}}{r_t}.$$

- As $N \to \infty$, $\frac{r_t}{N} \overset{a.s.}{\to} Z_t(x_t)$ and

$$\frac{r_{B_{U_t}}}{N} \overset{a.s.}{\to} \int_{\{\mathcal{X}_{t+1:T}, \mathcal{U}_{t:T-1} | u_t \in B_{U_t}\}} \exp\left(-\frac{C_{t:T}(x_{t:T}, u_{t:T-1})}{\lambda}\right) \times R(dx_{t+1:T}, du_{t:T-1} | x_t)$$

- $\Pr\{u_t \in B_{U_t} | \{x_{t:T}(i), u_{t:T-1}(i)\}_{i=1}^{N}\} \overset{a.s.}{\to} Q^*_{U_t|X_t}(B_{U_t} | x_t)$.

Outline

Simualtions

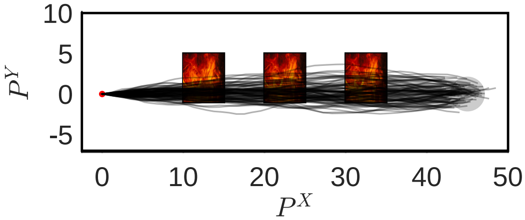


Figure: Paths under R, $\Pr^{\text{safe}} = 0.04$

- Agent's dynamics: unicycle model

$$P^X_{t+1} = P^X_t + S_t \cos \Theta_t h \qquad P^Y_{t+1} = P^Y_t + S_t \sin \Theta_t h$$
$$S_{t+1} = S_t + A_t h \qquad \Theta_{t+1} = \Theta_t + \Omega_t h$$

- Start: origin, Goal: disk of radius G^R centered at $[G^X \; G^Y]^\top$

Simualtions

- Reference policy:

$$R_{U_t|X_t}(\cdot|x_t) = \frac{\exp\left[-\frac{1}{2}(u_t - \overline{u}_t)^\top \Sigma_t^{-1}(u_t - \overline{u}_t)\right]}{\sqrt{(2\pi)^2|\Sigma_t|}},$$

where $\overline{u}_t$ is designed using a proportional controller.

Simualtions

- Reference policy:

$$R_{U_t|X_t}(\cdot|x_t) = \frac{\exp\left[-\frac{1}{2}(u_t - \overline{u}_t)^\top \Sigma_t^{-1}(u_t - \overline{u}_t)\right]}{\sqrt{(2\pi)^2|\Sigma_t|}},$$

  where $\overline{u}_t$ is designed using a proportional controller.
- Cost function:

$$C_{0:T}(X_{0:T}, U_{0:T-1}) = \sum_{t=0}^{T} \mathbb{1}_{[P_t^X \; P_t^Y]^\top \in \mathcal{X}^{\text{fire}}}$$

Simualtions

- Reference policy:

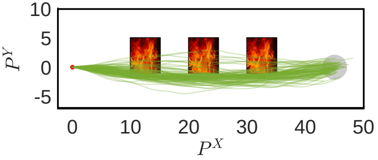
$$R_{U_t|X_t}(\cdot|x_t) = \frac{\exp\left[-\frac{1}{2}(u_t - \overline{u}_t)^\top \Sigma_t^{-1}(u_t - \overline{u}_t)\right]}{\sqrt{(2\pi)^2|\Sigma_t|}},$$

  where $\overline{u}_t$ is designed using a proportional controller.
- Cost function:

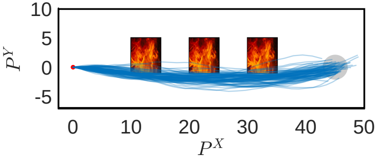
$$C_{0:T}(X_{0:T}, U_{0:T-1}) = \sum_{t=0}^{T} \mathbb{1}_{[P_t^X \ P_t^Y]^\top \in \mathcal{X}^{\mathrm{fire}}}$$

- Number of samples: 10^5

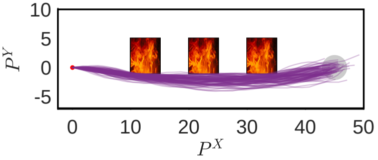
Simualtions



(a) $\lambda = 3$, $\mathrm{Pr}^{\mathrm{safe}} = 0.48$

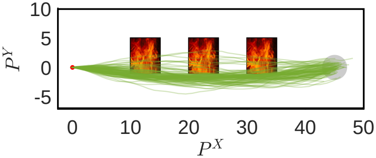


(b) $\lambda = 2$, $\mathrm{Pr}^{\mathrm{safe}} = 0.62$

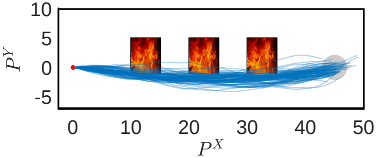


(c) $\lambda = 0.5$, $\mathrm{Pr}^{\mathrm{safe}} = 0.94$

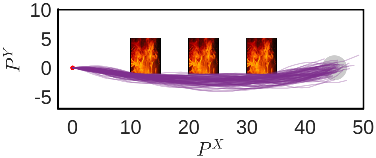
Simualtions



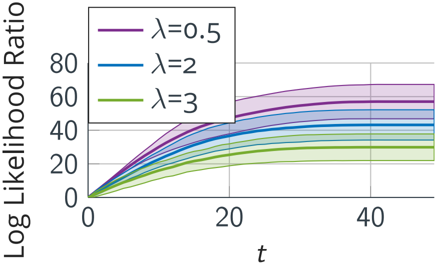
(a) $\lambda = 3$, $\Pr^{\text{safe}} = 0.48$



(b) $\lambda = 2$, $\Pr^{\text{safe}} = 0.62$



(c) $\lambda = 0.5$, $\Pr^{\text{safe}} = 0.94$



Outline

Conclusion

- Presented a deception problem under supervisory control for continuous-state discrete-time stochastic systems.

[1] Patil et al. "Sample Complexity of Discrete-Time Path-Integral Control," submitted to ECC 2024

Conclusion

- Presented a deception problem under supervisory control for continuous-state discrete-time stochastic systems.
- Formalized the synthesis of an optimal deceptive policy as a KL control problem.

[1] Patil et al. "Sample Complexity of Discrete-Time Path-Integral Control," submitted to ECC 2024

Conclusion

- Presented a deception problem under supervisory control for continuous-state discrete-time stochastic systems.
- Formalized the synthesis of an optimal deceptive policy as a KL control problem.
- Proposed a simulator-driven algorithm to compute optimal deceptive actions online via MC sampling.

[1] Patil et al. "Sample Complexity of Discrete-Time Path-Integral Control," submitted to ECC 2024

Conclusion

- Presented a deception problem under supervisory control for continuous-state discrete-time stochastic systems.
- Formalized the synthesis of an optimal deceptive policy as a KL control problem.
- Proposed a simulator-driven algorithm to compute optimal deceptive actions online via MC sampling.

Check out the paper for more details and results.

[1] Patil et al. "Sample Complexity of Discrete-Time Path-Integral Control," submitted to ECC 2024

Conclusion

- Presented a deception problem under supervisory control for continuous-state discrete-time stochastic systems.
- Formalized the synthesis of an optimal deceptive policy as a KL control problem.
- Proposed a simulator-driven algorithm to compute optimal deceptive actions online via MC sampling.

Check out the paper for more details and results.

- Future work:
  - Deception problem for continuous-time stochastic systems.
  - Sample complexity analysis of path integral approach to solve KL control problems[1].

[1] Patil et al. "Sample Complexity of Discrete-Time Path-Integral Control," submitted to ECC 2024