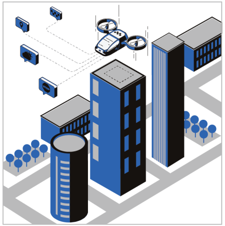
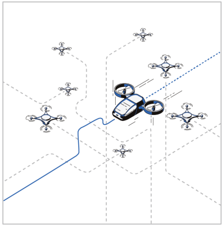
LP Solutions for Stochastic Optimal Control Problems via Hilbert Space Embeddings of Distributions

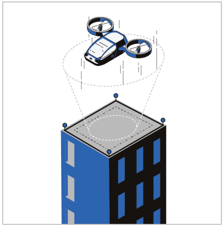
Adam Thorpe

University of New Mexico

2022

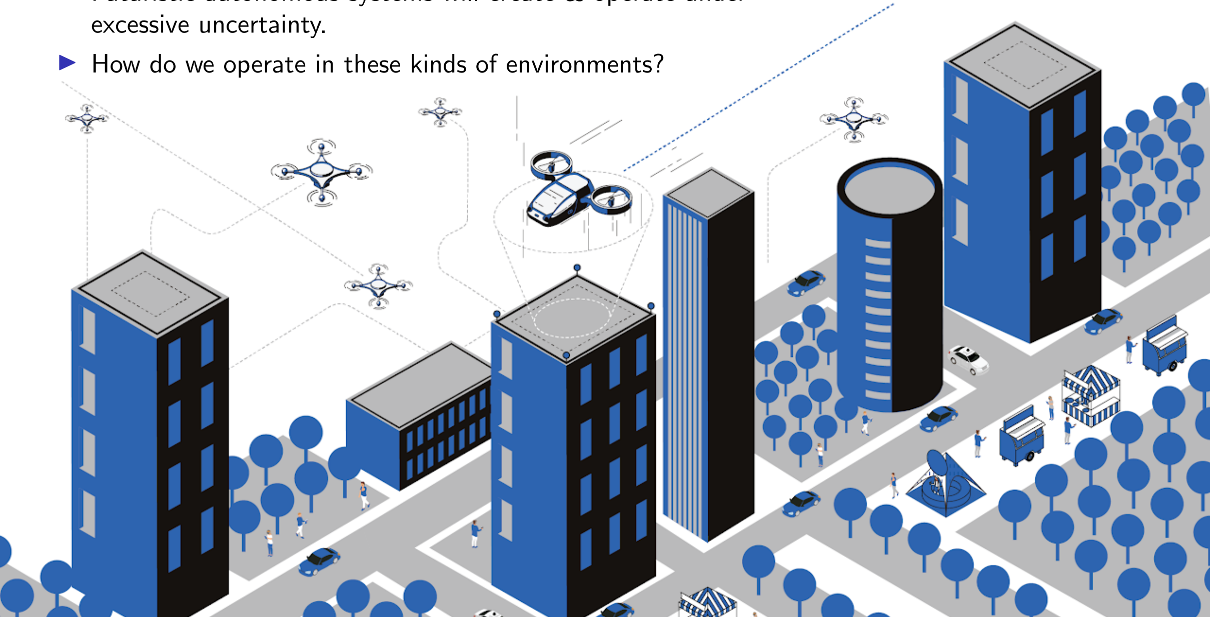




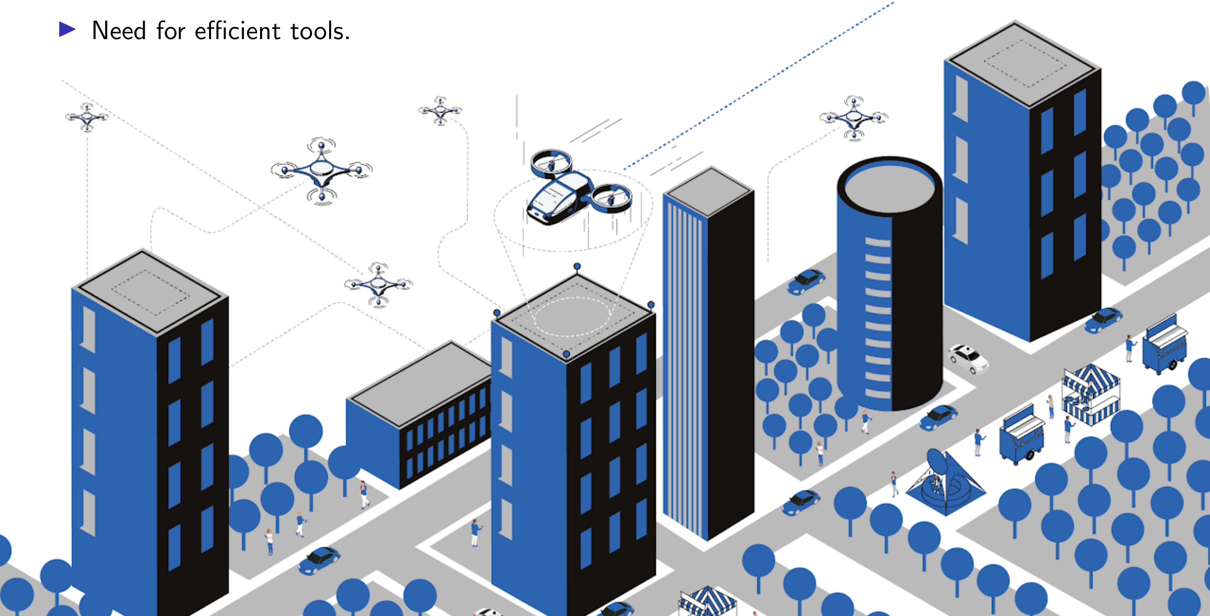


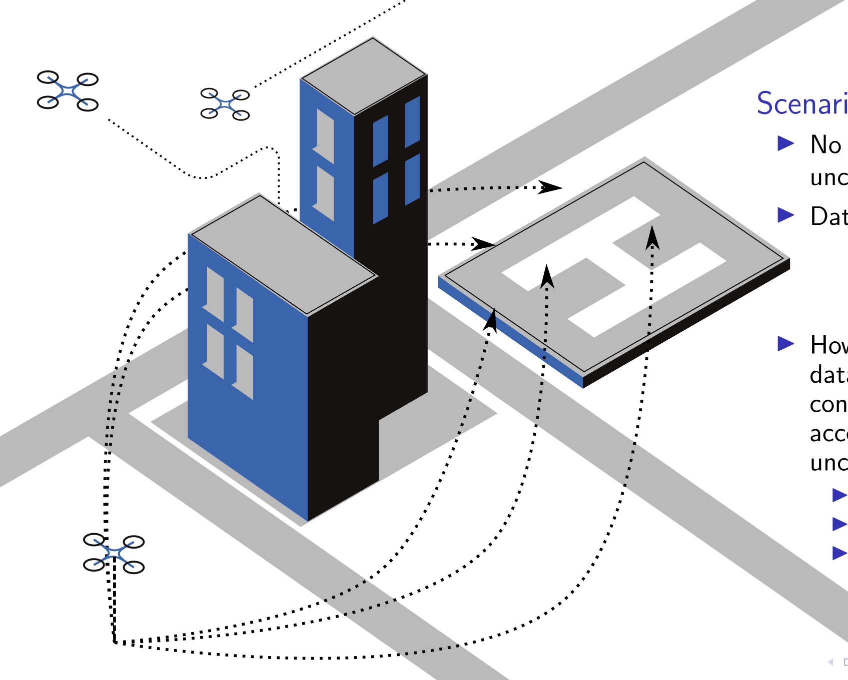
NASA ULI THE UNIVERSITY OF NEW MEXICO

- Futuristic autonomous systems will create & operate under excessive uncertainty.
- How do we operate in these kinds of environments?



- **Data-driven** solutions are needed.
- Need for efficient tools.



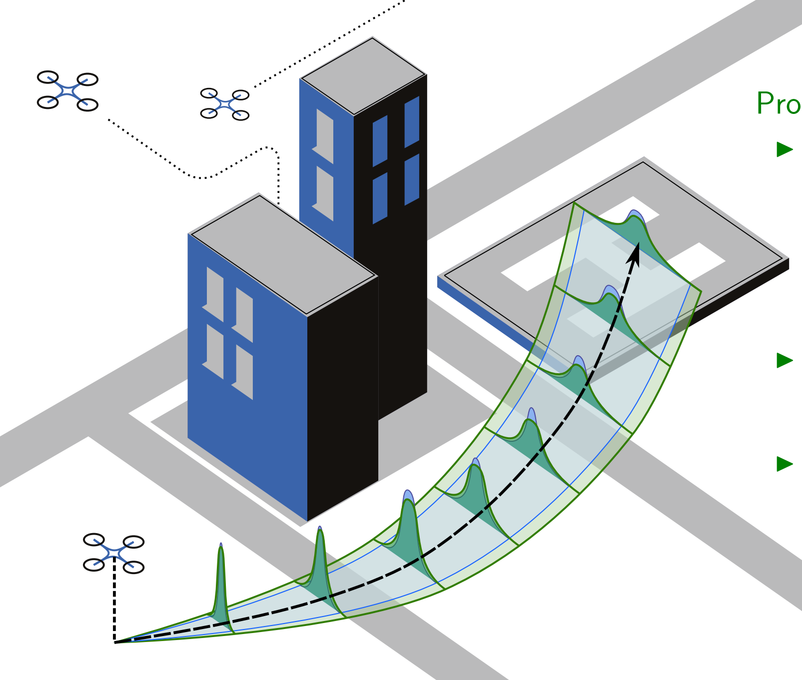


Scenario:

- No knowledge of dynamics or uncertainty.
- Data is available.

- How do we efficiently solve data-driven stochastic optimal control problems while accounting for real-world uncertainty?
  - Complex mechanical systems.
  - Complex environments.
  - Humans.

NASA ULI THE UNIVERSITY OF NEW MEXICO



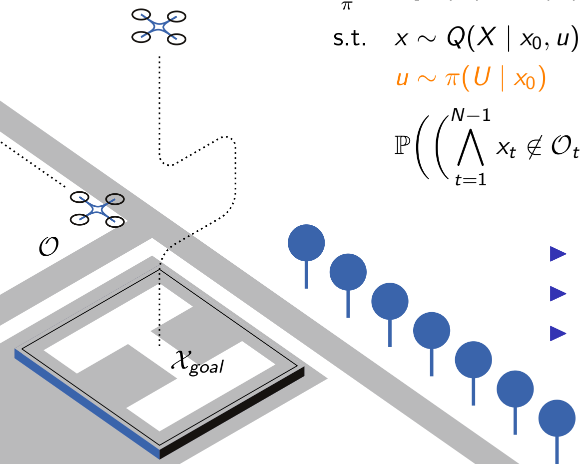
Proposed approach:

- Project data into a high-dimensional function space known as a reproducing kernel Hilbert space (RKHS).
- Can compute empirical approximations of distributions in an RKHS.
- Stochastic optimal control problem can be viewed as an LP.

Stochastic Optimal Control

Unknown

$$
\begin{aligned}
\min_{\pi} \quad & \mathbb{E}[\ell^x(X) + \ell^u(U)] \\
\text{s.t.} \quad & x \sim Q(X \mid x_0, u) \\
& u \sim \pi(U \mid x_0) \\
& \mathbb{P}\left(\left(\bigwedge_{t=1}^{N-1} x_t \notin \mathcal{O}_t \right) \wedge (x_N \in \mathcal{X}_{goal}) \right) \geq 1 - \delta
\end{aligned}
$$



- Intractable since Q is unknown.
- From Ono et al., 2016, we seek a *stochastic* policy.
- Amenable to convexity and other properties.

Related Work

Notable

Sutter, 2017
Martinelli, 2022
Bhattacharyya, 2020
Chua, 2018
Choi et al., 2020

Model-Predictive Control

Mesbah, 2016
Rosolia & Borrelli, 2017

Kernel Embeddings

Song, 2009, 2010
Grünewalder, 2012
Nishiyama, 2012

Chance-Constrained Optimization

Ono, et al., 2016
Schmerling & Pavone, 2017

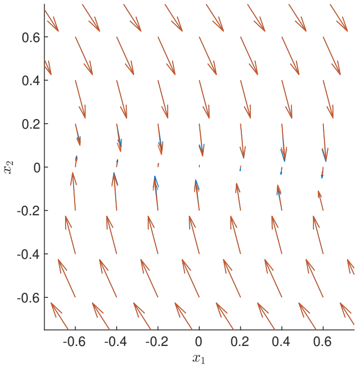
Gaussian Processes & Koopman Ops.

Rasmussen & Williams, 2010
Deisenroth, 2015
Koller, 2018
Lew, 2021
Abraham & Murphey, 2019

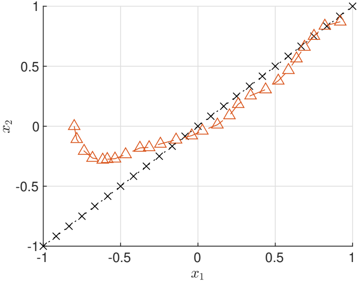
Kernel Methods

Koppel, et al., 2018, 2020
Thorpe & Oishi, 2020, 2021
Thorpe, Lew, Oishi, Pavone, 2022 (submitted)
Thorpe & Oishi, 2022 (to appear)

Related Work on Unconstrained Control



- Very good performance.
- Low error vs. model-based.



- Dynamic programming.
- Nonlinear dynamics.
- Non-Gaussian disturbances.
- Dual solution.

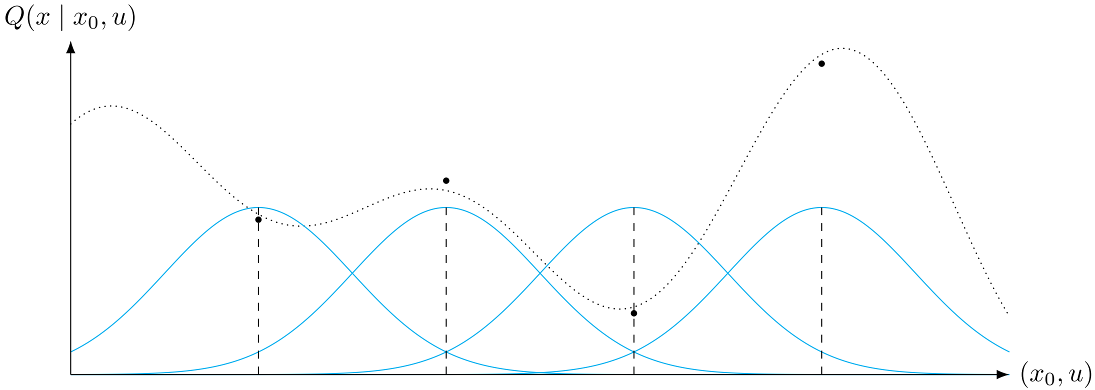
Assumptions

- Dataset is available.
- Functions in RKHS.
- Kernel measurable and bounded.

- Constrained problems are more difficult.

Thorpe, Oishi, "Stochastic Optimal Control via Hilbert Space Embeddings of Distributions," CDC 2021

Kernel Approximation

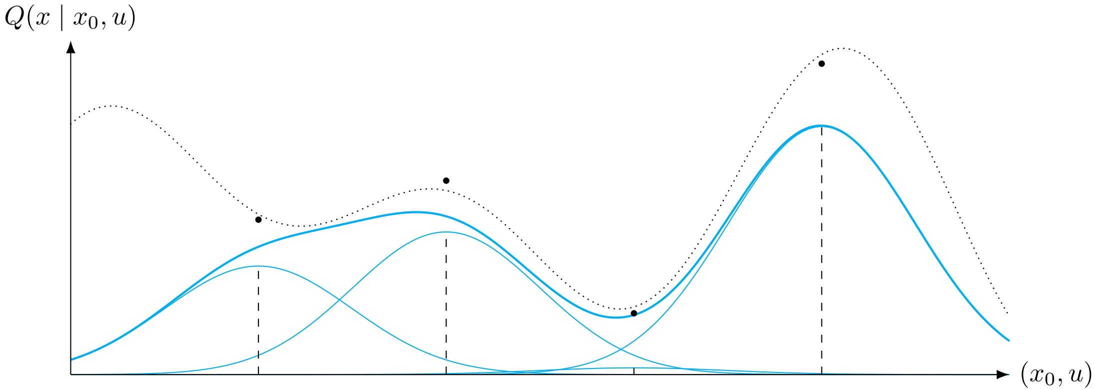


- Given: dataset $\mathcal{D} = \{(x^i, x_0^i, u^i)\}_{i=1}^M$ taken i.i.d. from Q.
- A kernel is a *positive definite function*, e.g.

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$

NASA ULI THE UNIVERSITY OF NEW MEXICO

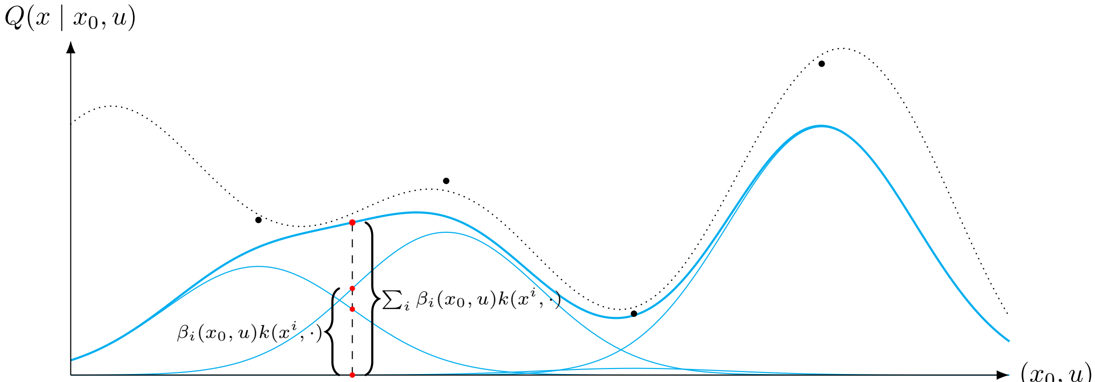
Kernel Approximation



- Solution is a linear combination.
- Compute weights using a regularized-least squares problem:

$$\hat{m} = \arg\min_{f \in \mathscr{Q}} \frac{1}{M} \sum_{i=1}^{M} \| k(x^i, \cdot) - f(x_0^i, u^i) \|_{\mathscr{H}}^2 + \lambda \| f \|_{\mathscr{Q}}^2$$

Kernel Approximation



- Can approximate functions (and expectations) using *reproducing property*:

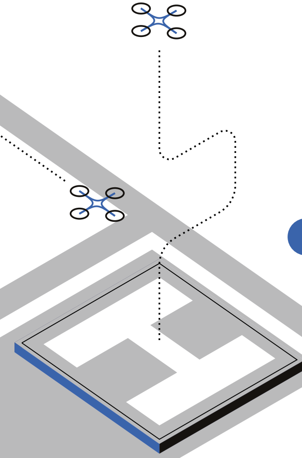
$$\mathbb{E}_{x \sim Q(\cdot|x_0,u)}[f(x)] \approx \langle f, \hat{m}(x_0, u) \rangle$$

- To evaluate, we compute the weighted sum of kernel functions.
- Related to mean component of $GP(\mu, \Sigma)$.

RKHS Reformulation

$$\min_{p(x_0)} \quad \langle p(x_0), \langle m(x_0, u), \ell^x \rangle + \ell^u \rangle$$

$$\text{s.t.} \quad \langle p(x_0), \langle m(x_0, u), \mathbf{1}_{\mathcal{O}_1^c \times \cdots \times \mathcal{O}_{N-1}^c \times \mathcal{X}_{goal}} \rangle\rangle \geq 1 - \delta$$



- **Goal**: compute $\hat{m}(x_0, u) \approx m(x_0, u)$, find $p(x_0)$.

Computational Challenge

- RKHS may be infinite-dimensional.
  - Thus, we search in a finite subspace.

$$p(x_0) = \sum_{j=1}^{P} \gamma_j(x_0) k(\tilde{u}^j, \cdot)$$

$\mathbb{R}$ Coefficients

Finite support

NASA ULI THE UNIVERSITY OF NEW MEXICO

Step 1: Stochastic Kernel Approximation

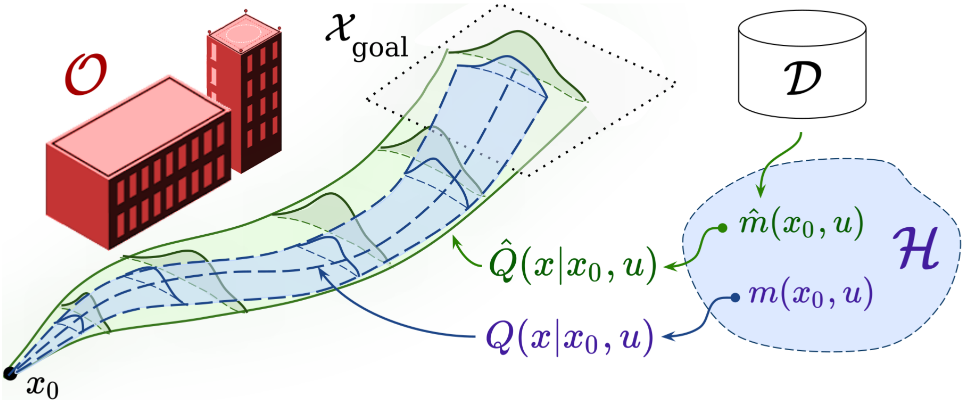


Figure reproduced with permission from Thomas Lew (thomas.lew@stanford.edu)

Step 2: Stochastic Policy

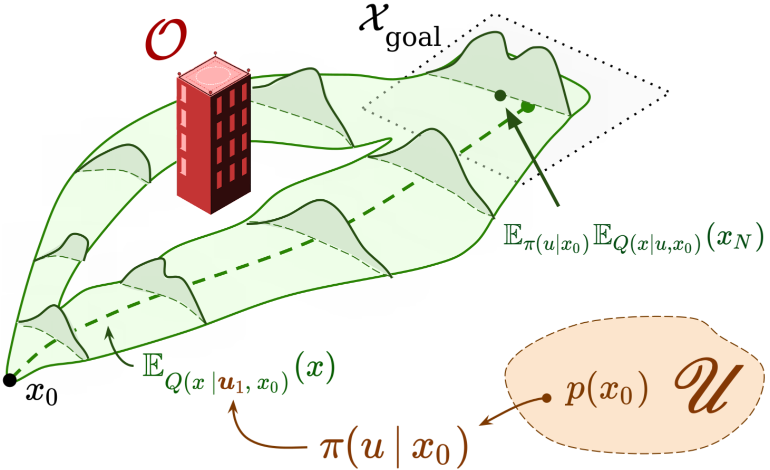


Figure reproduced with permission from Thomas Lew (thomas.lew@stanford.edu)

NASA ULI THE UNIVERSITY OF NEW MEXICO

Structure of Policy

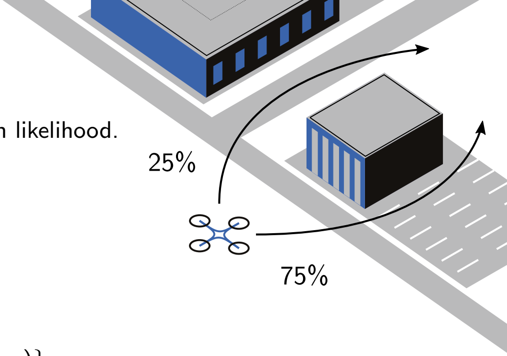
- Optimal stochastic policy is *mixed* (Ono, 2016).
  - Means we choose between controls with a certain likelihood.

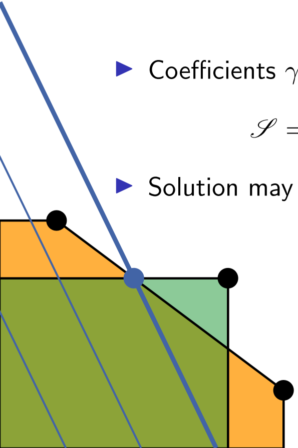
$$p(x_0) = \sum_{j=1}^{P} \gamma_j(x_0) k(\tilde{u}^j, \cdot)$$

- Coefficients $\gamma(x_0) \in \mathbb{R}^P$ in probability simplex.

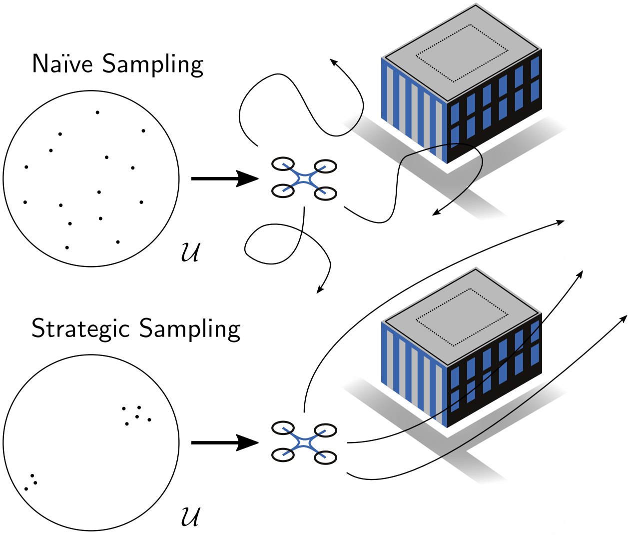
$$\mathscr{S} = \{\gamma(x_0) \in \mathbb{R}^P \mid \mathbf{1}^\top \gamma(x_0) = 1, 0 \preceq \gamma(x_0)\}$$

- Solution may be on an "edge" of the simplex, i.e. mixed.





Sampling



- Dataset is key to obtain a good result.
- Naïve sampling insufficient.

Strategic Sampling

- Strategic sampling needed in practice.
- E.g. using a PD controller.
- We obtain controls which are good candidates.

NASA ULI THE UNIVERSITY OF NEW MEXICO

Approximate RKHS Problem

$$\min_{p(x_0)} \quad \langle p(x_0), \langle \hat{m}(x_0, u), \ell^x \rangle + \ell^u \rangle$$
$$\text{s.t.} \quad \langle p(x_0), \langle \hat{m}(x_0, u), \mathbf{1}_{\mathcal{O}_1^c \times \cdots \times \mathcal{O}_{N-1}^c \times \mathcal{X}_{goal}} \rangle \rangle \geq 1 - \delta$$
$$\mathbf{1}^\top \gamma(x_0) = 1$$
$$0 \preceq \gamma(x_0)$$

$$\min_{x} \quad c^\top x$$
$$\text{s.t.} \quad Ax \geq 1 - \delta$$
$$\mathbf{1}^\top x = 1$$
$$0 \preceq x$$

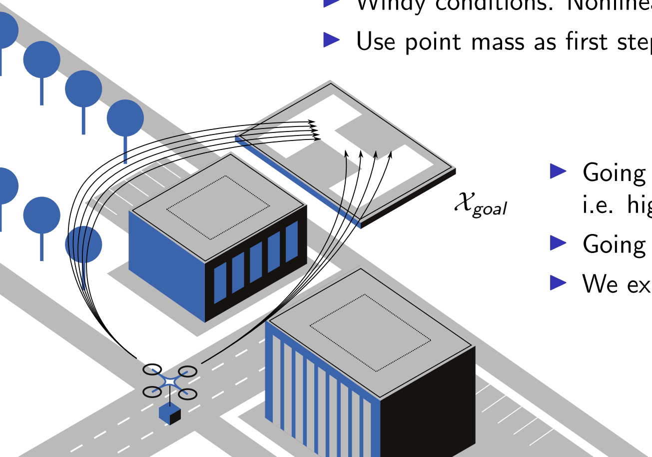
- **Approximate problem is an LP.**
- Can solve using off-the-shelf solvers using interior point or simplex methods.

Demonstration (Quadrotor with Uncertain Mass)

Scenario

Quadrotor carrying a payload around obstacles in windy conditions.

- Uncertain payload mass.
- Windy conditions. Nonlinear drift.
- Use point mass as first step for comparison.

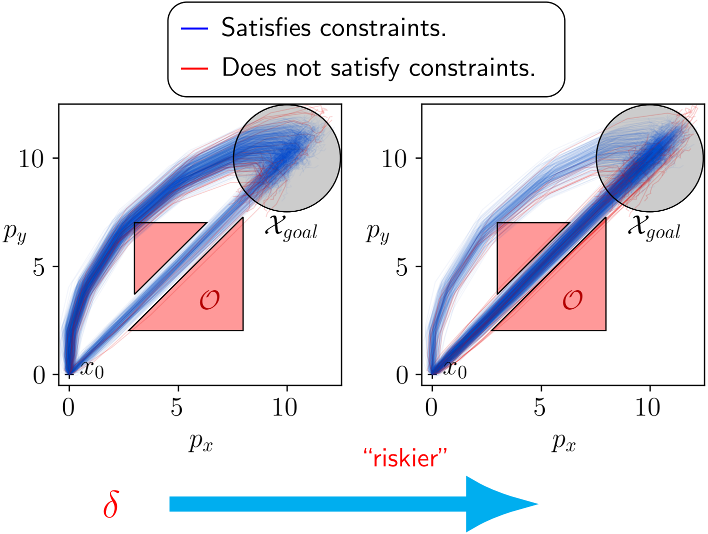


- Going between obstacles has higher risk, i.e. higher chance of collision.
- Going around has higher cost.
- We expect a policy that mixes the two paths.

NASA ULI THE UNIVERSITY OF NEW MEXICO

Results

- Higher δ leads to a policy that has higher likelihood of choosing risky middle corridor.
- Sample size: $M, P = 2500$
- Time horizon: $N = 15$
- Generating controls: $\approx$ 5s
- Computing LP: $\approx$ 100ms
- Scales with sample size: $\mathcal{O}(M^3)$



Tolerable probability of failure δ	0.05	0.20
Empirical probability of failure	0.029	0.039

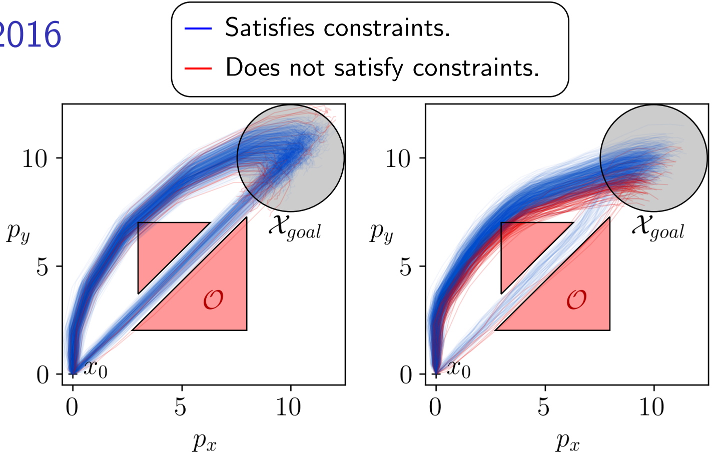
NASA ULI THE UNIVERSITY OF NEW MEXICO

Comparison with Ono et al., 2016

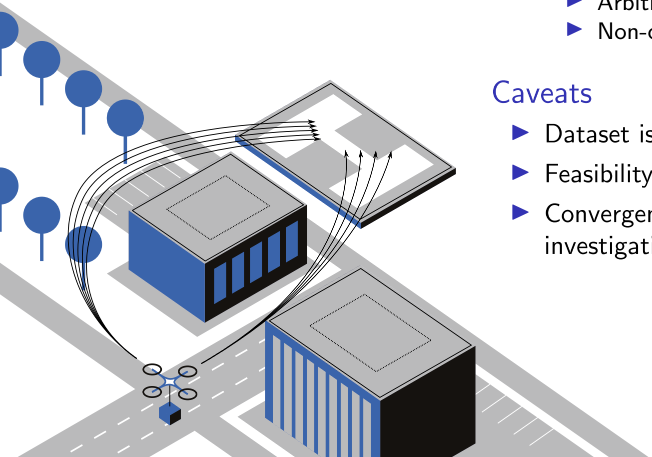
- Boole's inequality.
- Lagrangian relaxation.
- Bisection technique.

- Difficult dynamics:

$$x_{t+1} = Ax_t + \frac{1}{m}Bu_t - \alpha d(x_t) + w_t$$



	Kernel Approach	Ono 2016
Tolerable probability of failure δ	0.05	0.05
Empirical probability of failure	0.029	0.102



Advantages

- LP reformulation.
- Mixed policies.
- Accommodates:
  - Arbitrary disturbances.
  - Non-convex cost & constraints.

Caveats

- Dataset is key to obtaining a good result.
- Feasibility depends on data.
- Convergence to optimal policy is still under investigation.

Thank you!

Contact: ajthor@unm.edu

Code: https://github.com/ajthor/socks

SOCKS: Python toolbox for Stochastic Optimal Control using Kernel Methods

The NASA University Leadership initiative (Grant #80NSSC20M0163) provided funds to assist the authors with their research, but this article solely reflects the opinions and conclusions of its authors and not any NASA entity.

RKHS Embeddings

- Define a positive definite kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.
- Moore-Aronszajn theorem: there exists a unique RKHS $\mathscr{H}$ corresponding to k.
  - Reproducing property: $f(x) = \langle f, k(x, \cdot)\rangle$
  - Kernel trick: $k(x, x') = \langle k(x, \cdot), k(x', \cdot)\rangle$
  - Every $f \in \mathscr{H}$ is a linear combination of kernel functions.

$$f = \sum_i \alpha_i k(x_i, \cdot) \qquad f(x) = \langle f, k(x, \cdot)\rangle = \sum_i \alpha_i k(x_i, x)$$

- Riesz lemma: $\mathbb{E}[f(X)]$ is *linear* $\implies \exists m \in \mathscr{H}$ such that $\mathbb{E}[f(x)] = \langle f, m\rangle$.

NASA ULI THE UNIVERSITY OF NEW MEXICO

1. Using dataset $\mathcal{D} = \{(x_0^i, u^i, x^i)\}_{i=1}^M$, estimate:

$$\mathbb{E}[f(x)] := \int_{\mathcal{X}} f(x) Q(\mathrm{d}x \mid x_0, u) \approx \langle f, \hat{m}(x_0, u) \rangle = \boldsymbol{f}^\top (\Psi \Psi^\top + \lambda M I)^{-1} \Psi k(x_0, \cdot) l(u, \cdot),$$

   where $\boldsymbol{f} = [f(x^1), \ldots, f(x^M)]^\top$ and $\Psi = [k(x_0^1, \cdot) l(u^1, \cdot), \ldots, k(x_0^M, \cdot) l(u^M, \cdot)]^\top$.

2. Given $\{\tilde{u}^j\}_{j=1}^P$,

$$\mathbb{E}[g(u)] := \int_{\mathcal{U}} g(u) \pi(\mathrm{d}u \mid x_0) = \langle g, p(x_0) \rangle = \left\langle g, \sum_{j=1}^P \gamma_j(x_0) l(\tilde{u}^j, \cdot) \right\rangle,$$

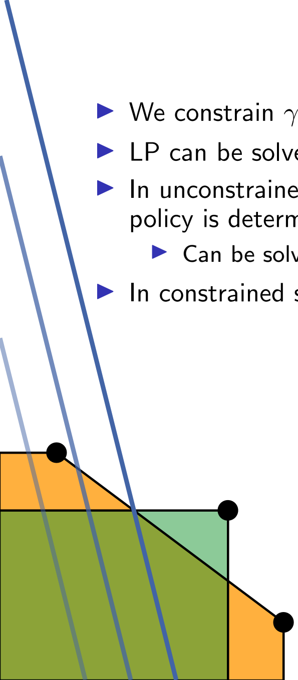
   where $\gamma(x_0) \in \mathbb{R}^P$.

3. Compute:

$$\begin{aligned} \mathbb{E}_{u \sim \pi(\cdot \mid x_0)}[\mathbb{E}_{x \sim Q(\cdot \mid x_0, u)}[f(x)]] &= \int_{\mathcal{U}} \int_{\mathcal{X}} f(x) Q(\mathrm{d}x \mid x_0, u) \pi(\mathrm{d}u \mid x_0) \\ &\approx \boldsymbol{f}^\top (\Psi \Psi^\top + \lambda M I)^{-1} \Psi \Upsilon^\top k(x_0, \cdot) \gamma(x_0) \\ &= \boldsymbol{f}^\top W K(x_0) \gamma(x_0) \end{aligned}$$

- We constrain $\gamma(x_0)$ to be in the probability simplex $\mathscr{S} = \{x \in \mathbb{R}^P \mid \mathbf{1}^\top x = 1, 0 \preceq x\}$.
- LP can be solved easily, e.g. via interior-point or simplex methods.
- In unconstrained setting, the solution is at a vertex of the probability simplex, meaning policy is deterministic.
  - Can be solved efficiently via the Lagrangian dual.
- In constrained setting, the solution may be on an "edge", meaning the policy is mixed.

$$\begin{aligned} \min_{\gamma(x_0) \in \mathbb{R}^P} \quad & c^\top \gamma(x_0) \\ \text{s.t.} \quad & A\gamma(x_0) \geq 1 - \delta \\ & \mathbf{1}^\top \gamma(x_0) = 1 \\ & 0 \preceq \gamma(x_0) \end{aligned}$$



- Let $\mathscr{H}$ be an RKHS with kernel k and $\mathscr{Q}$ be a vector-valued RKHS of functions on $\mathcal{X} \times \mathcal{U}$ mapping to $\mathscr{H}$. Let k be bounded by $\rho < \infty$, and let υ be a σ-admissible loss function with respect to $\mathscr{Q}$. Then the learning algorithm given by

$$\hat{m} = \arg\min_{f \in \mathscr{Q}} \frac{1}{M} \sum_{i=1}^{M} \upsilon(f, (y_i, x_i, u_i)) + \lambda \|f\|^2_{\mathscr{Q}},$$

  has uniform stability α with respect to υ with $\alpha \leq \frac{\sigma^2 \rho^2}{2\lambda M}$.
- Let A be an algorithm with uniform stability α with respect to a loss function υ such that $0 \leq \upsilon(\hat{m}, (x, u, y)) \leq B$, for all $(x, u, y) \in \mathcal{Z}$ and all sets $\mathcal{S}$. Then for any $M \geq 1$ and any $\delta \in (0, 1)$ the following bounds hold with probability $1 - \delta$ of the random draw of the sample $\mathcal{S}$:

$$R(\hat{m}) \leq R_{\mathcal{S}}(\hat{m}) + 2\alpha + (4M\alpha + B)\sqrt{\frac{\log(1/\delta)}{2M}}.$$

- Thus, we have that for any $M \geq 1$ and any $\delta \in (0, 1)$, with probability $1 - \delta$, the risk R is bounded by:

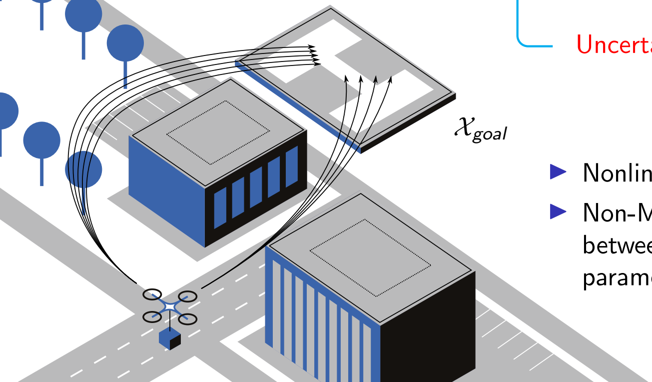
$$R(\hat{m}) \leq R_{\mathcal{S}}(\hat{m}) + \frac{\sigma^2 \rho^2}{\lambda M} + \left(\frac{2\sigma^2 \rho^2}{\lambda} + \rho\right)\sqrt{\frac{\log(1/\delta)}{2M}}.$$

Example Dynamics

$$x_{t+1} = Ax_t + Bu_t + d(x_t) + w_t,$$

$$A = \begin{bmatrix} 1 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T_s \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B = \frac{1}{m}\begin{bmatrix} T_s^2/2 & 0 \\ T_s & 0 \\ 0 & T_s^2/2 \\ 0 & T_s \end{bmatrix}, \quad d(x_t) = -\alpha \begin{bmatrix} T_s^2|v_x|v_x/2 \\ T_s|v_x|v_x \\ T_s^2|v_y|v_y/2 \\ T_s|v_y|v_y \end{bmatrix}$$

Uncertain parameters



- Nonlinear due to drift term.
- Non-Markovian due to the temporal correlation between the state trajectory x and the uncertain parameters.