Stochastic Optimal Control & Safety Via Kernel Embeddings

A Data-Driven Approach

Adam Thorpe

University of New Mexico

2022

Motivation

- Modern systems operate in uncertain environments.
- Must deal with strict operating constraints, safety constraints, and humans.

The New York Times

Uber's Self-Driving Cars Were Struggli
Before Arizona Crash

SAN FRANCISCO — Uber's robotic vehicle project was not living up to expectations months before a self-driving car operated by the company struck and killed a woman in Tempe, Ariz.

The Washington Post

'Full Self-Driving' clips show owner of Teslas fighting for control, and experts see deep flaws

SAN FRANCISCO — In one video, a Tesla tries to drive down some light-rail tracks. In another, a Tesla fails to stop for a pedestrian in a crosswalk. And at one point, the most advanced driver-assistance product available to consumers appears to slam into a bike lane bollard at 11 mph.

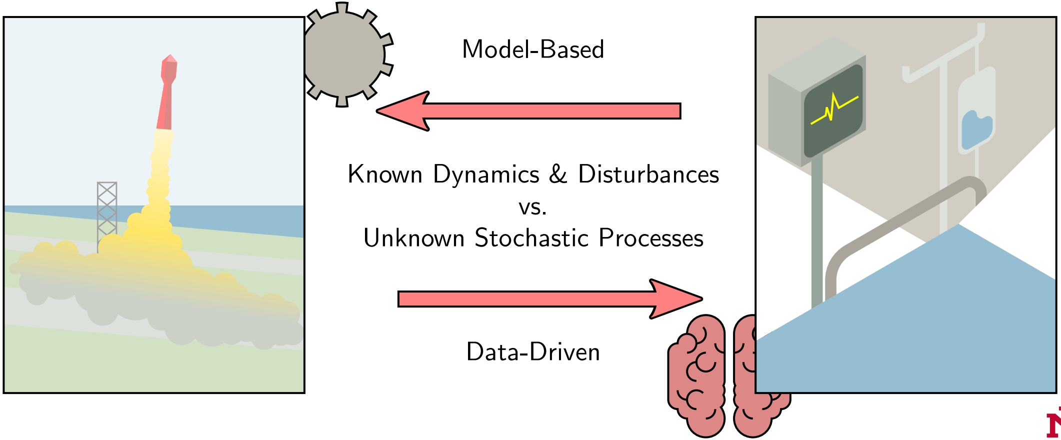
- Unknown dynamics & stochastic processes.
- Complex mechanical systems & dynamics.
- Unforeseen human factors.

Images courtesy of: NASA, The National Archives, and Ford Motor Company from USA, CC BY 2.0, via Wikimedia Commons

THE UNIVERSITY OF NEW MEXICO

Data-Driven vs. Model-Based Control

- Autonomy will only become more prevalent.
- Motivates the need for **data-driven** techniques that enable control and assess safety.



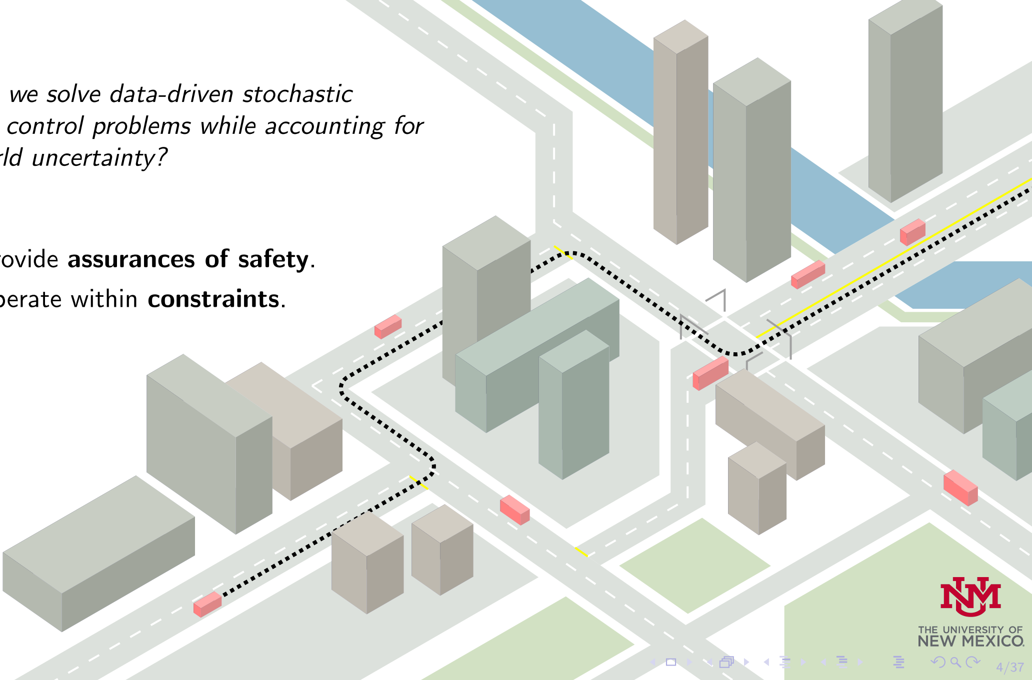
THE UNIVERSITY OF NEW MEXICO

Problem

- *How do we solve data-driven stochastic optimal control problems while accounting for real-world uncertainty?*

Challenges

- Must provide **assurances of safety**.
- Must operate within **constraints**.



THE UNIVERSITY OF NEW MEXICO

Data-Driven Control & Stochastic Reachability

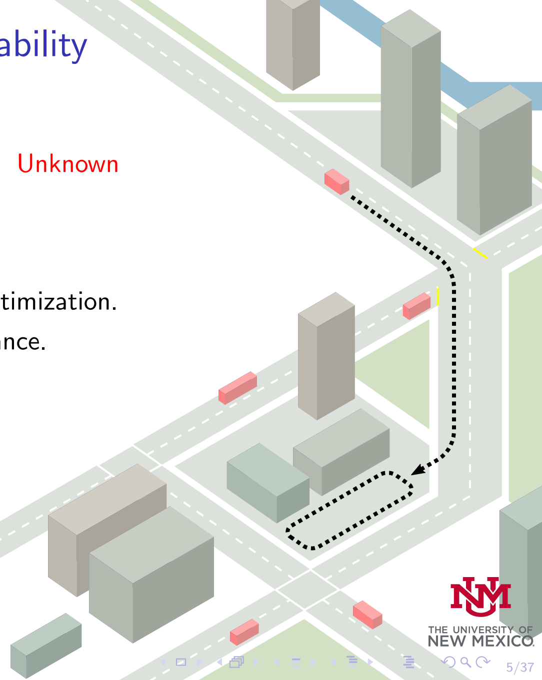
$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}[g(x, u)] \\
\text{s.t.} \quad & x \sim Q(\cdot \mid x_0, u) \longleftarrow \text{Unknown} \\
& u \sim \pi(\cdot \mid x_0) \\
& \mathbb{P}(x \in F) \geq 1 - \delta
\end{aligned}
$$

- Difficult since Q is unknown, chance-constrained optimization.
- Model-based often limited to LTI, Gaussian disturbance.

Stochastic Reachability

$$
\begin{aligned}
\max_{\pi \in \Pi} \quad & \mathbb{E}\left[\left(\prod_{t=0}^{N-1} \mathbf{1}_{\mathcal{K}}(x_t)\right) \mathbf{1}_{\mathcal{T}}(x_N)\right] \\
\text{s.t.} \quad & x_{t+1} \sim Q(\cdot \mid x_t, u_t) \\
& u_t \sim \pi(\cdot \mid x_t)
\end{aligned}
$$

- Multiplicative cost.



Our Proposed Approach

- Project data collected from Q into a high-dimensional function space known as a reproducing kernel Hilbert space (RKHS) $\mathscr{H}$.

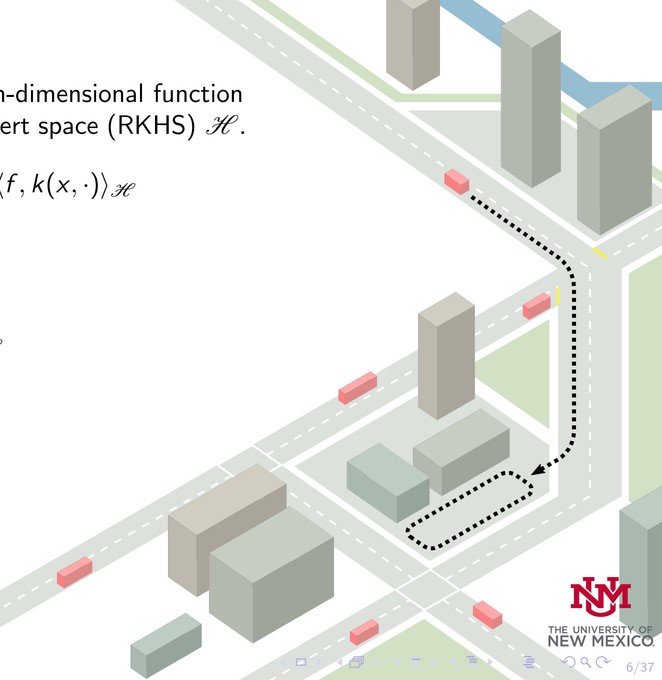
$$\text{Reproducing property: } f(x) = \langle f, k(x, \cdot) \rangle_{\mathscr{H}}$$

- Examples of Hilbert spaces: $\mathbb{R}^n$, L^2, ℓ^2
- Kernel embeddings of distributions:

$$\mathbb{E}[f(X)] = \langle f, m \rangle_{\mathscr{H}}$$

Advantages

- Natural fit for many control problems.
- Mathematically rigorous.
- Nonparametric, i.e. few assumptions.
- **Stochastic control problems can be viewed as an LP.**



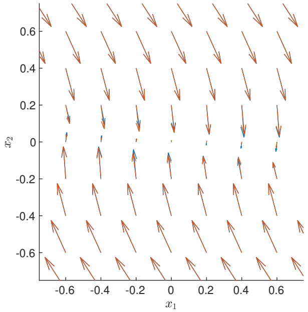
THE UNIVERSITY OF NEW MEXICO

Related Work

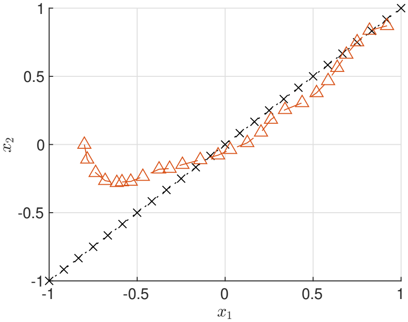
- Dynamic Programming & Optimal Control
  Bertsekas & Shreve 1978, Puterman 2014
- Chance-Constrained Control
  Lew et al. 2022, Lew, Bonalli, Pavone 2020, Ono, Chamie, Pavone, Acikmese 2016
- Stochastic Reachability
  Abate et al. 2008, Summers & Lygeros 2012, Thorpe & Oishi 2019
- Robust Control & Optimization
  Zhu, Jitkrittum, Diehl, Schölkopf 2021, Nemmour, Schölkopf, Zhu 2021
- Infinite LPs
  Martinelli, Gargiani, Lygeros 2021 & 2022, Sutter, Kamoutsi, Esfahani, Lygeros 2017
- Kernel-Based Control
  Song 2009, Grünewalder 2012, Thorpe & Oishi 2021, Thorpe, Lew, Oishi, Pavone 2022

THE UNIVERSITY OF NEW MEXICO

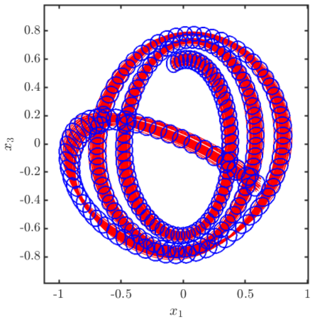
Prior Work on Unconstrained Control & Reachability



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



- Nonlinear systems.
- Non-Gaussian disturbances.
- Dynamic programming.
- **Constraints are more difficult.**



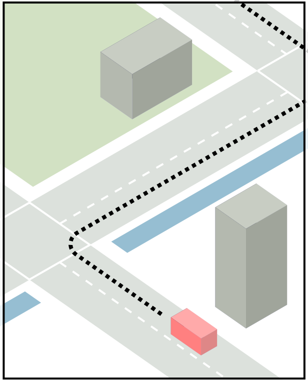
- Forward reachability.
- Learning-enabled components.

Thorpe, Oishi, CDC 2021
Thorpe, Ortiz, Oishi, L4DC 2021

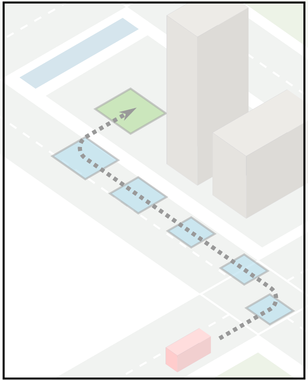
THE UNIVERSITY OF NEW MEXICO

Outline

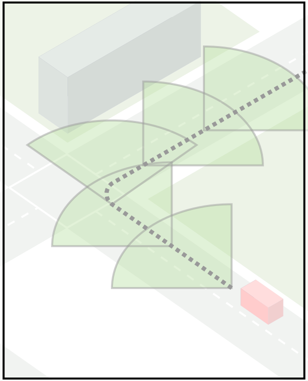
Stochastic Optimal Control



Stochastic Reachability



Future Work



THE UNIVERSITY OF NEW MEXICO

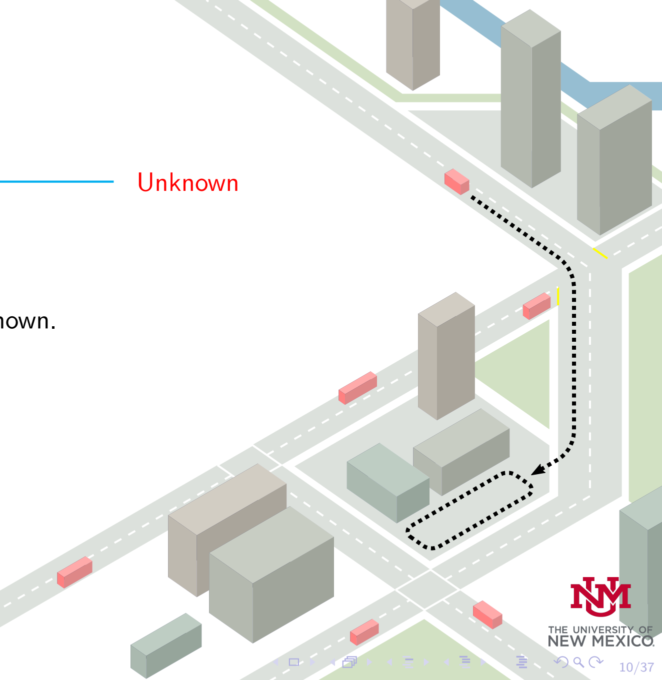
Stochastic Optimal Control

$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}[g(x, u)] \\
\text{s.t.} \quad & x \sim Q(\cdot \mid x_0, u) \longleftarrow \textcolor{red}{\text{Unknown}} \\
& u \sim \pi(\cdot \mid x_0) \\
& \mathbb{P}(x \in F) \geq 1 - \delta
\end{aligned}
$$

- Intractable since stochastic kernel is unknown.
- Data taken i.i.d. from Q is available.
- We seek a stochastic policy.
- Solution is non-trivial.

Challenges

- Computing expectations and probability.
- Policy synthesis.



THE UNIVERSITY OF NEW MEXICO

Stochastic Kernels

- A stochastic kernel κ from $(\mathcal{X}, \mathscr{B}_{\mathcal{X}})$ to $(\mathcal{Y}, \mathscr{B}_{\mathcal{Y}})$,

$$\kappa : \mathscr{B}_{\mathcal{Y}} \times \mathcal{X} \to [0, 1]$$

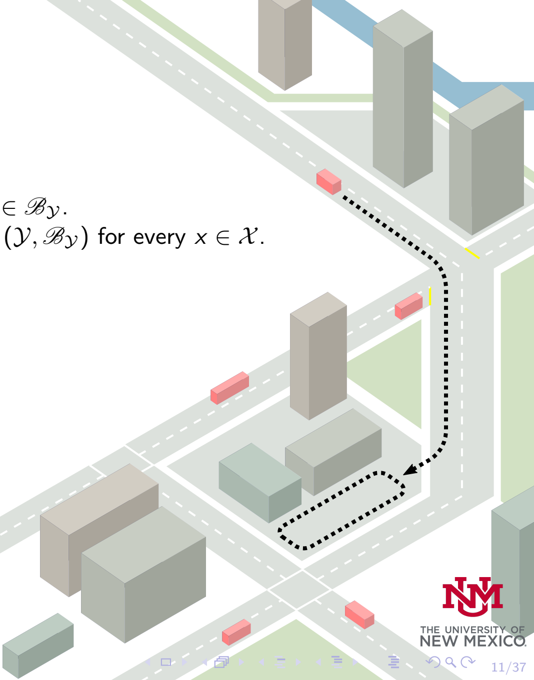
  1. The map $x \mapsto \kappa(B \mid x)$ is $\mathscr{B}_{\mathcal{X}}$-measurable for all $B \in \mathscr{B}_{\mathcal{Y}}$.
  2. The map $B \mapsto \kappa(B \mid x)$ is a probability measure on $(\mathcal{Y}, \mathscr{B}_{\mathcal{Y}})$ for every $x \in \mathcal{X}$.

- Defines an integral operator κf,

$$\kappa f(x) = \int_{\mathcal{Y}} f(y)\kappa(\mathrm{d}y \mid x)$$

Main Idea

- Expectations (integrals) are linear.
- For f in RKHS, can use Riesz theorem.



THE UNIVERSITY OF NEW MEXICO

System Definition Via Stochastic Kernels

- Discrete-time stochastic system,

$$x_{t+1} = f(x_t, u_t, w_t)$$

  states $x_t \in \mathcal{X}$, control inputs $u_t \in \mathcal{U}$, independent RVs w_t.

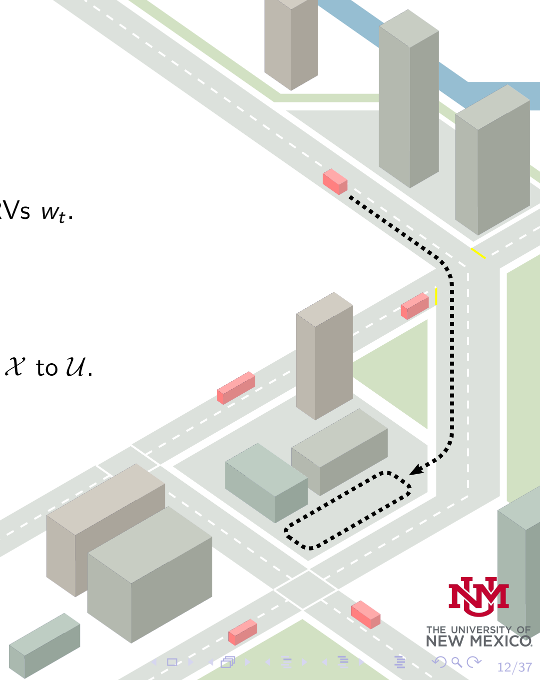
- Can express as a Markov control process $(\mathcal{X}, \mathcal{U}, Q)$.

$$x_{t+1} \sim Q(\cdot \mid x_t, u_t)$$

- A policy is a sequence of stochastic kernels π_t from $\mathcal{X}$ to $\mathcal{U}$.
- Can look at trajectories $x = \{x_t\}_{t=1}^{N}$, $u = \{u_t\}_{t=0}^{N-1}$.

Main Idea

- Embed Q in an RKHS.
- Estimate using data.
- Embed policy in an RKHS.



Stochastic Optimal Control

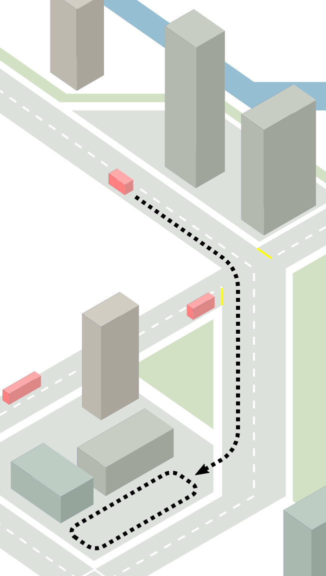
$$\begin{aligned}\min_{\pi \in \Pi} \quad & \int_{\mathcal{U}^N} \int_{\mathcal{X}^N} f_0(y) Q(\mathrm{d}y \mid x_0, u) \pi(\mathrm{d}u \mid x_0) \\ \text{s.t.} \quad & \int_{\mathcal{U}^N} \int_{\mathcal{X}^N} f_i(y) Q(\mathrm{d}y \mid x_0, u) \pi(\mathrm{d}u \mid x_0) \geq 1 - \delta, i = 1, \ldots, n\end{aligned}$$

- Assume $f_0, \ldots, f_n$ are in an RKHS $\mathscr{H}$ and $\int_{\mathcal{X}^N} f(y) Q(\mathrm{d}y \mid x_0, u) < \infty$,

$$\int_{\mathcal{X}^N} f(y) Q(\mathrm{d}y \mid x_0, u) = \langle f, m(x_0, u) \rangle_{\mathscr{H}} \approx \langle f, \hat{m}(x_0, u) \rangle_{\mathscr{H}}$$

Step 1: Compute an estimate $\hat{m}(x_0, u)$ of $m(x_0, u)$ in $\mathscr{H}$.

Step 2: Find policy representation $p(x_0)$ in RKHS $\mathscr{U}$.



THE UNIVERSITY OF NEW MEXICO

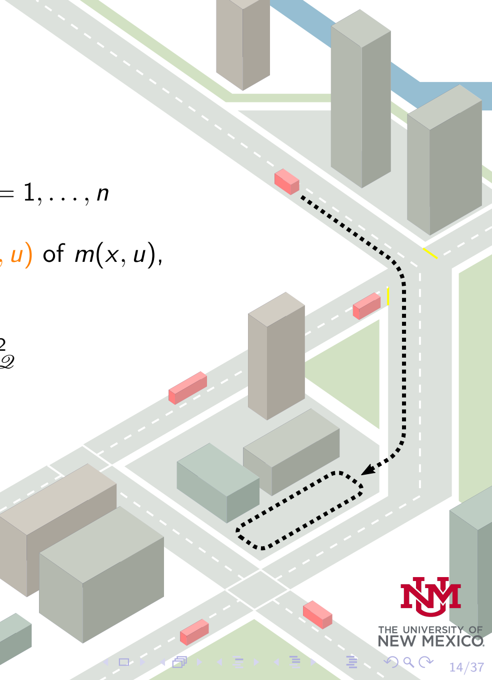
Step 1: Embedding Stochastic Kernel

$$\begin{aligned} \min_{p(x_0) \in \mathscr{U}} \quad & \langle \langle f_0, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, p(x_0) \rangle_{\mathscr{U}} \\ \text{s.t.} \quad & \langle \langle f_i, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, p(x_0) \rangle_{\mathscr{U}} \geq 1 - \delta, i = 1, \ldots, n \end{aligned}$$

- Using data $\mathcal{S} = \{(x_0^i, u^i, x^i)\}_{i=1}^M$, compute estimate $\hat{m}(x, u)$ of $m(x, u)$,

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

$$\langle f, \hat{m}(x_0, u) \rangle_{\mathscr{H}} \approx \int_{\mathcal{X}^N} f(y) Q(\mathrm{d}y \mid x_0, u)$$



THE UNIVERSITY OF NEW MEXICO

Step 2: Policy Synthesis

$$
\begin{aligned}
\min_{p(x_0) \in \mathscr{U}} \quad & \langle \langle f_0, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, p(x_0) \rangle_{\mathscr{U}} \\
\text{s.t.} \quad & \langle \langle f_i, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, p(x_0) \rangle_{\mathscr{U}} \geq 1 - \delta, i = 1, \ldots, n
\end{aligned}
$$

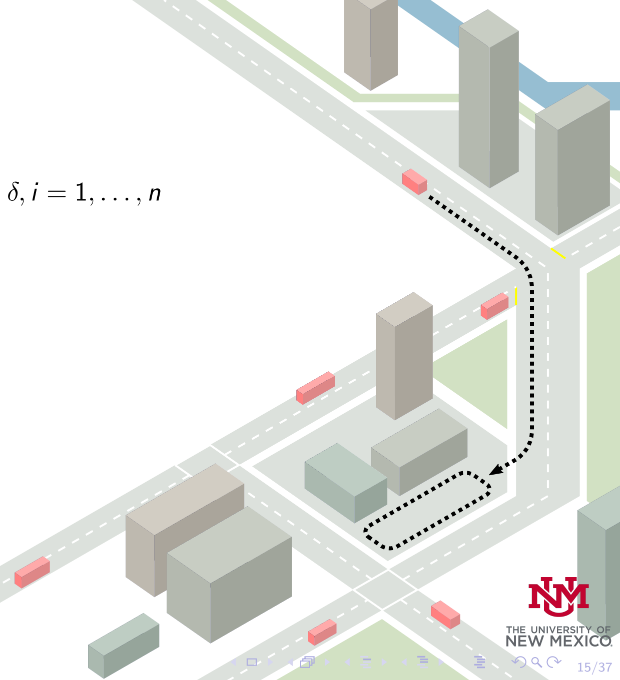
- Represent the policy in an RKHS $\mathscr{U}$.
- RKHS may be infinite-dimensional.
  - We search in a finite subspace.

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

$\mathbb{R}$ Coefficients

Finite Support

Thorpe, Lew, Oishi, Pavone, L4DC 2022
Thorpe, Oishi, CDC 2021



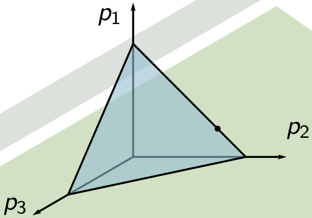
THE UNIVERSITY OF NEW MEXICO

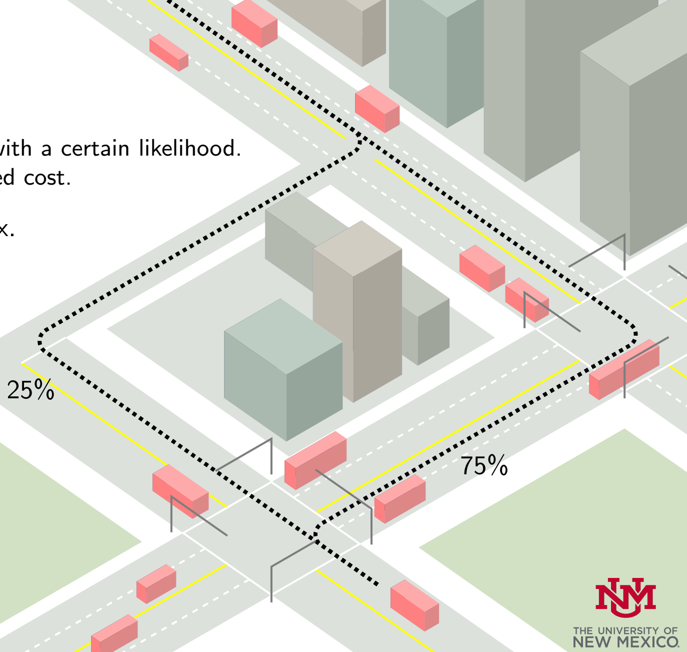
Structure of the Policy

- Optimal stochastic policy is mixed.
  - Means we choose between controls with a certain likelihood.
  - Mixed policy may have lower expected cost.
- Coefficients $\gamma(x_0)$ in probability simplex.

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

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



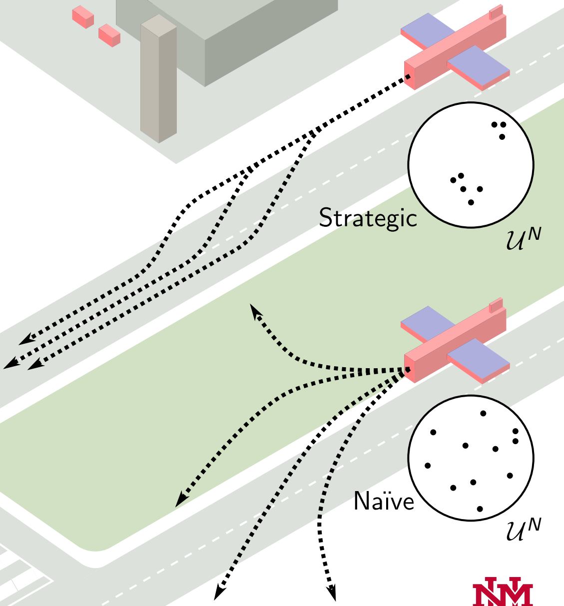


Strategic Sampling

- *How do we choose controls $\tilde{u}^j$?*

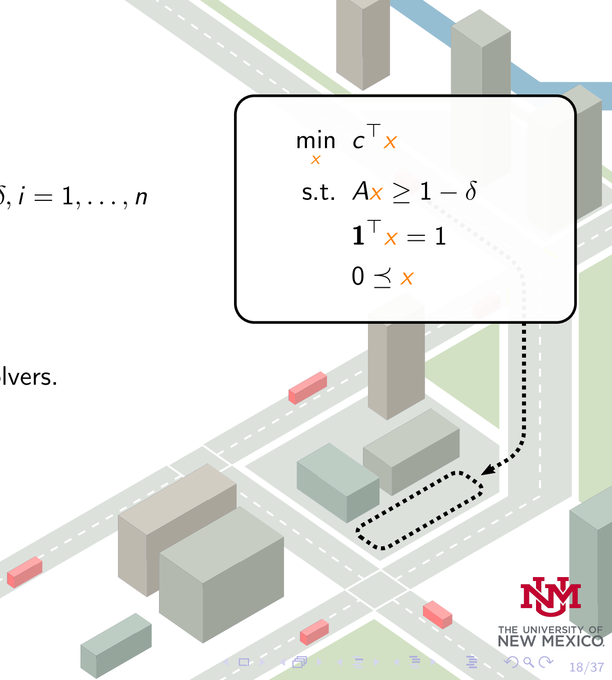
$$p(x_0) = \sum_{j=1}^{P} \gamma_j(x_0) k(\tilde{u}^j, \cdot)$$

- Naïve sampling insufficient.
- Strategic sampling needed in practice, e.g. using:
  - PD controller.
  - Motion primitives.
- We obtain good candidate controls.



Approximate Problem

$$
\begin{aligned}
\min_{\gamma(x_0)\in\mathbb{R}^P} \quad & \boldsymbol{f}_0^\top W \Psi \Upsilon^\top k(x_0, \cdot)\gamma(x_0) \\
\text{s.t.} \quad & \boldsymbol{f}_i^\top W \Psi \Upsilon^\top k(x_0, \cdot)\gamma(x_0) \geq 1 - \delta, i = 1, \ldots, n \\
& \mathbf{1}^\top \gamma(x_0) = 1 \\
& 0 \preceq \gamma(x_0)
\end{aligned}
$$



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

- **Approximate problem is an LP.**
- Can be solved efficiently using off-the-shelf solvers.
  - Interior point or simplex algorithms.

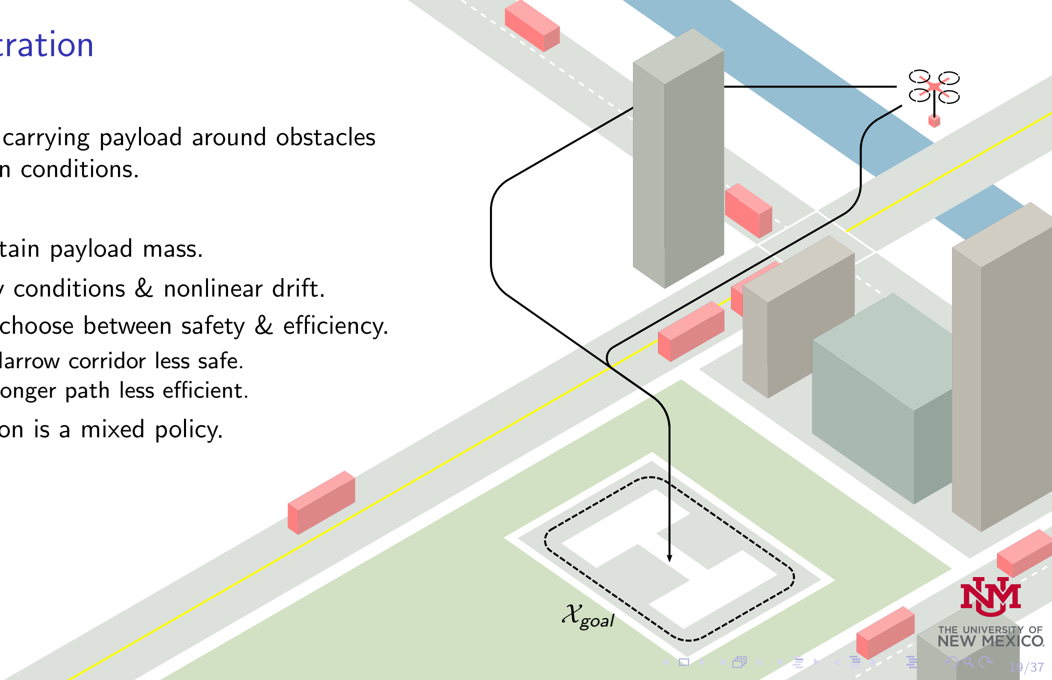
Thorpe, Lew, Oishi, Pavone, L4DC 2022
Thorpe, Oishi, CDC 2021

Demonstration

Scenario

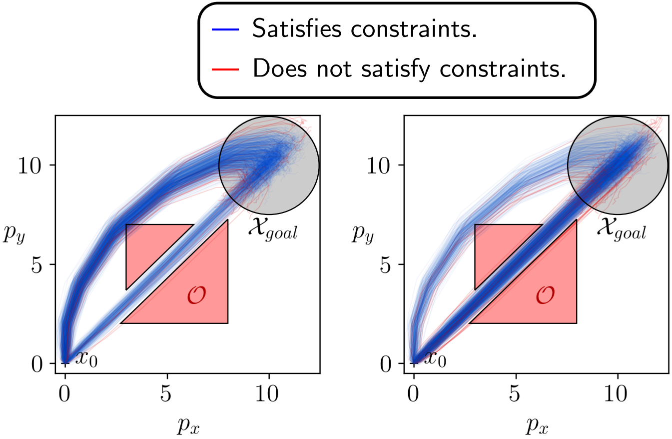
Quadrotor carrying payload around obstacles in uncertain conditions.

- Uncertain payload mass.
- Windy conditions & nonlinear drift.
- Must choose between safety & efficiency.
  - Narrow corridor less safe.
  - Longer path less efficient.
- Solution is a mixed policy.



THE UNIVERSITY OF NEW MEXICO

Results



- Sample size: $M, P = 2{,}500$
- Time horizon: $N = 15$

- Generating sample: $\approx$ 5s
- Generating controls: $\approx$ 5s
- Computing LP: $\approx$ 100 ms
- Scales with sample size: $\mathcal{O}(M^3)$

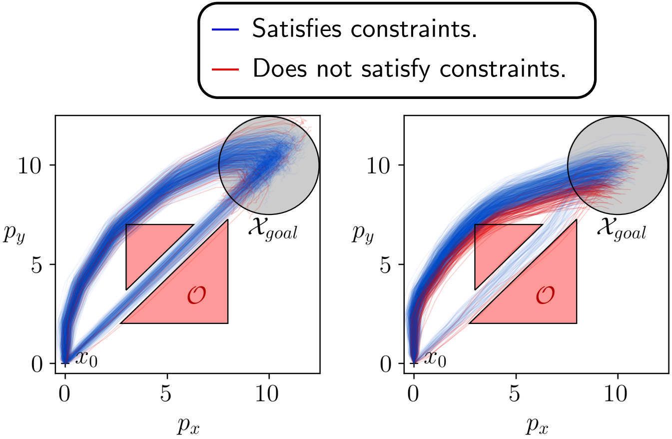
- Higher δ leads to a policy that has a higher chance of choosing "risky" middle corridor.

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

(Recall: chance constraint $\mathbb{P}(x \in F) \geq 1 - \delta$)

THE UNIVERSITY OF NEW MEXICO

Comparison With Ono 2016



- Boole's inequality.
- Lagrangian relaxation.
- Difficult dynamics:

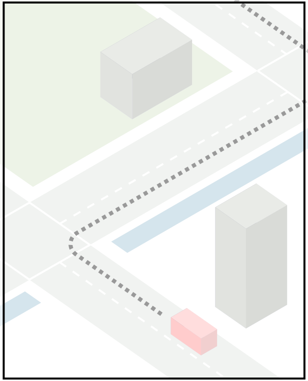
$$x_{t+1} = Ax_t + \frac{1}{m}Bu_t - \alpha d(x_t) + w_t$$

- Parameters m, α are uncertain.
- $d(x_t)$ is nonlinear drift term.
- Existing methods do not satisfy chance constraints.

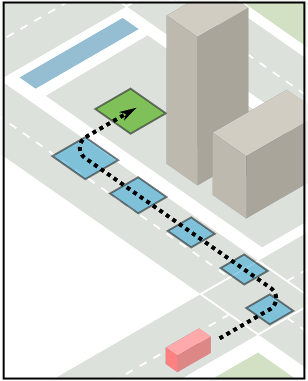
	Our Approach	Ono 2016
Tolerable probability of failure δ	0.05	0.05
Empirical probability of failure	0.029	0.102

Outline

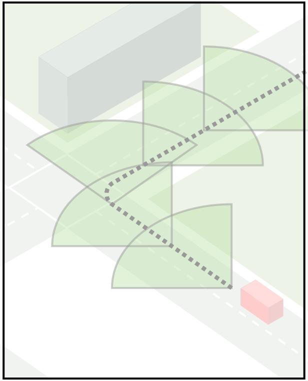
Stochastic Optimal Control



Stochastic Reachability



Future Work



THE UNIVERSITY OF NEW MEXICO

Stochastic Reachability (Terminal-Hitting Time Problem)

- Goal: Compute "safety probabilities," i.e. the likelihood that a system following a fixed policy $\pi = \{\pi_0, \pi_1, \dots \pi_{N-1}\} \in \Pi$ will reach a target set $\mathcal{T}$ at $t = N$ while remaining within a safe set $\mathcal{K}$ for all $t < N$.

$$r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T}) = \mathbb{P}(x_N \in \mathcal{T} \wedge x_t \in \mathcal{K}, \forall t = 0, 1, \dots, N-1) = \mathbb{E}_{x_0}^{\pi}\left[\left(\prod_{t=0}^{N-1} \mathbf{1}_{\mathcal{K}}(x_t)\right)\mathbf{1}_{\mathcal{T}}(x_N)\right]$$

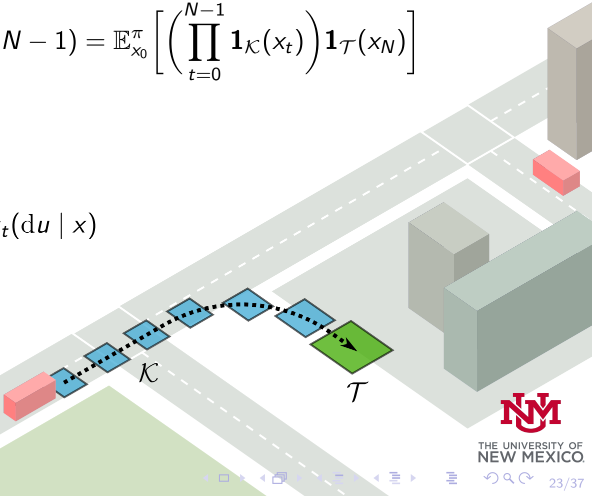
$$V_N^{\pi}(x) = \mathbf{1}_{\mathcal{T}}(x)$$

$$V_t^{\pi}(x) = \mathbf{1}_{\mathcal{K}}(x)\int_{\mathcal{U}}\int_{\mathcal{X}} V_{t+1}^{\pi}(y) Q(\mathrm{d}y \mid x, u)\pi_t(\mathrm{d}u \mid x)$$

$$\Downarrow$$

$$V_0^{\pi}(x_0) = r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T})$$

- Intractable since Q is unknown.



Maximally Safe Policies

- Find a policy $\pi^* = \{\pi_0^*, \pi_1^*, \ldots, \pi_{N-1}^*\} \in \Pi$ such that for all $x \in \mathcal{X}$,

$$\pi_t^* = \arg\sup_{\pi_t} \mathbf{1}_{\mathcal{K}}(x) \int_{\mathcal{U}} \int_{\mathcal{X}} V_{t+1}^*(y) Q(\mathrm{d}y \mid x, u) \pi_t(\mathrm{d}u \mid x)$$

$$V_N^*(x) = \mathbf{1}_{\mathcal{T}}(x)$$

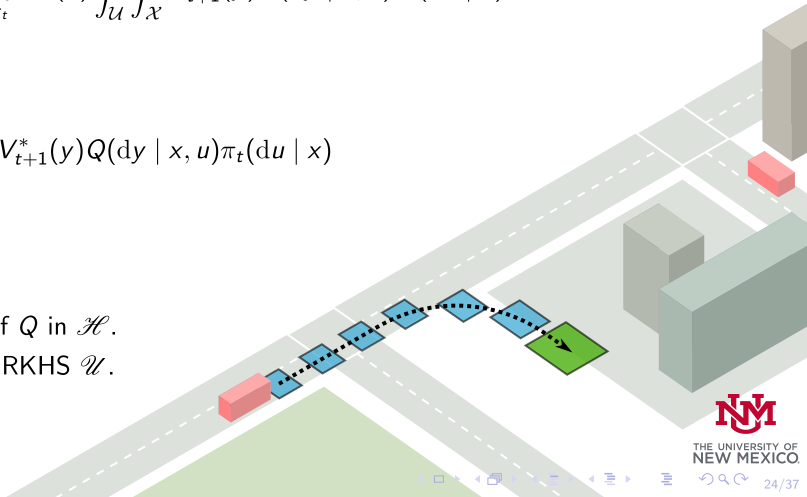
$$V_t^*(x) = \sup_{\pi_t} \mathbf{1}_{\mathcal{K}}(x) \int_{\mathcal{U}} \int_{\mathcal{X}} V_{t+1}^*(y) Q(\mathrm{d}y \mid x, u) \pi_t(\mathrm{d}u \mid x)$$

$$\Downarrow$$

$$V_0^*(x_0) = r_{x_0}^{\pi^*}(\mathcal{K}, \mathcal{T})$$

Step 1: Compute estimate $\hat{m}$ of Q in $\mathscr{H}$.

Step 2: Represent policy π^* in RKHS $\mathscr{U}$.



THE UNIVERSITY OF NEW MEXICO

Approximate Stochastic Reachability

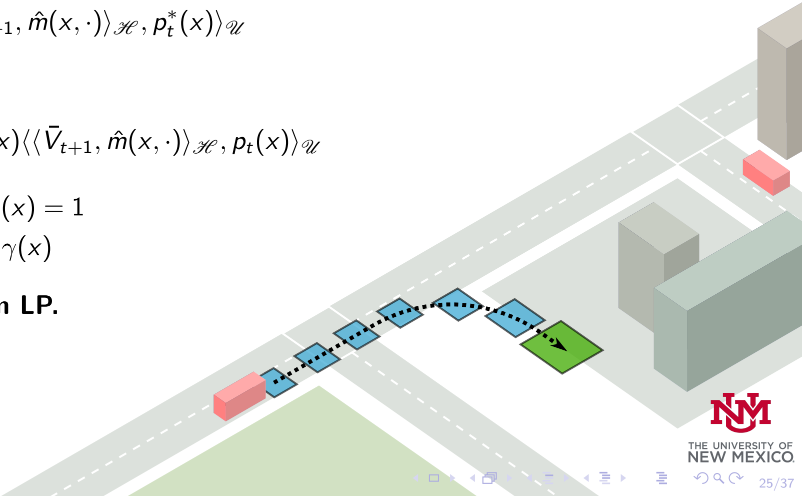
- Use kernel embeddings to estimate $\hat{m}(x,u) \in \mathscr{H}$ and $p_0^*(x), p_1^*(x), \dots, p_{N-1}^*(x) \in \mathscr{U}$.

$$\bar{V}_N(x) = \mathbf{1}_{\mathcal{T}}(x)$$
$$\bar{V}_t(x) = \mathbf{1}_{\mathcal{K}}(x) \langle \langle \bar{V}_{t+1}, \hat{m}(x,\cdot) \rangle_{\mathscr{H}}, p_t^*(x) \rangle_{\mathscr{U}}$$

- At each time step,

$$\begin{aligned} p_t^*(x) = \arg \sup_{p_t(x)} \quad & \mathbf{1}_{\mathcal{K}}(x) \langle \langle \bar{V}_{t+1}, \hat{m}(x,\cdot) \rangle_{\mathscr{H}}, p_t(x) \rangle_{\mathscr{U}} \\ \text{s.t.} \quad & \mathbf{1}^\top \gamma(x) = 1 \\ & 0 \preceq \gamma(x) \end{aligned}$$

- **Approximate problem is an LP.**

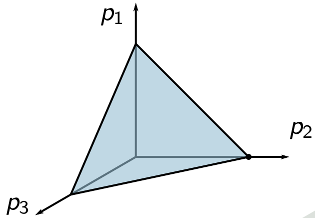


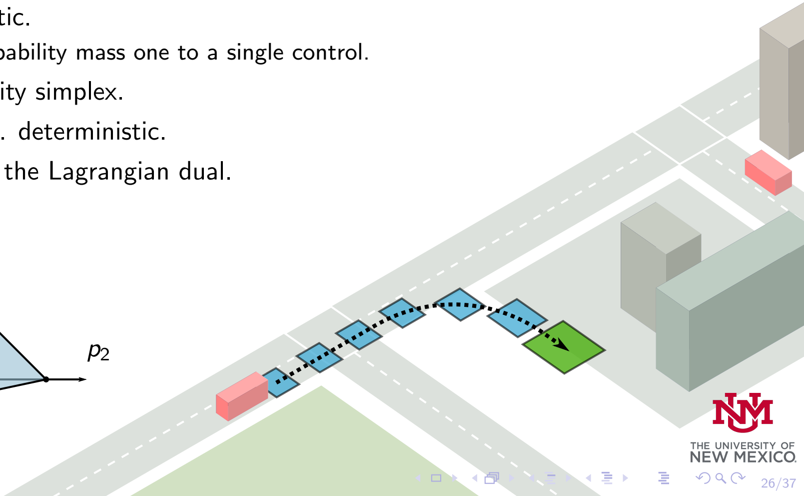
Thorpe, Oishi, L-CSS 2019

Structure of the Policy

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

- Optimal policy is deterministic.
  - Means policy assigns probability mass one to a single control.
- Coefficients $\gamma(x)$ in probability simplex.
- Solution is at a “vertex”, i.e. deterministic.
- Can be solved efficiently via the Lagrangian dual.

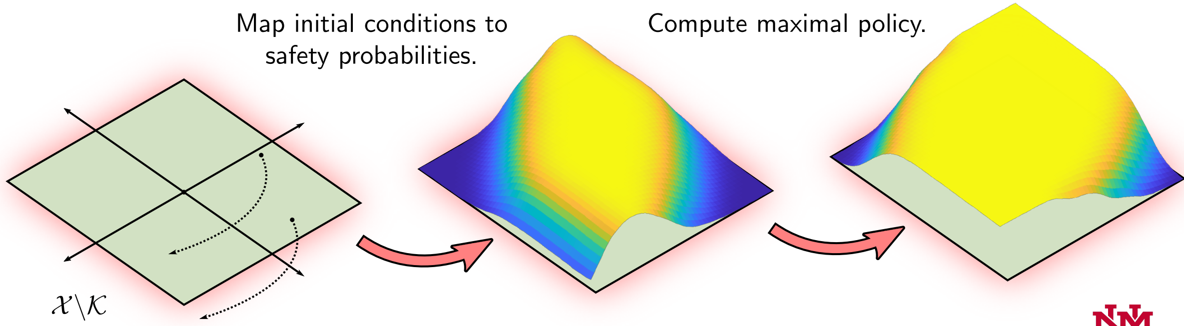




Demonstration

- Point mass system (double integrator).
  - Chosen for validation.
- Goal is to stay within a small region around the origin.

$$r^{\pi}_{x_0}(\mathcal{K}, \mathcal{T}) = \mathbb{P}(x_N \in \mathcal{T} \wedge x_t \in \mathcal{K}, \forall t = 0, 1, \ldots, N-1)$$



Unsafe trajectories leave safe set.

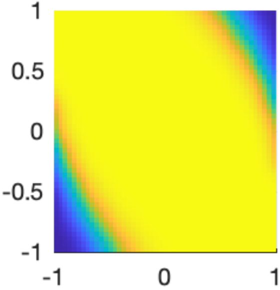
Results

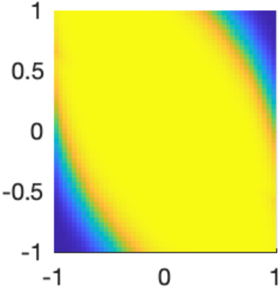
- Sample size: $M, P = 2{,}500$
- Time horizon: $N = 15$

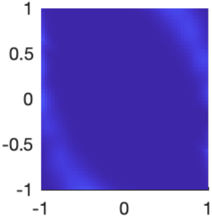
- Generating sample: $\approx$ 1s
- Computing LP: $\approx$ 100 ms
- Scales with sample size: $\mathcal{O}(M^3)$

- Maximum absolute error: $\approx$ 10%

- Warmer colors denote higher safety probability.
- Maximally safe policy has better performance.

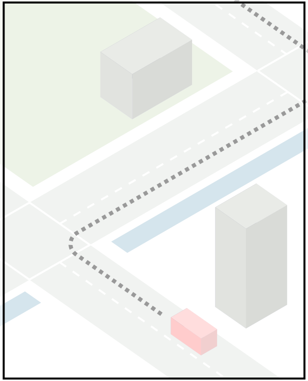




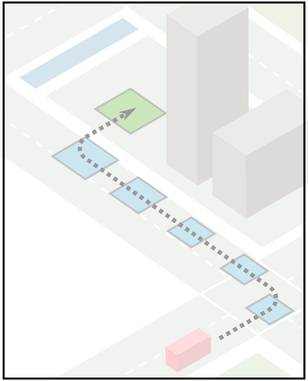


Outline

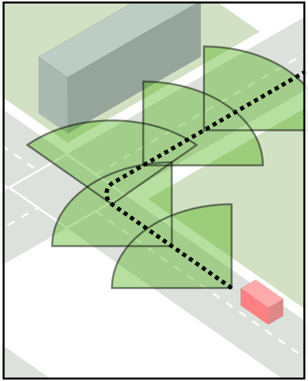
Stochastic Optimal Control



Stochastic Reachability



Future Work



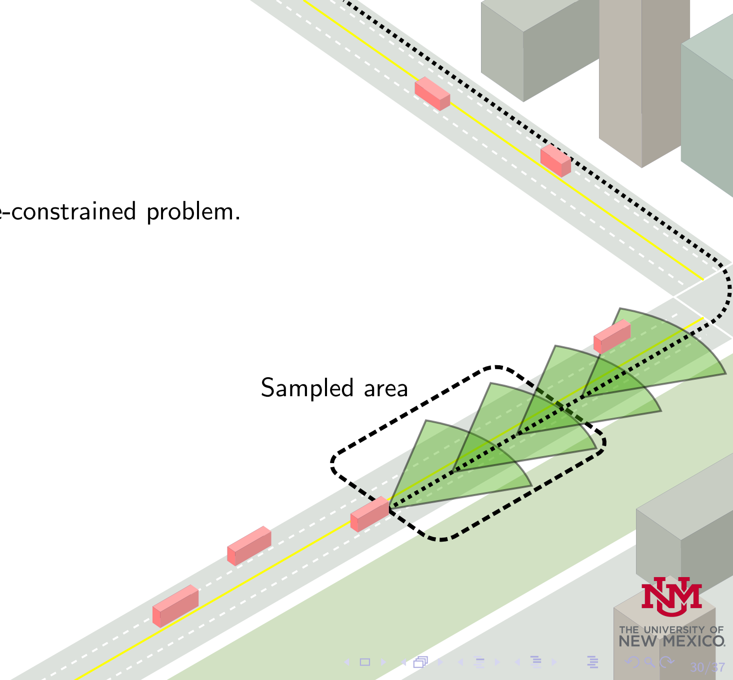
THE UNIVERSITY OF NEW MEXICO

Existing Challenges

- Out-of-sample generalization.
- Strategic sampling techniques.
- Incorporating feedback into chance-constrained problem.

Limitations

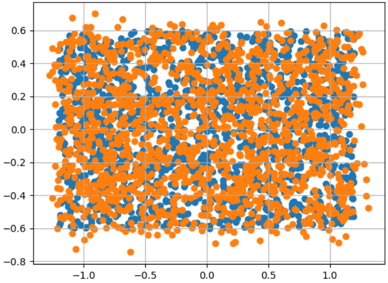
- Currently open-loop.
- Highly dependent on sample used.



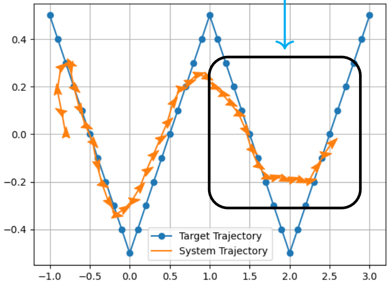
Out-Of-Sample Generalization

- Sample size: $M = 2{,}500$
- Time horizon: $N = 40$
- Computing LP: $\approx$ 150 ms

- Nonlinear system.
- Nonholonomic dynamics.
- Non-Gaussian disturbance.



Outside sampled region



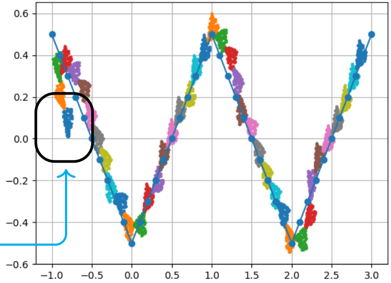
- Outside sampled region, control exhibits poor performance.
  - I.e. poor out-of-sample generalization.
- *Can we use known properties of the system to improve data-driven control performance?*
  - E.g., symmetry, knowledge of the vector field, $SE(n)$-**invariance**.

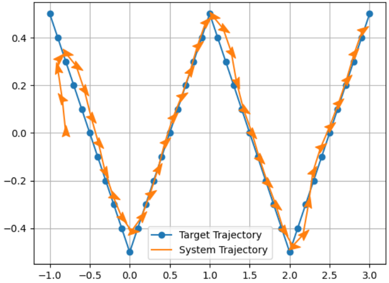
THE UNIVERSITY OF NEW MEXICO

Incorporating Dynamics

- Sample size: $M = 100$
- Time horizon: $N = 40$
- Computing LP: ≈ 10 ms

Sample in body frame





- Main idea: Collect sample in body frame and transform the sample as the system moves.
- Much better performance with smaller sample.
- Uses physics-informed learning, side information.

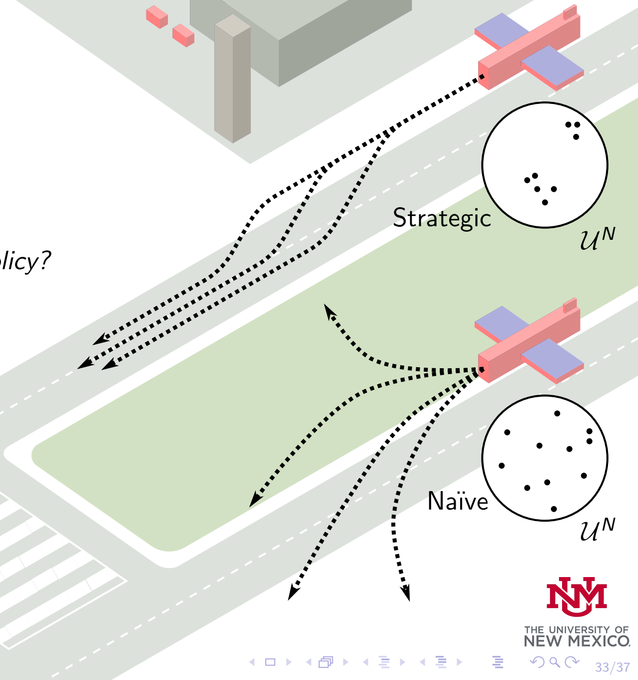
THE UNIVERSITY OF NEW MEXICO

Strategic Sampling

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

Finite Support

- *How do we choose the "support" of the policy?*
- Strategic sampling techniques needed.
  - Learning-based approaches.
  - Active sampling.
  - Motion primitives.
- Direct optimization possible?

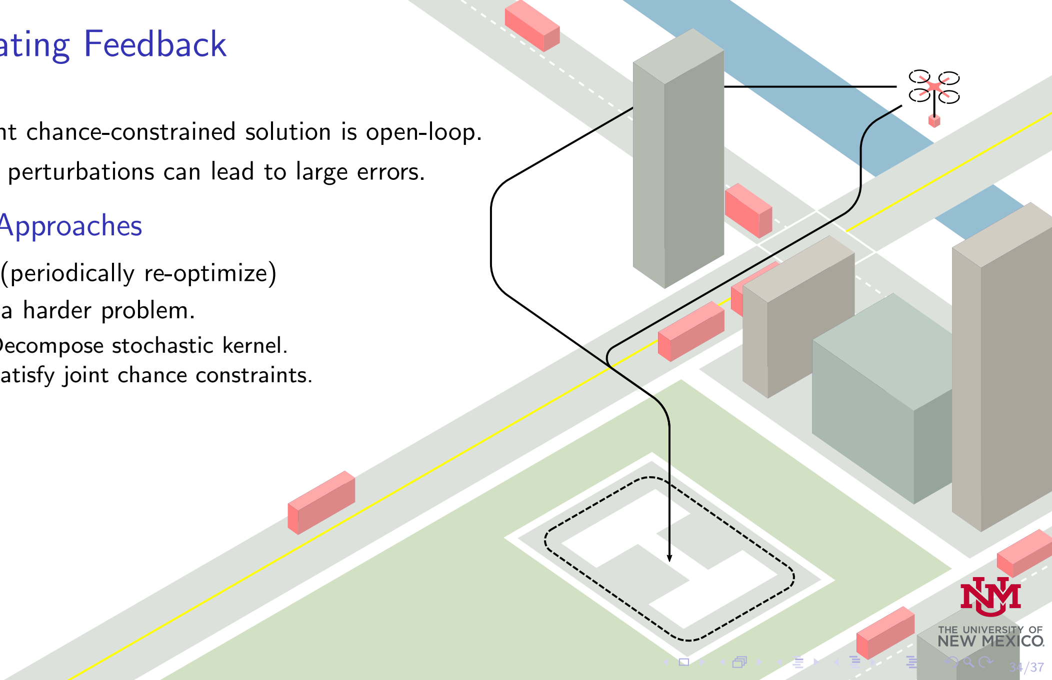


Incorporating Feedback

- Current chance-constrained solution is open-loop.
- Minor perturbations can lead to large errors.

Possible Approaches

- MPC (periodically re-optimize)
- Solve a harder problem.
  - Decompose stochastic kernel.
  - Satisfy joint chance constraints.



THE UNIVERSITY OF NEW MEXICO

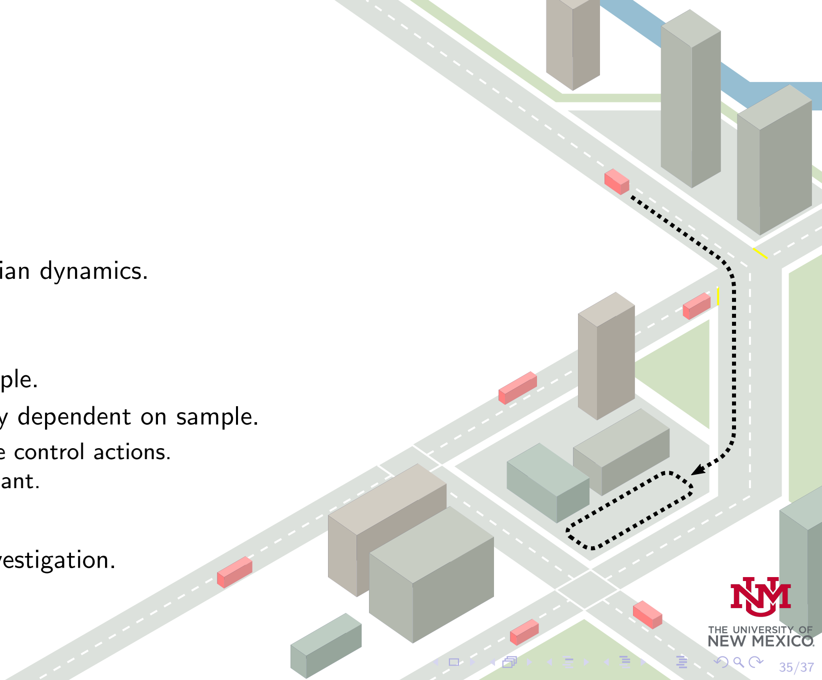
Conclusion

Advantages

- LP reformulation.
- Mixed policies.
- Arbitrary disturbances.
- Nonlinear and non-Markovian dynamics.

Current Limitations

- Feasibility depends on sample.
- Quality of solution is highly dependent on sample.
  - Need to properly choose control actions.
  - Generalization is important.
- Open-loop controllers.
- Convergence still under investigation.

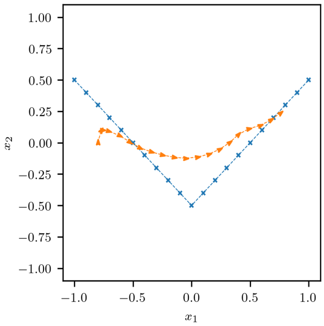


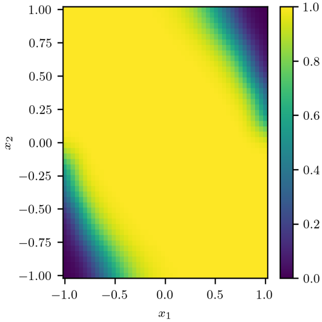
THE UNIVERSITY OF NEW MEXICO

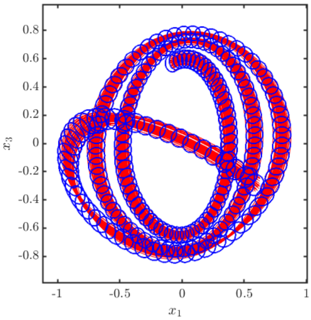
SOCKS: Stochastic Optimal Control Using Kernel Methods

- Stochastic Optimal Control
- Dynamic Programming
- Stochastic Reachability
- Forward Reachability

- Written in Python.
- Several examples & benchmarks.
- Binder integration.







https://github.com/ajthor/socks

Thorpe, Oishi, HSCC 2022

THE UNIVERSITY OF NEW MEXICO

Thank you!

Contact: ajthor@unm.edu

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

SOCKS: Python Toolbox for Stochastic Optimal Control using Kernel Methods

This material is based upon work supported by the National Science Foundation under NSF Grant Number CNS-1836900. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
This research was supported in part by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in this article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

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.

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

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

- 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 $\Psi = [k(x_0^1, \cdot) l(u^1, \cdot), \ldots, k(x_0^M, \cdot) l(u^M, \cdot)]^\top$.

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

- Compute:

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

THE UNIVERSITY OF NEW MEXICO

- 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.

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

THE UNIVERSITY OF NEW MEXICO

- 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

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

  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:

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

THE UNIVERSITY OF NEW MEXICO

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.

THE UNIVERSITY OF NEW MEXICO