Stochastic Optimal Control & Safety Via Kernel Embeddings

A Data-Driven Approach

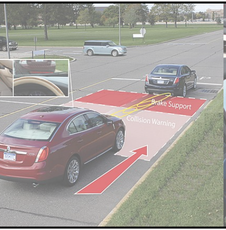
Adam Thorpe

University of New Mexico

2023

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.

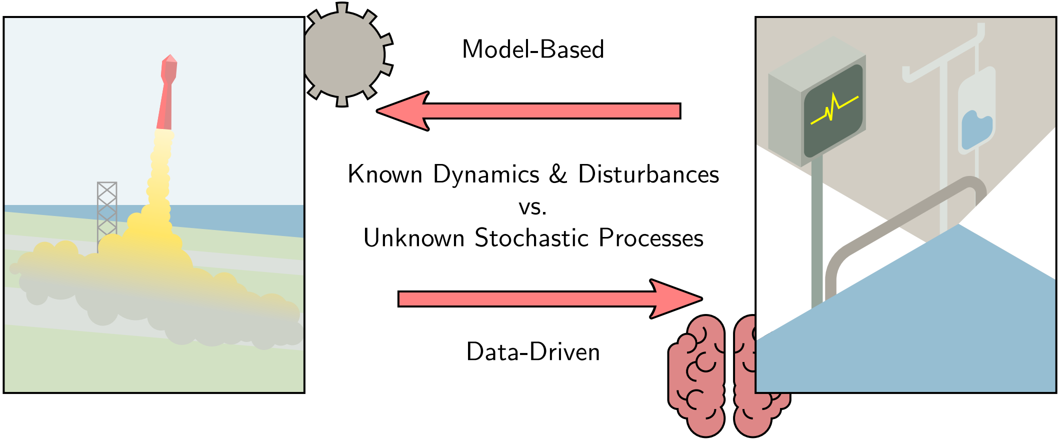
- Unknown dynamics & stochastic processes.
- Complex mechanical systems & dynamics.
- Unforeseen human factors.

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.



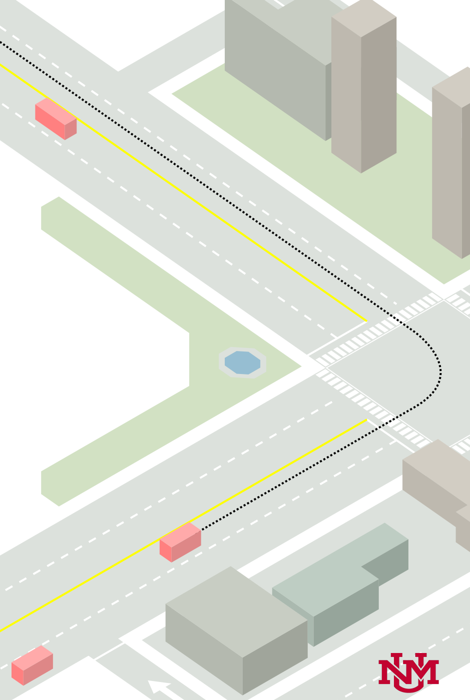
Stochastic Optimal Control

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

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

- Intractable since stochastic kernel is unknown – difficult!
- Model-based often limited to LTI, Gaussian disturbances.

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



THE UNIVERSITY OF NEW MEXICO

Our Proposed Approach

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

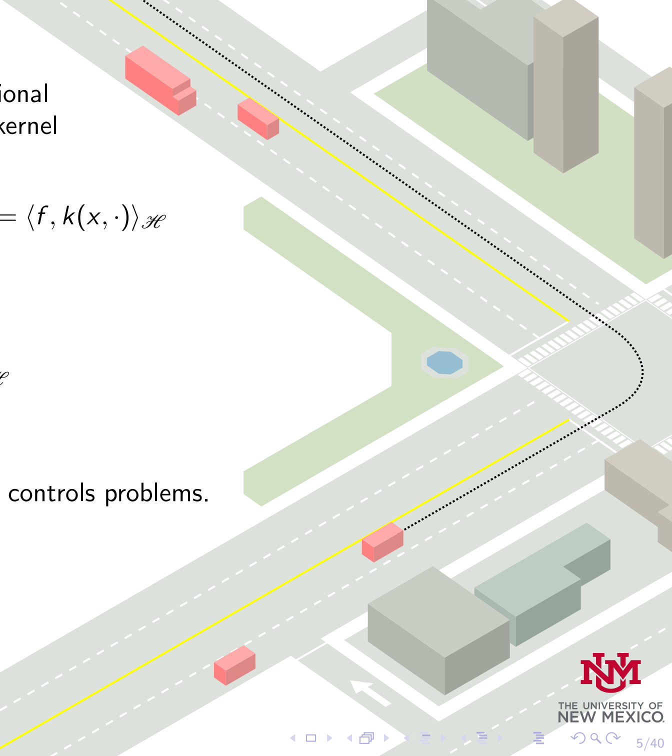
$$\text{Reproducing property:} \qquad f(x) = \langle f, k(x, \cdot) \rangle_{\mathscr{H}}$$

- 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

- Hilbert spaces are a natural fit for many controls problems.
- Nonparametric, i.e. few assumptions.
- Mathematically rigorous.



Thesis Contribution

1. I have developed **data-driven control techniques** using the theory of Hilbert spaces.
   1.1 Amenable to systems with nonlinear dynamics & arbitrary stochastic disturbances.
   1.2 Can solve stochastic optimal control problems as a linear program.

2. I have developed techniques to solve **approximate stochastic reachability** problems.
   2.1 Amenable to several stochastic reachability problems, including first-hitting time problem.
   2.2 I have developed **finite sample bounds for kernel-based techniques**.

3. I have developed techniques for **out-of-sample generalization**.
   3.1 Incorporating prior knowledge of the system dynamics or its properties.

THE UNIVERSITY OF NEW MEXICO

Contributions I

Journals

Thorpe, A. J., Ortiz, K. R., Oishi, M. M. K., State-Based Confidence Bounds for Data-Driven Stochastic Reachability Using Hilbert Space Embeddings. *Automatica,* 110146 (2022).

Vinod, A. P., **Thorpe, A. J.,** Olaniyi, P. A., Summers, T. H., Oishi, M. M. K., Sensor Selection for Dynamics-Driven User-Interface Design. *IEEE Transactions on Control Systems Technology,* 71–84 (2022).

Thorpe, A. J., Oishi, M. M. K., Model-Free Stochastic Reachability Using Kernel Distribution Embeddings. *IEEE Control Systems Letters,* 512–517 (2020).

Conferences

Thorpe, A. J., Djeumou, F., Neary, C., Oishi, M. M. K., Topcu, U., *Physics-Informed Kernel Embeddings: Integrating Prior System Knowledge with Data-Driven Control. 2023 IEEE Conference on Decision and Control* (Submitted) (2023).

Thorpe, A. J., Gonzales, J. A., Oishi, M. M. K., *Data-Driven Stochastic Optimal Control Using Kernel Gradients. 2023 American Control Conference* (Accepted) (2023).

Ortiz, K. R., **Thorpe, A. J.,** Perez, A., Luster, M., Pitts, B. J., Oishi, M., *Characterizing Within-Driver Variability in Driving Dynamics During Obstacle Avoidance Maneuvers. IFAC Workshop on Cyber-Physical & Human Systems* (2022).

THE UNIVERSITY OF NEW MEXICO

Contributions II

Thorpe, A., Lew, T., Oishi, M., Pavone, M., *Data-Driven Chance Constrained Control using Kernel Distribution Embeddings*. *Proceedings of The 4th Annual Learning for Dynamics and Control Conference* (2022), 790–802.

Thorpe, A. J., Oishi, M. M. K., *SOCKS: A Stochastic Optimal Control and Reachability Toolbox Using Kernel Methods*. *25th ACM International Conference on Hybrid Systems: Computation and Control* (2022).

Abate, A. *ARCH-COMP21 Category Report: Stochastic Models*. *8th International Workshop on Applied Verification of Continuous and Hybrid Systems (ARCH21)* (2021), 55–89.

Thorpe, A. J., Oishi, M. M. K., *Stochastic Optimal Control via Hilbert Space Embeddings of Distributions*. *2021 60th IEEE Conference on Decision and Control* (2021), 904–911.

Thorpe, A. J., Ortiz, K. R., Oishi, M. M. K., *Learning Approximate Forward Reachable Sets Using Separating Kernels*. *Proceedings of the 3rd Conference on Learning for Dynamics and Control* (2021), 201–212.

Thorpe, A. J., Ortiz, K. R., Oishi, M. M. K., *SReachTools Kernel Module: Data-Driven Stochastic Reachability Using Hilbert Space Embeddings of Distributions*. *2021 60th IEEE Conference on Decision and Control* (2021), 5073–5079.

Thorpe, A. J., Sivaramakrishnan, V., Oishi, M. M. K., *Approximate Stochastic Reachability for High Dimensional Systems*. *2021 American Control Conference* (2021), 1287–1293.

THE UNIVERSITY OF NEW MEXICO

Related Work

Dynamic Programming

- Puterman, M. L. (1994)
- Bertsekas, D. (1978)

Stochastic Reachability

- Summers, S. (2010)
- Abate, A. (2008)

Chance-Constrained Control

- Lew, T. (2020)
- Ono, M. (2015)

Kernel Embeddings

- Li, Z. (2022)
- Song, L. (2013)
- Grünewälder, S. (2012)
- Grünewälder, S. (2012)
- Nishiyama, Y. (2012)
- Berlinet, A. (2011)
- Song, L. (2010)
- Song, L. (2009)

Learning Methods

- Mauroy, A. (2020)
- Williams, C. K. (2006)

Infinite LPs

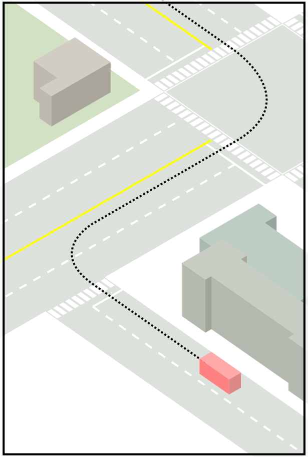
- Martinelli, A. (2022)
- Martinelli, A. (2022)

RKHS Control

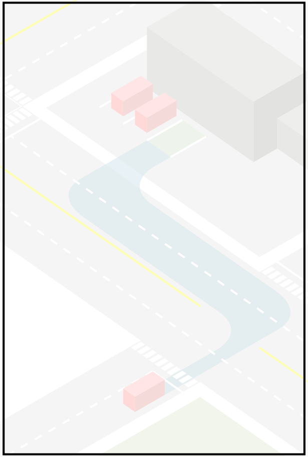
- Nemmour, Y. (2022)
- Vien, N. A. (2016)
- Lever, G. (2015)

Outline

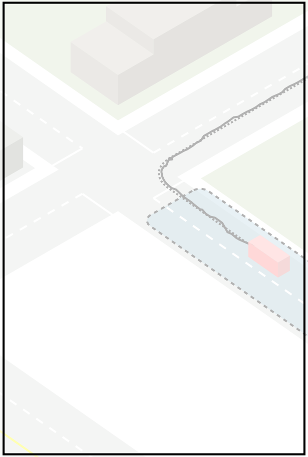
Stochastic Optimal Control



Stochastic Reachability



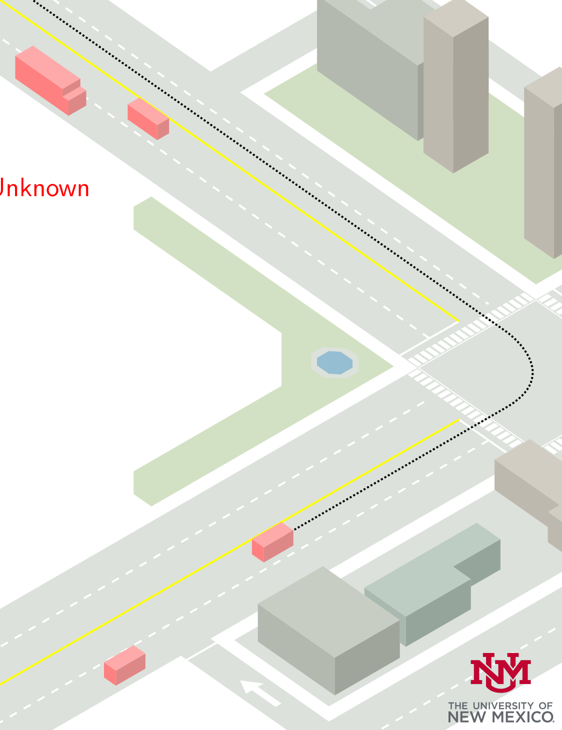
Using Prior Knowledge



THE UNIVERSITY OF NEW MEXICO

Stochastic Optimal Control

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



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

Challenges

- Computing expectations and probabilities.
- Policy synthesis.

[1] Here $x = \{x_1, \ldots, x_N\}$ and $u = \{u_0, \ldots, u_{N-1}\}$, policy is open loop.

Kernel Embeddings of Distributions

- Expectations (integrals) are linear operators.
- For $g \in \mathscr{H}$, the Riesz theorem guarantees the existence of the element $m(x_0, u) \in \mathscr{H}$ called the *kernel distribution embedding*, such that

$$\langle g, m(x_0, u) \rangle_{\mathscr{H}} = \mathbb{E}_{x \sim Q(\cdot | x_0, u)}[g(x)] = \int_{\mathcal{X}^N} g(x) Q(\mathrm{d}x \mid x_0, u).$$

- Empirical estimate $\hat{m}(x_0, u)$ of $m(x_0, u)$ using data,

Given: $\quad \mathcal{S} = \{(x_0^1, u^1, x^1), \ldots, (x_0^M, u^M, x^M)\}, \quad x^i \sim Q(\cdot \mid x_0^i, u^i)$

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

Main Idea

Step 1: Embed Q in an RKHS & estimate using data.

Step 2: Embed policy π in an RKHS.

THE UNIVERSITY OF NEW MEXICO

Data-Driven Stochastic Optimal Control Problem

- Write $\mathbb{P}(x \in F)$ as $\mathbb{E}[\mathbf{1}_F(x)]$, i.e. expectation of indicator function.
- Assume cost and constraints are in RKHS $\mathscr{H}$ and $\int_{\mathcal{X}^N} g(x) Q(\mathrm{d}x \mid x_0, u) < \infty$.

$$\begin{aligned} \min_{p(x_0) \in \mathscr{U}} \quad & \langle \langle g, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, p(x_0) \rangle_{\mathscr{U}} \\ \text{s.t.} \quad & \langle \langle \mathbf{1}_F, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, p(x_0) \rangle_{\mathscr{U}} \geq 1 - \delta \end{aligned}$$

Step 1: Embed Q in an RKHS $\mathscr{H}$ & estimate using data,

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

Step 2: Embed policy π in an RKHS $\mathscr{U}$,

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

Finite support

$\mathbb{R}$ coefficients

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

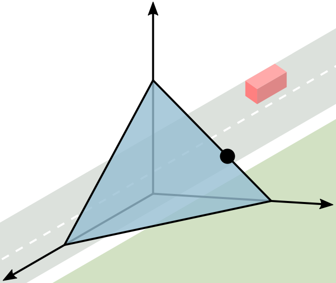
THE UNIVERSITY OF NEW MEXICO

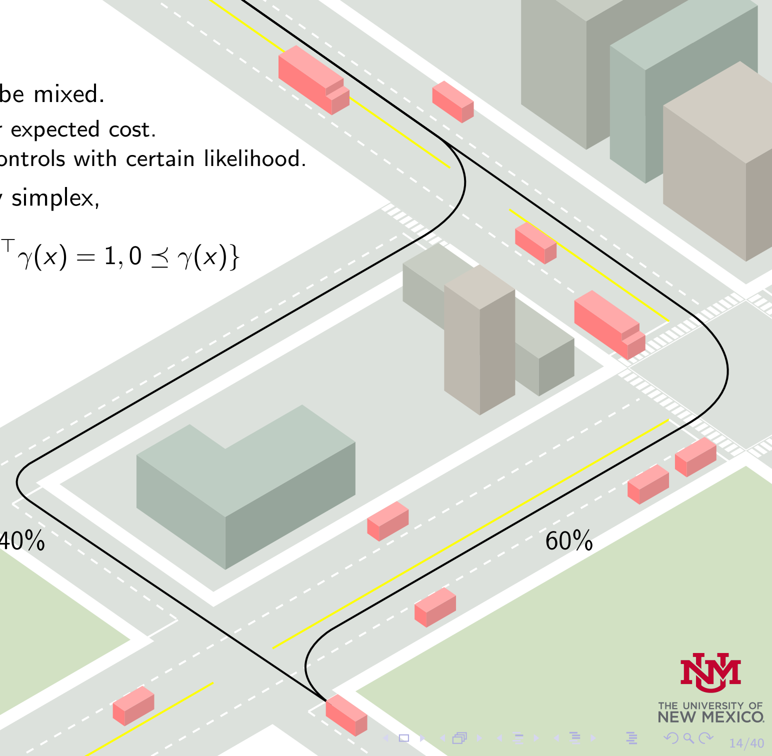
Structure of the Policy

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

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

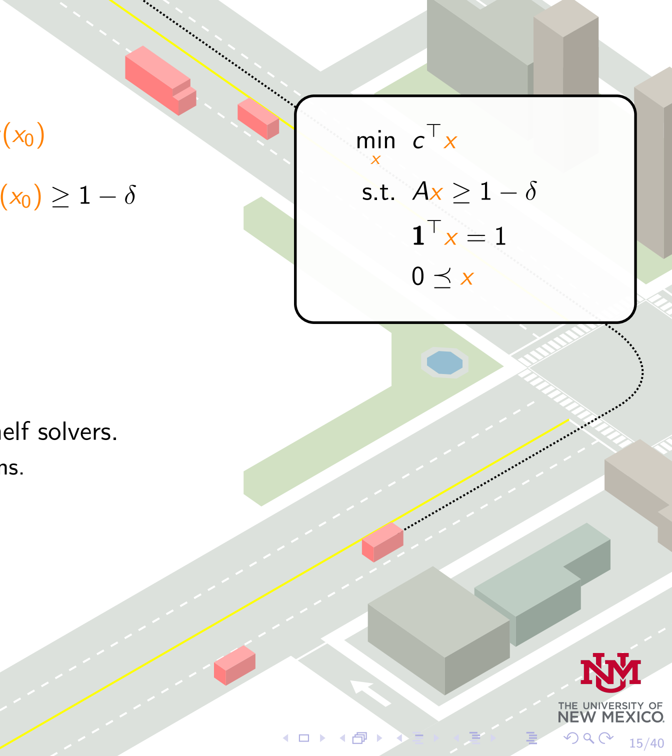
- Solution may be on an "edge" of the simplex, i.e. mixed.





Approximate Problem

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



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

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

However, ...

- One potential drawback:
  Must strategically choose support $\tilde{\omega}^j$.

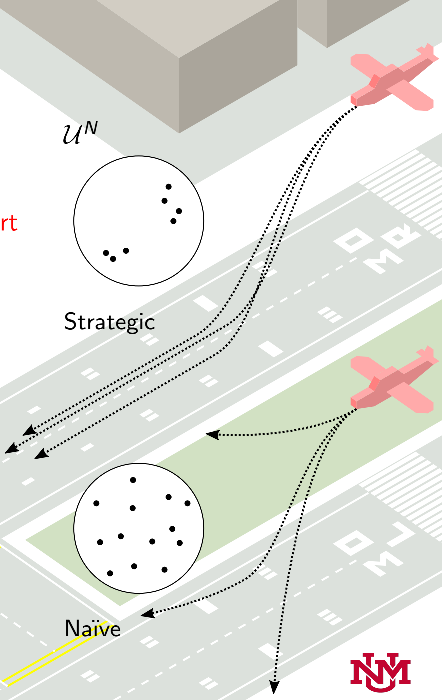
Strategic Sampling

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

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

Finite support

- Naïve sampling insufficient.
  - May fail to generate meaningful control sequences.
- Strategic sampling needed in practice, e.g. using:
  - PD controller.
  - Motion primitives.
- We obtain good candidate controls.



THE UNIVERSITY OF NEW MEXICO

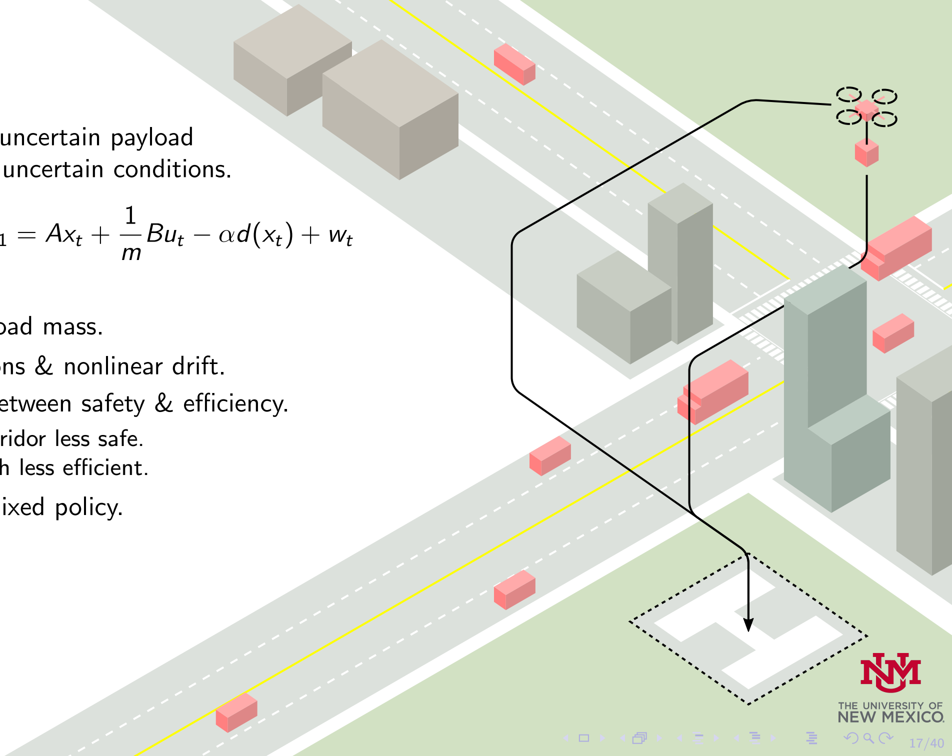
Demonstration

Scenario

Quadrotor carrying uncertain payload around obstacles in uncertain conditions.

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

- 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.

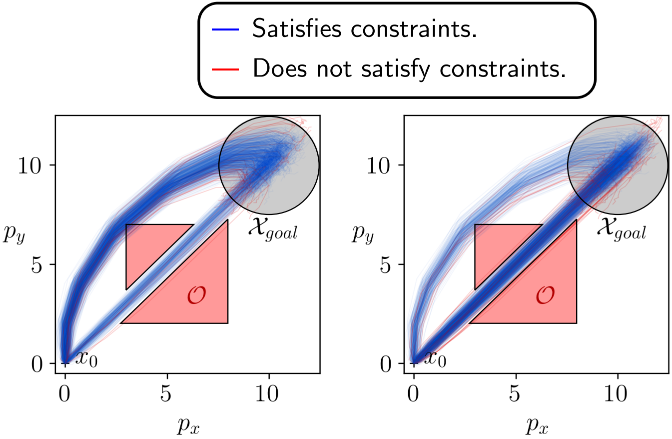


THE UNIVERSITY OF NEW MEXICO

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: $\mathcal{O}(M^3)$



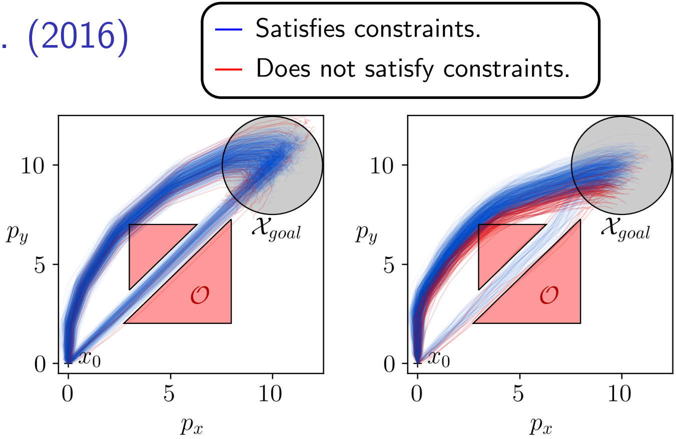
- Higher δ leads to a policy that has a higher chance of choosing "risky" middle corridor.

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

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

THE UNIVERSITY OF NEW MEXICO

Comparison with Ono et. al. (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 may 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

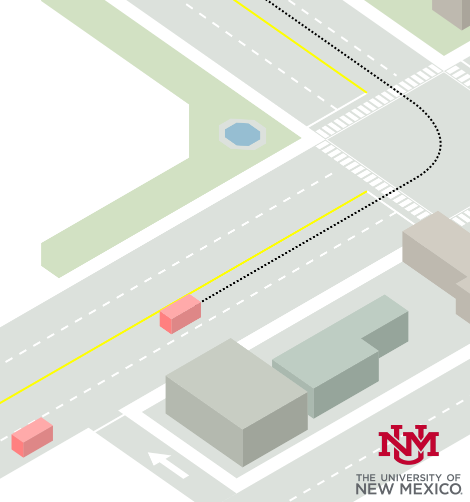
THE UNIVERSITY OF NEW MEXICO

Gradient Descent in an RKHS

- Consider data-driven problem with *convex* cost g:

$$\min_{u \in \mathcal{U}} \quad \langle \langle g, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, l(u, \cdot) \rangle_{\mathscr{U}}$$

given estimate $\hat{m}$, initial guess u_0
repeat
$\Delta u_n \leftarrow \langle \mathbf{g}^\top W \Psi k(x_0, \cdot), \partial^{1,0} l(u_n, \cdot) \rangle_{\mathscr{U}}$
choose step size η
$u_{n+1} \leftarrow u_n - \eta \Delta u_n$
until stopping criterion satisfied



Challenge

- Problem is generally non-convex if kernel l is non-convex.
- Use partial derivative reproducing property of RKHS,

$$\partial^1 h(u) = \langle h, \partial^{1,0} l(u, \cdot) \rangle_{\mathscr{U}}.$$

- Compute gradient of approx. expected cost w/rt. u as,

$$\langle \langle g, \hat{m}(x_0, \cdot) \rangle_{\mathscr{H}}, \partial^{1,0} l(u, \cdot) \rangle_{\mathscr{U}}.$$

- **Does not rely upon a sample-based approach.**

Demonstration

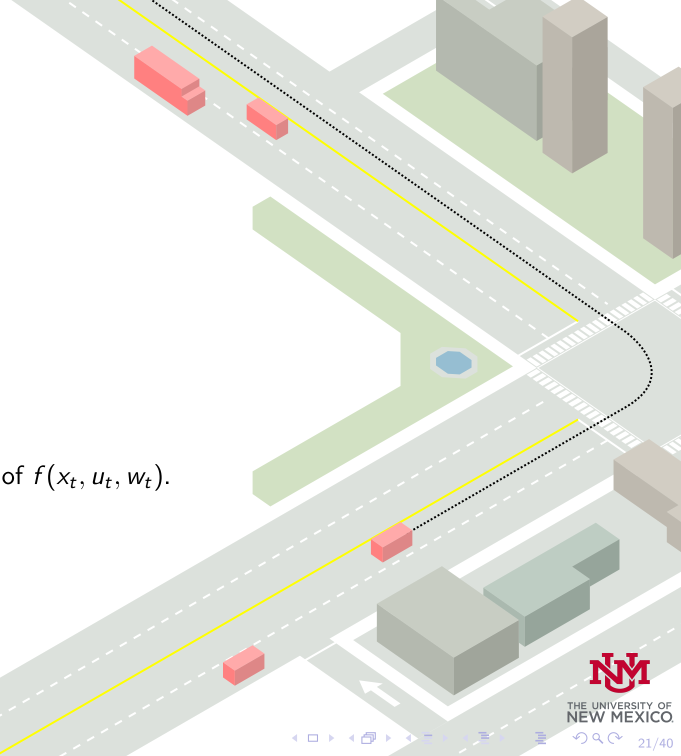
Scenario

Nonholonomic vehicle tracking
a target trajectory (waypoints).

$$\dot{x}_1 = u_1 \sin(x_3)$$
$$\dot{x}_2 = u_1 \cos(x_3)$$
$$\dot{x}_3 = u_2$$

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

- Unconstrained.
- Stochastic kernel Q is a representation of $f(x_t, u_t, w_t)$. (dynamics over a single time step)



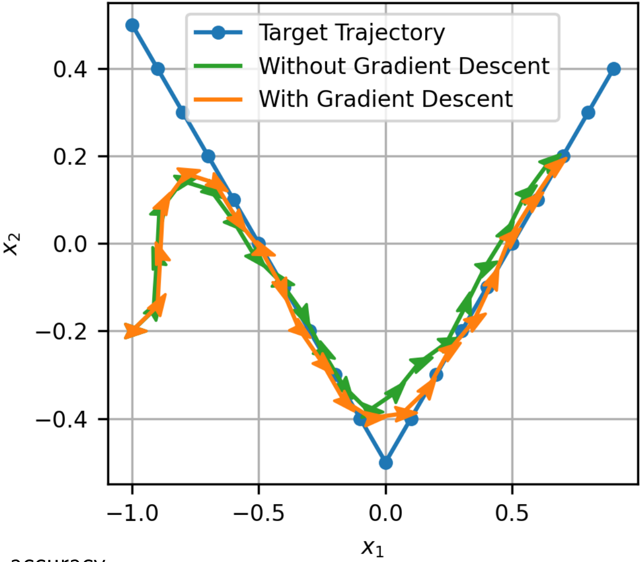
THE UNIVERSITY OF NEW MEXICO

Results

- Sample size: $M = 3{,}000$, $P = 210$
- Time horizon: $N = 20$

- Generating sample: $\approx$ 5s
- Generating controls: $\approx$ 5s
- Computing LP: $\approx$ 100 ms
- Computing GD Solution: $\approx$ 6 s
  - Step size: 0.1
  - Iterations: 100

- Gradient solution performs better, but with higher computational cost.



- Tradeoff between computation time & accuracy.
- Other gradient-based convex optimization techniques may be applicable.
  - E.g. Newton-Raphson, interior point.

THE UNIVERSITY OF NEW MEXICO

Dynamic Programming

- Consider the following optimization problem,

$$\min_{\pi \in \Pi} \quad \mathbb{E}\left[g_N(x_N) + \sum_{t=0}^{N-1} g_t(x_t, u_t) \right]$$

- Define the value functions $V_t^* : \mathcal{X} \to \mathbb{R}$, $t = 0, 1, \dots, N-1$,

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

  initialized with $V_N^*(x) = g_N(x)$.
- Solve for π backwards in time using dynamic programming.

Challenge

- Stochastic kernel Q is unknown.
- Curse of dimensionality.

[1]Note that Q here is over a single time step.

Approximate Dynamic Programming Using Kernels

Step 1: Embed Q in an RKHS & estimate using data,

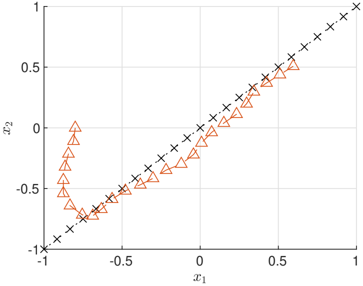
$$\langle \tilde{V}^*_{t+1}, \hat{m}(x,u) \rangle_{\mathscr{H}} \approx \int_{\mathcal{X}} \tilde{V}^*_{t+1}(y) Q(\mathrm{d}y \mid x, u).$$

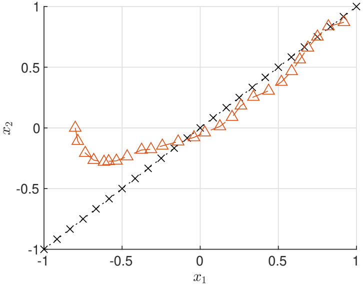
Step 2: Represent policy π as embeddings in an RKHS,

$$p_t(x) = \sum_{j=1}^{P} \gamma_j(x) l(\tilde{u}_j, \cdot).$$

Step 3: Solve for $\pi = \{p_0, p_1, \ldots, p_{N-1}\}$ by recursively approximating & substituting value functions,

$$\tilde{V}^*_N(x) = g_N(x)$$
$$\tilde{V}^*_t(x) = \min_{\gamma(x) \in \mathbb{R}^P} \langle g_t(x, \cdot), p_t(x) \rangle_{\mathscr{U}} + \langle \langle \tilde{V}^*_{t+1}, \hat{m}(x, \cdot) \rangle_{\mathscr{H}}, p_t(x) \rangle_{\mathscr{U}}$$



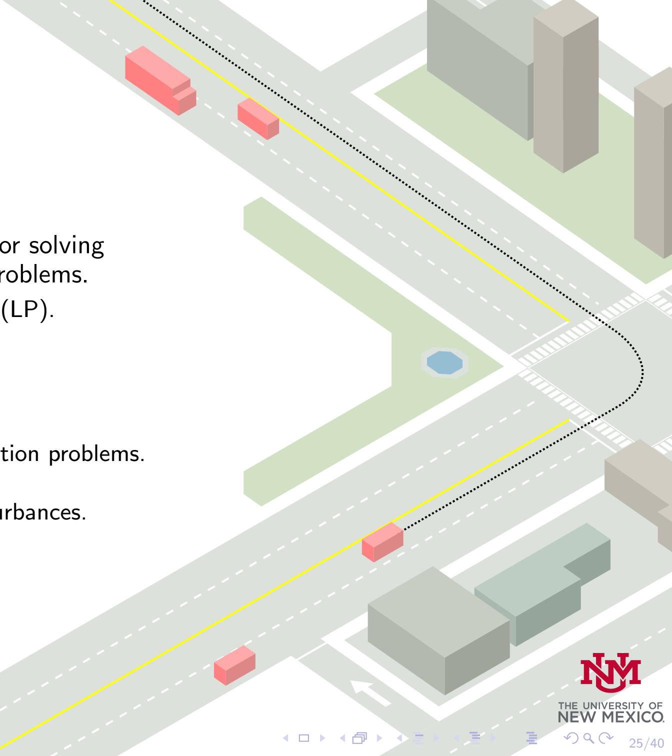


- **Can solve as an LP.**
- Complexity scales with sample size and is generally $\mathcal{O}(M^3)$, *not* with dimensionality of data.

Recap

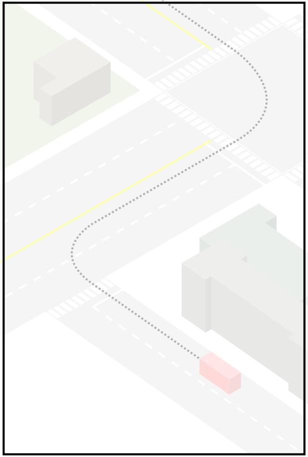
- We have demonstrated two techniques for solving data-driven stochastic optimal control problems.
    1. Compute a policy as a linear program (LP).
    2. Kernel gradient-based optimization.

- These techniques are amenable to:
    - Constrained & unconstrained optimization problems.
    - Dynamic programming.
    - Systems with arbitrary stochastic disturbances.
    - Computing mixed policies.

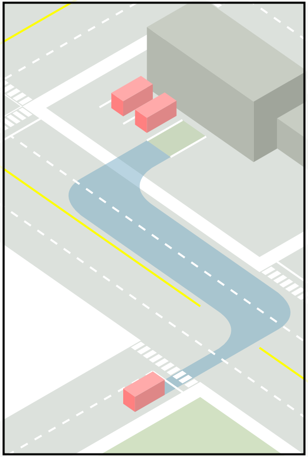


Outline

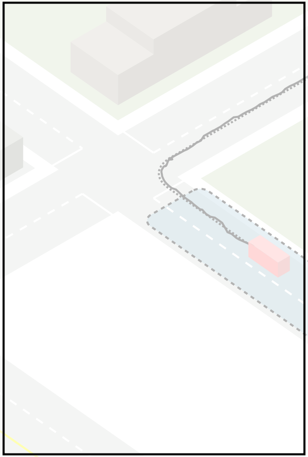
Stochastic Optimal Control



Stochastic Reachability



Using Prior Knowledge



Stochastic Reachability (Terminal-Hitting Time Problem)

- Safety probabilities: 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 $t < N$.

Multiplicative cost

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

- Define the value functions $V_t^{\pi} : \mathcal{X} \to \mathbb{R}$, $t = 0, 1, \dots, N-1$,

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

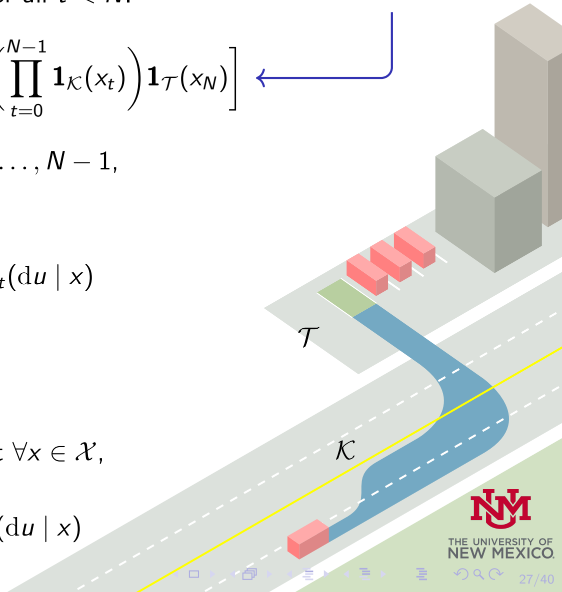
$$V_t^{\pi}(x) = \mathbf{1}_{\mathcal{K}}(x) \int_{\mathcal{U}} \int_{\mathcal{X}} V_{t+1}^{\pi}(y) Q(\mathrm{d}y \mid x, u) \pi_t(\mathrm{d}u \mid x)$$

$$\Downarrow$$

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

- Goal: find $\pi^* = \{\pi_0^*, \pi_1^*, \dots, \pi_{N-1}^*\} \in \Pi$ such that $\forall x \in \mathcal{X}$,

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



THE UNIVERSITY OF NEW MEXICO

Approximate Stochastic Reachability

- Can use similar procedure as before:

  Step 1: Embed Q in an RKHS & estimate using data,

$$\langle \tilde{V}_{t+1}^*, \hat{m}(x, u) \rangle_{\mathscr{H}} \approx \int_{\mathcal{X}} \tilde{V}_{t+1}^*(y) Q(\mathrm{d}y \mid x, u).$$

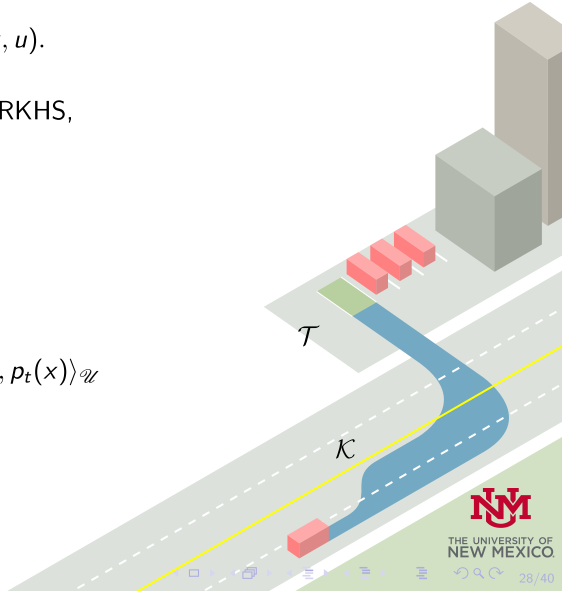
  Step 2: Represent policy π^* as embeddings in an RKHS,

$$p_t^*(x) = \sum_{j=1}^{P} \gamma_j(x) l(\tilde{u}_j, \cdot).$$

  Step 3: Solve for $p_0^*, p_1^*, \ldots, p_{N-1}^*$ by recursively approximating & substituting value functions,

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

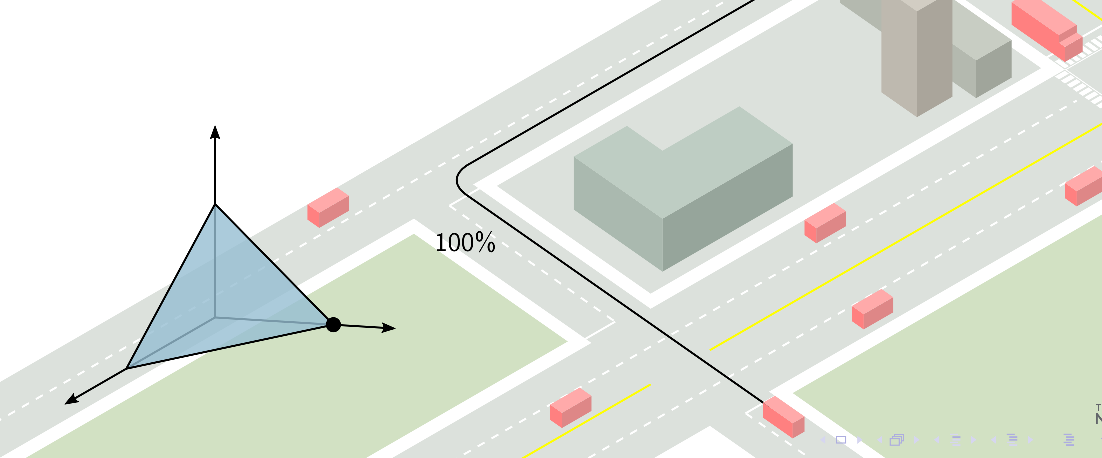
- **Can solve as an LP.**



THE UNIVERSITY OF NEW MEXICO

Structure of the Policy

- Optimal stochastic policy is deterministic.
  - Deterministic since problem is unconstrained.
  - Means policy assigns probability mass one to a single input.
- Coefficients $\gamma(x)$ in probability simplex $\mathscr{S}$.
- Solution is at a "vertex", i.e. deterministic.
- Can be solved efficiently via the Lagrangian dual.



THE UNIVERSITY OF NEW MEXICO

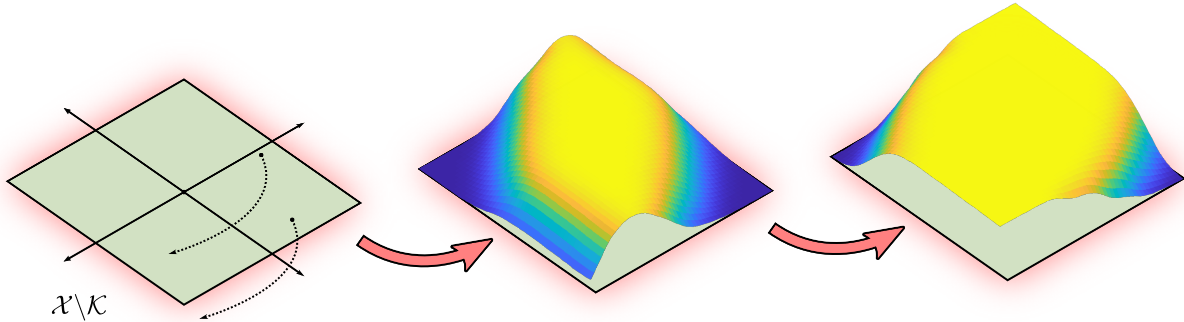
Demonstration

- Point mass system (double integrator).
  - Chosen for validation.

$$x_{t+1} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix} x_t + \begin{bmatrix} T_s^2/2 \\ T_s \end{bmatrix} u_t + w_t$$

- Goal is to stay within a small region around the origin.

$$r_{x_0}^{\pi}(\mathcal{K}, \mathcal{T}) = \mathbb{P}(x_N \in \mathcal{T} \wedge x_t \in \mathcal{K}, \forall t < N)$$



Map initial conditions to safety probabilities.

Compute maximal policy.

Results

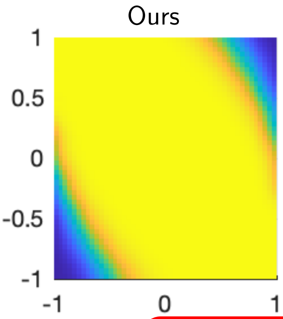
- Sample size: $M = 2{,}500$, $P = 1{,}000$
- Time horizon: $N = 15$

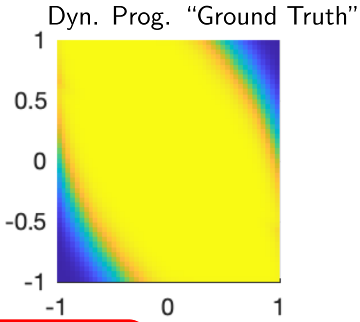
- Generating sample: $\approx$ 1s
- Computing LP: $\approx$ 100 ms
- Scales with sample size: $\mathcal{O}(M^3)$

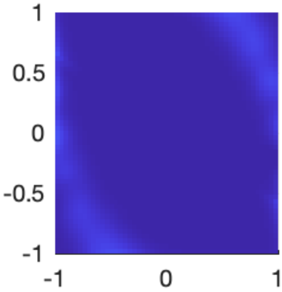
- Computing Dyn. Prog.: $\approx$ 30 s
- Dynamic program scales with grid size.

- Maximum absolute error: $\approx 10\%$

- Warmer colors denote higher safety probability.
- Maximally safe policy has better performance.
- Can also be used for the first-hitting time problem.
  - Traditionally very difficult to solve.
  - Probabilistic safe sets non-convex, meaning techniques using polytopic representations may fail.



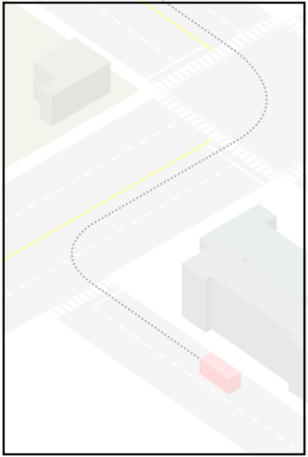




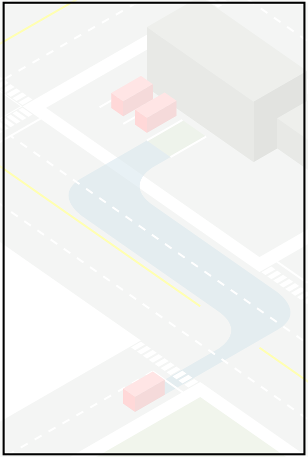
Absolute error.

Outline

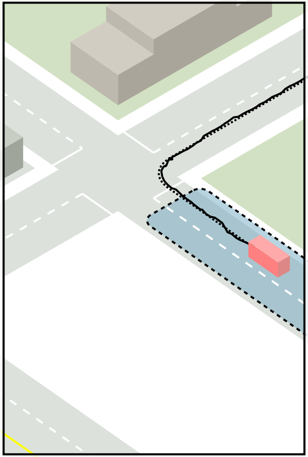
Stochastic Optimal Control



Stochastic Reachability



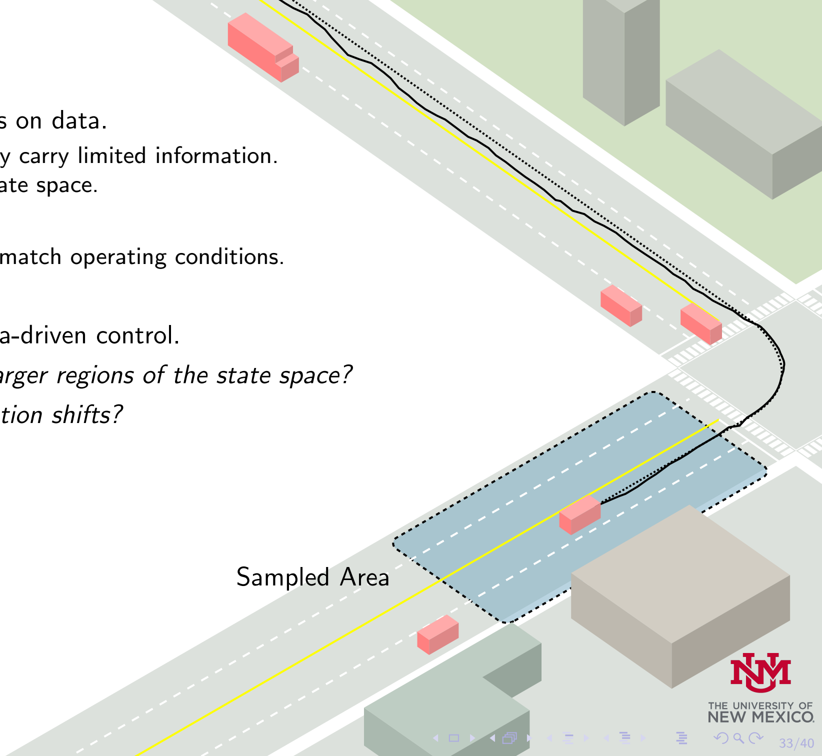
Using Prior Knowledge



Existing Challenges

- Quality of solution depends on data.
  - Finite samples inherently carry limited information.
  - May not cover entire state space.
- Out-of-distribution data.
  - Training data does not match operating conditions.

- Challenges apply to all data-driven control.
- *How do we generalize to larger regions of the state space?*
- *How do we handle distribution shifts?*

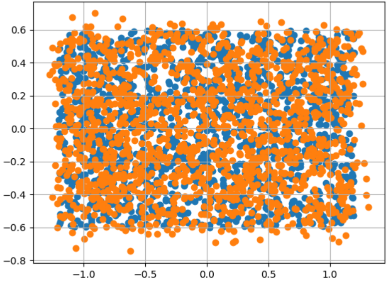


THE UNIVERSITY OF NEW MEXICO

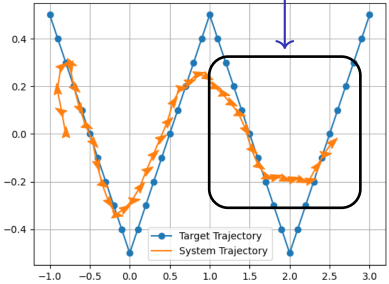
Out-Of-Sample Generalization

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

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



Outside sampled region



- Quality of solution depends on data.

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

Penalty term

- Penalizes solutions far from 0.
- Outside region we have data, control exhibits poor performance.

THE UNIVERSITY OF NEW MEXICO

Biased Regularization

- Given:
  - Known *approximate* dynamics: $x_{t+1} = \tilde{f}(x_t, u_t)$
  - Sample from *true* dynamics:

$$\mathcal{S} = \{(x^1, u^1, y^1), \dots, (x^M, u^M, y^M)\}, \quad y^i \sim Q(\cdot \mid x^i, u^i)$$

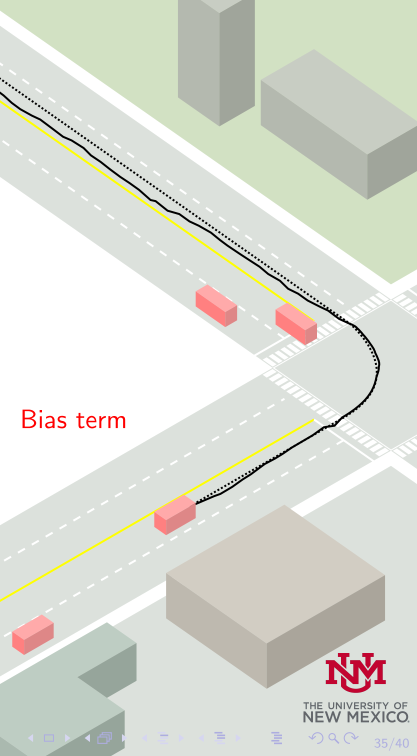
- Compute:

$$\underbrace{\mathbb{E}[g(y) - g(\tilde{f}(x, u))]}_{\text{Empirically estimate using data.}} + g(\tilde{f}(x, u))$$

- Biased regularization:

$$\hat{m} = \arg\min_{f \in \mathcal{V}} \frac{1}{2\lambda} \sum_{i=1}^{M} \| k(y^i, \cdot) - f(x^i, u^i) \|_{\mathscr{H}}^2 + \frac{1}{2} \| f \|_{\mathcal{V}}^2 \overbrace{- \langle f, f_0 \rangle_{\mathcal{V}}}^{\text{Bias term}}$$

- Define $f_0(x, u) = k(\tilde{f}(x, u), \cdot)$.
- Penalizes solutions far from f_0.

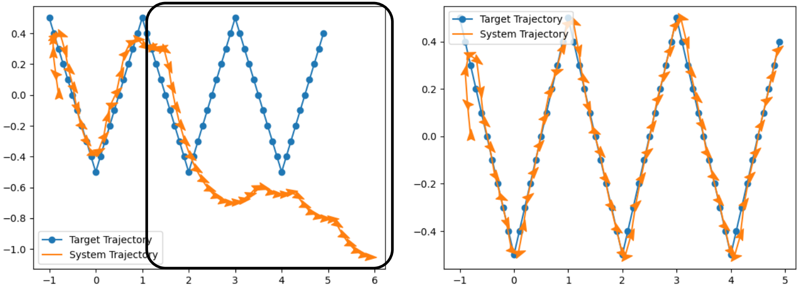


Comparison of Biased vs. Unbiased Regression

- *Can we use prior knowledge to improve data-driven control performance?*
- We often have at least approximate dynamical knowledge of the system.

Outside sampled region.

- Sample size: $M = 100$
- Time horizon: $N = 60$
- Generating sample: $\approx$ 1s
- Computing LP: $\approx$ 10 ms



- Main idea: Use approximate dynamics as our regression baseline.
- Much better generalization performance with smaller sample size.

THE UNIVERSITY OF NEW MEXICO

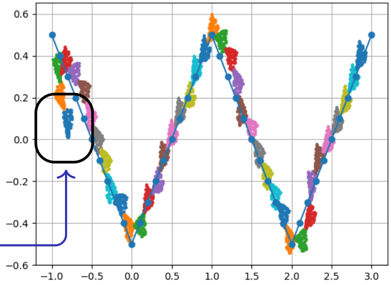
Exploiting Known System Properties

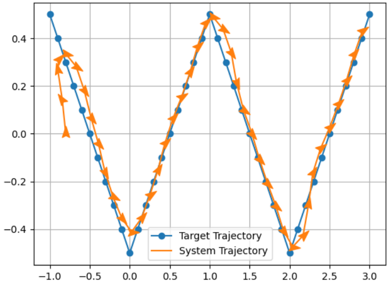
- *Can we use known properties of the dynamics to improve data-driven control performance?*
- E.g. symmetry, knowledge of the vector field, $SE(n)$-**invariance**.

- $SE(n)$-**invariance:** Dynamics are the same in all frames of reference.

- 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 evolves.
- Much better performance with smaller sample size.
- Uses physics-informed learning, side information.

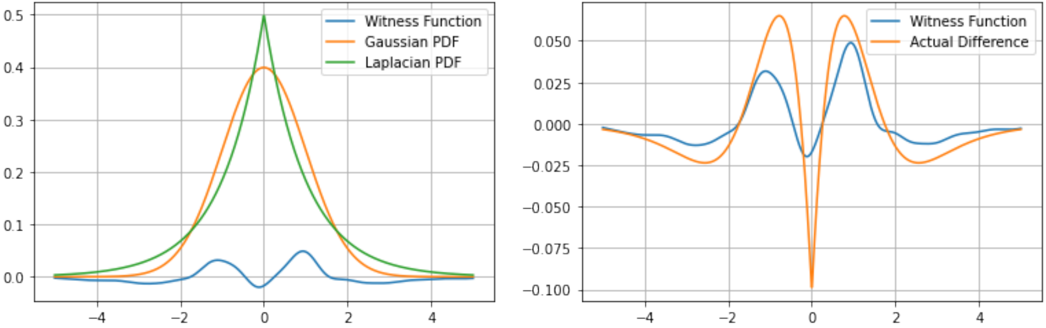
THE UNIVERSITY OF NEW MEXICO

Out-Of-Distribution (OOD) Detection

- *Given samples from two distributions $\mathbb{P}$ and $\mathbb{Q}$, how can we determine if $\mathbb{P}$ and $\mathbb{Q}$ differ?*
- Maximum mean discrepancy:

$$\mathrm{MMD}^2(\mathbb{P},\mathbb{Q}) = \|m_{\mathbb{P}} - m_{\mathbb{Q}}\|^2_{\mathscr{H}} = \left[\sup_{\|f\|_{\mathscr{H}}\le 1} \left(\mathbb{E}_{\mathbb{P}}[f(X)] - \mathbb{E}_{\mathbb{Q}}[f(Y)]\right)\right]^2$$

- Compare the distance between *empirical* kernel distribution embeddings in an RKHS.
- Example: Gaussian & Laplacian distribution (same mean & variance).

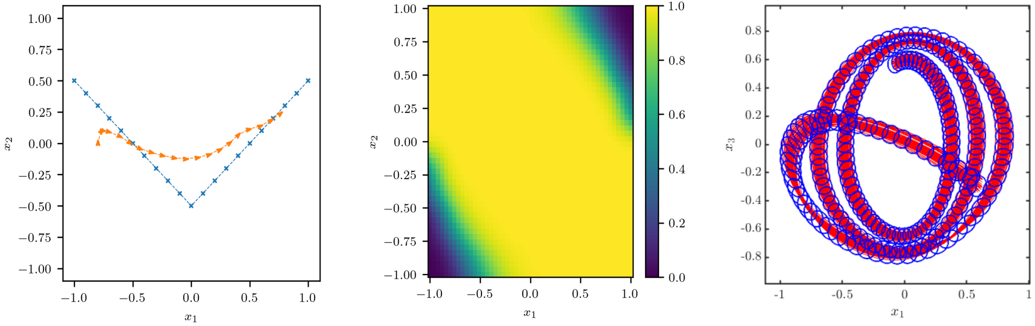


THE UNIVERSITY OF NEW MEXICO

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`

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.

Finite Sample Bounds

- Finite sample bounds via bounded variation.

Theorem

For any value function $V_t \in \mathscr{H}$, given $\delta/2 \in (0,1)$, with probability $1-\delta/2$, the difference between the true and empirical expectation of the value functions is bounded by:

$$|\langle V_t, m(x,u) - \hat{m}(x,u)\rangle_{\mathscr{H}}| \leq 2\sqrt{\mathrm{tr}(\boldsymbol{\beta}^\top \Psi \Psi^\top \boldsymbol{\beta}} + 3\sqrt{\frac{\rho^2 \log(2/\delta)}{2\lambda^2 M}} \tag{1}$$

- Outline of proof:

  Step 1: Bound the worst-case difference between the true and empirical expectation.

  Step 2: Use McDiarmid's inequality & the matrix inversion lemma, and bound the variation of changing a single observation by $\rho/(\lambda M)$, where $\rho < \infty$ is an upper bound on the kernel.

  Step 3: Use a symmetrization argument and a *ghost sample* (an independent copy of $\mathcal{S}$), and the definition of the dual norm in the RKHS.

  Step 4: Then, via simple substitution, we obtain the bound in (1).

THE UNIVERSITY OF NEW MEXICO

Finite Sample Bounds (Continued)

- Finite sample bounds via algorithmic stability.
- Bound the difference between risk & empirical risk:

$$R(\hat{m}) = \int_{\mathcal{X}} \|k(y,\cdot) - \hat{m}(x,u)\|^2_{\mathscr{H}} Q(\mathrm{d}y \mid x,u)$$

$$R_{\mathcal{S}}(\hat{m}) = \frac{1}{M}\sum_{i=1}^{M} \|k(y_i,\cdot) - \hat{m}(x_i,u_i)\|^2_{\mathscr{H}} + \lambda\|\hat{m}\|^2_{\mathcal{V}}$$

Theorem

The RLS algorithm has uniform stability $\alpha \leq \sigma^2\rho^2/(2\lambda M)$ *with respect to the loss function* $0 \leq \|k(y,\cdot) - \hat{m}(x,u)\|^2_{\mathscr{H}} \leq \rho$ *for all* $(x,u,y) \in \mathcal{Z}$ *and all sets* $\mathcal{S} = \mathcal{Z}^M$. *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* $\mathcal{S}$,

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

[1] Note that better bounds have been developed in Li, et. al. (2022) that show finite sample bounds of conditional embeddings and derive a minimax optimal rate of $\mathcal{O}(M^{-1/2})$.

Embeddings of Distributions

- Given a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $\exists$ a corresponding RKHS $\mathscr{H}$ of functions from $\mathcal{X}$ to $\mathbb{R}$.
- Expectations are linear in the function argument:

$$\mathbb{E}[f(y)] = \int_{\mathcal{X}} f(y)\mathbb{P}(\mathrm{d}y).$$

- Assuming $f \in \mathscr{H}$, by the Riesz theorem, there exists an element $m \in \mathscr{H}$ such that

$$\begin{aligned}
\langle f, m \rangle_{\mathscr{H}} &= \left\langle f, \int_{\mathcal{X}} k(y, \cdot)\mathbb{P}(\mathrm{d}y) \right\rangle_{\mathscr{H}} \\
&= \int_{\mathcal{X}} \langle f, k(y, \cdot)\rangle_{\mathscr{H}} \mathbb{P}(\mathrm{d}y) \\
&= \int_{\mathcal{X}} f(y)\mathbb{P}(\mathrm{d}y)
\end{aligned}$$

- Kernel k must be measurable and bounded to ensure integral exists.
- Also applicable to conditional distributions, stochastic kernels.