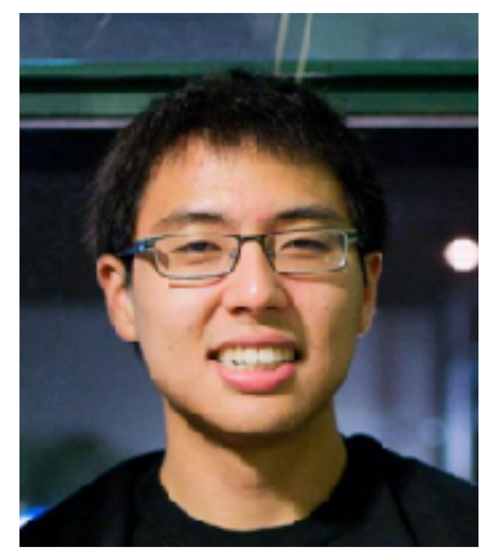
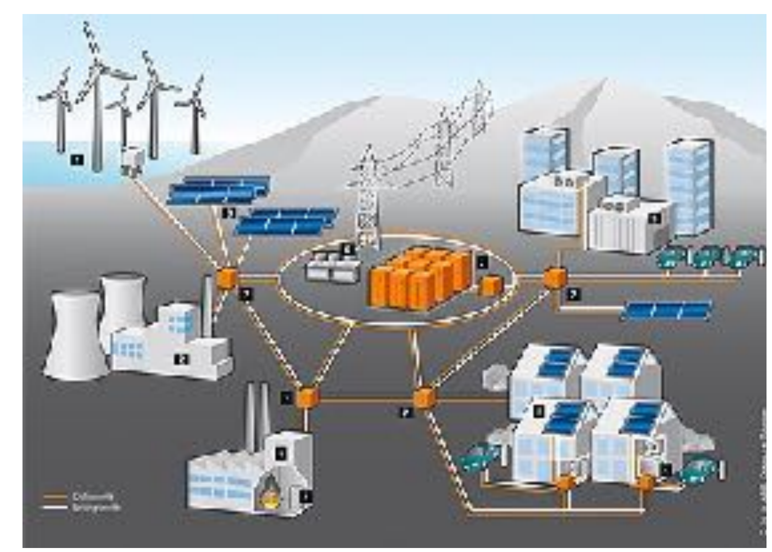
learning to control the linear quadratic regulator

Benjamin Recht
University of California, Berkeley

Collaborators

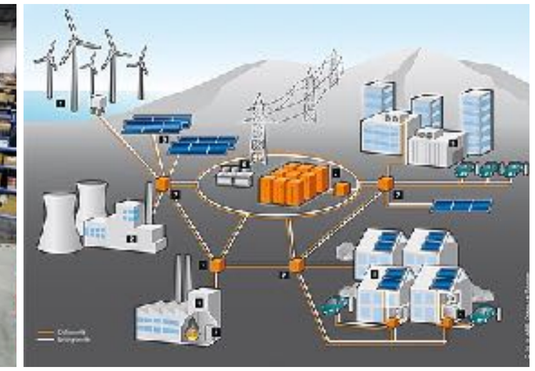


Joint work with Sarah Dean, Horia Mania, Nikolai Matni, Max Simchowitz, and Stephen Tu.



trustable, scalable, predictable





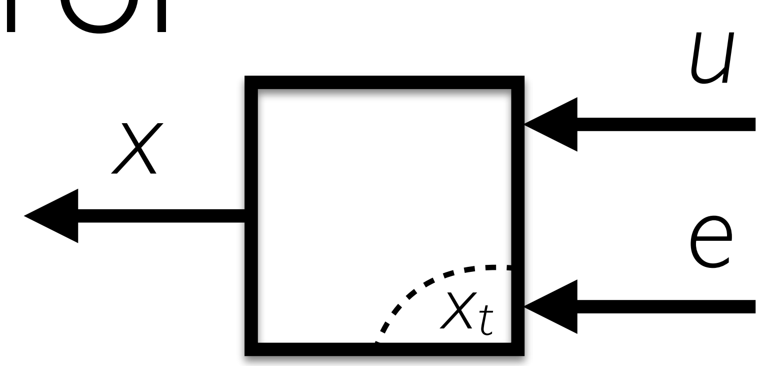
What are the fundamental limits of learning systems that interact with the physical environment?

How well must we understand a system in order to control it?

theoretical foundations

- statistical learning theory
- robust control theory
- core optimization

Optimal control



$$\begin{array}{ll} \text{minimize} & \mathbb{E}_{\mathrm{e}}\left[\sum_{t=1}^{T} C_t(x_t, u_t)\right] \\ \text{s.t.} & x_{t+1} = f_t(x_t, u_t, \mathrm{e}_t) \\ & u_t = \pi_t(\tau_t) \end{array}$$

C_t is the *cost.* If you maximize, it's called a *reward.*

x_t is the state, u_t is the input, e_t is a noise process

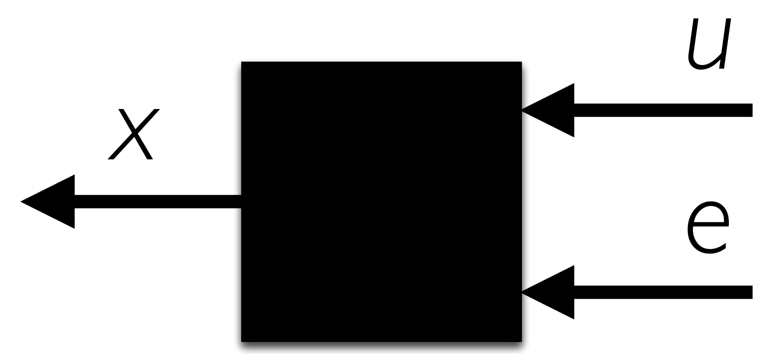
f_t is the state-transition function

$\tau_t = (u_1, \ldots, u_{t-1}, x_0, \ldots, x_t)$ is an observed *trajectory*

$\pi_t(\tau_t)$ is the *policy.* This is the <u>optimization decision variable</u>.

Learning to control

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}_e \left[\sum_{t=1}^{T} C_t(x_t, u_t) \right] \\ \text{s.t.} \quad & x_{t+1} = f_t(x_t, u_t, e_t) \\ & u_t = \pi_t(\tau_t) \end{aligned}$$



C_t is the *cost.* If you maximize, it's called a *reward.*

x_t is the state, u_t is the input, e_t is a noise process

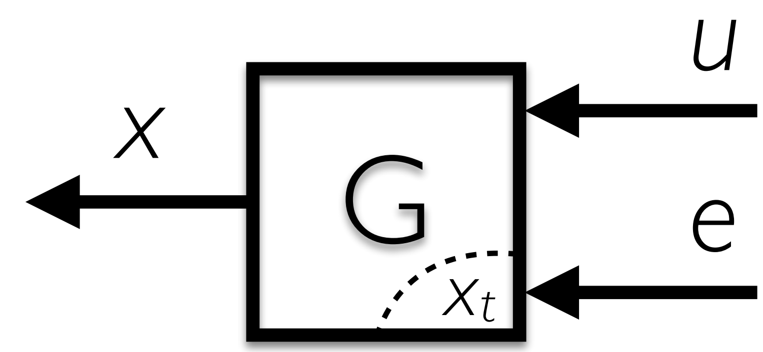
f_t is the state-transition function *unknown!*

$\tau_t = (u_1, \ldots, u_{t-1}, x_0, \ldots, x_t)$ is an observed *trajectory*

$\pi_t(\tau_t)$ is the *policy.* This is the optimization decision variable.

Perennial challenge: how to perform optimal control when the system is unknown?

RL Triopoly



approximate dynamic programming

minimize $\mathbb{E}_e\left[\sum_{t=1}^{T} C_t(x_t, u_t)\right]$

s.t. $x_{t+1} = f_t(x_t, u_t, e_t)$ model-based

$u_t = \pi_t(\tau_t)$ direct policy search

How to solve optimal control when the model *f* is unknown?

- **Model-based:** fit model from data
- **Model-free**
  - **Approximate dynamic programming:** estimate cost from data
  - **Direct policy search:** search for actions from data

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_e \left[\sum_{t=1}^{T} C_t(x_t, u_t) \right] \\ \text{s.t.} & x_{t+1} = f(x_t, u_t, e_t) \\ & u_t = \pi_t(\tau_t) \end{array}$$

Model-based RL

Collect some simulation data. Should have $\boxed{x_{t+1} \approx \varphi(x_t, u_t) + \nu_t}$

Fit dynamics with *supervised learning:* $\boxed{\hat{\varphi} = \arg\min_{\varphi} \sum_{t=0}^{T-1} ||x_{t+1} - \varphi(x_t, u_t)||^2}$

Solve approximate problem:

$$\boxed{\begin{array}{ll} \text{minimize} & \mathbb{E}_\omega \left[\sum_{t=1}^{T} C_t(x_t, u_t) \right] \\ \text{s.t.} & x_{t+1} = \varphi(x_t, u_t) + \omega_t \\ & u_t = \pi(\tau_t) \end{array}}$$

"Simplest" Example: LQR

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\ \text{s.t.} & x_{t+1} = A x_t + B u_t + e_t \end{array}$$

- Optimization simplicity
- Elegant Dynamic Programming solutions
- Exact solution for baseline
- Natural robustness
- Broadly applicable as is
- Core of many MPC and nonlinear control methods
- Useful model for sensorimotor modeling

"Simplest" Example: LQR

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\ \text{s.t.} & x_{t+1} = A x_t + B u_t + e_t \end{array}$$

Oracle: You can generate N trajectories of length T.

Challenge: Build a controller with smallest error with fixed sampling budget (N x T).

What is the optimal estimation/design scheme?

How many samples are needed for near optimal control?

Model-based LQR

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\ \text{s.t.} & x_{t+1} = A x_t + B u_t + e_t \end{array}$$

Collect some simulation data. Will have $\boxed{x_{t+1} = A x_t + B u_t + e_t}$

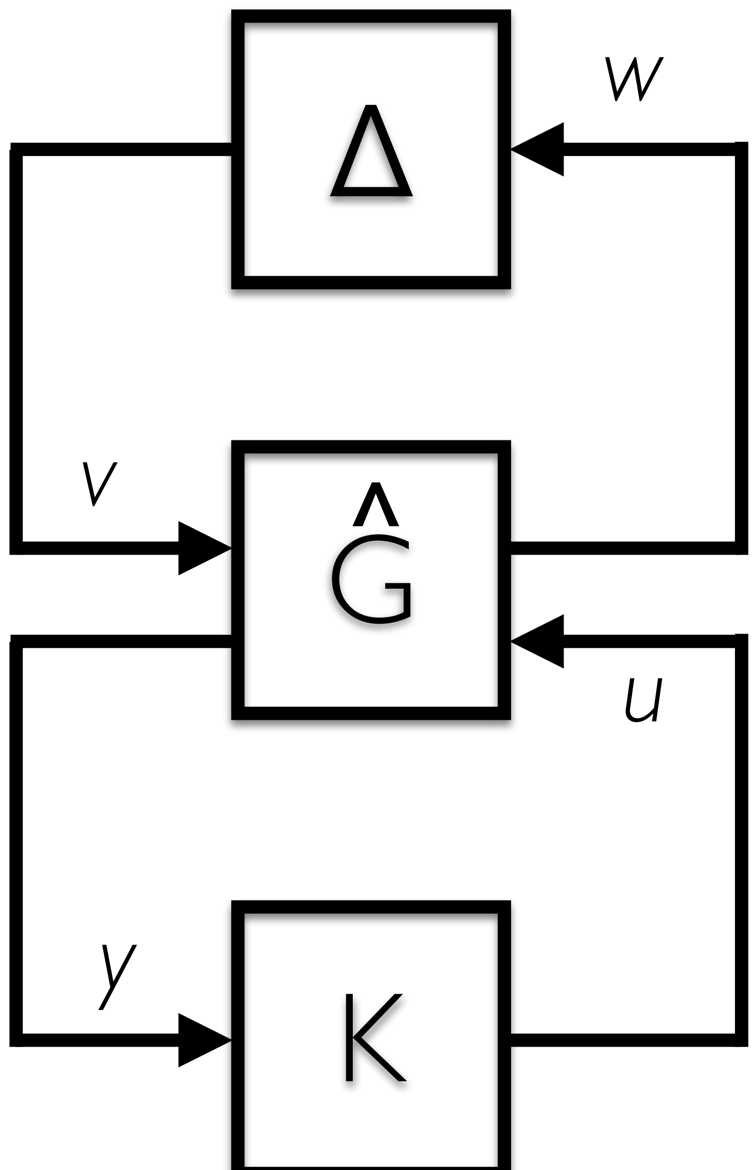
Fit dynamics with *supervised learning:*

$$\boxed{\text{minimize}_{(A,B)} \quad \sum_{i=1}^{T} \|x_{i+1} - A x_i - B u_i\|^2}$$

Solve approximate problem:

$$\boxed{\begin{array}{ll} \text{minimize} & \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\ \text{s.t.} & x_{t+1} = \hat{A} x_t + \hat{B} u_t + \omega_t \end{array}}$$

Coarse-ID control



High dimensional stats bounds the error

Coarse-grained model is trivial to fit

Design robust control for feedback loop

Robust certainty equivalence.

"Simple" Example: LQR

$$\text{minimize} \quad \lim_{T\to\infty} \mathbb{E}\left[\tfrac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \qquad x \in \mathbb{R}^d$$

$$\text{s.t.} \quad x_{t+1} = A x_t + B u_t + e_t \qquad u \in \mathbb{R}^p$$

Gaussian noise

How many samples are needed to Estimate (A,B)? (A stable)

Run an experiment for T steps with random input. Then

$$\text{minimize}_{(A,B)} \quad \sum_{i=1}^{T} \|x_{i+1} - A x_i - B u_i\|^2$$

If $T \geq \tilde{O}\left(\frac{\sigma^2 (d+p)}{\lambda_{\min}(\Lambda_c)\epsilon^2}\right)$ where $\Lambda_c = A \Lambda_c A^* + BB^*$

controllability Gramian

then $\|A - \hat{A}\| \leq \epsilon$ and $\|B - \hat{B}\| \leq \epsilon$ w.h.p.

Similar result for non-stable A.

[Dean, Mania, Matni, R.,Tu, 2017]

[Mania, Jordan, R., Simchowitz, Tu, 2018]

"Simple" Example: LQR

"Obvious strategy": Estimate $(\hat{A}, \hat{B})$, build control $u_t = \hat{K}x_t$

$$\begin{array}{ll} \underset{u}{\text{minimize}} & \sup_{\|\Delta_A\|_2 \le \epsilon_A,\ \|\Delta_B\|_2 \le \epsilon_B} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t \\ \text{s.t.} & x_{t+1} = (\hat{A} + \Delta_A)x_t + (\hat{B} + \Delta_B)u_t \end{array}$$

Solving an SDP relaxation of this robust control problem yields

$$\frac{J(\hat{K}) - J_\star}{J_\star} \le C\,\Gamma_{\mathrm{cl}} \left(\lambda_{\min}(\Lambda_c)^{-1/2} + \|K_\star\|_2\right) \sqrt{\frac{\sigma^2(d+p)}{T}} \quad \text{w.h.p.}$$

$$\Lambda_c = A\Lambda_c A^* + BB^*$$

controllability Gramian

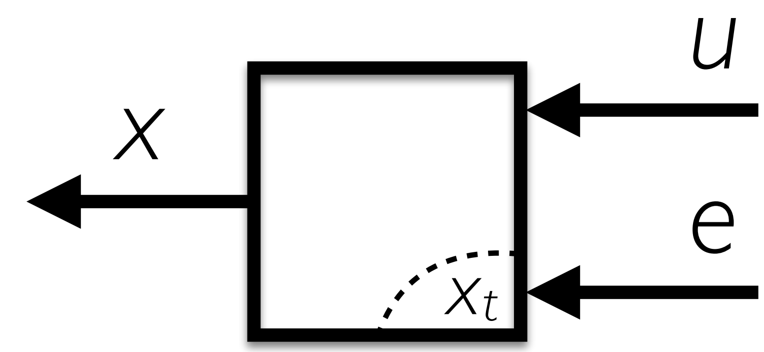
$$\Gamma_{\mathrm{cl}} := \|(zI - A - BK_\star)^{-1}\|_{\mathcal{H}_\infty}$$

closed loop gain

This also tells you when your cost is finite!

[Dean, Mania, Matni, R., Tu 2017]

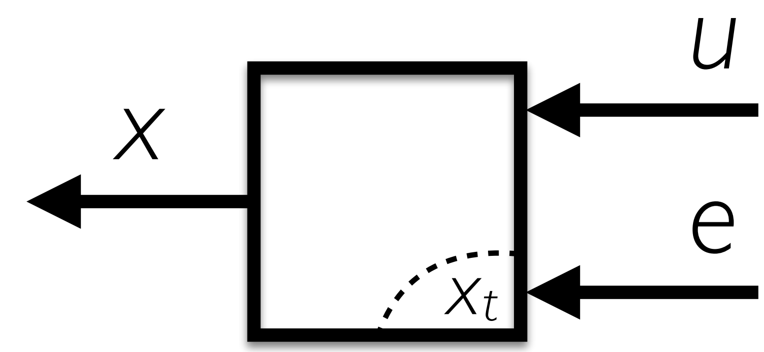
$$
\begin{aligned}
\underset{u}{\text{minimize}} \quad & \lim_{T\to\infty} \mathbb{E}\left[\tfrac{1}{T}\textstyle\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\
\text{s.t.} \quad & x_{t+1} = A x_t + B u_t + e_t
\end{aligned}
$$



Key to formulation:
Write u as LTI function of disturbance. (Disturbance feedback)

$$
\boxed{u_t = \sum_{k=1}^{t} \Phi_u[k] e_{t-k}}
$$

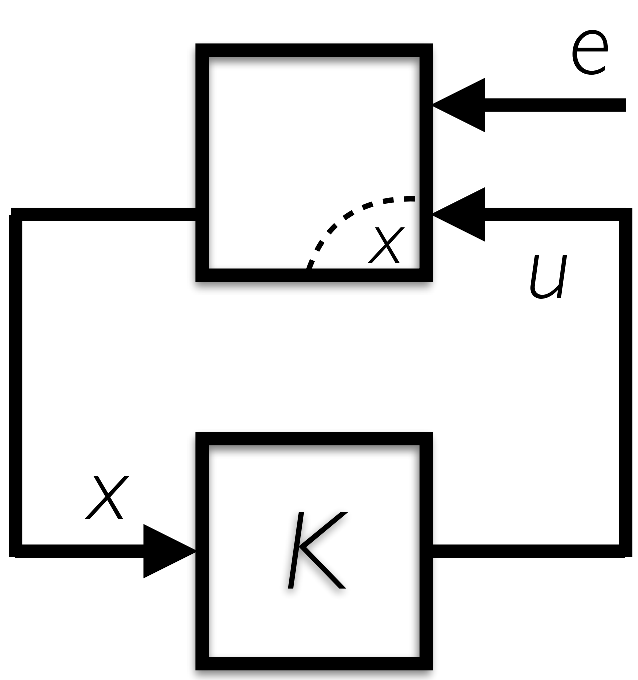
$$\underset{u}{\text{minimize}} \quad \lim_{T\to\infty} \mathbb{E}\left[\tfrac{1}{T}\textstyle\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right]$$
$$\text{s.t.} \quad x_{t+1} = Ax_t + Bu_t + e_t$$



Key to formulation:
Write *u* as LTI function of disturbance. (Disturbance feedback)

Then *x* is a linear function of the disturbance as well.

$$\boxed{\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t} \begin{bmatrix} \Phi_x[k] \\ \Phi_u[k] \end{bmatrix} e_{t-k}}$$



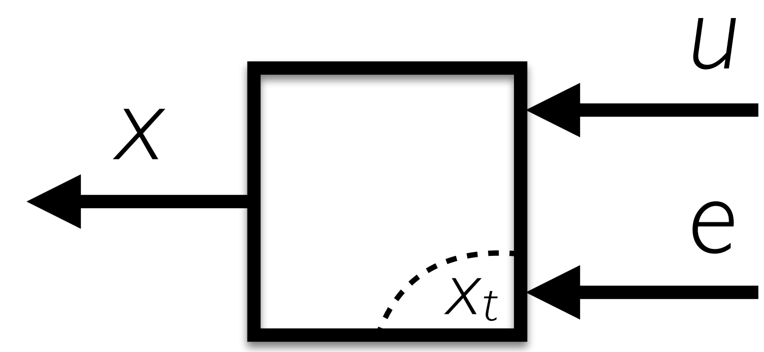
$$K = \Phi_u \Phi_x^{-1}$$

$$x \xrightarrow{\Phi_x^{-1}} e \xrightarrow{\Phi_u} u$$

In closed loop, can't decouple these boxes: consider the mapping from disturbance to both signals.

$$\underset{u}{\text{minimize}} \quad \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right]$$

$$\text{s.t.} \quad x_{t+1} = Ax_t + Bu_t + e_t$$



Key to formulation:
Write u as LTI function of disturbance. (Disturbance feedback)

Then x is a linear function of the disturbance as well.

$$\boxed{\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t} \begin{bmatrix} \Phi_x[k] \\ \Phi_u[k] \end{bmatrix} e_{t-k}}$$

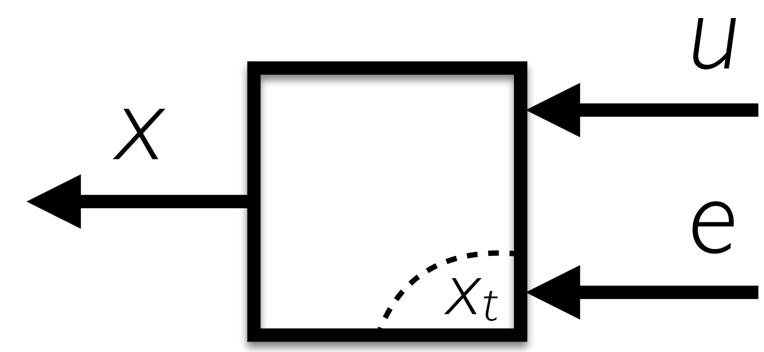
$$\mathbb{E}\left[x_t^* Q x_t\right] = \sigma^2 \sum_{k=1}^{t} \operatorname{Tr}(\Phi_x[k]^* Q \Phi_x[k]) \qquad \mathbb{E}\left[u_t^* R u_t\right] = \sigma^2 \sum_{k=1}^{t} \operatorname{Tr}(\Phi_u[k]^* R \Phi_u[k])$$

Dynamic equality constraint implies:

$$\boxed{z\Phi_x e = A\Phi_x e + B\Phi_u e + e}$$

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I$$

$$
\begin{aligned}
\underset{u}{\text{minimize}} \quad & \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\
\text{s.t.} \quad & x_{t+1} = A x_t + B u_t + e_t
\end{aligned}
$$



Key to formulation:
Write u as LTI function of disturbance. (Disturbance feedback)

Then x is a linear function of the disturbance as well.

$$
\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t} \begin{bmatrix} \Phi_x[k] \\ \Phi_u[k] \end{bmatrix} e_{t-k}
$$

$$
\begin{aligned}
\underset{\Phi}{\text{minimize}} \quad & \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2}^2 \\
\text{s.t.} \quad & \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I
\end{aligned}
$$

System Level Synthesis

Suppose $(A, B) = (\hat{A} + \Delta_A, \hat{B} + \Delta_B)$ (i.e., nominal + error). Note that if

Note that if $$\begin{bmatrix} zI - \hat{A} & -\hat{B} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I$$

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I + \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} =: I + \Delta$$

And hence $$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} (I + \Delta)^{-1} = I$$

Satisfying nominal constraints results in true system responses:

$$\begin{bmatrix} \tilde{\Phi}_x \\ \tilde{\Phi}_u \end{bmatrix} = \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} (I + \Delta)^{-1}$$

Key to formulation:
Write (x,u) as linear function of disturbance

$$\boxed{\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t} \begin{bmatrix} \Phi_x[k] \\ \Phi_u[k] \end{bmatrix} e_{t-k}}$$

$$\begin{array}{ll} \underset{\Phi}{\text{minimize}} & \underset{\|\Delta_A\|_2 \le \epsilon_A,\ \|\Delta_B\|_2 \le \epsilon_B}{\sup} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2}^2 \\ \text{s.t.} & \begin{bmatrix} zI - (\hat{A} + \Delta_A) & -(\hat{B} + \Delta_B) \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I \end{array}$$

Push robustness into cost.

$$\begin{array}{ll} \underset{\Phi}{\text{minimize}} & \underset{\|\Delta_A\|_2 \le \epsilon_A,\ \|\Delta_B\|_2 \le \epsilon_B}{\sup} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} (I + \Delta)^{-1} \right\|_{\mathcal{H}_2}^2 \\ \text{s.t.} & \begin{bmatrix} zI - \hat{A} & -\hat{B} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I \end{array}$$

SLS Formulation of Robust LQR

Key to formulation:
Write (x,u) as linear function of disturbance

$$\boxed{\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t} \begin{bmatrix} \Phi_x[k] \\ \Phi_u[k] \end{bmatrix} e_{t-k}}$$

$$\underset{\gamma \in [0,1)}{\text{minimize}} \; \frac{1}{1-\gamma} \quad \min_{\Phi_x, \Phi_u} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2}$$

s.t.

$$\begin{bmatrix} zI - \hat{A} & -\hat{B} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I$$

$$\left\| \begin{bmatrix} \epsilon_A \Phi_x \\ \epsilon_B \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \frac{\gamma}{\sqrt{2}}$$

- Approximately solvable by SDP for fixed γ
- Binary search over γ to find optimal solution

SLS Formulation of Robust LQR

$$\text{minimize}_{X,Z,W,\gamma} \quad \frac{1}{(1-\gamma)^2}\left\{\text{Trace}(QW_{11}) + \text{Trace}(RW_{22})\right\}$$

$$\text{subject to} \quad \begin{bmatrix} X & X & Z^* \\ X & W_{11} & W_{12} \\ Z & W_{21} & W_{22} \end{bmatrix} \succeq 0$$

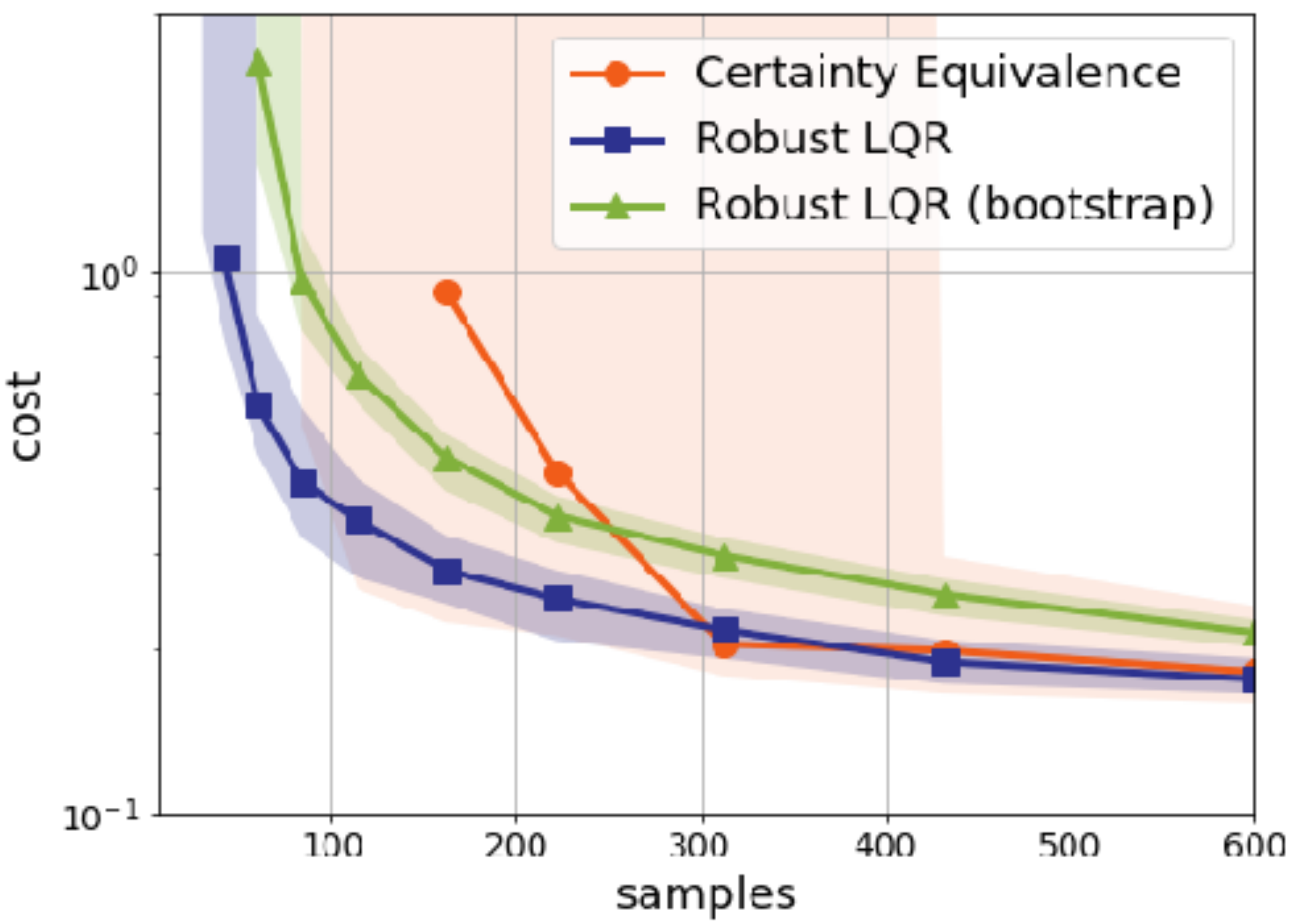
$$\begin{bmatrix} X - I & \hat{A}X + \hat{B}Z & 0 & 0 \\ (\hat{A}X + \hat{B}Z)^* & X & \epsilon_A X & \epsilon_B Z^* \\ 0 & \epsilon_A X & \alpha\gamma^2 I & 0 \\ 0 & \epsilon_B Z & 0 & (1-\alpha)\gamma^2 I \end{bmatrix} \succeq 0\,.$$

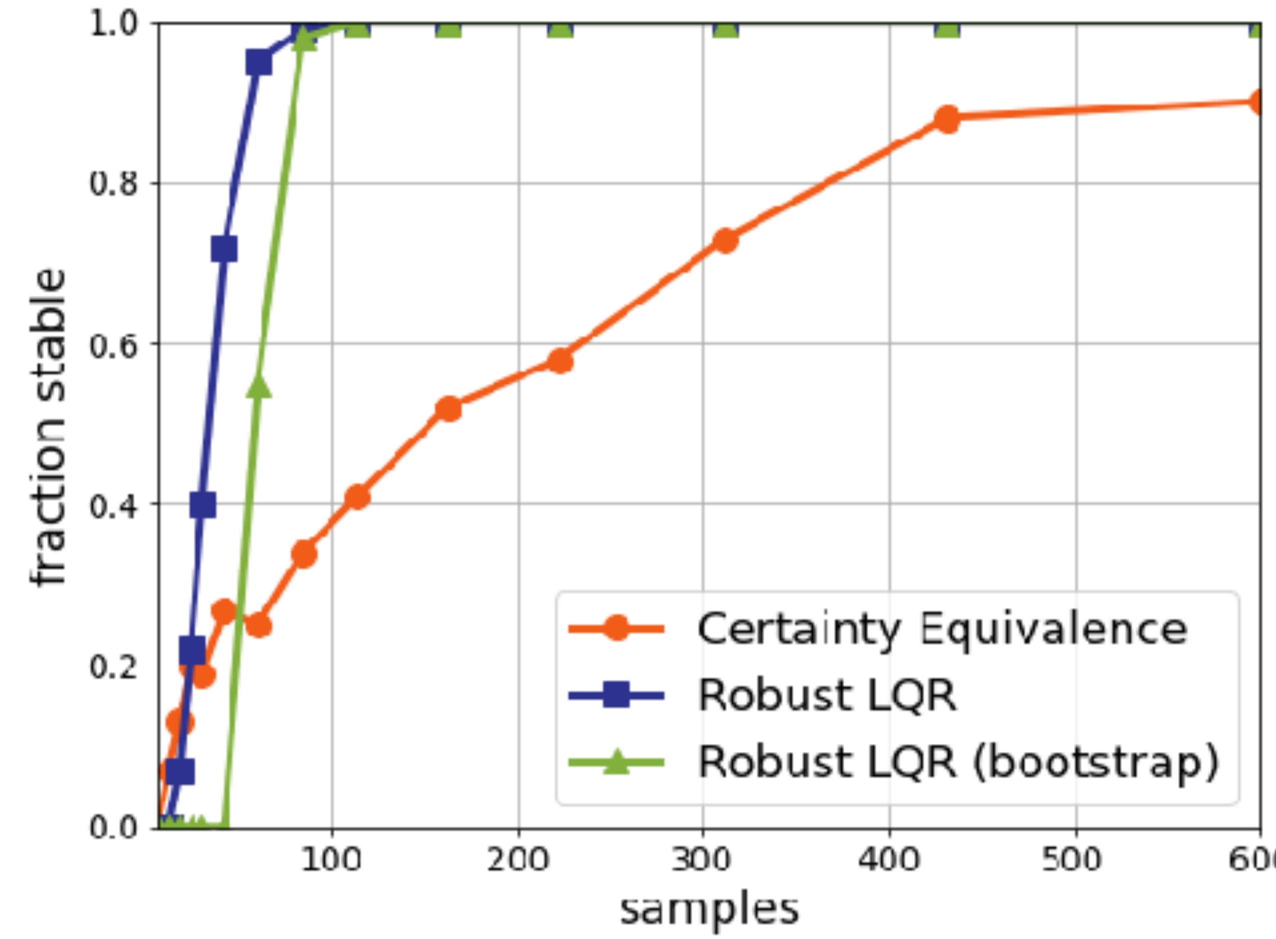
- Solvable by SDP for fixed γ
- Binary search over γ to find optimal solution
- Optimal controller is $K=-ZX^{-1}$

Why robust?

$$x_{t+1} = \begin{bmatrix} 1.01 & 0.01 & 0 \\ 0.01 & 1.01 & 0.01 \\ 0 & 0.01 & 1.01 \end{bmatrix} x_t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u_t + e_t$$

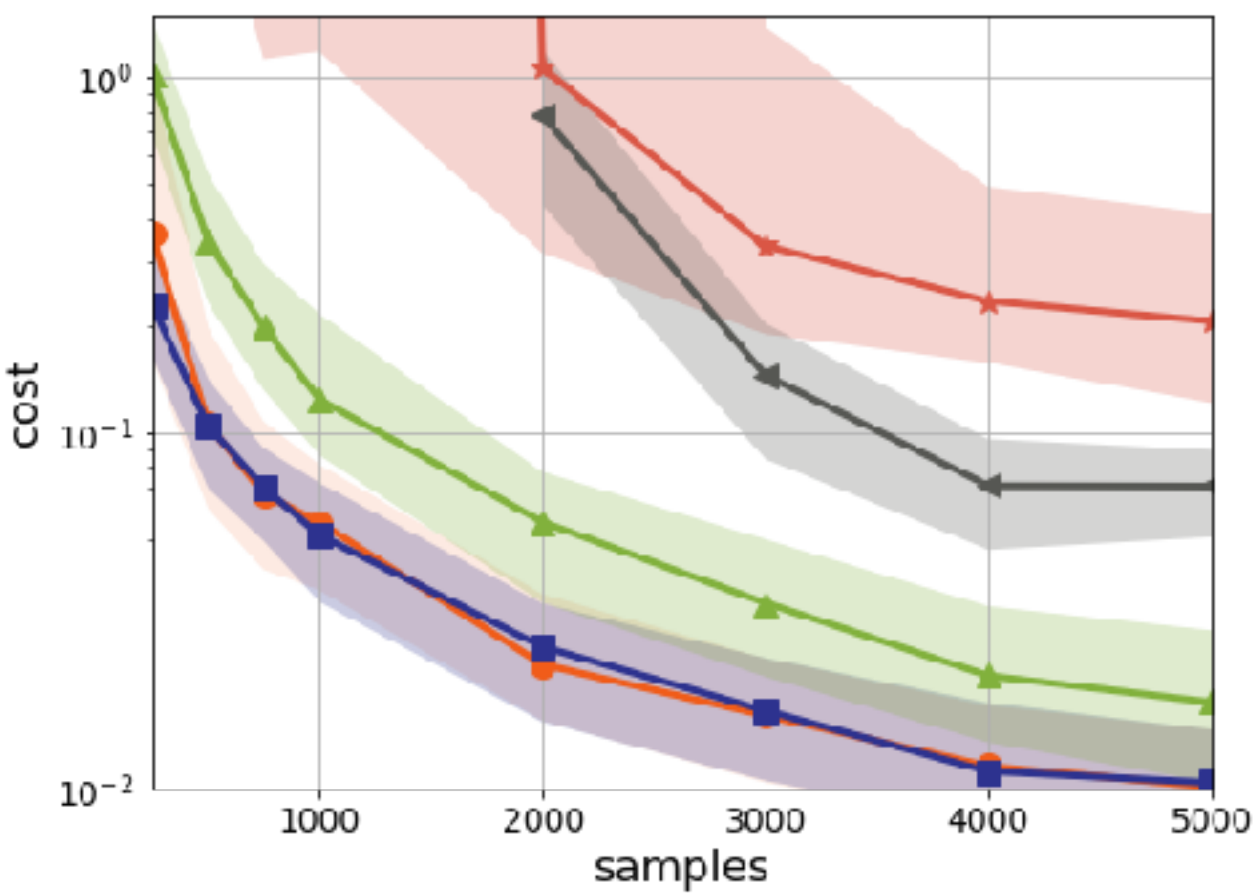
Slightly unstable system, system ID tends to think some nodes are stable

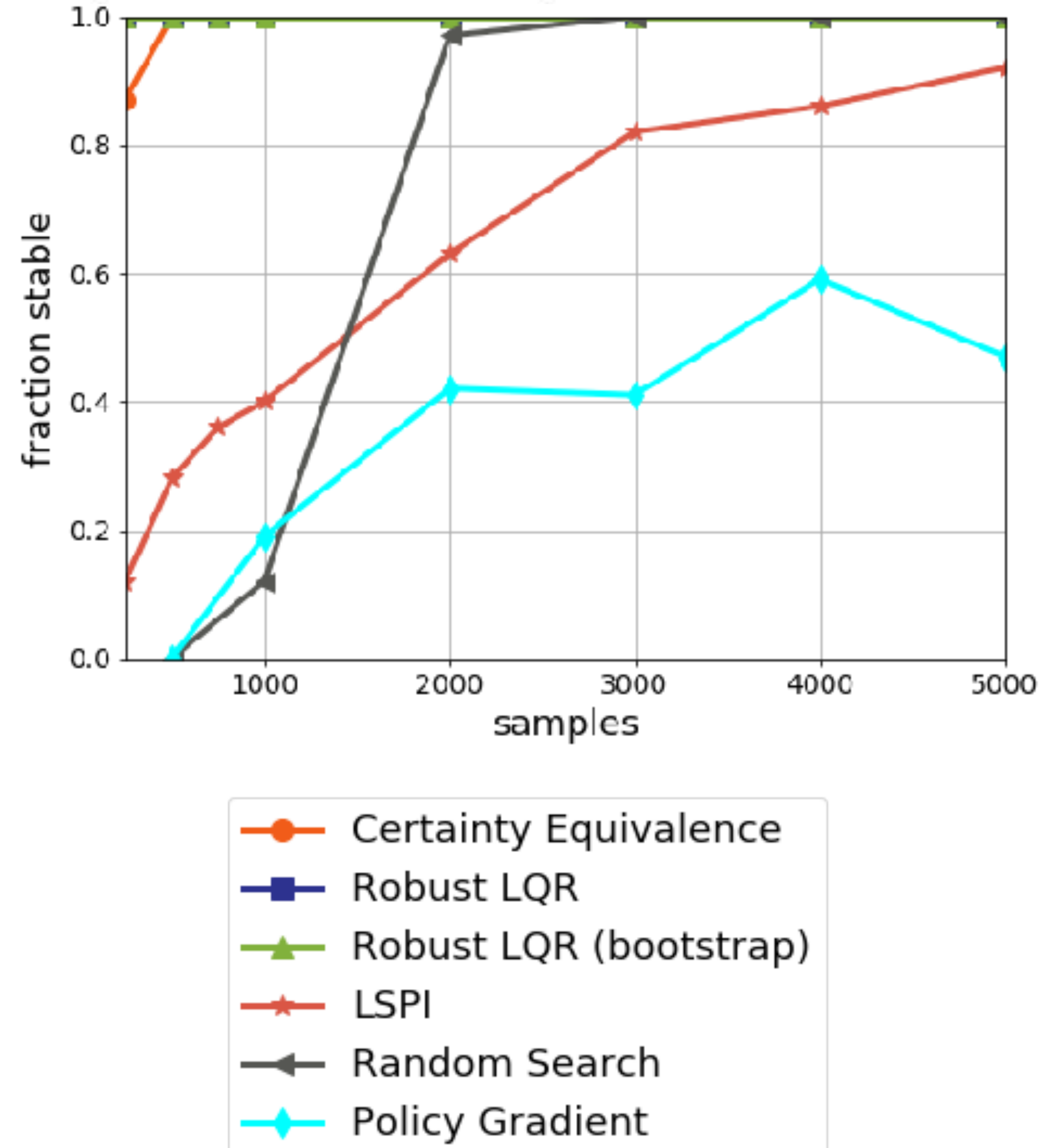




Certainty equivalence may yield unstable controller

Robust synthesis yields stable controller





Model-free performs worse than model-based

Adaptive LQR

$$\begin{array}{ll} \underset{u}{\text{minimize}} & \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\ \text{s.t.} & x_{t+1} = Ax_t + Bu_t + e_t \end{array}$$

Oracle: You can generate one trajectory of length T.

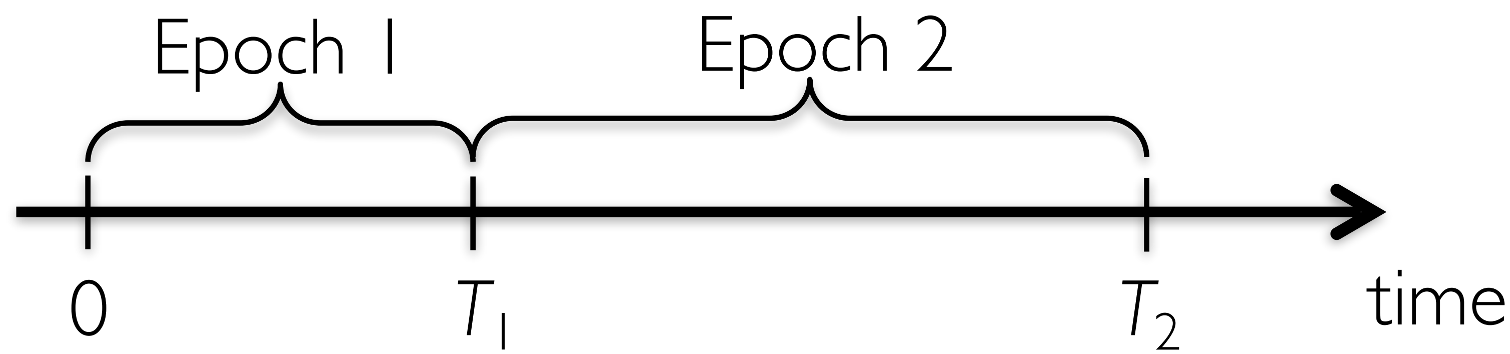
Challenge: Build a controller online with smallest error at every time step.

$$\text{minimize } R(T) := \sum_{t=1}^{T} \left[x_t^* Q x_t + u_t^* R u_t - J_\star\right]$$

What is the optimal exploration/exploitation scheme?

SLS for Adaptive LQR

At every T_i, do:



1. $$(\hat{A}^{(i)}, \hat{B}^{(i)}) = \arg\min_{(A,B)} \sum_{t \in E_i} \|x_{t+1} - Ax_t - Bu_t\|^2$$

2. $$\mathbf{K}^{(i)} = \mathrm{RobustSLS}\left(\hat{A}^{(i)}, \hat{B}^{(i)}, \epsilon_A^{(i)}, \epsilon_B^{(i)}\right)$$

3. $$\mathbf{u}^{(i)} = \mathbf{K}^{(i)}\mathbf{x}^{(i)} + \eta^{(i)}$$

probing noise (shrinks with T)

[Dean, Mania, Matni, Recht, Tu, NIPS 2018]

Sharp bounds from time-series data:

OLS

Set $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2 I\right)$ $\Rightarrow$ $\left\| \begin{bmatrix} \hat{A} - A \\ \hat{B} - B \end{bmatrix} \right\| = \tilde{O}\left(\frac{1}{\sigma_\eta T^{1/2}} \right)$

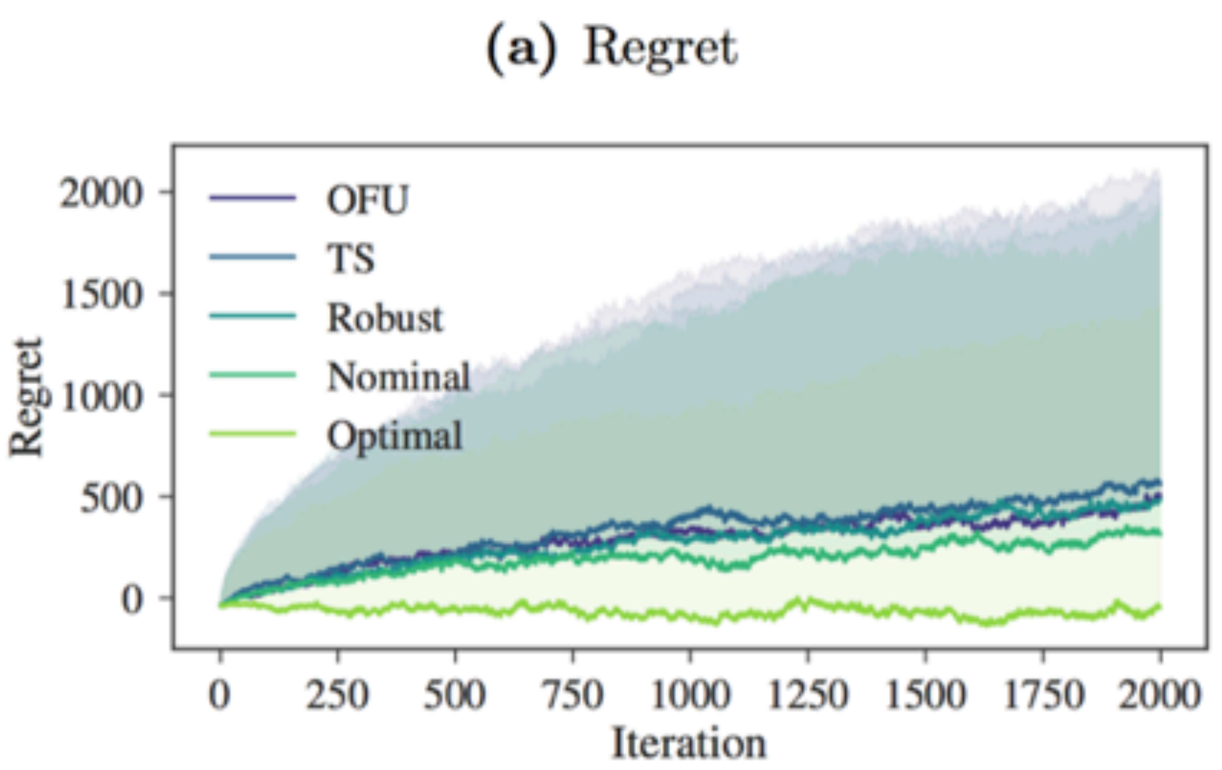
[Simchowitz, Mania, Tu, Jordan, Recht, COLT 2018]

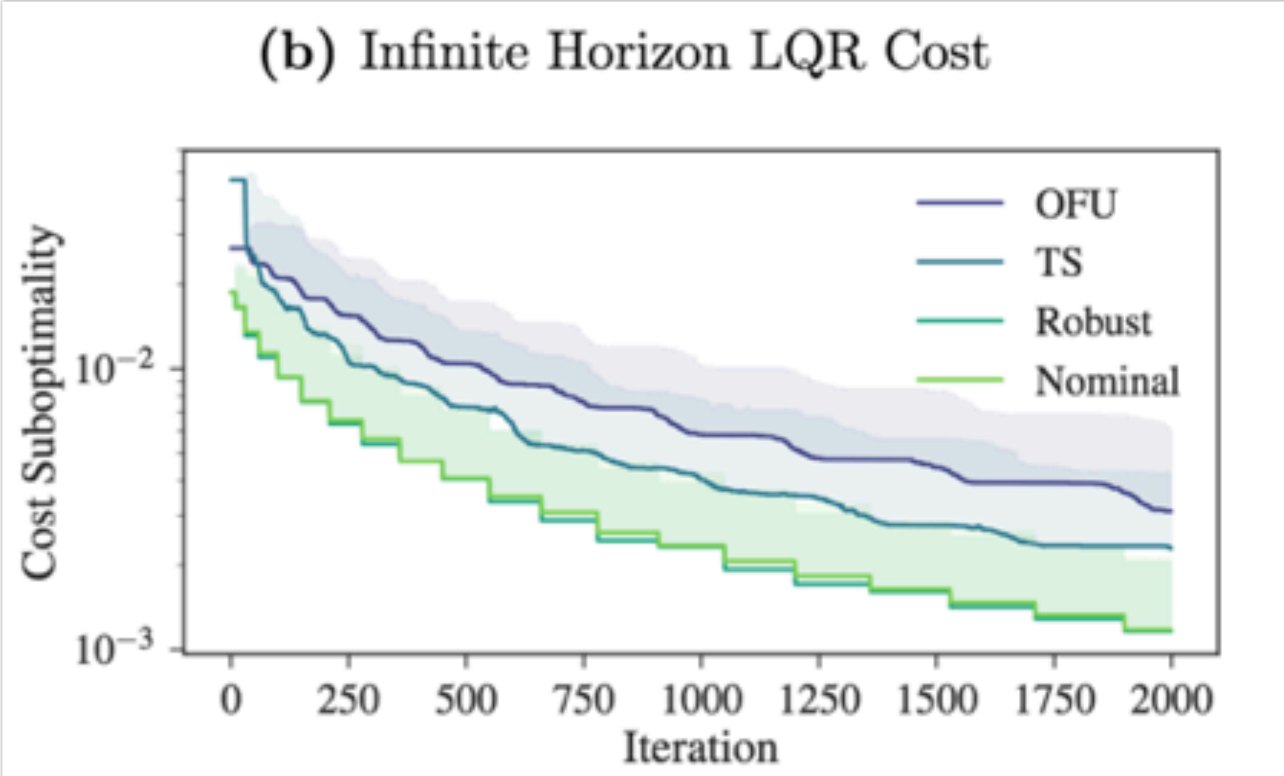
Explore vs. exploit:

$\tilde{O}\left(\frac{T^{1/2}}{\sigma_\eta} \right)$ **+** $\tilde{O}\left(\sigma_\eta^2 T\right)$ $\Rightarrow$ $\sigma_2 = C_\eta T^{-\frac{1}{3}}$

Model Mismatch **Excitation**

[Dean, Mania, Matni, Recht, Tu, NIPS 2018]





[Dean, Mania, Matni, Recht, Tu, NIPS 2018]

Safe exploration

$$\begin{array}{ll} \underset{u}{\text{minimize}} & \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^* Q x_t + u_t^* R u_t\right] \\ \text{s.t.} & x_{t+1} = Ax_t + Bu_t + e_t \end{array}$$

$$\text{minimize} \quad \frac{1}{1-\gamma}\left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2}$$

$$\text{s.t.} \quad \begin{bmatrix} zI - \hat{A} & -\hat{B} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I$$

Robust Dynamics

$$\left\| \begin{bmatrix} \epsilon_{A,2}\Phi_x \\ \epsilon_{B,2}\Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \frac{\gamma}{\sqrt{2}}, \quad \left\| \begin{bmatrix} \epsilon_{A,\infty}\Phi_x \\ \epsilon_{B,\infty}\Phi_u \end{bmatrix} \right\|_{\mathcal{L}_1} \leq \tau$$

$$F_j \Phi_x x_0 + \frac{\sigma_e}{1-\tau}\|F_j\Phi_x[t:1]\|_1 \leq b_j \;\; \forall j, t$$

Robustness to constraints

[Dean, Tu, Matni, Recht 2018]

Safe exploration

$$\text{minimize} \quad \frac{1}{1-\gamma}\left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} \right\|_{\mathcal{H}_2}$$

$$\text{s.t.} \quad \begin{bmatrix} zI - \hat{A} & -\hat{B} \end{bmatrix} \begin{bmatrix} \Phi_x \\ \Phi_u \end{bmatrix} = I$$

Robust Dynamics

$$\left\| \begin{bmatrix} \epsilon_{A,2}\Phi_x \\ \epsilon_{B,2}\Phi_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \frac{\gamma}{\sqrt{2}}, \quad \left\| \begin{bmatrix} \epsilon_{A,\infty}\Phi_x \\ \epsilon_{B,\infty}\Phi_u \end{bmatrix} \right\|_{\mathcal{L}_1} \leq \tau$$

$$F_j\Phi_x x_0 + \frac{\sigma_e}{1-\tau}\|F_j\Phi_x[t:1]\|_1 \leq b_j \quad \forall j, t$$

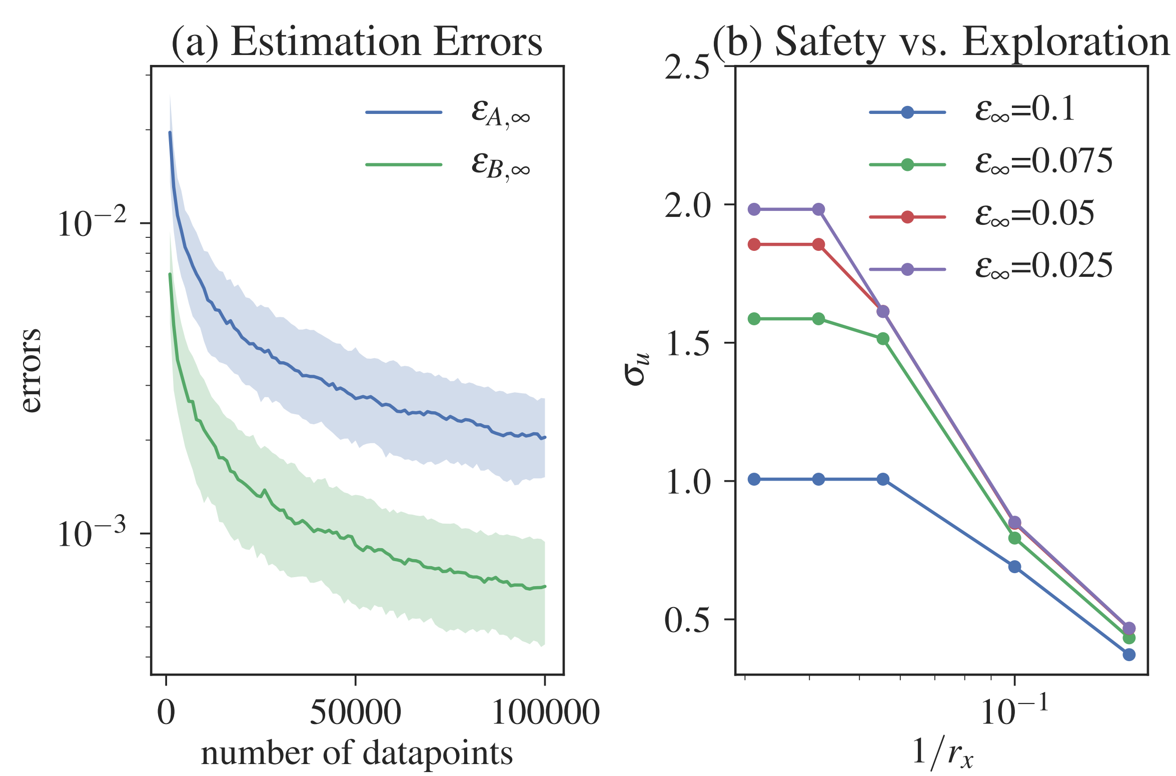
Robustness to constraints

Enables exploration with safety:

$$\mathbf{u} = \mathbf{K}\mathbf{x} + \eta$$

Robustly Synthesized

Probing



[Dean, Tu, Matni, Recht 2018]

So far…

- Model based methods seem to perform better than model free ones in theory and practice.
- The field needs more baselines!
- Simple algorithms seem to be surprisingly competitive.
- Analysis of time series is harder than it appears.

Next Steps

- Nonlinear models and constraints via learned ILQR.
- Learning about uncertain environments.
- Model mismatch: what happens when the model is wrong? Improper learning.
- Implementing in test-beds.

References

- `argmin.net`
- "On the Sample Complexity of the Linear Quadratic Regulator." S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. `arXiv:1710.01688`
- "Non-asymptotic Analysis of Robust Control from Coarse-grained Identification." S. Tu, R. Boczar, A. Packard, and B. Recht. `arXiv:1707.04791`
- "Least-squares Temporal Differencing for the Linear Quadratic Regulator" S. Tu and B. Recht. In submission to ICML 2018. `arXiv:1712.08642`
- "Learning without Mixing." H. Mania, B. Recht, M. Simchowitz, and S. Tu. In submission to COLT 2018. `arXiv:1802.08334`
- "Simple random search provides a competitive approach to reinforcement learning." H. Mania, A. Guy, and B. Recht. `arXiv:1803.07055`
- "Regret Bounds for Robust Adaptive Control of the Linear Quadratic Regulator." S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu. `arXiv:1805.09388`
- "A Tour of Reinforcement Learning: The View from Continuous Control." B. Recht. `arXiv:1806.09460`

`https://people.eecs.berkeley.edu/~brecht/publications.html`