What do we know about learning and the Linear Quadratic Regulator?

Nikolai Matni

EECS, UC Berkeley

joint work with Sarah Dean, Horia Mania, Stephen Tu, and Ben Recht

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

Tesla Says Autopilot Was Engaged in Fatal Crash Under Investigation in California

Vehicle's system shows driver had hands off the wheel for six seconds before striking highway divider

Las Vegas' self-driving bus crashes in first hour of service

Google AI looks at rifles and sees helicopters

Street sign hack fools self-driving cars

Data-driven methods need guarantees of stability, performance, robustness, safety

Robust Control and Learning?

Machine Learning

uses data to

reduce uncertainty

more data
→ better models/predictions

probabilistic guarantees

Robust Control

uses feedback to

mitigate uncertainty

better models/predictions
→ better performance

worst-case guarantees

Can ML and RC be combined so that we **safely achieve** more data → better performance?

Two main takeaways

Robustness is key

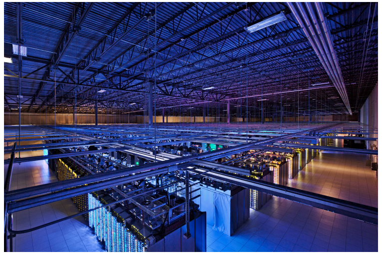
Robustness matters in practice and makes theory tractable

Robust and optimal control as optimization

System Level Synthesis

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

Cooling a server farm



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

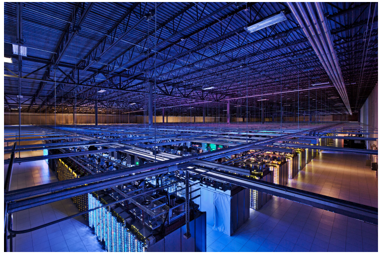
temperature deviations

cooling

“noise”

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

Cooling a server farm

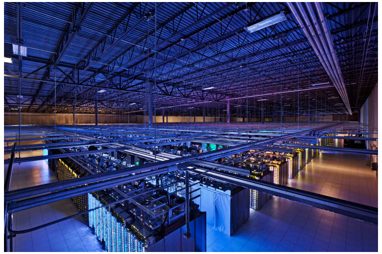


$$x_{t+1} = \begin{bmatrix} 1.01 & 0.01 & 0 \\ 0.01 & 1.01 & 0.01 \\ 0 & 0.01 & 0.01 \end{bmatrix} x_t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u_t + \delta_t$$

Goal: $$\min_{u_0, u_1, \ldots} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}\left[\underbrace{x_t^T Q x_t}_{\text{don't burn servers}} + \underbrace{u_t^T R u_t}_{\text{don't burn \$\$\$}} \right]$$

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

Cooling a server farm



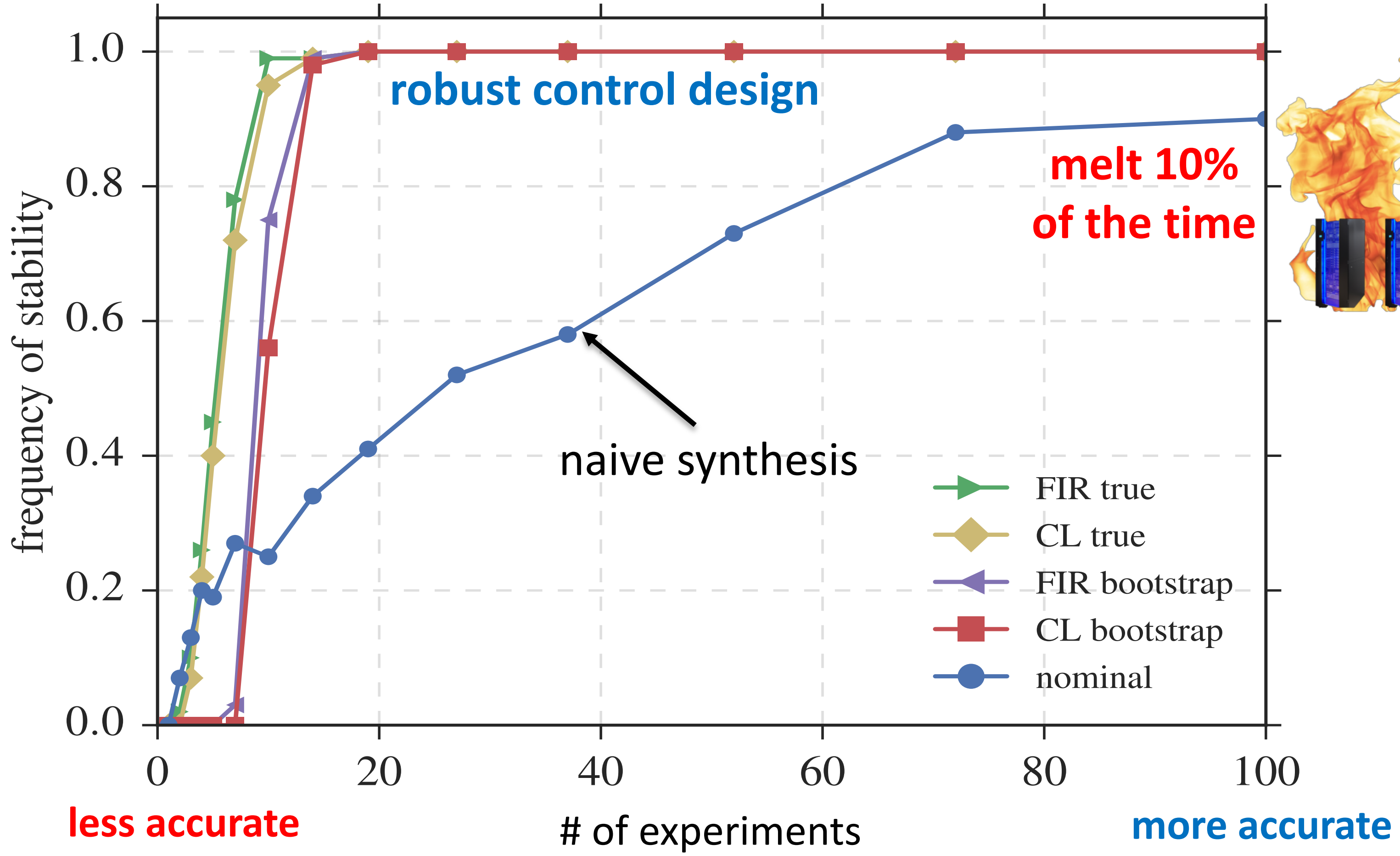
$$x_{t+1} = \begin{bmatrix} 1.01 & 0.01 & 0 \\ 0.01 & 1.01 & 0.01 \\ 0 & 0.01 & 0.01 \end{bmatrix} x_t + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u_t + \delta_t$$

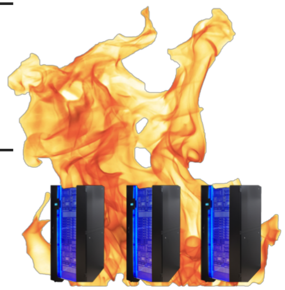
Known dynamics: $u_t = K_t x_t$ **feedback based policy**

Unknown dynamics?

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

Frequency of Servers NOT melting (stability)





[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

The Linear Quadratic Regulator

$$\min_{u} \quad \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T} x_t^\top Q x_t + u_t^\top R u_t\right]$$
$$\text{s.t.} \quad \underbrace{x_{t+1}}_{\text{system state}} = \underbrace{A x_t}_{\text{autonomous dynamics}} + \underbrace{B u_t}_{\text{actuation}} + \underbrace{\delta_t}_{\text{disturbance}}$$

state-feedback

$$u_t = K x_t$$

Closed form solution for known *(A,B)*

Fundamental problem in control theory

(linearize nonlinear systems, MPC)

The offline learning LQR problem

$$\min_{u} \quad \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T} x_t^\top Q x_t + u_t^\top R u_t\right]$$
$$\text{s.t.} \quad x_{t+1} = Ax_t + Bu_t + \delta_t$$

unknown

state-feedback

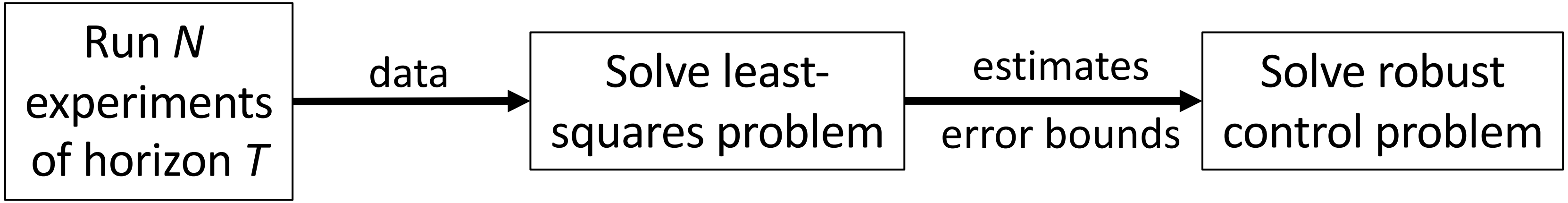
$$u_t = Kx_t$$

Obvious strategy:

run some experiments to estimate *(A,B)*, then compute a controller

Question: how many samples are needed for near optimal control?

The Coarse-ID control pipeline



High-dimensional statistics/bootstrapping

System Level Synthesis

Theorem

With probability $1-\delta$, for N sufficiently large, the synthesized controller is stabilizing and achieves the relative performance bound

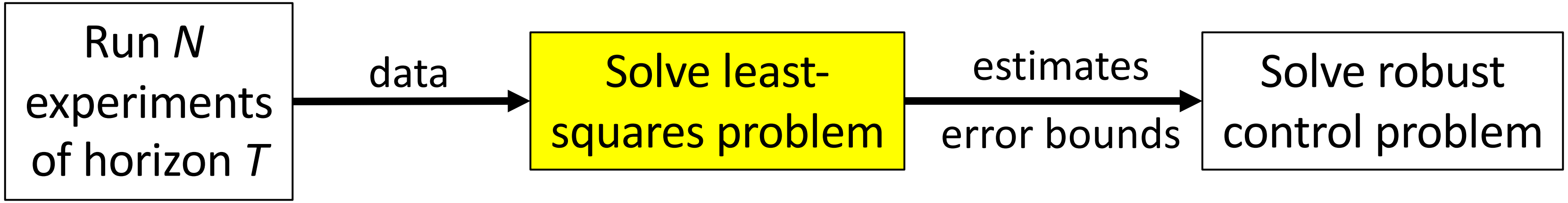
$$\frac{\widehat{J}-J_\star}{J_\star} \leq \mathcal{O}\left(\mathcal{C}(A,B,Q,R,T)\sqrt{\frac{(n+p)\log(1/\delta)}{N}}\right)$$

\# of states

\# of inputs

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

The Coarse-ID control pipeline



High-dimensional statistics/bootstrapping

System Level Synthesis

Theorem

With probability $1-\delta$, for N sufficiently large, the synthesized controller is stabilizing and achieves the relative performance bound

$$\frac{\widehat{J} - J_\star}{J_\star} \leq \mathcal{O}\left(\mathcal{C}(A,B,Q,R,T) \sqrt{\frac{(n+p)\log(1/\delta)}{N}} \right)$$

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

How easy is it to identify a system?

Run *N* experiments for *T* steps with random input. Then

$$\min_{(A,B)} \sum_{i=1}^{N} \left\| x_{T+1}^{(i)} - A x_T^{(i)} - B u_T^{(i)} \right\|_2^2$$

Controllability Gramian

If $N \geq \tilde{O}\left(\frac{\sigma_w^2}{\sigma_u^2} \frac{(n+p)}{\lambda_{\min}(\Lambda_c)} \frac{1}{\epsilon^2} \right)$ where $\Lambda_c \approx A \Lambda_c A^* + B B^*$

least excitable mode

then $\|A - \hat{A}\| \leq \epsilon$ and $\|B - \hat{B}\| \leq \epsilon$

How easy is it to identify a ***stable*** system?

Run ***1*** experiment for *T* steps with random input. Then

$$\min_{(A,B)} \sum_{t=1}^{T} \|x_{t+1} - Ax_t - Bu_t\|_2^2$$

Controllability Gramian

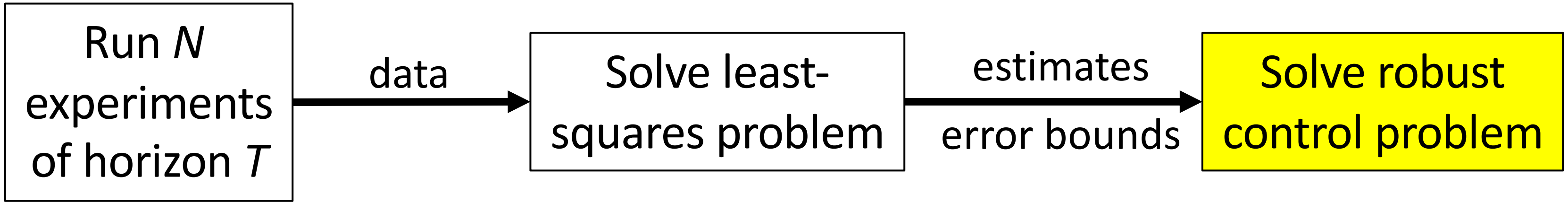
If $T \geq \tilde{O}\left(\frac{\sigma_w^2}{\sigma_u^2} \underbrace{\frac{(n+p)}{\lambda_{\min}(\Lambda_c)}}_{} \frac{1}{\epsilon^2}\right)$ where $\Lambda_c = A\Lambda_c A^* + BB^*$

least excitable mode

then $\|A - \hat{A}\| \leq \epsilon$ and $\|B - \hat{B}\| \leq \epsilon$

[Simchowitz et al, COLT 2018]

The Coarse-ID control pipeline



High-dimensional statistics/bootstrapping

System Level Synthesis

Theorem

With probability $1-\delta$, for N sufficiently large, the synthesized controller is stabilizing and achieves the relative performance bound

$$\frac{\widehat{J}-J_\star}{J_\star} \leq \mathcal{O}\left(\mathcal{C}(A,B,Q,R,T)\sqrt{\frac{(n+p)\log(1/\delta)}{N}}\right)$$

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

Working with system responses

$$x_{t+1} = Ax_t + Bu_t + \delta_t$$
$$u_t = Kx_t$$

End-to-end (closed loop) system responses

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t} \begin{bmatrix} (A+BK)^{t-k} \\ K(A+BK)^{t-k} \end{bmatrix} \delta_{k-1} =: \sum_{k=1}^{t} \begin{bmatrix} \Phi_x(t-k) \\ \Phi_u(t-k) \end{bmatrix} \delta_{k-1}$$

[Wang, **Matni**, Doyle TAC 2018, submitted]

Working with system responses

$$\mathbb{E}[x_t^\top Q x_t] = \boxed{\sum_{k=1}^{t} \mathrm{Tr}[(A+BK)^{t-k}]^\top Q (A+BK)^{t-k}} = \boxed{\sum_{k=1}^{t} \mathrm{Tr}\Phi_x(k)^\top Q \Phi_x(k)}$$

$$\mathbb{E}[u_t^\top R u_t] = \boxed{\sum_{k=1}^{t} \mathrm{Tr}[K(A+BK)^{t-k}]^\top R K (A+BK)^{t-k}} = \boxed{\sum_{k=1}^{t} \mathrm{Tr}\Phi_u(k)^\top R \Phi_u(k)}$$

finite dimensional but **non-convex**

infinite dimensional but **convex**

how do we constrain system responses so that they are achievable?

[Wang, **Matni**, Doyle TAC 2018, submitted]

Working with system responses

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t} \begin{bmatrix} (A+BK)^{t-k} \\ K(A+BK)^{t-k} \end{bmatrix} \delta_{k-1} =: \sum_{k=1}^{t} \begin{bmatrix} \Phi_x(t-k) \\ \Phi_u(t-k) \end{bmatrix} \delta_{k-1}$$

and sufficient

A simple necessary condition

$$\begin{aligned} \Phi_x(t+1) &= (A+BK)^t \\ &= (A+BK)\Phi_x(t) \\ &= A\Phi_x(t) + BK\Phi_x(t) \\ &= A\Phi_x(t) + B\Phi_u(t) \end{aligned}$$

[Wang, **Matni**, Doyle TAC 2018, submitted]

LQR via system responses

$$
\begin{array}{ll}
\min\limits_{u} & \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T} x_t^\top Q x_t + u_t^\top R u_t\right] \\
\text{s.t.} & x_{t+1} = Ax_t + Bu_t + \delta_t
\end{array}
$$

[Wang, **Matni**, Doyle TAC 2018, submitted]

LQR via system responses

$$\begin{aligned} \min_{\Phi_x,\Phi_u} \quad & \textstyle\sum_{t=0}^{\infty} \mathrm{Tr}\left[\Phi_x(t)^\top Q \Phi_x(t) + \Phi_u(t)^\top R \Phi_u(t)\right] \\ \text{s.t.} \quad & \Phi_x(t+1) = A\Phi_x(t) + B\Phi_u(t),\ \Phi_x(1) = I \end{aligned}$$

[Wang, **Matni**, Doyle TAC 2018, submitted]

LQR via system responses

Equivalent formulation: why bother?

$$\min_{\mathbf{\Phi_x}, \mathbf{\Phi_u}} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & \\ & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_2}$$

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

To achieve desired response set $\boldsymbol{u} = \boldsymbol{\Phi_u \Phi_x^{-1} x}$

[Wang, **Matni**, Doyle TAC 2018, submitted]

Robust system responses

achievability

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

$$\boldsymbol{u = \Phi_u \Phi_x^{-1} x}$$

response

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi_x} \\ \boldsymbol{\Phi_u} \end{bmatrix} \boldsymbol{\delta}$$

[Wang, **Matni**, Doyle TAC 2018, submitted]

Robust system responses

robust **achievability**

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \hat{\mathbf{\Phi}}_{\mathbf{x}} \\ \hat{\mathbf{\Phi}}_{\mathbf{u}} \end{bmatrix} = I + \mathbf{\Delta}$$

unknown

$$A \approx \hat{A},\; B \approx \hat{B}$$

$$\boldsymbol{u} = \hat{\mathbf{\Phi}}_{\boldsymbol{u}} \hat{\mathbf{\Phi}}_{\boldsymbol{x}}^{-1} \boldsymbol{x}$$

actual **response**

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{\Phi}}_{\mathbf{x}} \\ \hat{\mathbf{\Phi}}_{\mathbf{u}} \end{bmatrix} \overbrace{(I + \mathbf{\Delta})^{-1}}^{\text{effect of uncertainty}} \delta$$

[**Matni**, Wang, Anderson CDC 2017]

Robust SLS LQR problem

$$A = \hat{A} + \Delta_A, \; B = \hat{B} + \Delta_B, \; \|\Delta_A\|_2 \le \epsilon_A, \|\Delta_B\|_2 \le \epsilon_B$$

Let $\mathbf{K}$ stabilize $(\hat{A}, \hat{B})$, and $(\hat{\boldsymbol{\Phi}}_{\boldsymbol{x}}, \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}})$ be its system response.

Then $\mathbf{K}$ achieves the following cost on the true system (A, B):

$$J(A, B, \mathbf{K}) := \left\| \underbrace{\begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} \\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}} \end{bmatrix}}_{\text{nominal performance}} \left(I + \underbrace{\begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\Phi}}_{\boldsymbol{x}} \\ \hat{\boldsymbol{\Phi}}_{\boldsymbol{u}} \end{bmatrix}}_{\hat{\boldsymbol{\Delta}} \text{ effect of uncertainty}} \right)^{-1} \right\|_{\mathcal{H}_2}$$

Stable if $\|\hat{\boldsymbol{\Delta}}\| < 1$

Robust SLS LQR problem

$$A = \hat{A} + \Delta_A, \;\; B = \hat{B} + \Delta_B, \;\; \|\Delta_A\|_2 \le \epsilon_A, \|\Delta_B\|_2 \le \epsilon_B$$

$$\min_{\mathbf{\Phi}_x, \mathbf{\Phi}_u} \overbrace{\max_{\Delta_A, \Delta_B} J(A, B, \mathbf{K})}^{\text{robust performance}} \le \min_{\mathbf{\Phi}_{\boldsymbol{x}}, \mathbf{\Phi}_{\boldsymbol{u}}} \overbrace{J(\hat{A}, \hat{B}, \mathbf{K})}^{\text{nominal performance}} \overbrace{f(\|\hat{\mathbf{\Delta}}\|)}^{\text{effect of uncertainty}}$$

$$\text{s.t. } \mathcal{A}(\mathbf{\Phi}, \Delta_A, \Delta_B) = 0$$

achievability

$$\text{s.t. } \hat{\mathcal{A}}(\mathbf{\Phi}) = 0$$

nominal achievability

$$\|\hat{\mathbf{\Delta}}\| < 1$$

robust stability

Robust SLS LQR problem

$$\min_{\gamma\in(0,1),\mathbf{\Phi}_x,\mathbf{\Phi}_u} \overbrace{\frac{1}{1-\gamma}}^{\text{effect of uncertainty}} \overbrace{\left\| \begin{bmatrix} Q^{1/2} & 0 \\ 0 & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_2}}^{\text{nominal LQR cost}} \quad \text{s.t.}$$

nominal achievability $\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} = I \,, \quad \sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi}_x \\ \varepsilon_B \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \gamma \,,$

robust stability

But this is infinite dimensional!

Option 1: FIR truncation

$$\mathbf{\Phi_x} = \sum_{t=1}^{T} \Phi_x(t), \;\; \mathbf{\Phi_u} = \sum_{t=1}^{T} \Phi_u(t)$$

Frobenius norm

$$\min_{\gamma\in(0,1),\mathbf{\Phi}_x,\mathbf{\Phi}_u} \frac{1}{1-\gamma} \left\| \begin{bmatrix} Q^{1/2} & 0 \\ 0 & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_2} \quad \text{s.t.}$$

affine constraint

$$\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} = I\,, \quad \sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi}_x \\ \varepsilon_B \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \gamma(1 - C\rho^T),$$

SDP

***T* large enough**

$$\|A\Phi_x(T) + B\Phi_u(T)\| \leq C\rho^T$$

$O(T^3)$ complexity

Theorem: Gap between FIR and infinite dimensional solution decays as *O(*ρ^T*)*

Option 2: Common Lyapunov heuristic

$$
\begin{array}{ll}
\text{minimize}_{X,Z,W,\gamma} & \frac{1}{(1-\gamma)^2}\left\{\mathrm{Trace}(QW_{11}) + \mathrm{Trace}(RW_{22})\right\} \\
\text{subject to} & \begin{bmatrix} X & X & Z^* \\ X & W_{11} & W_{12} \\ Z & W_{21} & W_{22} \end{bmatrix} \succeq 0 \\
& \begin{bmatrix} X - I & \hat{A}X + \hat{B}Z & 0 & 0 \\ (\hat{A}X + \hat{B}Z)^* & X & \epsilon_A X & \epsilon_B Z^* \\ 0 & \epsilon_A X & \alpha\gamma^2 I & 0 \\ 0 & \epsilon_B Z & 0 & (1-\alpha)\gamma^2 I \end{bmatrix} \succeq 0\,.
\end{array}
$$

No provable guarantees, but works well in practice and is *much* faster to solve

Robust SLS LQR problem

$$\min_{\gamma\in(0,1),\mathbf{\Phi}_x,\mathbf{\Phi}_u} \overbrace{\frac{1}{1-\gamma}}^{\text{effect of uncertainty}} \overbrace{\left\| \begin{bmatrix} Q^{1/2} & 0 \\ 0 & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_2}}^{\text{nominal LQR cost}} \quad \text{s.t.}$$

nominal achievability

$$\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_x \\ \mathbf{\Phi}_u \end{bmatrix} = I\,, \quad \sqrt{2}\left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi}_x \\ \varepsilon_B \mathbf{\Phi}_u \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \gamma\,,$$

robust stability

infinite dimensional with tractable solutions

End-to-end sample complexity bounds

Theorem

With probability $1-\delta$, for N sufficiently large, the synthesized controller is stabilizing and achieves the relative performance bound

$$\frac{\hat{J}-J_\star}{J_\star} \leq C\underbrace{\Gamma_{cl}}_{\text{robustness}}\left(\underbrace{\lambda_{\min}(\Lambda_c)}_{\text{excitability}}^{-\frac{1}{2}} + \underbrace{\|K_\star\|}_{\text{difficulty to control}}{}_2\right)\sqrt{\frac{\sigma^2(n+p)\log(1/\delta)}{N}}$$

Closed Loop Robustness

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

Controllability Gramian

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

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

End-to-end sample complexity bounds

Theorem

With probability $1-\delta$, for N sufficiently large, the synthesized controller is stabilizing and achieves the relative performance bound

$$\frac{\hat{J}-J_\star}{J_\star} \leq C\underbrace{\Gamma_{cl}}_{\text{robustness}}\left(\underbrace{\lambda_{\min}(\Lambda_c)}_{\text{excitability}}{}^{-\frac{1}{2}} + \underbrace{\|K_\star\|}_{\text{difficulty to control}}{}_2\right)\sqrt{\frac{\sigma^2(n+p)\log(1/\delta)}{N}}$$

Hard to estimate

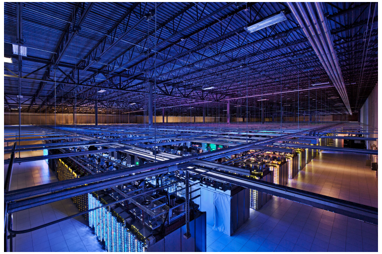
Control insensitive to mismatch

Easy to estimate

Control very sensitive to mismatch

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

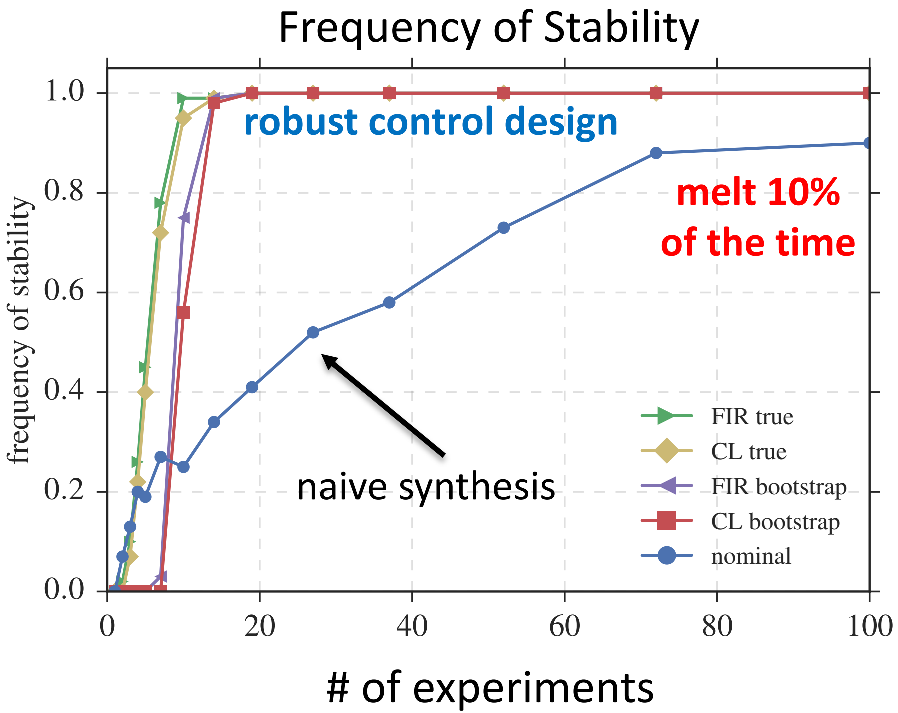
Cooling a server farm



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

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

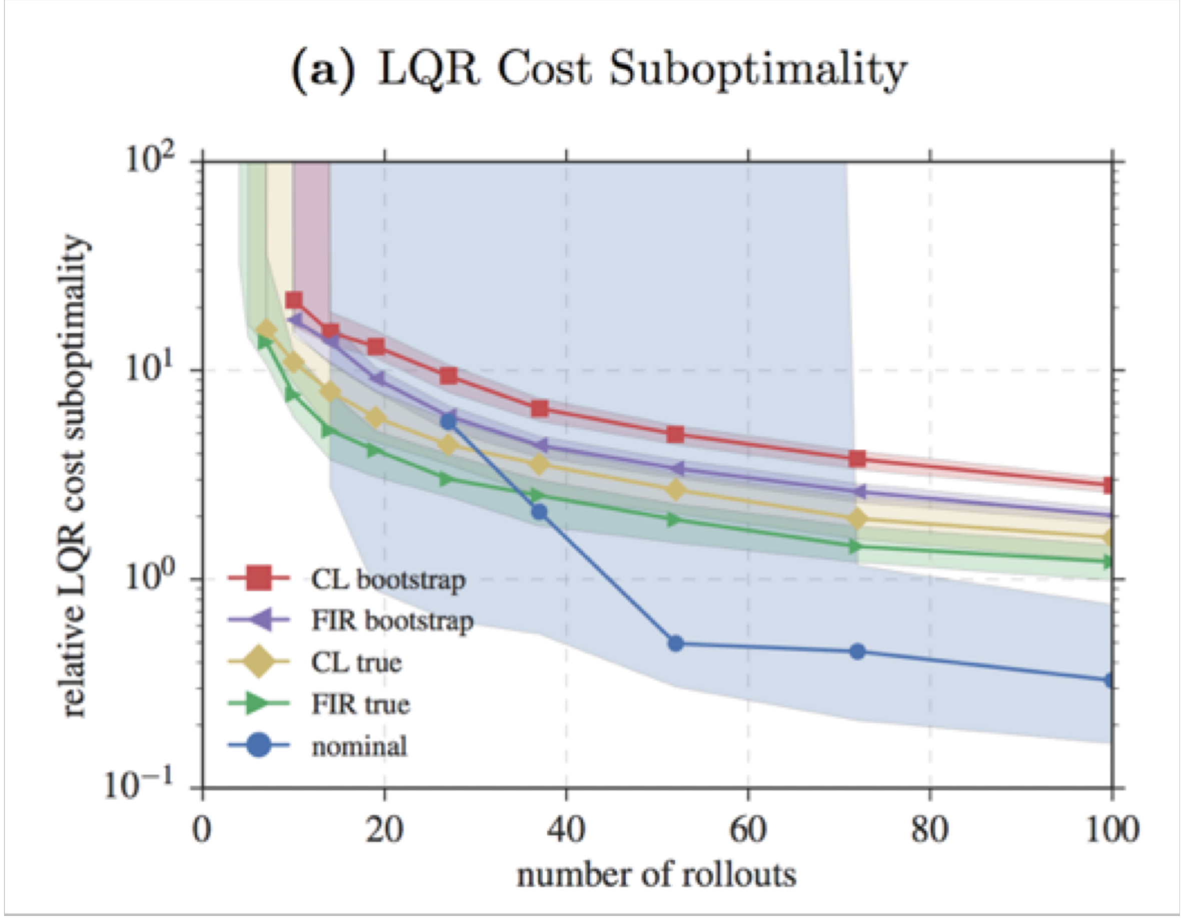
[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]



Least-squares estimate may yield unstable controller

Robust synthesis yields stable controller

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]



Robust synthesis reduces variance

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

Two main takeaways

Robustness is key

Robustness matters in practice and makes theory tractable

Robust and optimal control as optimization

System Level Synthesis

[Dean, Mania, **Matni**, Recht, Tu, FoCM 2018, submitted]

Extensions: LQR++

Safety

state/input constraints can be incorporated naturally

Adaptive online algorithm

can be incorporated into online algorithm with $O(T^{2/3})$ regret

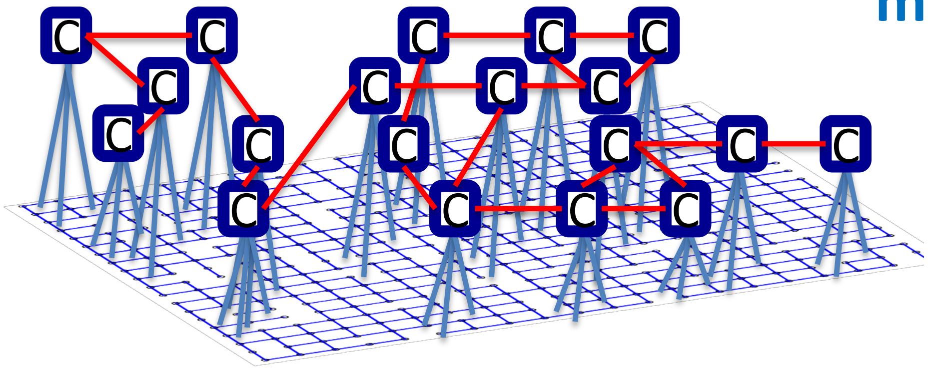
Large-scale adaptive control

Structure can be exploited and enforced

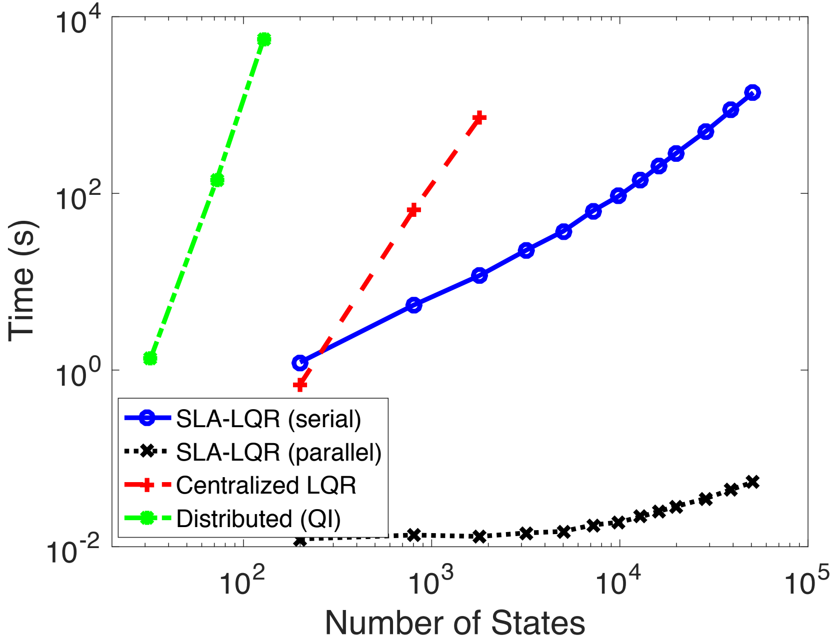
Nonlinear?

Still looking for the right approach/parameterization…

An aside on distributed control



System Level Synthesis: maximally exploit system structure



I can't find good dynamics models!

[Wang, **Matni**, Doyle, ACC 2017, TAC 2019]

References

- Y.-S. Wang, N. Matni, and J. C. Doyle, A system level approach to controller synthesis, IEEE TAC 2019, To Appear.

- S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, On the sample complexity of the linear quadratic regulator, FoCM 2018, Accepted s.t. minor revisions.

- S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, Regret Bounds for Robust Adaptive Control of the Linear Quadratic Regulator, NeurIPS 2018.

A brief (and incomplete) review

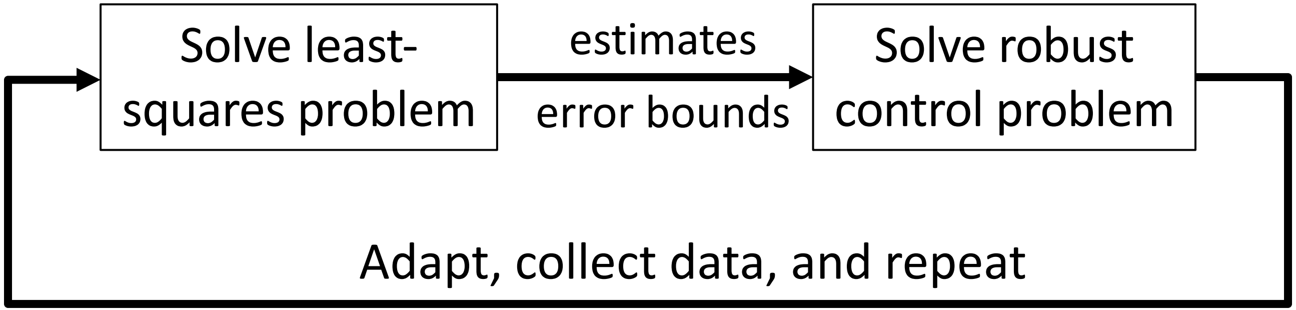
- **Learning Linear Systems**
  - Hardt, Ma, Recht, 2016: descent learns stable linear systems, strong assumptions
  - Hazan, Singh, Zhang, 2017: polynomial time algorithm, symmetry assumption

- **Probably Approximately Correct (PAC)**
  - Fietcher 1997: discounted costs, many assumptions on contractivity, some bugs in proof.

- **Optimism in the Face of Uncertainty (OFU)**
  - Abbas-Yadkori and Szepesvári, 2011: regret exponential in the dimension, no guarantee of parameter convergence, OFU NP-hard subroutine.
  - Faradonbeh, Tewari, Michailidis, 2017: address issues mentioned above except for OFU NP-hard subroutine.

- **Thompson Sampling**
  - Ouyang, Gagrani, Jain, 2017: replace OFU subroutine with random sampling approach, strong assumptions on uniform stability (contractivity).

Adaptive control as regret minimization



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

Line of work initiated by Abbasi-Yadkori and Szepesvari in 2011

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

Adaptive control as regret minimization

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

$\textbf{Theorem}$: With probability at least $1 - \delta$,

$$R(T) = \tilde{O}(T^{\frac{2}{3}})$$

Guaranteed stability throughout and identification of true system parameters

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

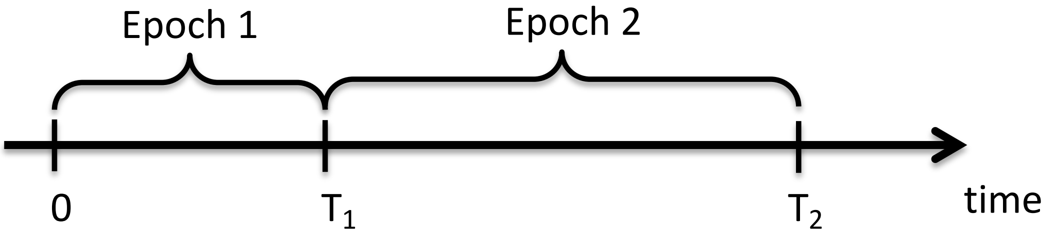
Adaptive control as regret minimization

Approach	Regret	Type	"Cheat"
Robust SLS	$O(T^{2/3})$	High probability	"Small" initial uncertainty
Thompson Sampling [Abeille, Lazariz 17]	$O(T^{2/3})$	High probability	Contractivity
Thompson Sampling [Ouyang, Gagrani, Jain 17]	$O(T^{1/2})$	Expectation	Pair-wise stability (contractivity)
OFU [Farabondeh, Tewari, Michalidis, 17]	$O(T^{1/2})$	High probability	NP-hard subroutine
LSTD [Abbasi-Yadkori, Lazic, Szepesvari, 18]	$O(T^{3/4})$	High probability	Contractivity

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

Adaptive control as regret minimization

At every T_i do:



- $(\hat{A}^{(i)}, \hat{B}^{(i)}) = \underset{(A,B)}{\operatorname{argmin}} \sum_{t \in E_i} \|x_{t+1} - Ax_t - Bu_t\|_2^2$
- $\mathbf{K}^{(i)} = \mathrm{RobustSLS}(\hat{A}^{(i)}, \hat{B}^{(i)}, \underline{\epsilon^{(i)}})$ **sharp bounds from time-series data?**
- $\mathbf{u}^{(i)} = \underline{\mathbf{K}^{(i)}\mathbf{x}} + \underline{\eta^{(i)}}$ **explore vs. exploit?**

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

Sharp bounds from time-series data:

Set $\eta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\eta^2 I) \overset{\text{OLS}}{\Longrightarrow} \left\| \begin{matrix} \hat{A} - A \\ \hat{B} - B \end{matrix} \right\| \leq \tilde{O}\left(\frac{1}{\sigma_\eta T^{\frac{1}{2}}} \right)$

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

Explore vs. exploit:

Model Mismatch **Excitation**

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

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

Adaptive control as regret minimization

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

Theorem: With probability at least $1-\delta$,

$$R(T) = \tilde{O}(T^{\frac{2}{3}})$$

Guaranteed stability throughout and identification of true system parameters

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

Lower bounds for epsilon-greedy approach

$$\mathbf{u}^{(i)} = \underbrace{\mathbf{K}^{(i)}\mathbf{x}}_{\text{exploit}} + \underbrace{\eta^{(i)}}_{\text{explore}} \qquad \eta_t^{(i)} \sim \mathcal{N}(0, \sigma_{\eta,i}^2 I)$$

exploit vs. explore

$$\sigma_{\eta,i}^2 = \tilde{O}(T_i^{-\alpha})$$

Regret lower bounded by:

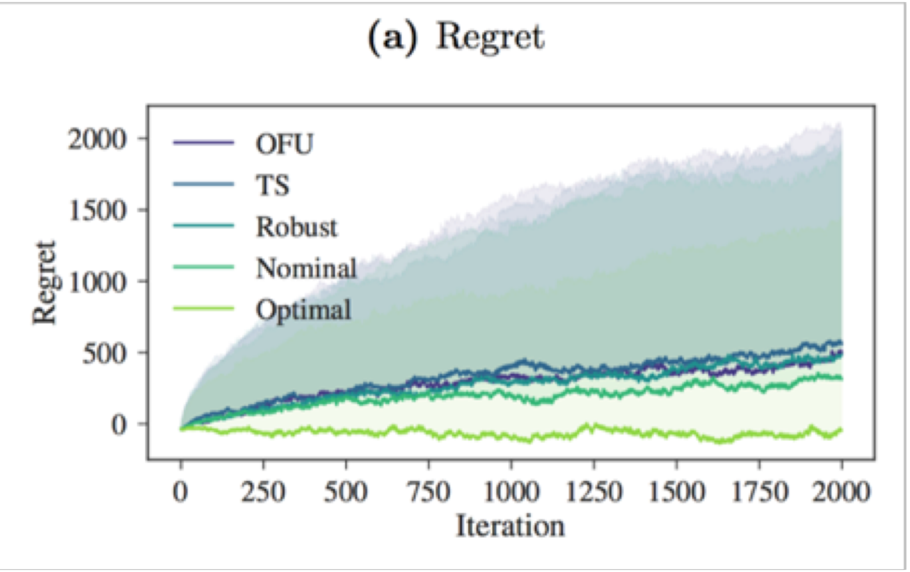
$$\sum_{t=1}^{T} \mathbb{E}\left[x_t^\top Q x_t + u_t^\top R u_t - J_\star\right] \geq \tilde{\Omega}(T^{1-\alpha})$$

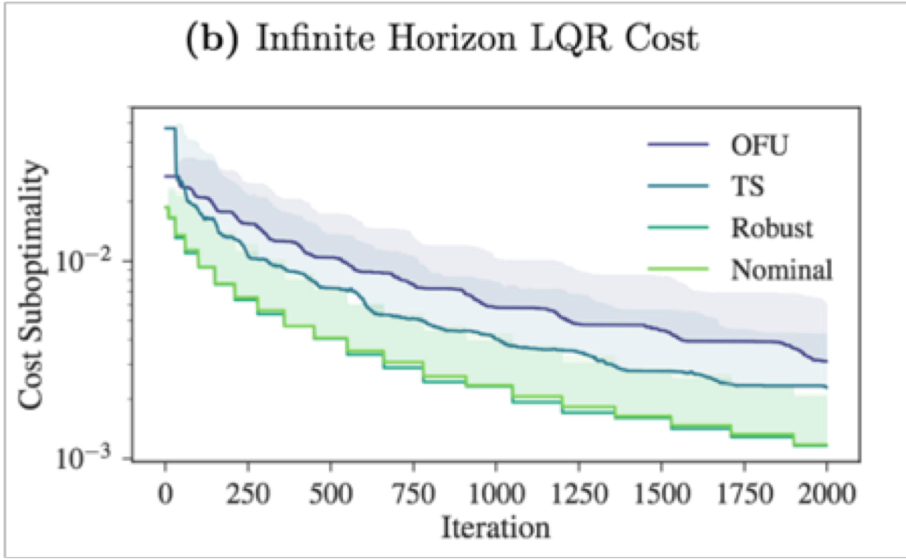
Estimation error of ϵ incurs $\tilde{\Omega}(\epsilon^{-2})$ regret

$$\left\| \begin{matrix} \hat{A} - A_\star \\ \hat{B} - B_\star \end{matrix} \right\| \leq \epsilon \implies T \geq \tilde{\Omega}\left(\epsilon^{-\frac{2}{1-\alpha}}\right) \Rightarrow R(T) \geq \tilde{\Omega}\left(\epsilon^{-2}\right)$$

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

Adaptive control as regret minimization





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