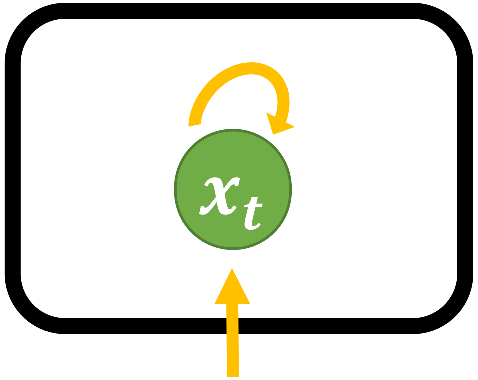
Online Control with Adversarial Disturbances

Naman Agarwal

Google AI Princeton

Joint Work with
Brian Bullins, Elad Hazan, Sham Kakade, Karan Singh

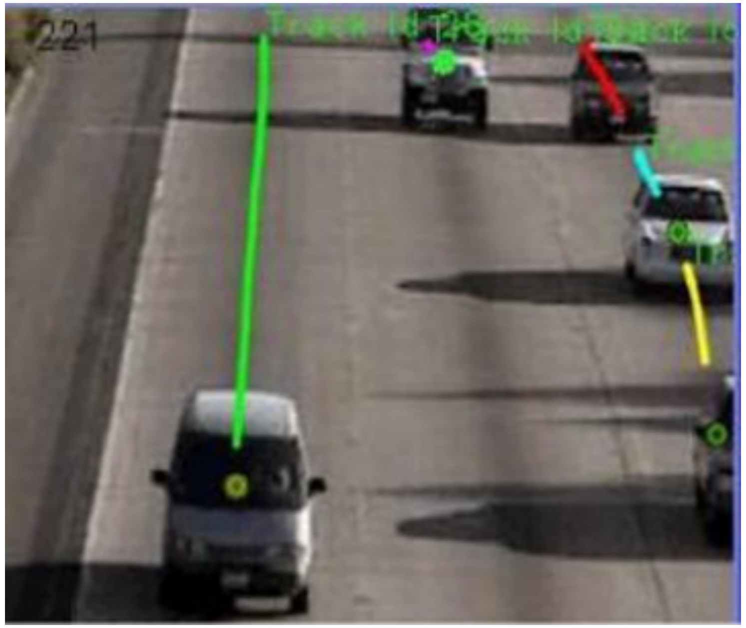
Dynamical Systems with Control

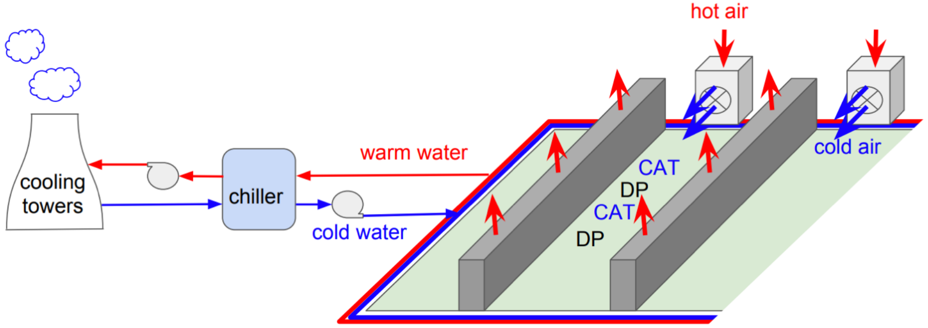


u_t

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

- Robotics
- Autonomous Vehicles
- Data Center Cooling





[Cohen et al ‘18]

Our Setting

x_t : State
u_t : Control

Robustly Control a Noisy Linear Dynamical System

$$x_{t+1} = Ax_t + Bu_t + w_t$$

- **Known Dynamics**
- **Fully Observable State**

Our Setting

x_t : State
u_t : Control

Robustly Control a Noisy Linear Dynamical System

$$x_{t+1} = Ax_t + Bu_t + w_t$$

- **Known Dynamics**
- **Fully Observable State**

Disturbance w_t adversarially chosen ($||w_t|| \leq 1$)

Our Setting

x_t : State
u_t : Control

Robustly Control a Noisy Linear Dynamical System

$$x_{t+1} = Ax_t + Bu_t + w_t$$

- **Known Dynamics**
- **Fully Observable State**

Disturbance w_t adversarially chosen ($||w_t|| \leq 1$)

Minimize Costs - $\sum c_t(x_t, u_t)$

- **Online and Adversarial**
- **General Convex Function**

Our Setting

x_t : State
u_t : Control

Robustly Control a Noisy Linear Dynamical System

$$x_{t+1} = Ax_t + Bu_t + w_t$$

- **Known Dynamics**
- **Fully Observable State**

Disturbance w_t adversarially chosen ($||w_t|| \leq 1$)

Minimize Costs - $\sum c_t(x_t, u_t)$

- **Online and Adversarial**
- **General Convex Function**

vs. Linear Quadratic Regulator (LQR):

Adversarial vs Random Disturbance

Online, Convex Costs vs Known Quadratic Loss

Goal – Minimize Regret

- Fixed Time horizon - T
- Produce actions $u_1, u_2 \ldots u_T$ to minimize ***regret*** w.r.t *best* in hindsight

$$\sum_{t=1}^{T} c_t(x_t, u_t) \; - \; \min_K \left(\sum_{t=1}^{T} c_t(x_t(K), K x_t(K)) \right)$$

Goal – Minimize Regret

- Fixed Time horizon - T
- Produce actions $u_1, u_2 \dots u_T$ to minimize ***regret*** w.r.t *best* in hindsight

$$\sum_{t=1}^{T} c_t(x_t, u_t) \; - \; \min_K \left(\sum_{t=1}^{T} c_t(x_t(K), K x_t(K)) \right)$$

u_t only knows $w_1 \dots w_t$

Best Linear Policy knowing $w_1 \dots w_T$

Optimal for LQR

Goal – Minimize Regret

- Fixed Time horizon - T
- Produce actions $u_1, u_2 \dots u_T$ to minimize ***regret*** w.r.t *best* in hindsight

$$\sum_{t=1}^{T} c_t(x_t, u_t) \; - \; \min_K \left(\sum_{t=1}^{T} c_t(x_t(K), K x_t(K)) \right)$$

u_t only knows $w_1 \dots w_t$

Best Linear Policy knowing $w_1 \dots w_T$

Optimal for LQR

Counterfactual Regret – $x_t(K)$ depends on K

Previous work: H_∞ Control

- min-max problem, worst case perturbation:

$$\min_{\mathrm{u}} \max_{w_{1:T}} \sum_t c(x_t, u(w_{t-1}, \ldots w_0))$$

- Disturbance $w_{1:T}$ adversarially chosen

Previous work: H_∞ Control

- min-max problem, worst case perturbation:

$$\min_{u} \max_{w_{1:T}} \sum_t c(x_t, u(w_{t-1}, \ldots w_0))$$

- Disturbance $w_{1:T}$ adversarially chosen

Compute

- Closed form: Quadratics
- Difficult for general costs

Adaptivity

- H_∞ is Pessimistic
- Regret: adapts to favorable sequence

Main Result

Efficient Online Algorithm: $u_1 \ldots u_T$ s.t.

$$\sum_{t=1}^{T} c_t(x_t, u_t) \; - \; \min_{K \in stable} \left(\sum_{t=1}^{T} c_t(x_t, Kx_t)\right) \; \leq O(\sqrt{T})$$

- **Convexity through Improper Relaxation**
- **Efficient → Polynomial in system parameters, logarithmic in T**

Outline of the approach

1. Improper Learning:
   Can we even figure out the best in hindsight policy?
   ”relaxed” policy class: Next Control a linear function of previous w_t

2. Strong Stability ⇒
   error feedback policy: learn change to action via ”small horizon” of previous disturbances.

3. Small Horizon ⇒
   Efficient Reduction to Online Convex Optimization (OCO) with memory [Anava et al.]

Thank You!

For more details
please visit the Poster
Pacific Ballroom #155

namanagarwal@google.com