A Regret Minimization Approach to Iterative Learning Control

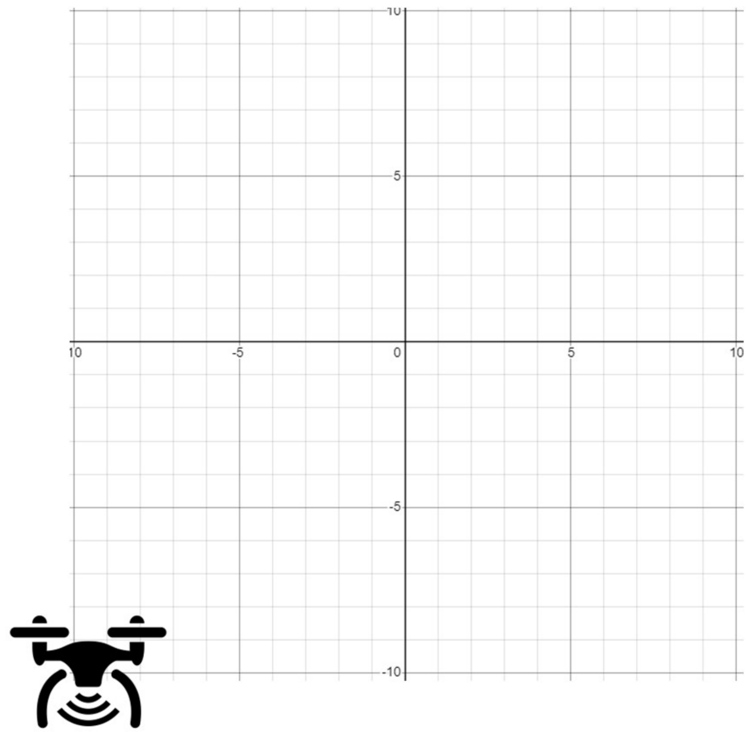
Naman Agarwal, Elad Hazan, Anirudha Majumdar, Karan Singh

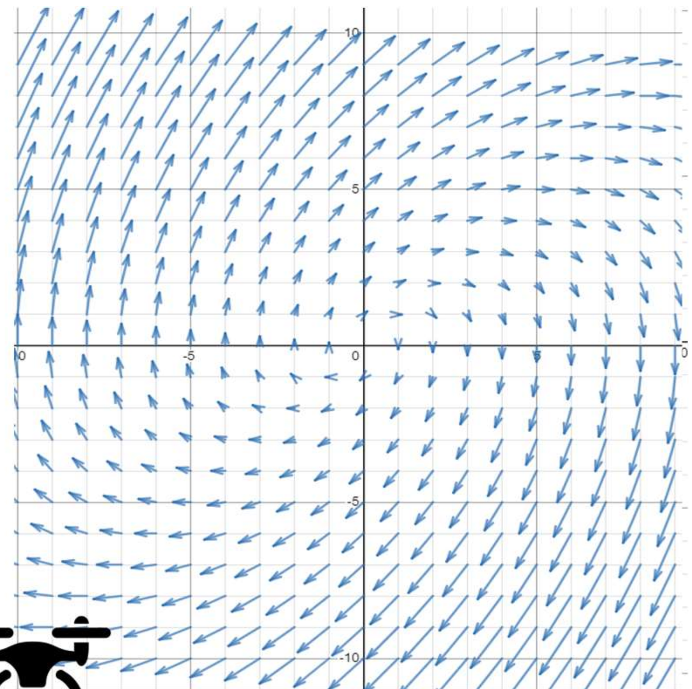



ICML | 2021

PRINCETON UNIVERSITY

Microsoft Research


10
5
10
-5
0
5
10
-5
-10

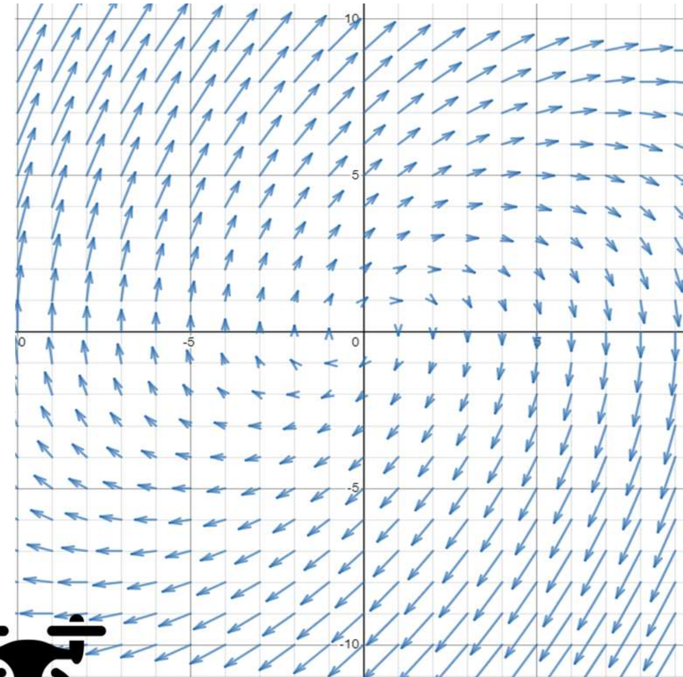


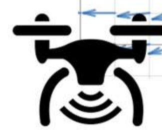
10
5
-5
-10
0

Wide-angle Perspective

Learn an (adaptive) policy

+ given an approximate model
+ subject to changing, unknown perturbations
+ in a *handful* of episodes





Why?

+ Arises in real-world applications
+ Sandboxed setup for sim2real, meta-learning, policy transfer

Problem Setting

Episode 1 of T

State

Action

Learner has a model $f(x,u)$

Timestep 1 of H

Play action u_h

$x_{h+1} = f(x_h, u_h) + w_h$

Suffer $c(x_h, u_h)$

Episodic $Cost_t = \sum_h c_h(x_h, u_h)$

Compare:

Iterative LQR > *no perturbations*

Iterative LQG > *Gaussian perturbations*

Model Predictive Control > *One-shot*

Iterative Learning Control > *Same setup (here)*

Unknown, Nonstationary Perturbations

(***arbitrary***, *no dist. assumption)*

(changes every ***step****, every* ***episode****)*

Objective: *Planning Regret*

Inter-episodic learning
Best Overall Open-Loop Plan

Intra-episodic learning
Instance-optimal Adaptation

$$\sum_{t=1}^{T} Cost_t(Alg) \approx \min_{U^*_{1:H}} \sum_{t=1}^{T} \min_{\pi^*_t} Cost_t(U^*_{1:H} + \pi^*_t)$$

Main Result

For time-varying linear dynamical system

Subject to arbitrary perturbation

An efficient gradient-based algorithm

$$\frac{1}{TH}\left(\sum_{t=1}^{T} Cost_t(Alg) - \min_{U^*_{1:H}} \sum_{t=1}^{T} \min_{\pi^*_t} Cost_t(U^*_{1:H} + \pi^*_t)\right) \leq \frac{1}{\sqrt{T}} + \frac{1}{\sqrt{H}}$$

Inter-episodic learning
Best Overall Open-Loop Plan

Intra-episodic learning
Instance-optimal Adaptation

Preview: *Nested Online Convex Opt.*

Biconvex $f_{t,h}(x, y)$.

Episode 1 of T

Timestep 1 of H

Choose parameter x, y

Update $y = y - \eta_y \nabla_y f_{t,h}(x, y)$

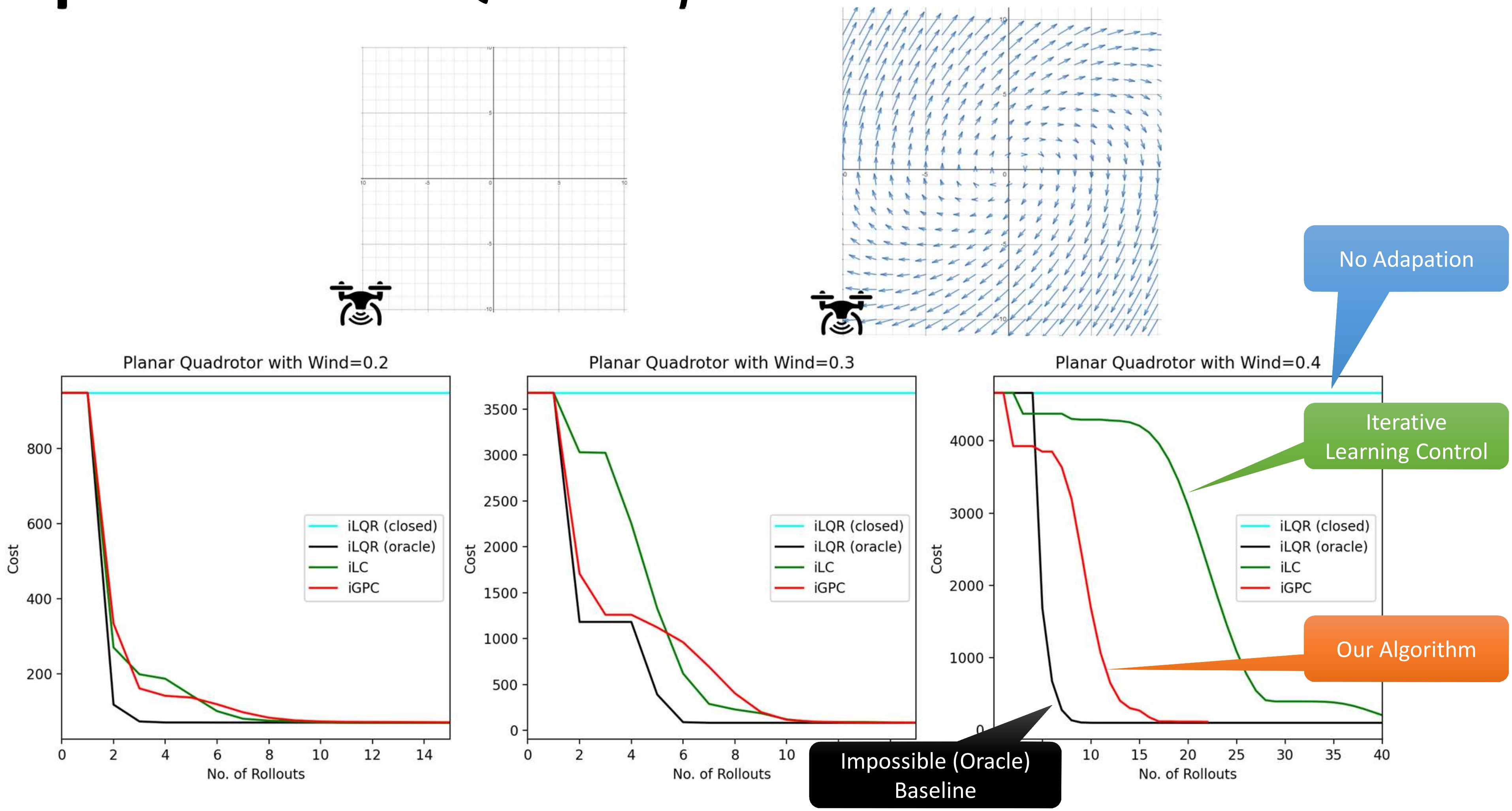
Update x $= x - \eta_x \nabla_x f_{t,h}(x, y)$

$$\frac{1}{TH}\left(\sum_{t=1}^{T}\sum_{h=1}^{H} f_{t,h}(x_t, y_{t,h}) - \min_{x^*}\sum_{t=1}^{T}\min_{y_t^*}\sum_{h=1}^{H} f_{t,h}(x^*, y_t^*)\right) \leq \frac{R_x}{\sqrt{T}} + \frac{R_y}{\sqrt{H}}$$

<u>Captures Initialization-based Meta Learning:</u>

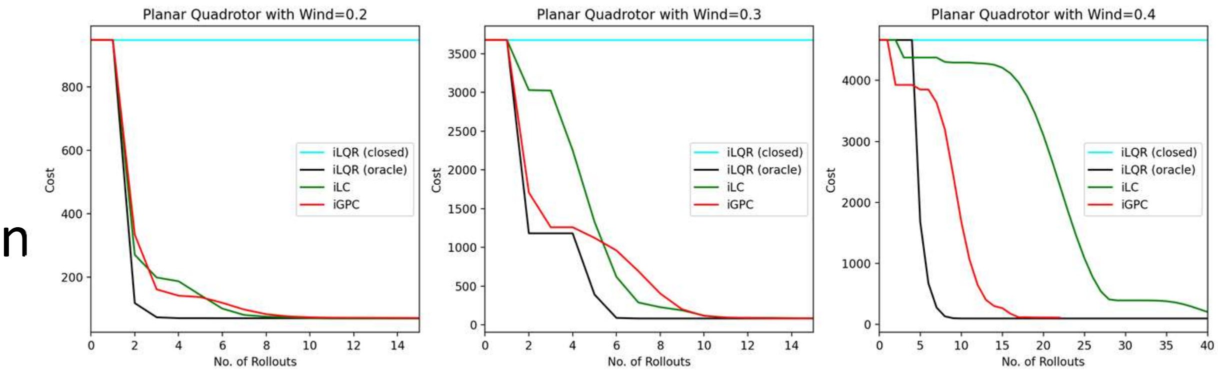
x, y share the same space and choose $x + y$

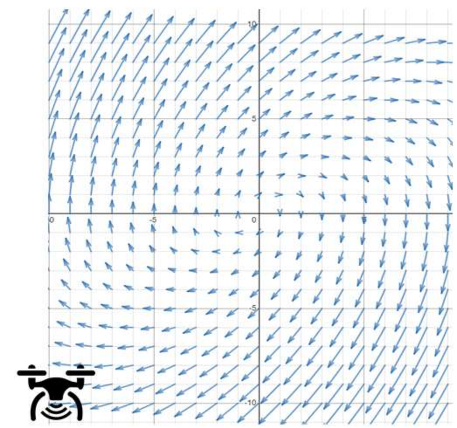
Experiment 1: *Quadcopter in Wind*



Learn an (adaptive) <u>policy</u>

+ given an <u>approximate model</u>
+ subject to arbitrary perturbation
+ in a *handful* of <u>episodes</u>





Intra-episodic learning

Instance-optimal Adaptation

Inter-episodic learning

Best Overall Open-Loop Plan

$$\frac{1}{TH}\left(\sum_{t=1}^{T} Cost_t(Alg) - \min_{U^*_{1:H}} \sum_{t=1}^{T} \min_{\pi^*_t} Cost_t(U^*_{1:H} + \pi^*_t)\right) \leq \frac{1}{\sqrt{T}} + \frac{1}{\sqrt{H}}$$