Learning-Augmented Linear Quadratic Control

with untrusted predictions

Tongxin Li
School of Data Science
The Chinese University of Hong Kong (Shenzhen)

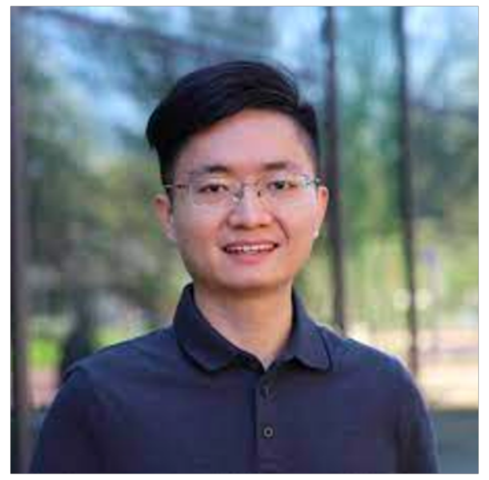
Acknowledgements

Ruixiao Yang

MIT

Guannan Qu

CMU

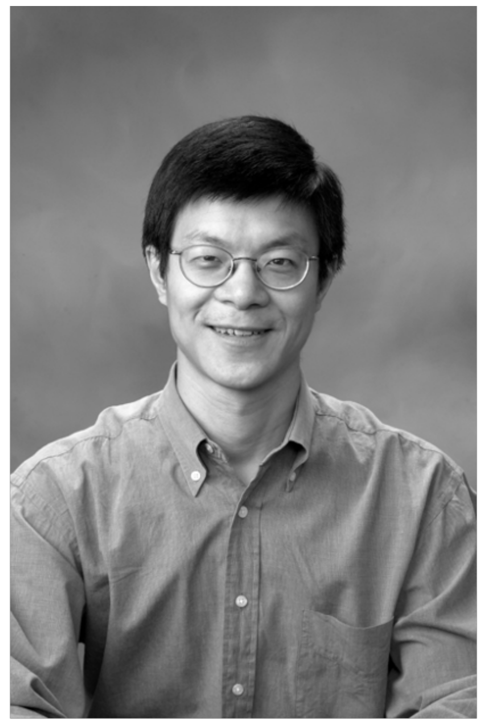


Guanya Shi

CMU

Chenkai Yu

Columbia



Steven Low

Caltech

Adam Wierman

Caltech

What is “Learning-Augmented”?

AI/ML Methods/Predictions

Classical Methods

AlphaGo

Suggested Moves

Image sources: DeepMind, Quartz, Insider

OpenAI

Suggested Strategies

Image sources: Dota2, OpenAI

What is “Learning-Augmented”?

***Blackbox** AI/ML Methods/**Imperfect** Predictions*

Classical Methods

CITY A.M.

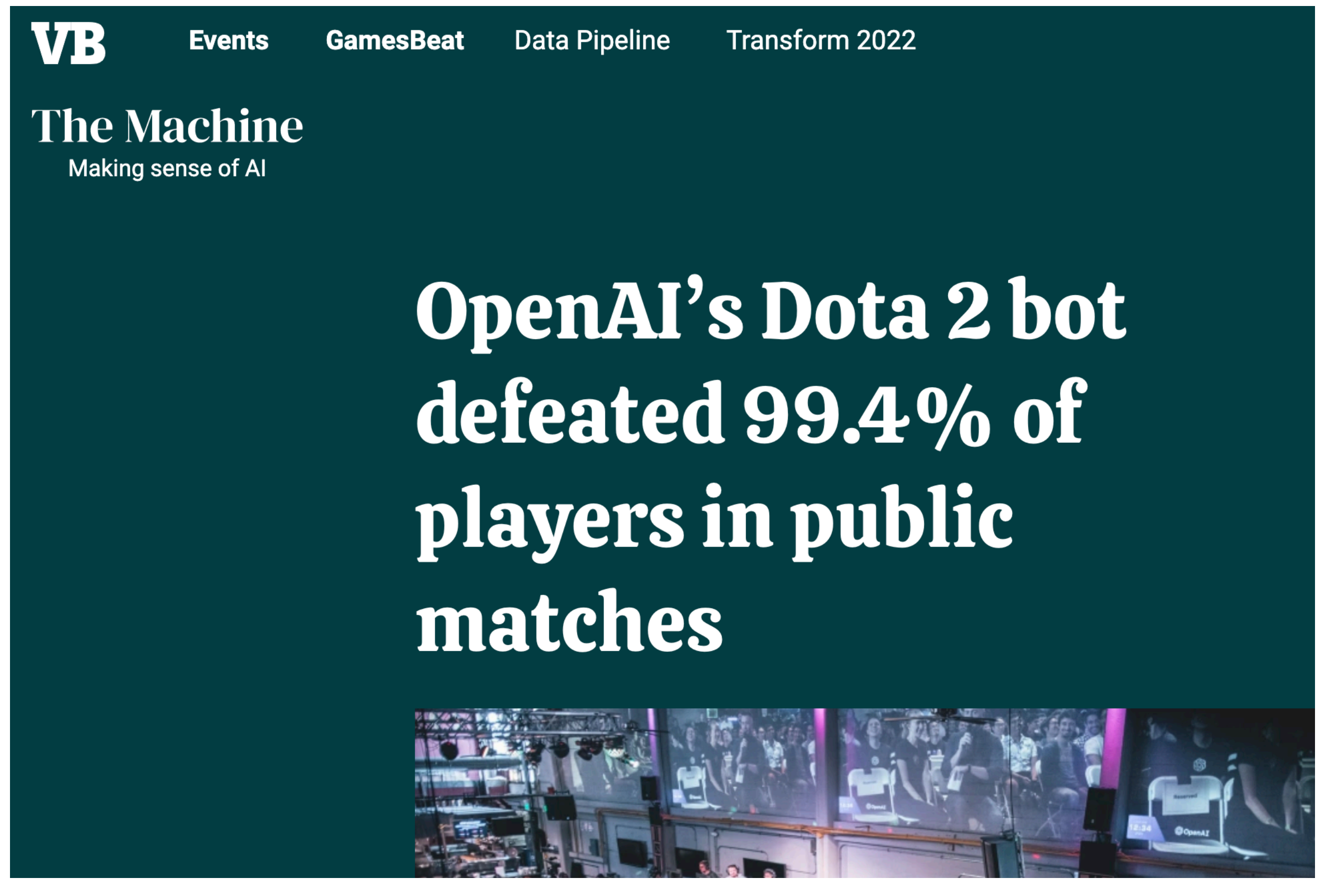
SUNDAY 13 MARCH 2016 9:45 AM

AlphaGo fourth Go game result: Google DeepMind’s artificial intelligence just made a mistake and lost to human player Lee Sedol

LYNSEY BARBER

Suggested Moves

Image sources: DeepMind, Quartz, Insider



VB Events GamesBeat Data Pipeline Transform 2022

The Machine

Making sense of AI

OpenAI’s Dota 2 bot defeated 99.4% of players in public matches

Suggested Strategies

Image sources: Dota2, OpenAI

Defeated by $0.6\,\%$ using adversarial strategies

Why Need "Learning-Augmented"?

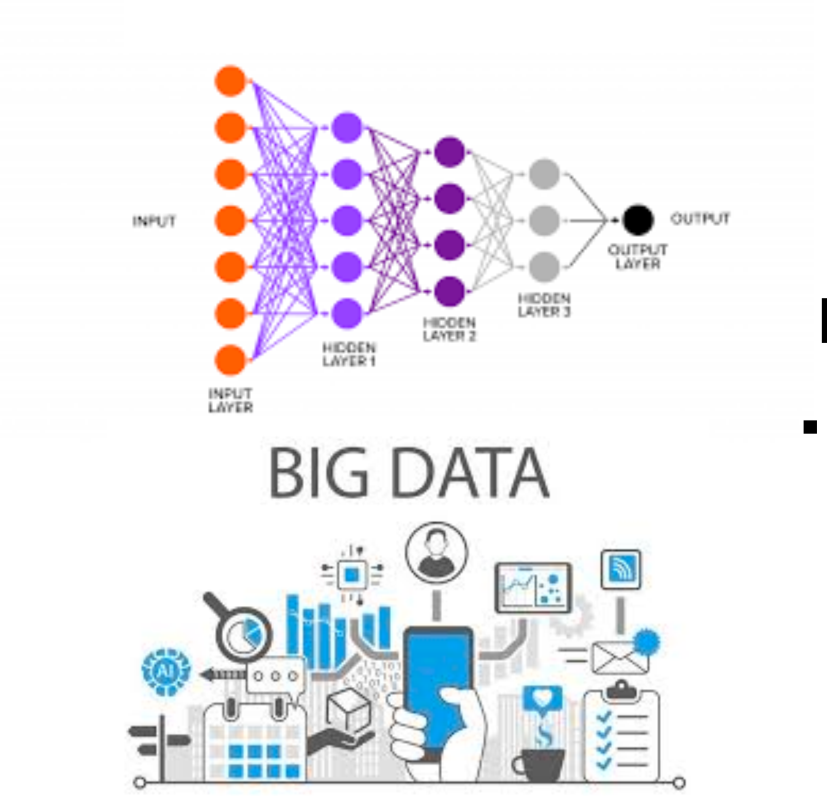
Blackbox *AI/ML Methods/Imperfect Predictions*

- *Safety* issues (lack of worst-case performance guarantees)
- *Scalability and privacy* issues

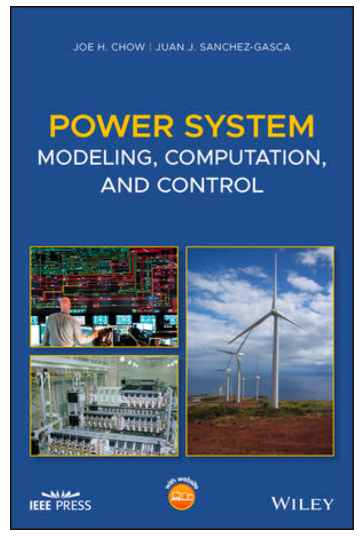
Classical Methods

- Do not take advantage of data and experience
- Need model approximations (e.g., Nonlinear to linear, AC-OPF to DC-OPF)
- Emphasize "too much" on worst-case performance
- Worse performance on-average

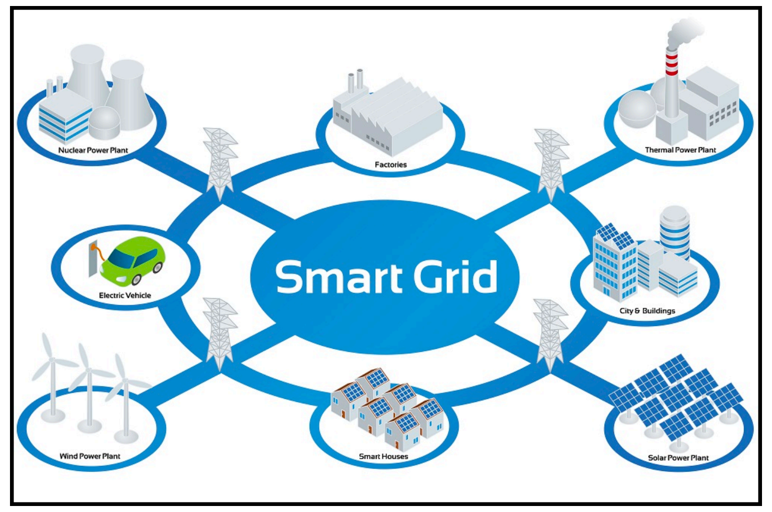
Taking Advantage of Classical Methods and AI ...

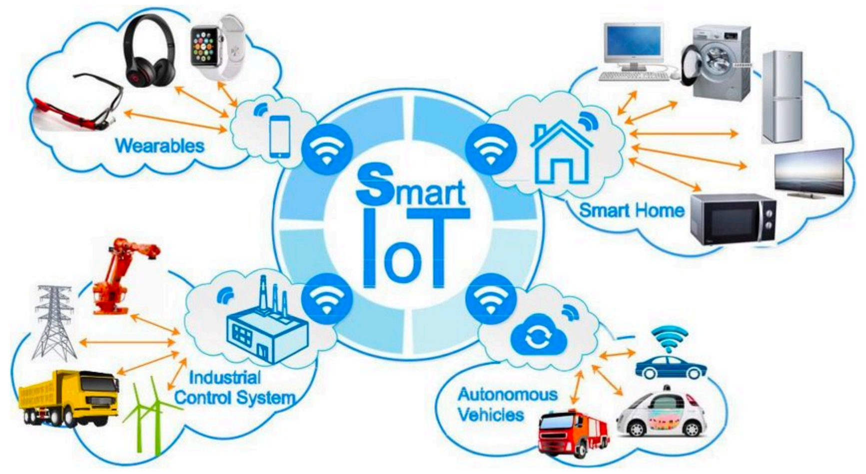


Predictions/Advice



=





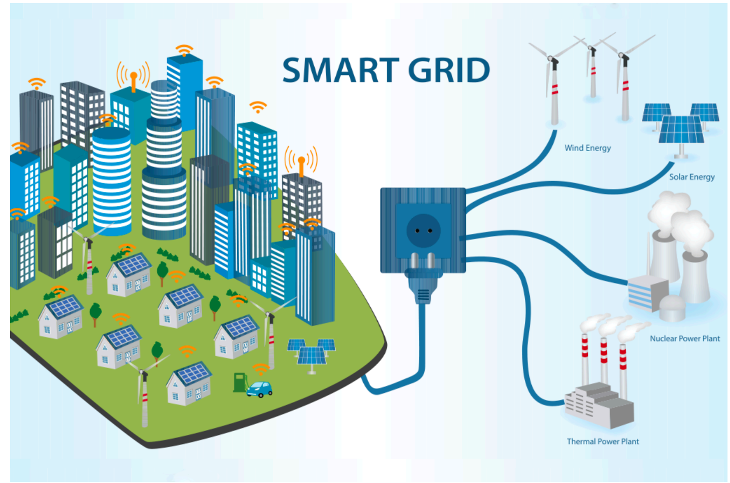


Image sources: Medium, Begum et. al.

Learning-Augmented Algorithms in Cyber-Physical Systems

Theoretical CS Problems	Imperfect Predictions	
Ski-rental	Number of Skiing Days	[Wei et. al. NeurIPS 2020] [Purohit et. al. NeurIPS 2018]
Secretary Problem	Maximum Price	[Antoniadis et. al. NeurIPS 2020]
Online Bipartite Matching	Adjacent Edge-weights	
	Black-box AI/ML Advice	
Convex Body Chasing	Suggested Actions	[Christianson et. al. COLT 2022]
Online Subset Sum	Decision	[Xu et. al. Journal of Global Optimization 2022]
Online Set Cover	Predicted Covering	[Bamas et. al. NeurIPS 2020]
…	…	

Classical Cyber-Physical Systems	Imperfect Predictions/Black-box AI/ML Advice		
Linear Quadratic Control	**Perturbations**	**[Li et. al. SIGMETRICS 2022]**	**Today's Topic**
Control with Nonlinear Dynamics	Blackbox Policy	[Li et. al. Preprint 2022]	
Two-controller System	Flexibility Advice	[Li et. al. e-Energy 2020a] [Li et. al. TSG 2021] [Li et. al. SIGMETRICS 2021]	
Learning-Augmented Decision-Making and Control			
…			

More in this seminar!

Our Problem: AI Predictions in Control

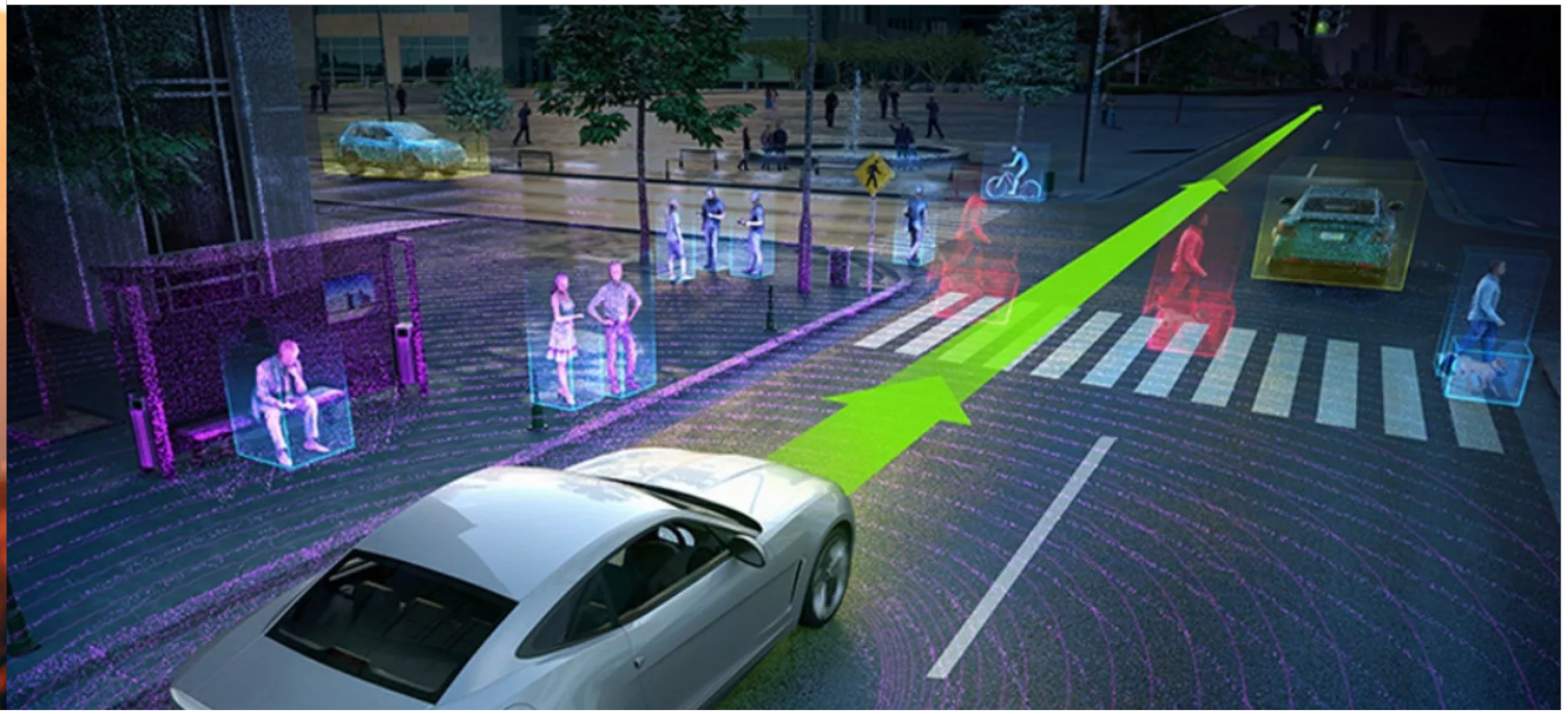
Predictions are widely available

Model Predictive Control (MPC)

Power grids

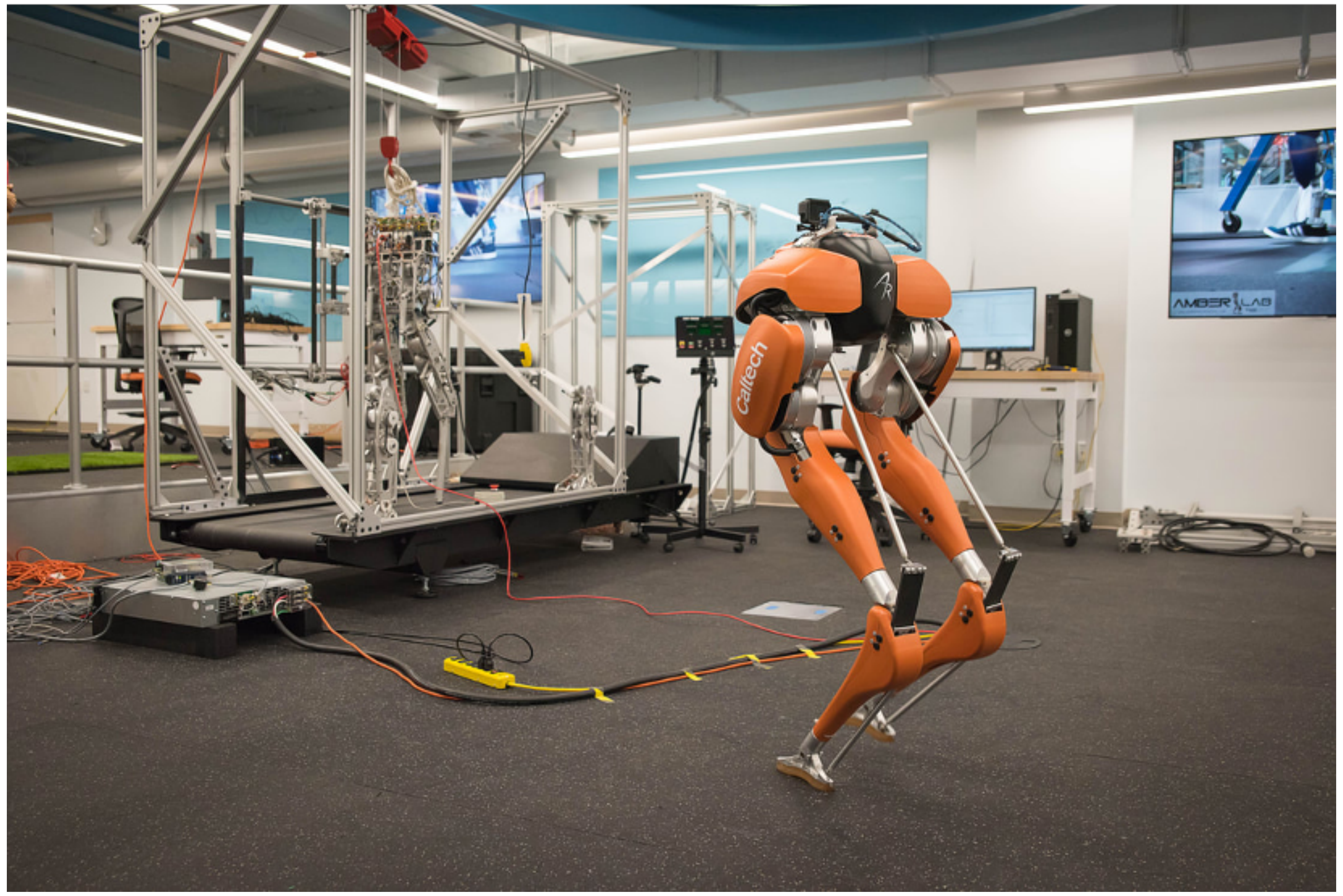
Drone control

Autonomous driving

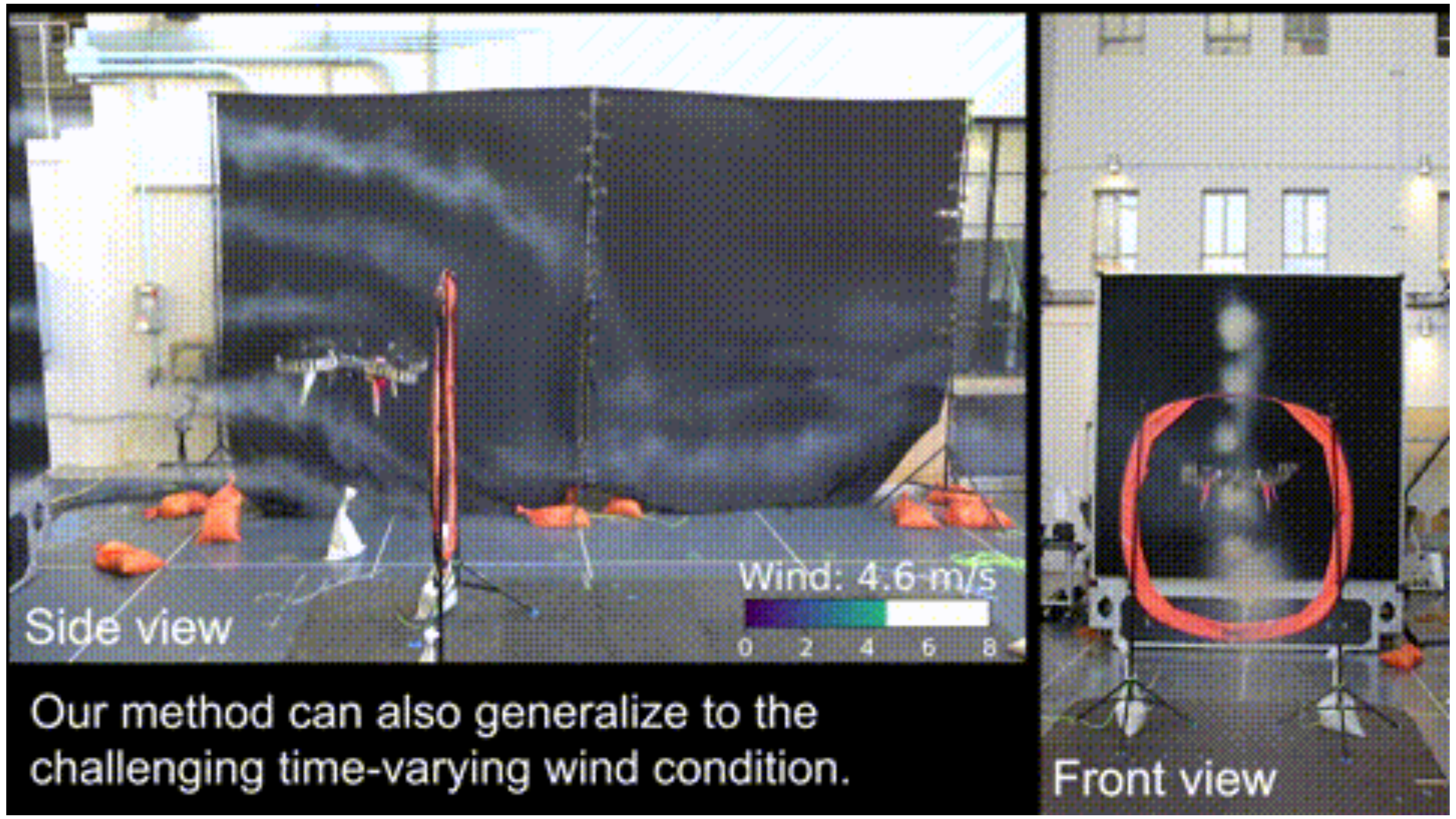


Having predictions is necessary

Robot walking



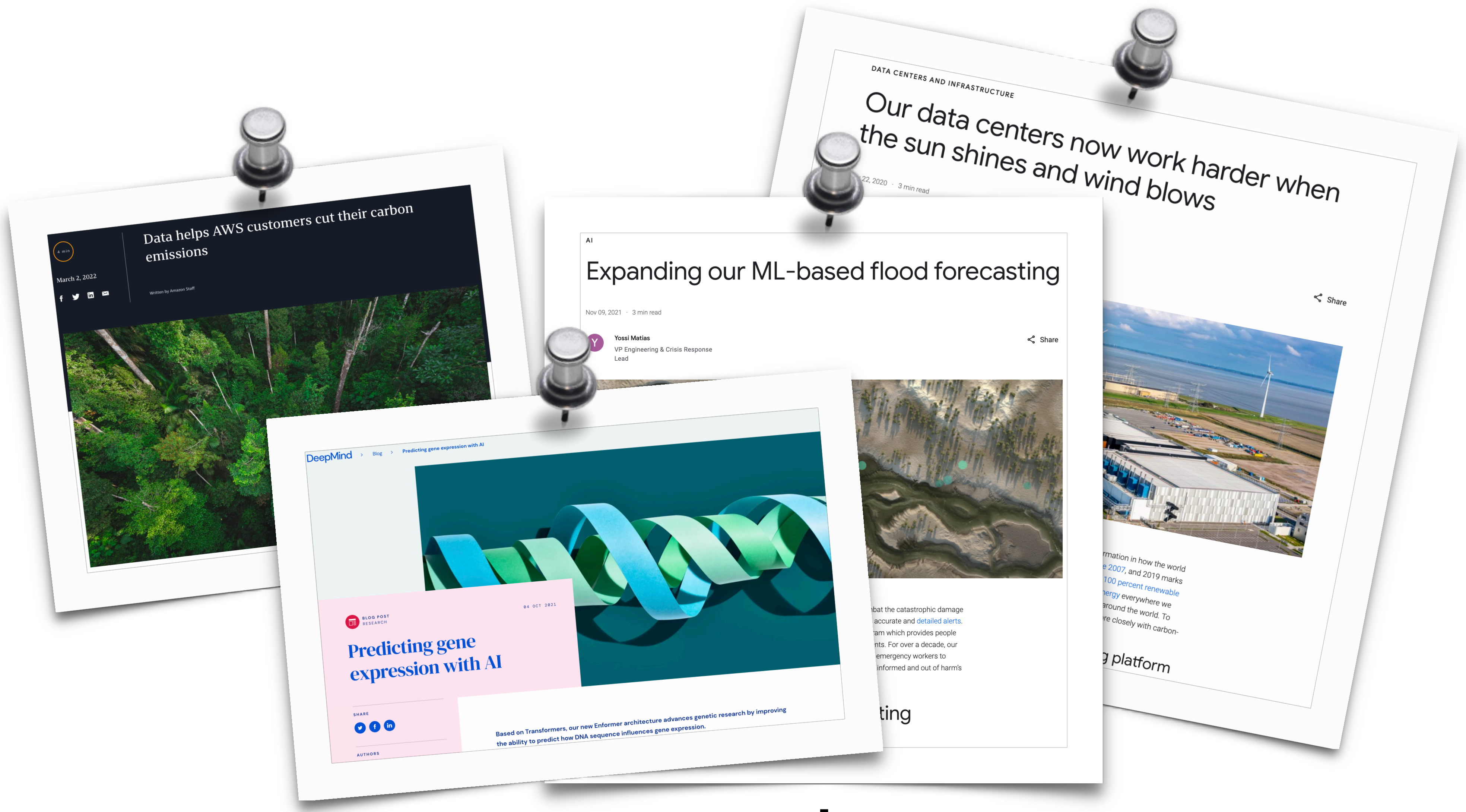
Drone flying

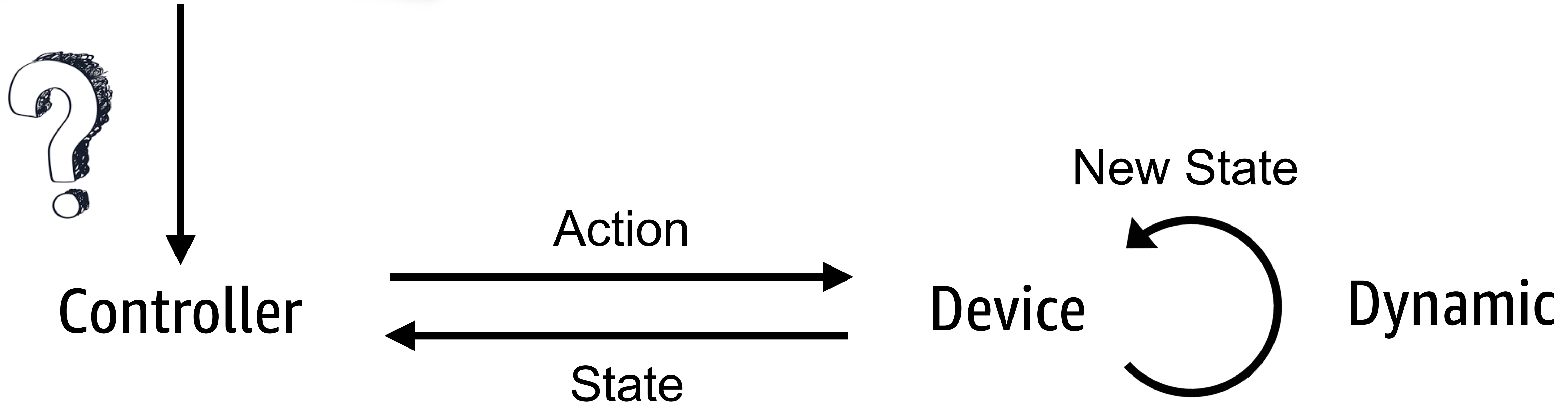


Our Problem: AI Predictions in Control

When Control meets Learning ...

Black-box AI predictions/advice can either be GOOD or BAD



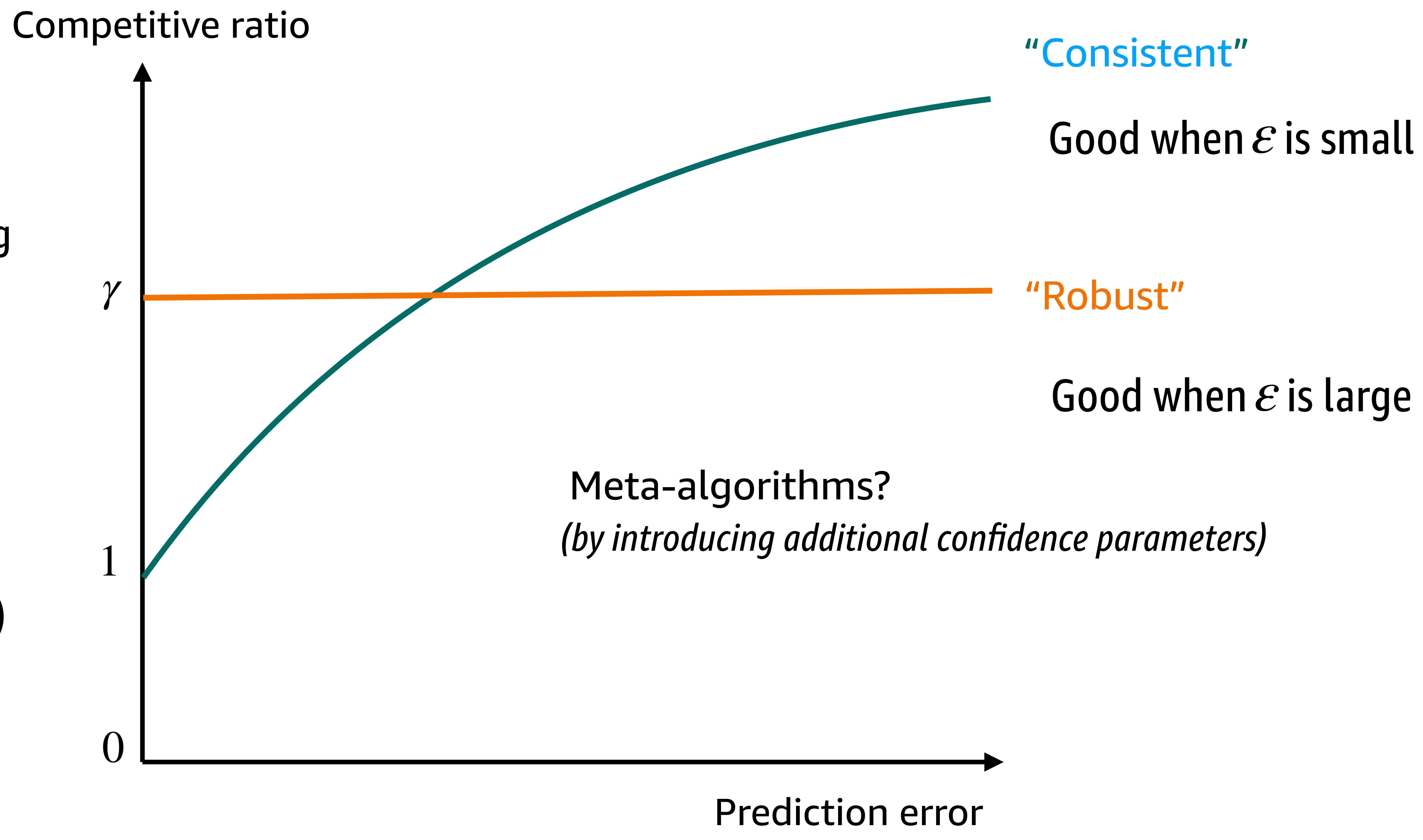


General Goal of Learning-Augmented Algorithms

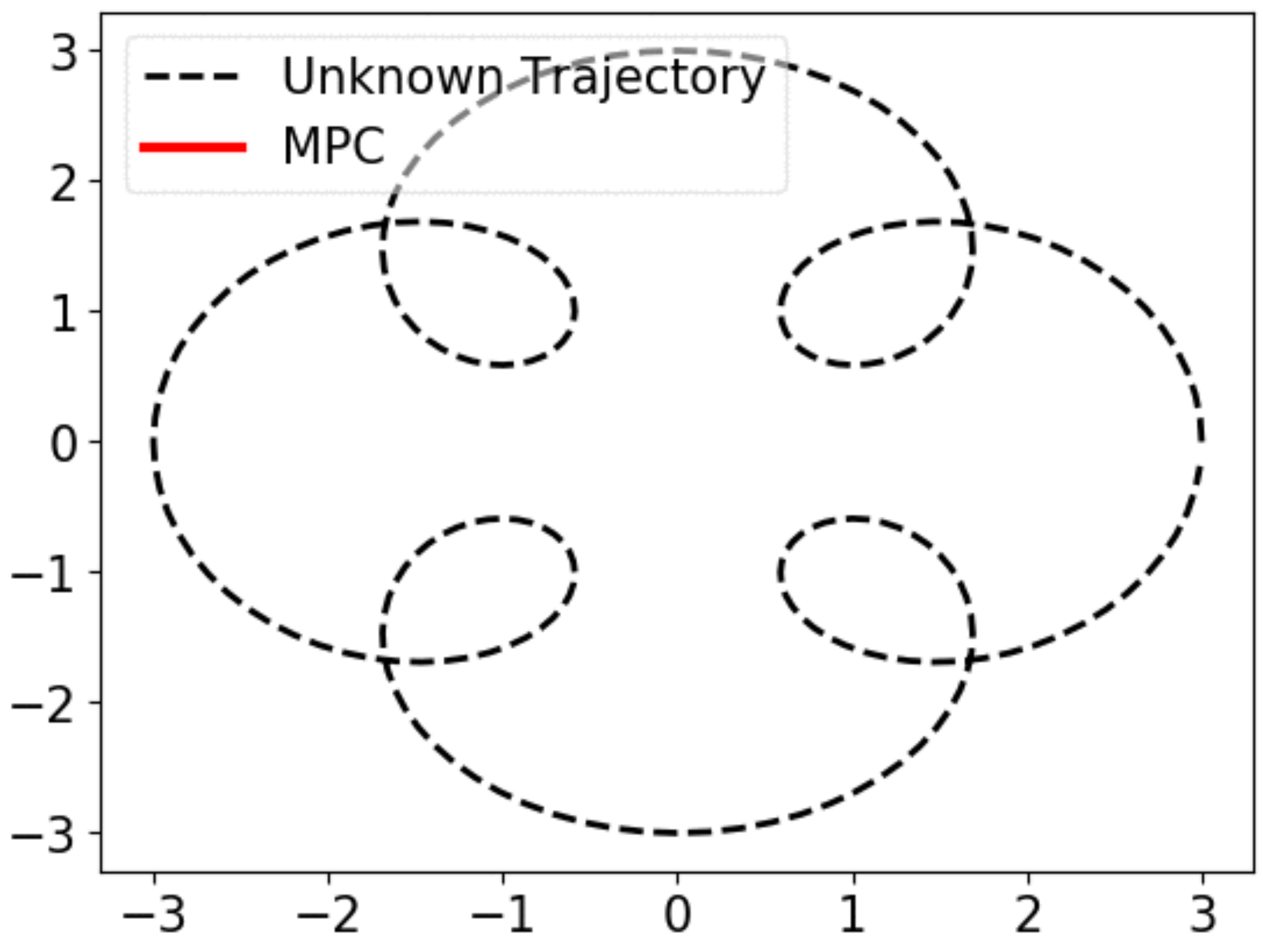
Learning-Augmented Online Algorithms

- Ski-rental
- Online matching
- Non-clairvoyant scheduling
- Metrical task systems
- Secretary problem
- Convex body chasing
- ...
- **Linear control (This work)**

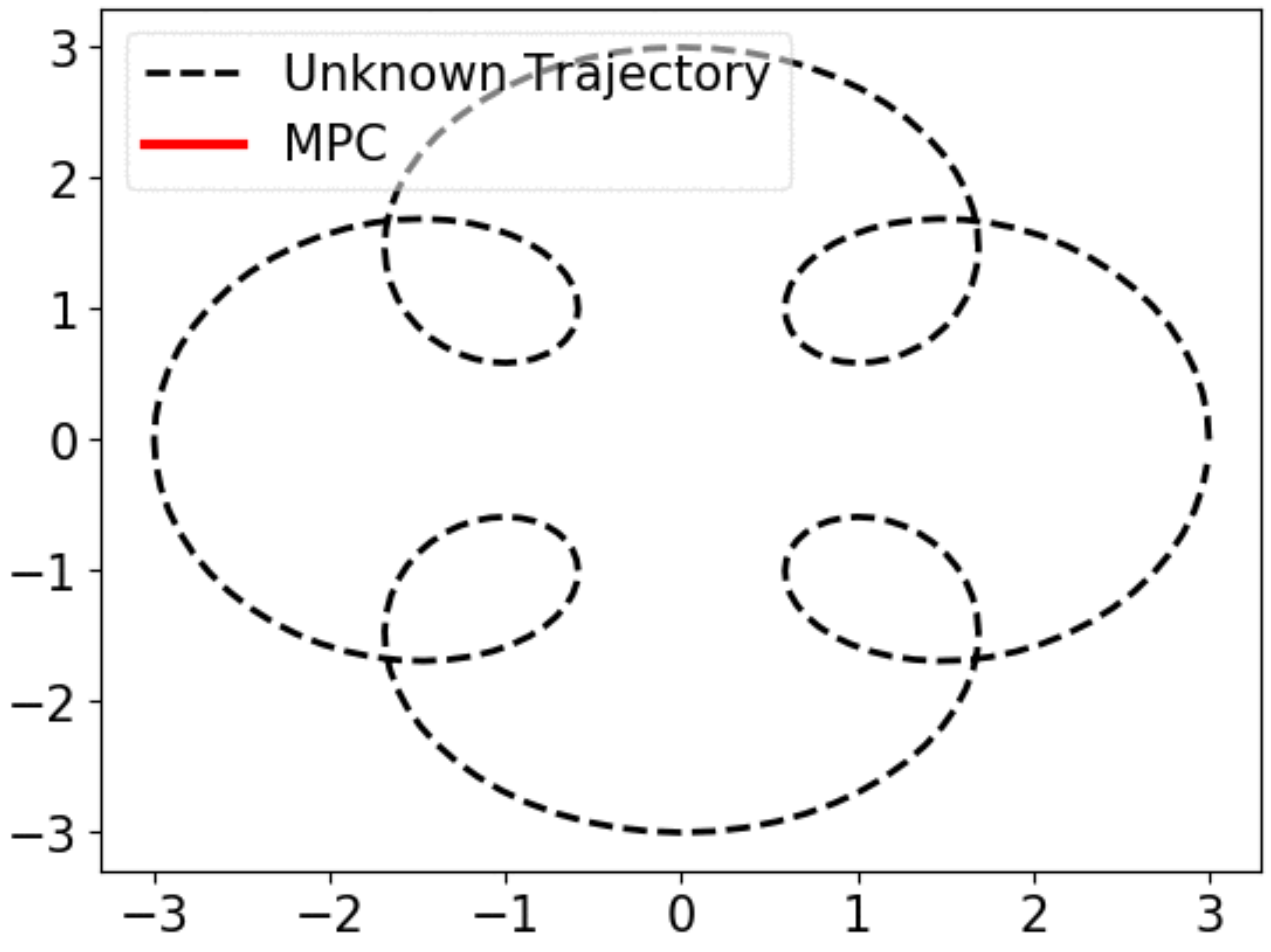
“Consistency” vs “Robustness” Trade-offs



MPC with Untrusted Predictions ...



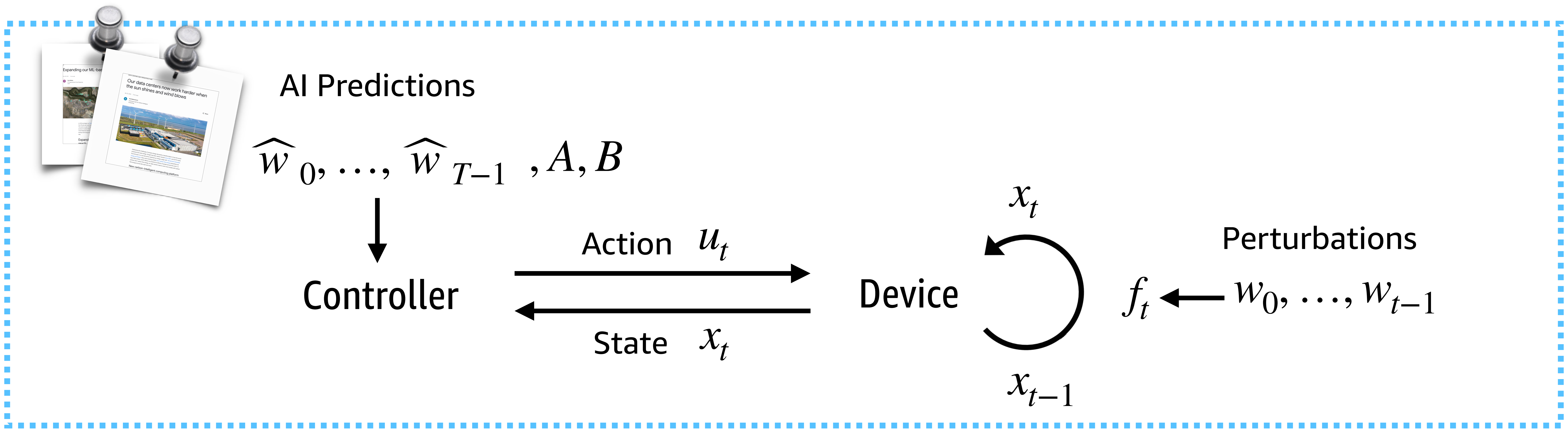
Prefect Predictions



Untrusted Predictions

Linear Quadratic Control

System



Dynamics	*Cost*	*Predictions*
$x_{t+1} = f_t(x_t, u_t) = Ax_t + Bu_t + w_t$	$\sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_f x_T$	$\widehat{w}_0, \ldots, \widehat{w}_{T-1}$

- The system is stabilizable
- $Q, R, Q_f \succ 0$

[2005, Mayne et al.] Robust Model Predictive Control of Constrained Linear Systems with Bounded Disturbances
[2019, Lopez et al.] Dynamic Tube MPC for Nonlinear Systems
[2022, Bujarbaruah et al.] Robust MPC for Linear Systems with Parametric and Additive Uncertainty: A Novel Constraint Tightening Approach

Cannot actively adapt depending on predictions

Performance Benchmark

Goal: Find an online algorithm with good *Competitive Ratio* CR regardless of *prediction error* ε

Idea:
- Be *conservative* if ε is large
- Be *aggressive* if ε is small

$$\mathrm{CR} := \max_{\varepsilon \geq 0} \mathrm{CR}(\varepsilon)$$

$$\mathrm{CR}(\varepsilon) := \max_{\mathbf{w}, \widehat{\mathbf{w}} : d(\mathbf{w}, \widehat{\mathbf{w}}) \leq \varepsilon} \frac{\mathrm{ALG}(\varepsilon)}{\mathrm{OPT}}$$

$\mathrm{ALG}(\varepsilon) :=$ Cost induced by an *online algorithm* with prediction error ε

$\mathrm{OPT} :=$ Optimal cost knowing $w_0, \ldots, w_{t-1}$ in hindsight

Prediction Error

Goal: Find an online algorithm with good *Competitive Ratio* CR regardless of *prediction error* ε

$$\varepsilon := \sum_{t=0}^{T-1} \left\| \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P(w_t - \widehat{w}_t) \right\|^2$$

$P :=$ Solution of DARE

$$F := A - BK = A - B(R + B^\top PB)^{-1}B^\top PA$$

Prediction error measures "how good the ML predictions are"

Recall in ski-rental, ε is the ℓ_1-norm $\eta := |y - b|$

Prediction Error

Goal: Find an online algorithm with good *Competitive Ratio* CR regardless of *prediction error* ε

$$\varepsilon := \sum_{t=0}^{T-1} \underbrace{\left\| \sum_{\tau=t}^{T-1} \left(F^{\top}\right)^{\tau-t} P(w_t - \widehat{w}_t) \right\|^2}_{\text{weighted sum}}$$

Why is it a "weighted sum"?

Quick Answer:

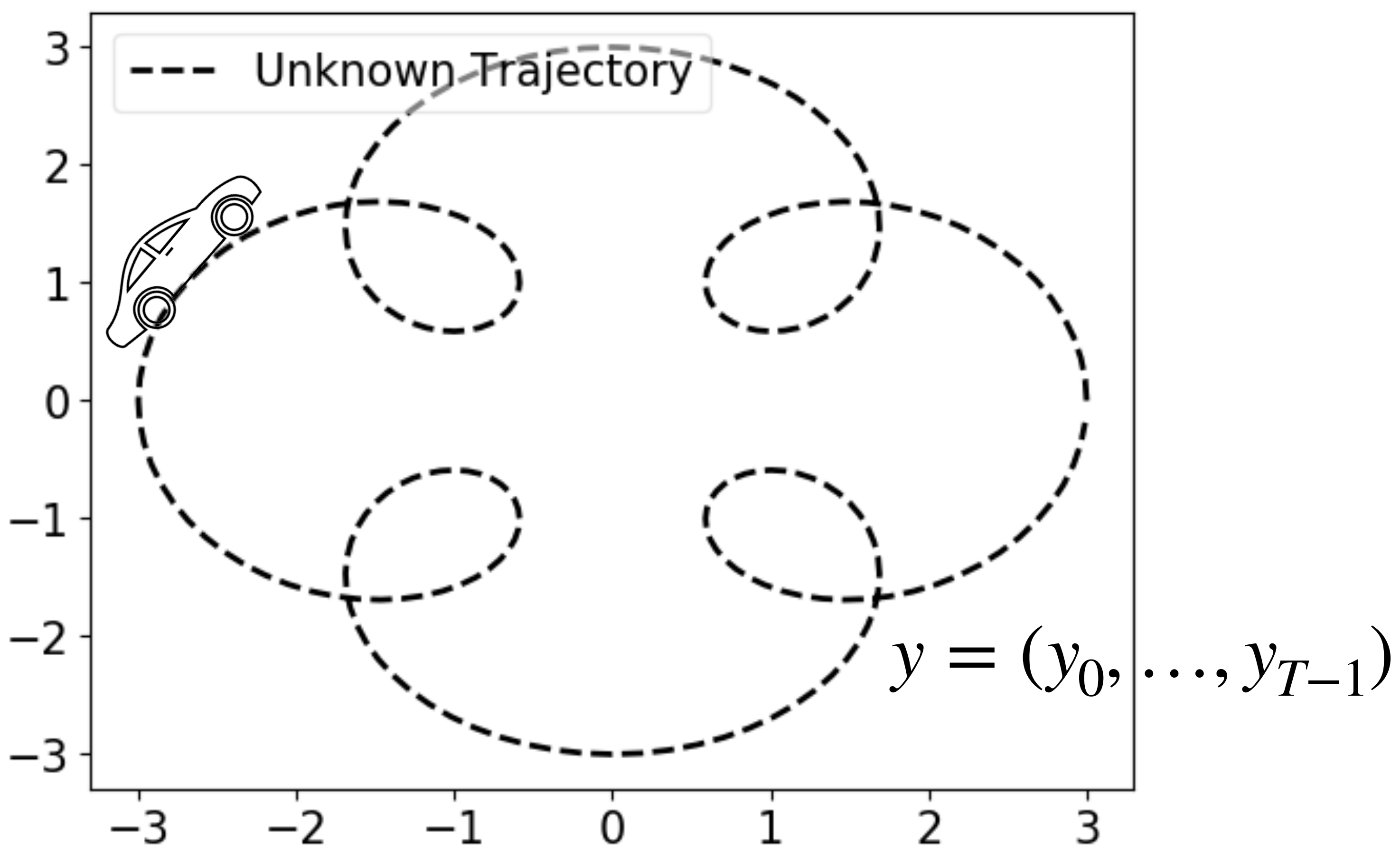
- Simplify expressions in our analysis

More fundamental Answers:

- Impact decays exponentially
- Meaning: "Error in the actions" due to perturbation-prediction mismatch

Toy Examples

Robot Tracking

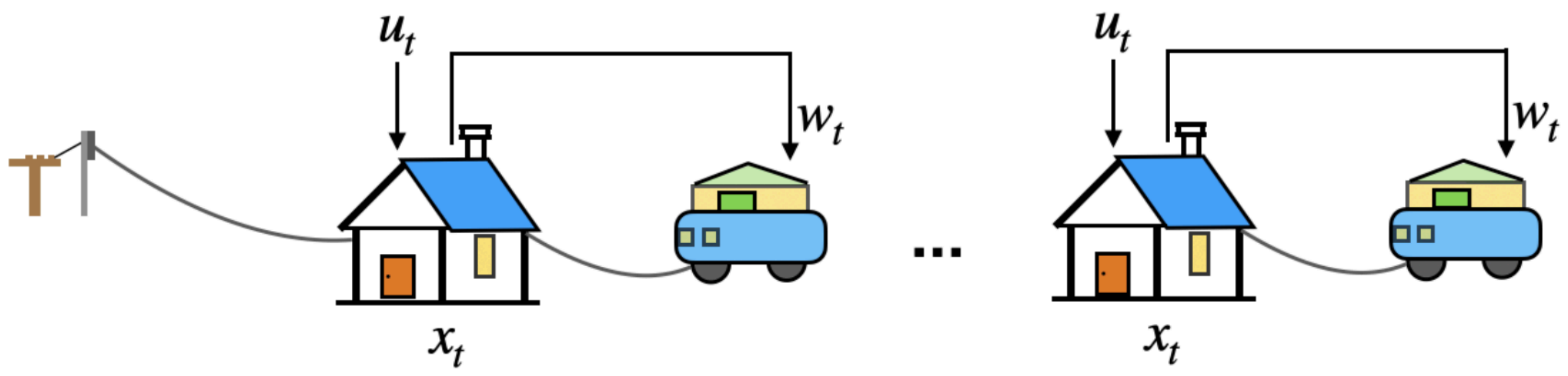


velocity $v_{t+1} = v_t + \Delta u_t \quad (x_t := p_t - y_t)$

location $p_{t+1} = p_t + \alpha v_t \quad (w_t := A y_t - y_{t+1})$

$$\begin{bmatrix} x_{t+1} \\ v_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \alpha & 0 \\ 0 & 1 & 0 & \alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ v_t \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \Delta & 0 \\ 0 & \Delta \end{bmatrix} u_t + w_t$$

Adaptive Battery Buffered EV Charging



$$w_t(i) := \begin{cases} E & \text{EV arrives at time } t \text{ \& charge at } i \\ 0 & \text{otherwise} \end{cases}$$

$x_t :=$ undelivered energy

$u_t :=$ charging allocation

$$x_{t+1} = I x_t - D u_t + w_t$$

Model Predictive Control

(**MPC** as a widely used control policy **...)**

$$u_t = \widehat{\pi}(x_t) := \mathrm{argmin}_{(u_t,\ldots,u_{T-1})} \left(\sum_{\tau=t}^{T-1} (x_\tau^\top Q x_\tau + u_\tau^\top R u_\tau) + x_T^\top P x_T \right)$$

Good when ε is small

$$x_{\tau+1} = Ax_\tau + Bu_\tau + \widehat{w}_\tau, \forall \tau = t, \ldots, T-1.$$

(Explicit Expressions [2020 Yu et al.])

$$\widehat{\pi}(x_t) = -(R + B^\top P B)^{-1} B^\top \left(PAx_t + \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P \widehat{w}_\tau \right)$$

[2020 Yu et al.] The power of predictions in online control, NeurIPS, 2020

Taking benefit of Two Policies ...

$$\widehat{\pi}(x_t) = -(R + B^\top P B)^{-1} B^\top \left(P A x_t + \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P \widehat{w}_\tau \right)$$ Good when ε is small

$$\overline{\pi}(x_t) = -(R + B^\top P B)^{-1} B^\top P A x_t = -K x_t$$ *Drop the predictions* Good when ε is large

(Optimal linear controller for LQR with Gaussian perturbations)

MPC Policy + Linear Policy

How about a convex combination?

$$\lambda\,\widehat{\pi}(x_t) + (1-\lambda)\overline{\pi}(x_t)$$

↑

Trust Parameter

λ-Confident Control

"1-confident"

$$\widehat{\pi}(x_t) = -(R + B^\top PB)^{-1}B^\top \left(PAx_t + \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P\hat{w}_\tau \right)$$

"λ-confident"

$$\lambda\widehat{\pi}(x_t) + (1-\lambda)\overline{\pi}(x_t) = -(R + B^\top PB)^{-1}B^\top \left(PAx_t + \lambda \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P\hat{w}_\tau \right)$$

Trust parameter

"0-confident"

$$\overline{\pi}(x_t) = -(R + B^\top PB)^{-1}B^\top PAx_t = -Kx_t$$

λ-Confident Control

"λ-confident" $\pi(x_t) = \lambda\widehat{\pi}(x_t) + (1-\lambda)\overline{\pi}(x_t) = -(R + B^\top PB)^{-1}B^\top\left(PAx_t + \lambda\sum_{\tau=t}^{T-1}\left(F^\top\right)^{\tau-t}P\hat{w}_\tau\right)$

Trust parameter

(Equivalent to)

$$\pi(x_t) := \text{argmin}_{(u_t,\ldots,u_{T-1})}\left(\sum_{\tau=t}^{T-1}(x_\tau^\top Qx_\tau + u_\tau^\top Ru_\tau) + x_T Px_T\right)$$

$$x_{\tau+1} = Ax_\tau + Bu_\tau + \lambda\widehat{w}_\tau, \forall\tau = t, \ldots, T-1.$$

Trust parameter

Competitive Ratio Results

Theorem (Informal; SIGMETRICS '22) “Meta Theorem”

Under model assumptions, with a fixed trust parameter $\lambda > 0$ *, the* *λ-confident algorithm* *has a worst-case competitive ratio of at most*

$$\mathrm{CR}(\varepsilon) \leq 1 + 2\|H\| \min\left\{\left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C}\right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W}\right)\right\}$$

where OPT *is the optimal cost*

$C > 0$ *is some constant*

$$\varepsilon := \sum_{t=0}^{T-1} \left\| \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P\left(w_\tau - \hat{w}_\tau\right) \right\|^2 \qquad H := B(R + B^\top P B)^{-1} B^\top$$

$$\overline{W} := \sum_{t=0}^{T-1} \left\| \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P \hat{w}_\tau \right\|^2$$

Competitive Ratio Results

Theorem (Informal; SIGMETRICS '22) "Meta Theorem"

Under model assumptions, with a fixed trust parameter $\lambda > 0$, the λ-confident algorithm has a worst-case competitive ratio of at most

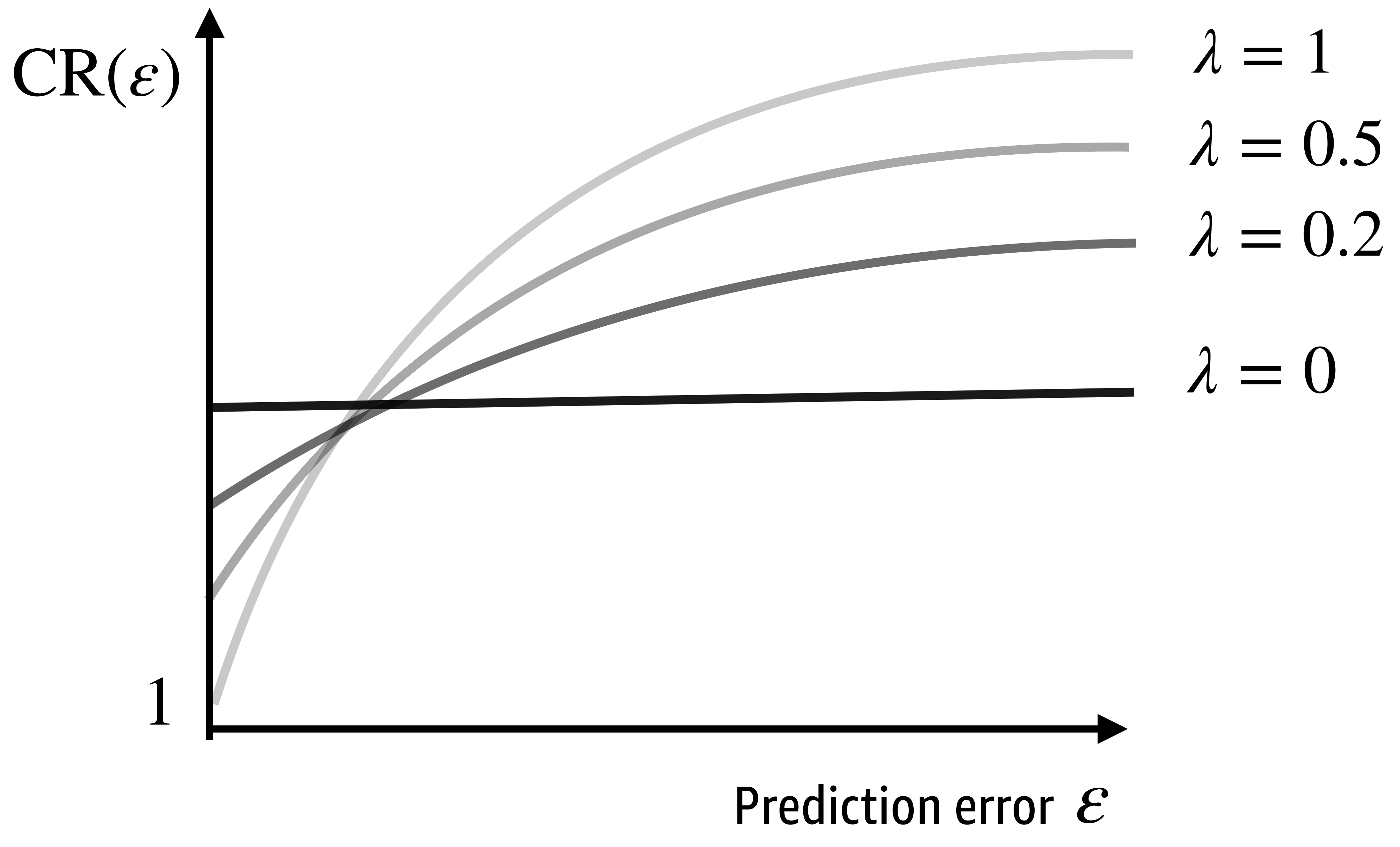
$$\mathsf{CR}(\varepsilon) \leq 1 + 2\|H\| \min\left\{\left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C}\right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W}\right)\right\}$$

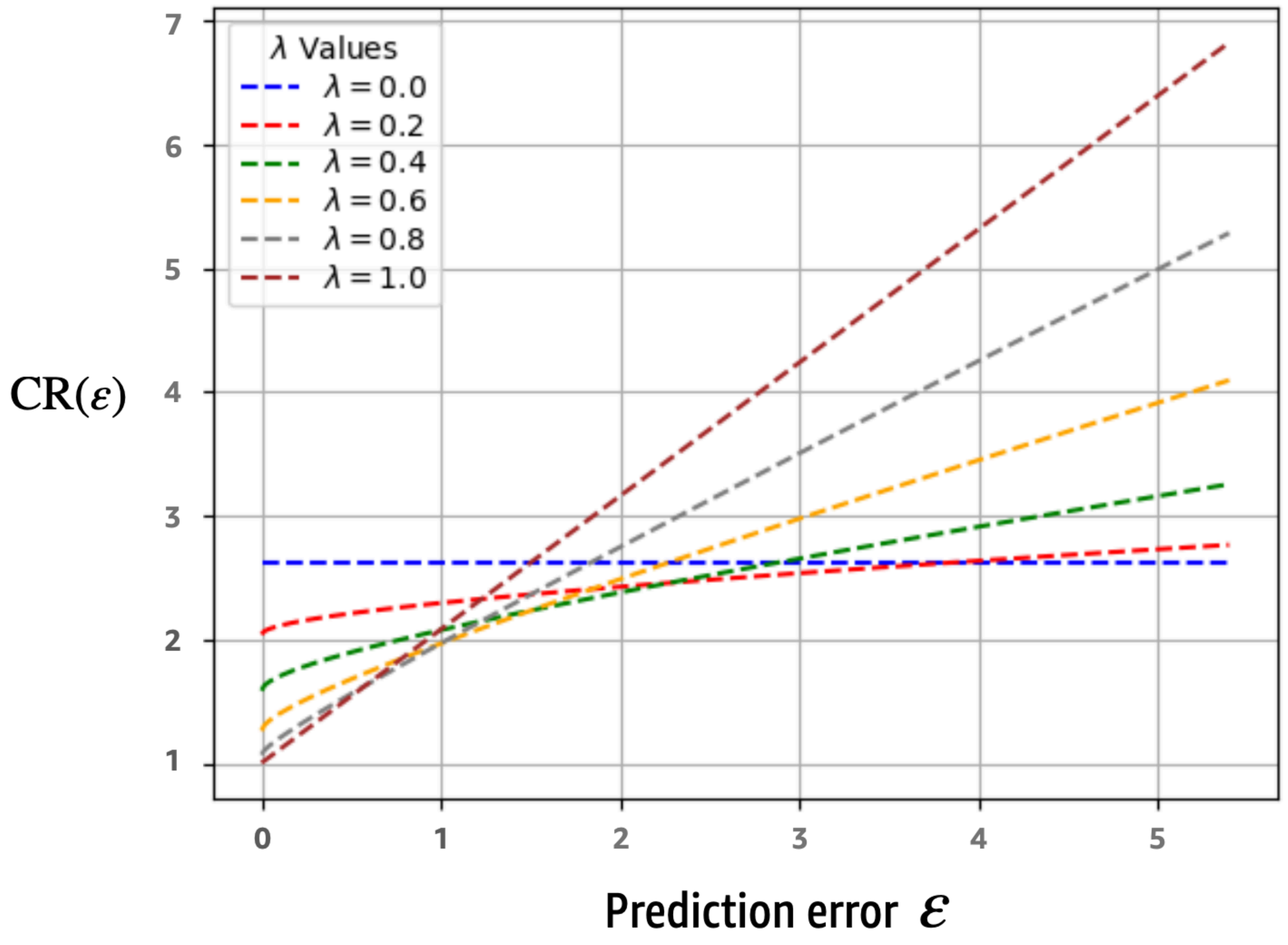
- Establish a trade-off between "robustness" and "consistency"
- Useful in the proof of the main results

Li, T., Yang, R., Qu, G., Shi, G., Yu. C., Wierman, A. & Low, S. H. Robustness and Consistency in Linear Quadratic Control with Untrusted Predictions. *SIGMETRICS 2022*.

Varying Trust Parameter λ

"Consistency" vs "Robustness" Tradeoffs in Learning-Augmented Control





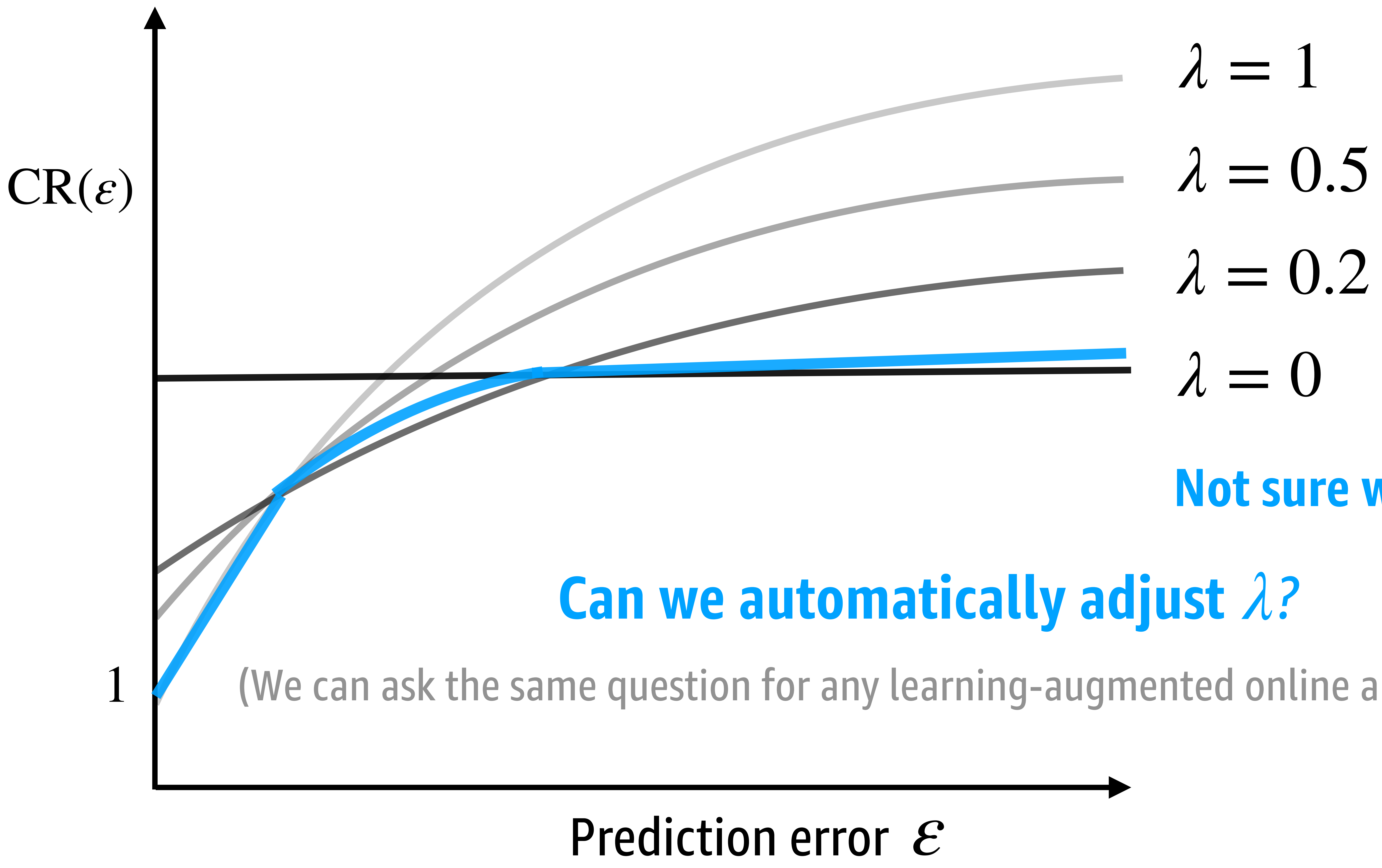
$$\mathrm{CR}(\varepsilon) \leq 1 + 2\|H\| \left(\frac{\lambda^2}{\mathsf{OPT}} \varepsilon + \frac{(1-\lambda)^2}{C} \right)$$

- When ε is large, the linear component dominates
- Selecting different λ realizes different trade-offs

Online Learning for λ

Issue: Prediction error ε is not known a priori

Goal: Find an online algorithm with good *Competitive Ratio regardless of prediction error* ε



Not sure what λ is optimal!

Can we automatically adjust λ?

(We can ask the same question for any learning-augmented online algorithms)

Online Learning for λ

Use time-varying trust parameters λ_t and learn them online!

Online Learning Approach

Quadratic function of λ

$$\lambda_t = \textbf{argmin}_\lambda \sum_{s=0}^{t-1} \left[\left(\sum_{\tau=s}^{t-1} (F^\top)^{\tau-s} P(w_\tau - \lambda \hat{w}_\tau) \right)^\top H \left(\sum_{\tau=s}^{t-1} (F^\top)^{\tau-s} P(w_\tau - \lambda \widehat{w}_\tau) \right) \right]$$

$\mathrm{ALG}_{t-1} - \mathrm{OPT}_{t-1}$ *"Optimize based on History"*

$$\Longrightarrow \quad \lambda_t = \frac{\sum_{s=0}^{t-1} \left(\eta(w; s, t-1)\right)^\top H \left(\eta(\widehat{w}; s, t-1)\right)}{\sum_{s=0}^{t-1} \left(\eta(\widehat{w}; s, t-1)\right)^\top H \left(\eta(\widehat{w}; s, t-1)\right)} \quad \textit{where} \quad \eta(w; s, t) := \sum_{\tau=s}^{t} (F^\top)^{\tau-s} P w_\tau$$

- "Follow-the-leader" design
- Only previously observed info is needed
- Computational complexity linear in T
- If $\widehat{w}$ and w are closer, λ_t is closer to 1

"Self-Tuning Control" Algorithm

For $t = 0,\ldots,T-1$

If $t = 0$ *Initialize* λ_0

Else Compute

$$\lambda_t = \frac{\sum_{s=0}^{t-1} \left(\eta(w; s, t-1)\right)^\top H \left(\eta(\hat{w}; s, t-1)\right)}{\sum_{s=0}^{t-1} \left(\eta(\hat{w}; s, t-1)\right)^\top H \left(\eta(\hat{w}; s, t-1)\right)}$$

where $\eta(w; s, t) := \sum_{\tau=s}^{t} \left(F^\top\right)^{\tau-s} P w_\tau$

Generate an action using the λ_t-confident algorithm

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

Competitive Ratio Bound for Self-tuning Control

Theorem (Informal; SIGMETRICS '22) “CR Theorem”

Under model assumptions, the competitive ratio of the self-tuning control algorithm is bounded by

$$\mathsf{CR}(\varepsilon) \leq 1 + 2\|H\| \frac{\varepsilon}{\mathsf{OPT} + C\varepsilon} + O\left(\frac{\left(\mu_{\mathsf{VAR}}(\mathbf{w}) + \mu_{\mathsf{VAR}}(\widehat{\mathbf{w}})\right)^2}{\mathsf{OPT}}\right).$$

dependency on ε

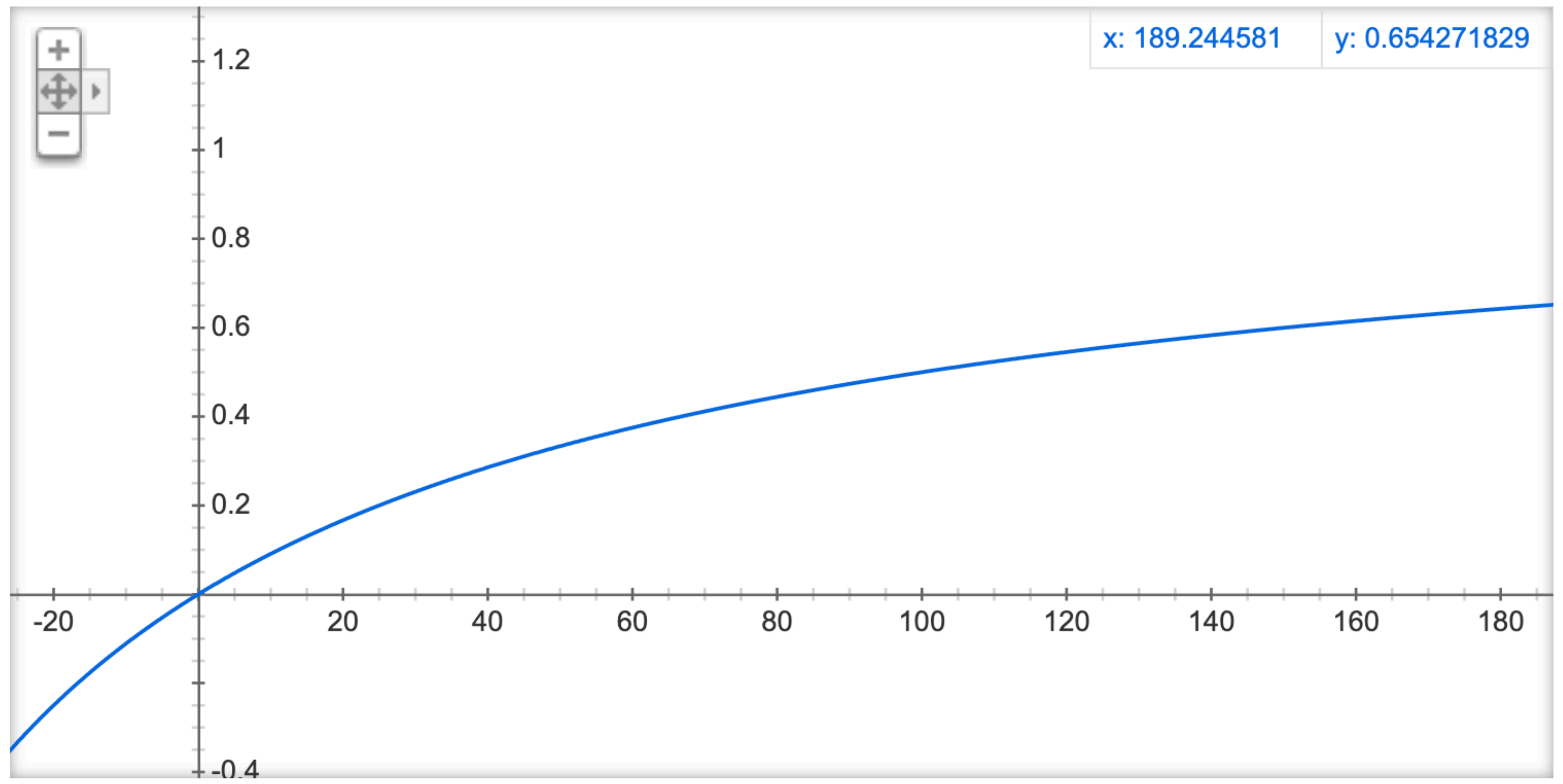
How fast $\mathbf{w}$ and $\widehat{\mathbf{w}}$ change over time

“maximal variation” (variation terms appear in many online learning literature)

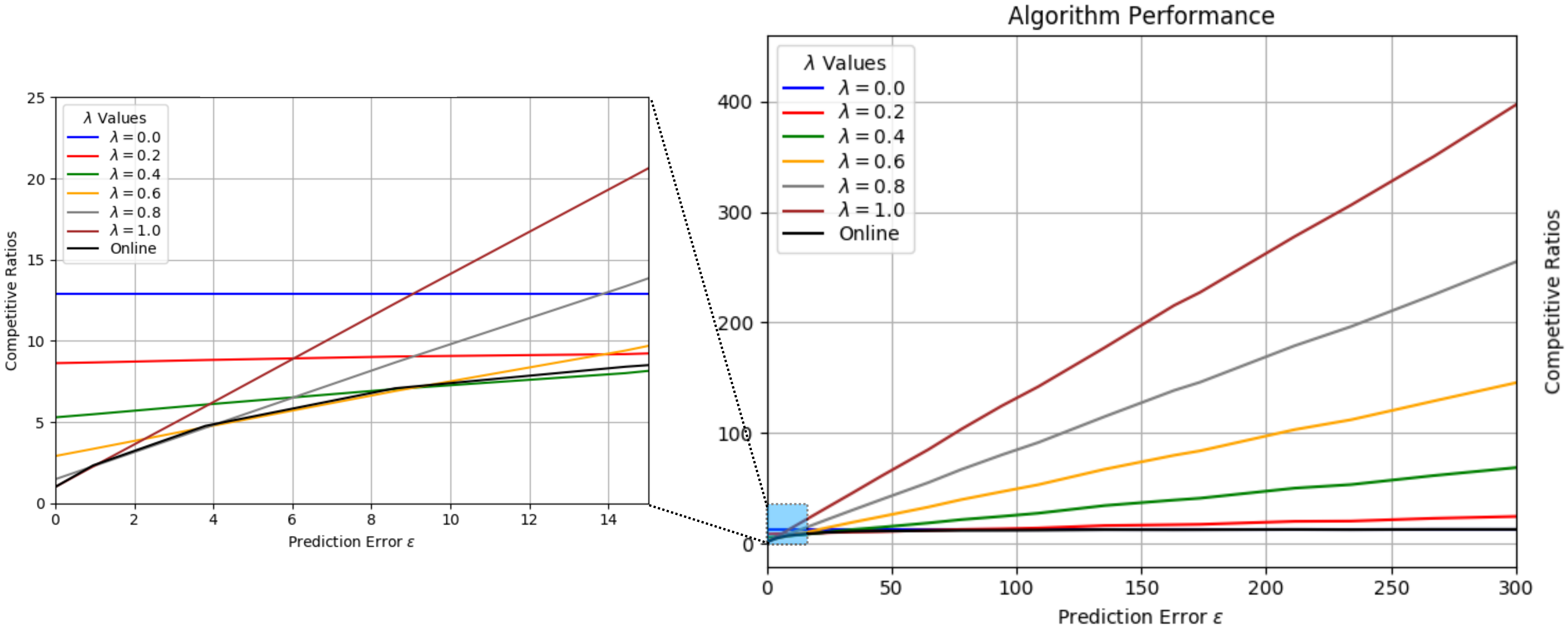
- $\mu_{\mathsf{VAR}}(\mathbf{x}) := \sum_{s=1}^{T-1} \max_{\tau=0,\ldots,s-1} \| x_\tau - x_{\tau+T-s} \|$

- When $\varepsilon = 0$, $\frac{\varepsilon}{\mathsf{OPT} + \varepsilon C} = 0$
- When $\varepsilon \to \infty$, $\frac{\varepsilon}{\mathsf{OPT} + \varepsilon C} \to \frac{1}{C}$ Bounded!

Graph for x/(100+x)



Case Study: Battery-Buffered EV Charging

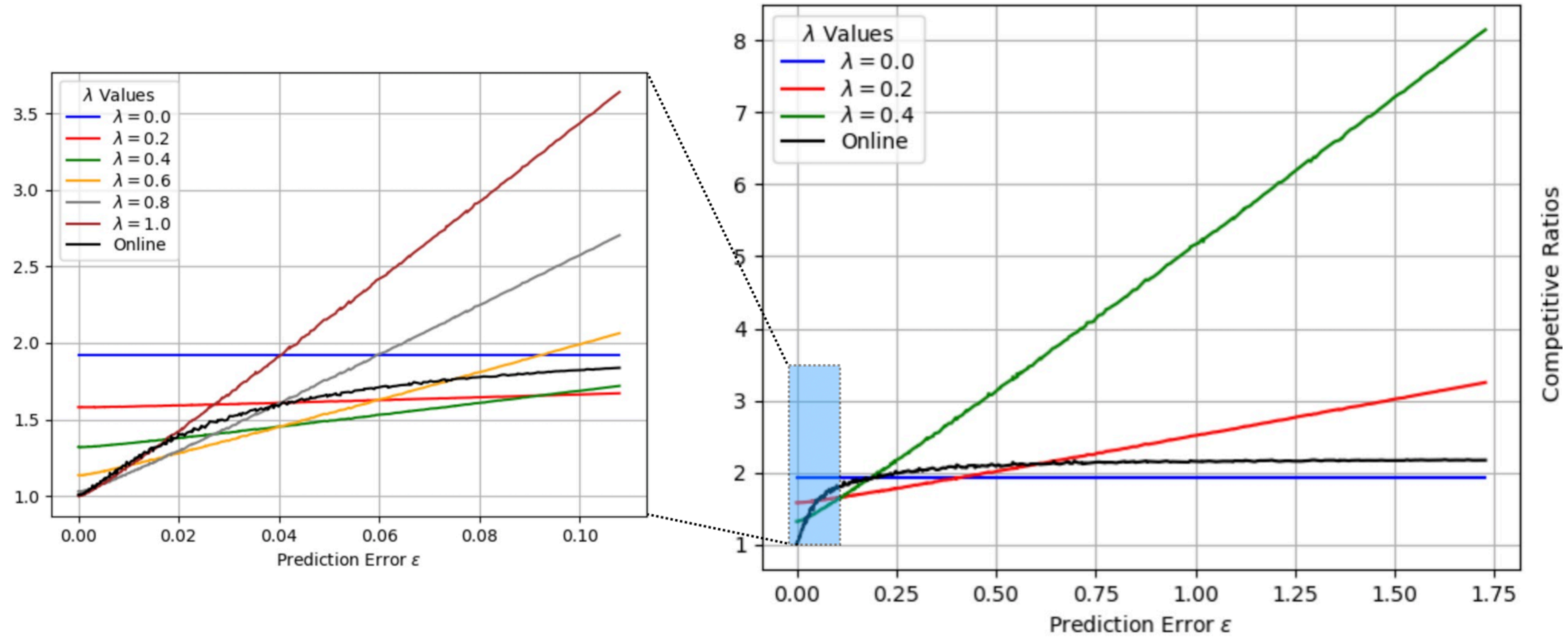


Main Results:

$$\mathrm{CR}(\varepsilon) \le 1 + O(\lambda^2 \varepsilon)$$

$$\mathrm{CR}(\varepsilon) \le 1 + \frac{O(\varepsilon)}{\Theta(1) + \Theta(\varepsilon)} + \text{Variation}$$

Case Study: Robot Tracking



Main Results:

$$\mathrm{CR}(\varepsilon) \leq 1 + O(\lambda^2 \varepsilon)$$

$$\mathrm{CR}(\varepsilon) \leq 1 + \frac{O(\varepsilon)}{\Theta(1) + \Theta(\varepsilon)} + \text{Variation}$$

Sketched Proof

Meta Theorem $\mathsf{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min \left\{ \left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C} \right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W} \right) \right\}$ **λ-Confident Control**

CR Theorem $\mathsf{CR}_{\text{self}}(\varepsilon) \leq 1 + 2\|H\| \frac{\varepsilon}{\mathsf{OPT} + C\varepsilon} + O\left(\frac{\left(\mu_{\mathsf{VAR}}(\mathbf{w}) + \mu_{\mathsf{VAR}}(\widehat{\mathbf{w}})\right)^2}{\mathsf{OPT}} \right)$ **Self-Tuning Control**

Sketched Proof

Meta Theorem $\mathsf{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min \left\{ \left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C} \right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W} \right) \right\}$

λ-Confident Control

$$\frac{\mathsf{ALG}(\lambda^*)}{\mathsf{OPT}} \leq 1 + 2\|H\| \frac{\varepsilon}{\mathsf{OPT} + \varepsilon C}$$

Optimize the upper bound over λ

Sketched Proof

Meta Theorem $\mathrm{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min \left\{ \left(\frac{\lambda^2}{\mathrm{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C} \right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathrm{OPT}}\overline{W} \right) \right\}$ **λ-Confident Control**

$$\frac{\mathrm{ALG}(\lambda^*)}{\mathrm{OPT}} \leq 1 + 2\|H\| \frac{\varepsilon}{\mathrm{OPT} + \varepsilon C}$$

Optimize the upper bound over λ

Regret Lemma $\mathrm{Regret} := \mathrm{ALG}(\lambda_0, \ldots, \lambda_{T-1}) - \mathrm{ALG}(\lambda^*)$

Want: $\mathrm{CR}_{\mathrm{self}}(\varepsilon) = \frac{\mathrm{ALG}(\lambda_0, \ldots, \lambda_{T-1})}{\mathrm{OPT}}$ (depends on ε; omitted)

Sketched Proof

Meta Theorem $\mathsf{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min \left\{ \left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C} \right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W} \right) \right\}$ **λ-Confident Control**

$$\frac{\mathsf{ALG}(\lambda^*)}{\mathsf{OPT}} \leq 1 + 2\|H\| \frac{\varepsilon}{\mathsf{OPT} + \varepsilon C}$$

Optimize the upper bound over λ

Regret Lemma Regret $:= \mathsf{ALG}(\lambda_0, \ldots, \lambda_{T-1}) - \mathsf{ALG}(\lambda^*)$

Want: $\mathsf{CR}_{\text{self}}(\varepsilon) = \frac{\mathsf{ALG}(\lambda_0, \ldots, \lambda_{T-1})}{\mathsf{OPT}}$ (depends on ε; omitted)

Sketched Proof

Meta Theorem $\mathsf{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min\left\{\left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C}\right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W}\right)\right\}$ **λ-Confident Control**

$$\frac{\mathsf{ALG}(\lambda^*)}{\mathsf{OPT}} \leq 1 + 2\|H\|\frac{\varepsilon}{\mathsf{OPT} + \varepsilon C}$$

Optimize the upper bound over λ

Regret Lemma $\text{Regret} \leq \|H\| \sum_{t=0}^{T-1} \left\| \left|\lambda_t - \lambda^*\right| \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P\hat{w}_\tau \right\|^2$

Want: $\mathsf{CR}_{\text{self}}(\varepsilon) = \frac{\mathsf{ALG}(\lambda_0, \ldots, \lambda_{T-1})}{\mathsf{OPT}}$ (depends on ε; omitted)

Sketched Proof

Meta Theorem $\mathsf{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min\left\{\left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C}\right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W}\right)\right\}$ **λ-Confident Control**

$$\frac{\mathsf{ALG}(\lambda^*)}{\mathsf{OPT}} \leq 1 + 2\|H\|\frac{\varepsilon}{\mathsf{OPT} + \varepsilon C}$$

Optimize the upper bound over λ

Regret Lemma $\text{Regret} \leq \|H\| \sum_{t=0}^{T-1} \left\| \left|\lambda_t - \lambda^*\right| \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P\hat{w}_\tau \right\|^2$

Need a convergence bound

Want: $\mathsf{CR}_{\text{self}}(\varepsilon) = \frac{\mathsf{ALG}(\lambda_0, \ldots, \lambda_{T-1})}{\mathsf{OPT}}$ (depends on ε; omitted)

Sketched Proof

Meta Theorem $\text{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min\left\{\left(\frac{\lambda^2}{\text{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C}\right), \left(\frac{1}{C} + \frac{\lambda^2}{\text{OPT}}\overline{W}\right)\right\}$ **λ-Confident Control**

$$\frac{\text{ALG}(\lambda^*)}{\text{OPT}} \leq 1 + 2\|H\| \frac{\varepsilon}{\text{OPT} + \varepsilon C}$$

Optimize the upper bound over λ

Regret Lemma $\quad \text{Regret} \leq \|H\| \sum_{t=0}^{T-1} \left\| \left|\lambda_t - \lambda^*\right| \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P\hat{w}_\tau \right\|^2$

Lemma: Convergence of λ_t $\quad \left|\lambda_t - \lambda^*\right| = O\left(\frac{\mu_{\text{Var}}(\mathbf{w}) + \mu_{\text{Var}}(\hat{\mathbf{w}})}{t}\right)$ **Online learning bound in terms of variations**

Want: $\text{CR}_{\text{self}}(\varepsilon) = \frac{\text{ALG}(\lambda_0, \ldots, \lambda_{T-1})}{\text{OPT}}$ (depends on ε; omitted)

Sketched Proof

λ-Confident Control

Meta Theorem $\mathsf{CR}_{\lambda-\text{confident}}(\varepsilon) \leq 1 + 2\|H\| \min\left\{\left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C}\right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W}\right)\right\}$

Regret Lemma $\mathsf{Regret} \leq \|H\| \sum_{t=0}^{T-1} \left\| \left|\lambda_t - \lambda^*\right| \sum_{\tau=t}^{T-1} \left(F^\top\right)^{\tau-t} P\hat{w}_\tau \right\|^2$

Lemma: Convergence of λ_t $\left|\lambda_t - \lambda^*\right| = O\left(\frac{\mu_{\mathsf{Var}}(\mathbf{w}) + \mu_{\mathsf{Var}}(\hat{\mathbf{w}})}{t}\right)$

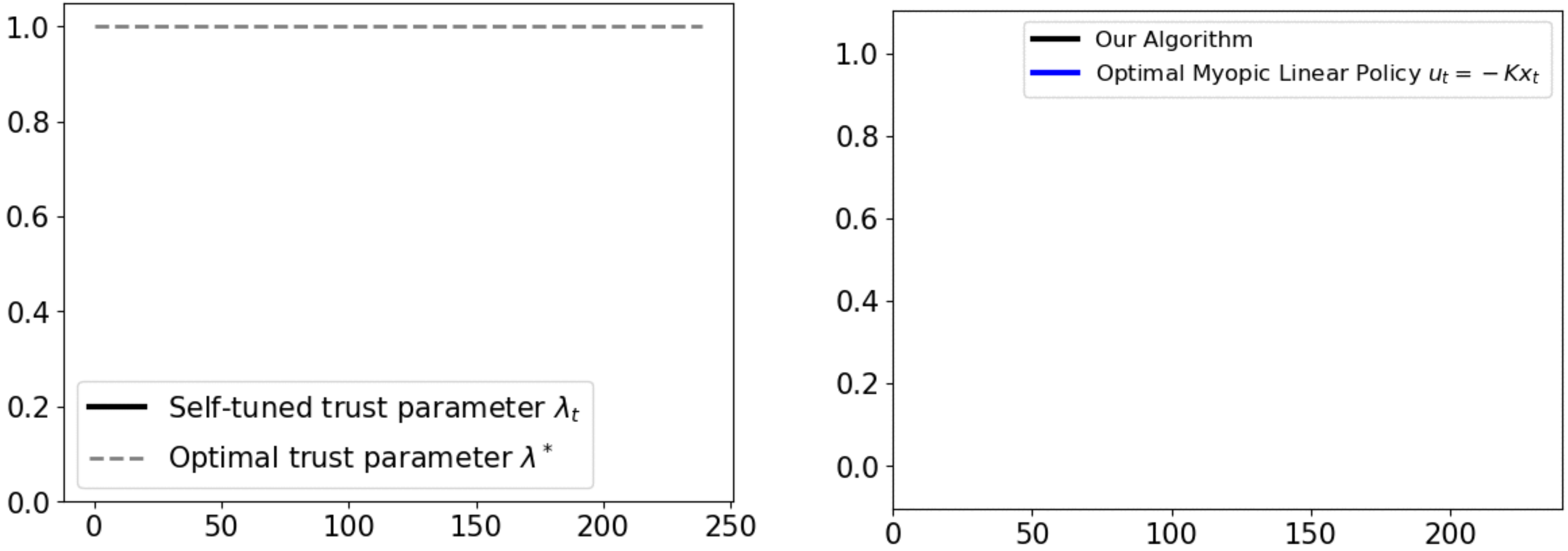
Self-Tuning Control

CR Theorem $\mathsf{CR}_{\text{self}}(\varepsilon) \leq 1 + 2\|H\| \frac{\varepsilon}{\mathsf{OPT} + C\varepsilon} + O\left(\frac{\left(\mu_{\mathsf{VAR}}(\mathbf{w}) + \mu_{\mathsf{VAR}}(\widehat{\mathbf{w}})\right)^2}{\mathsf{OPT}}\right)$

Verify the Convergence of Trust Parameters

Case Study: Robot Tracking

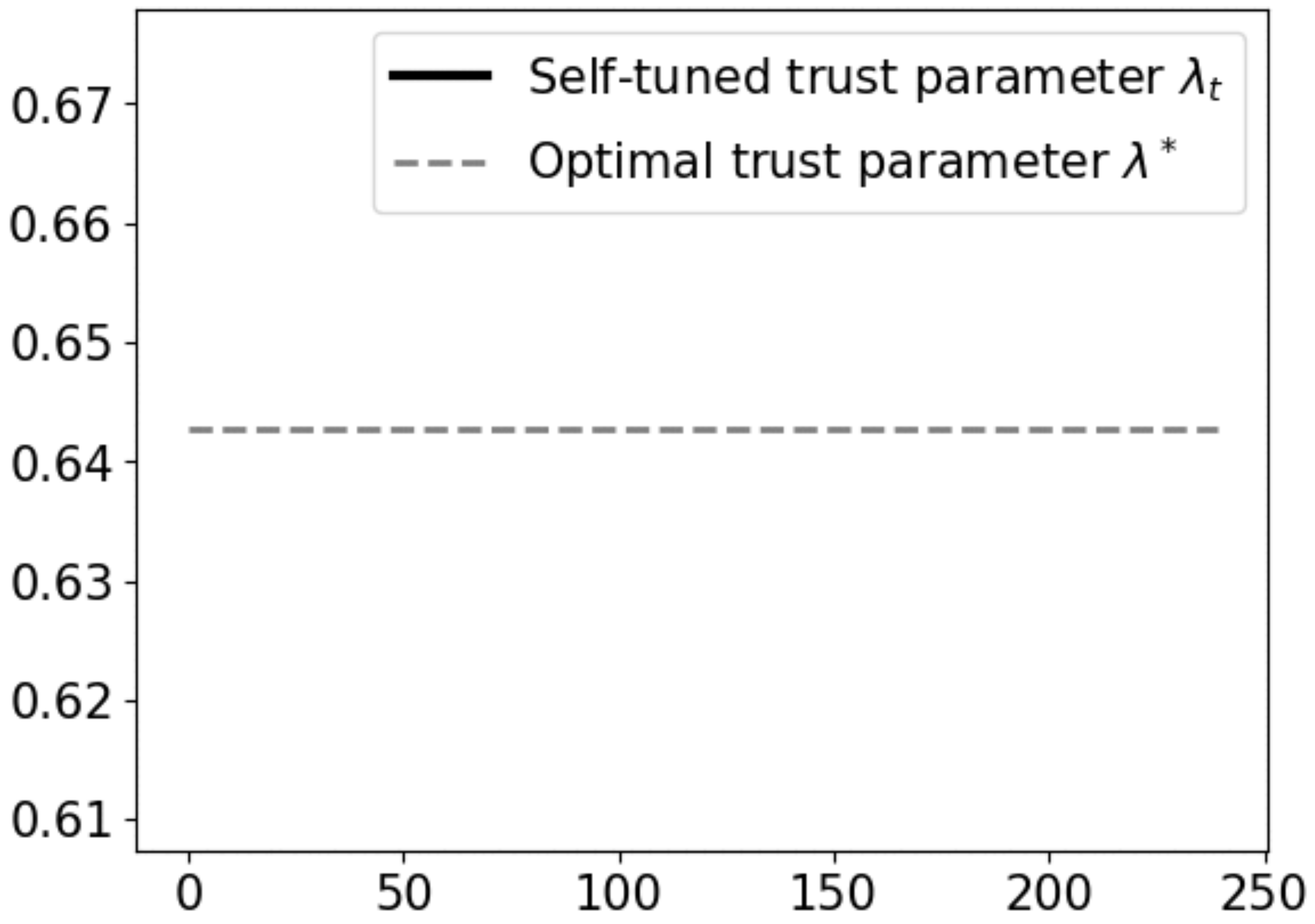
Low Error Case: Optimal $\lambda \approx 1$

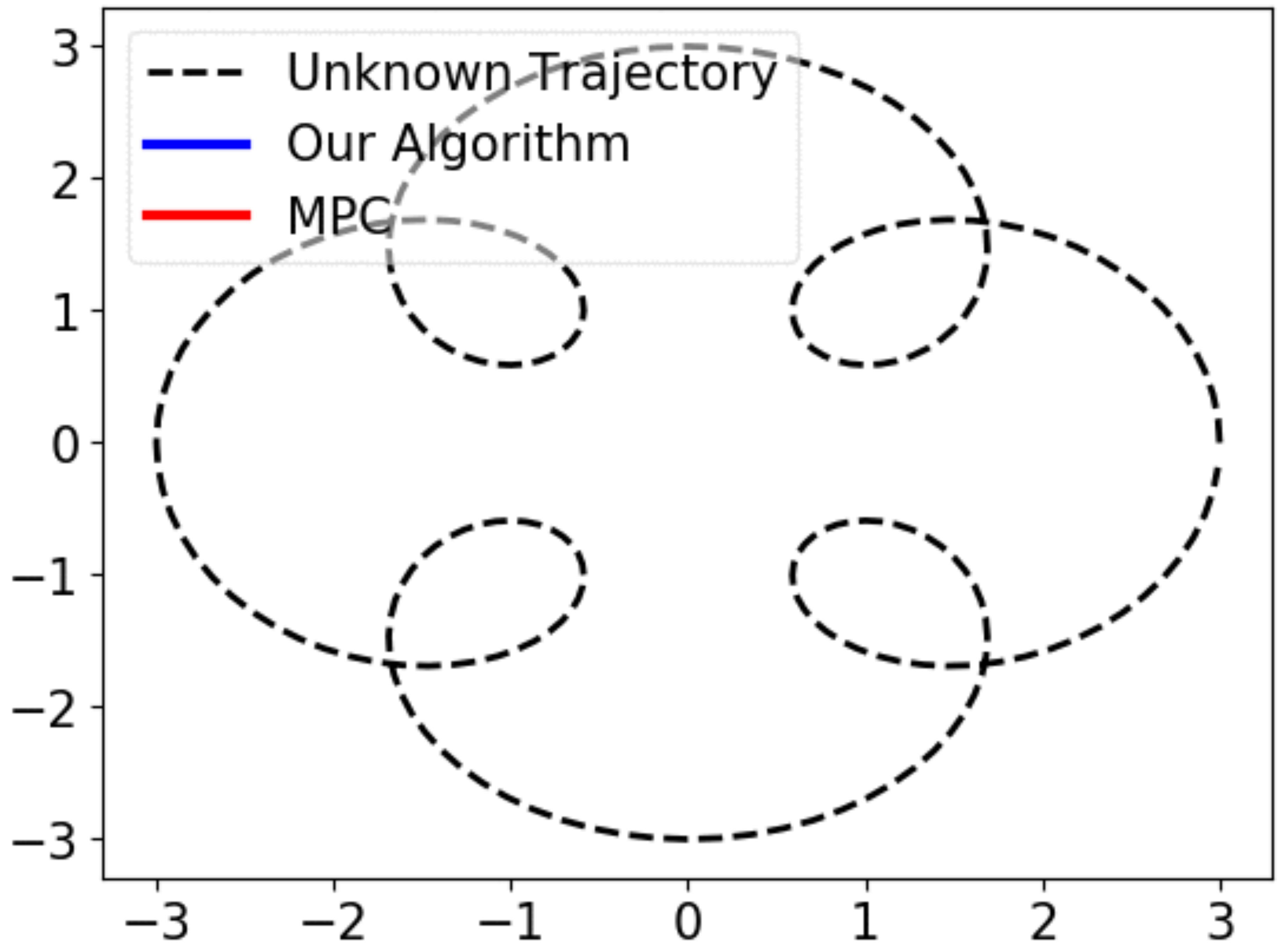


Verify the Convergence of Trust Parameters

Case Study: Robot Tracking

Medium Error Case: Optimal $0 < \lambda < 1$

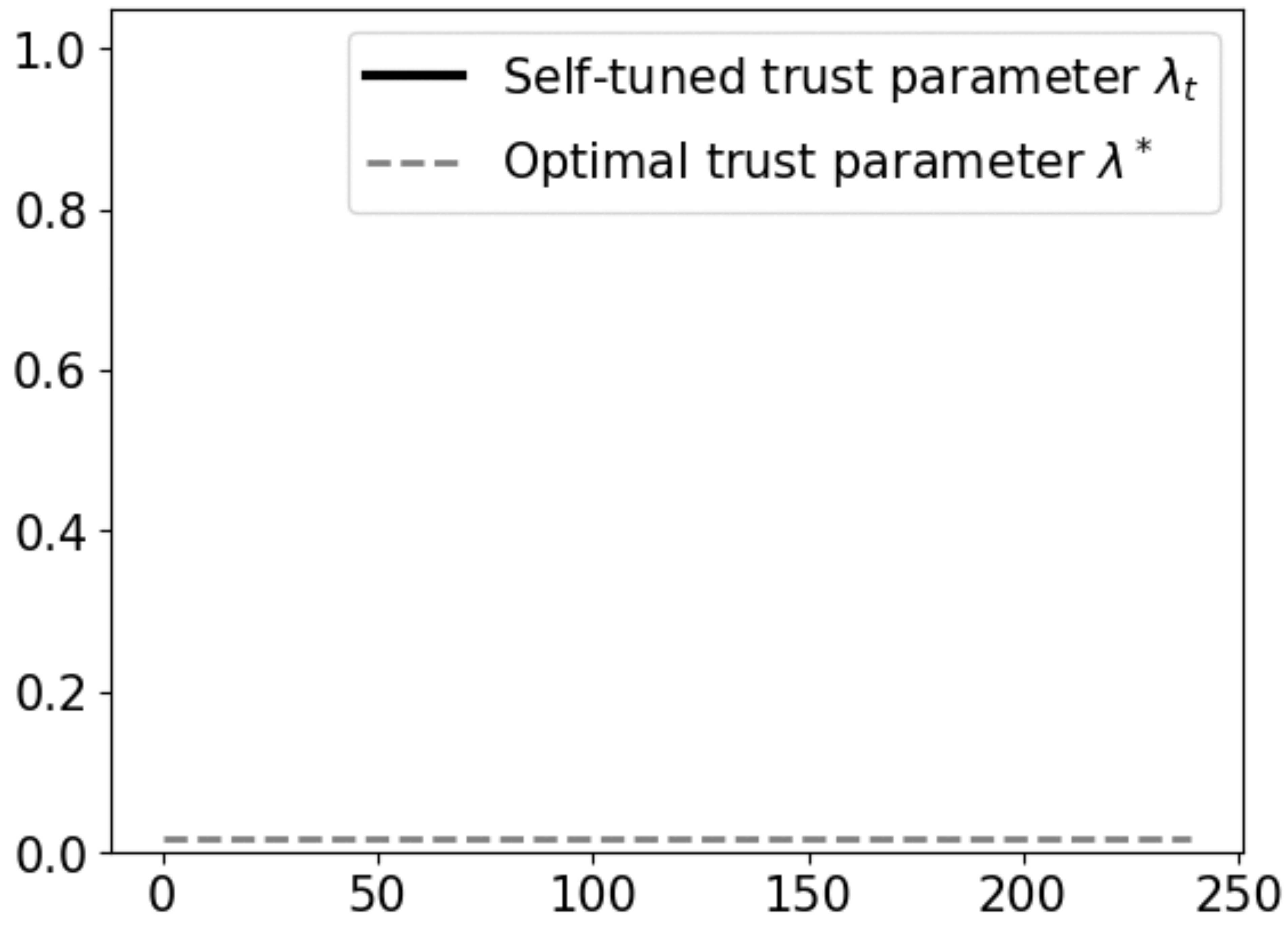


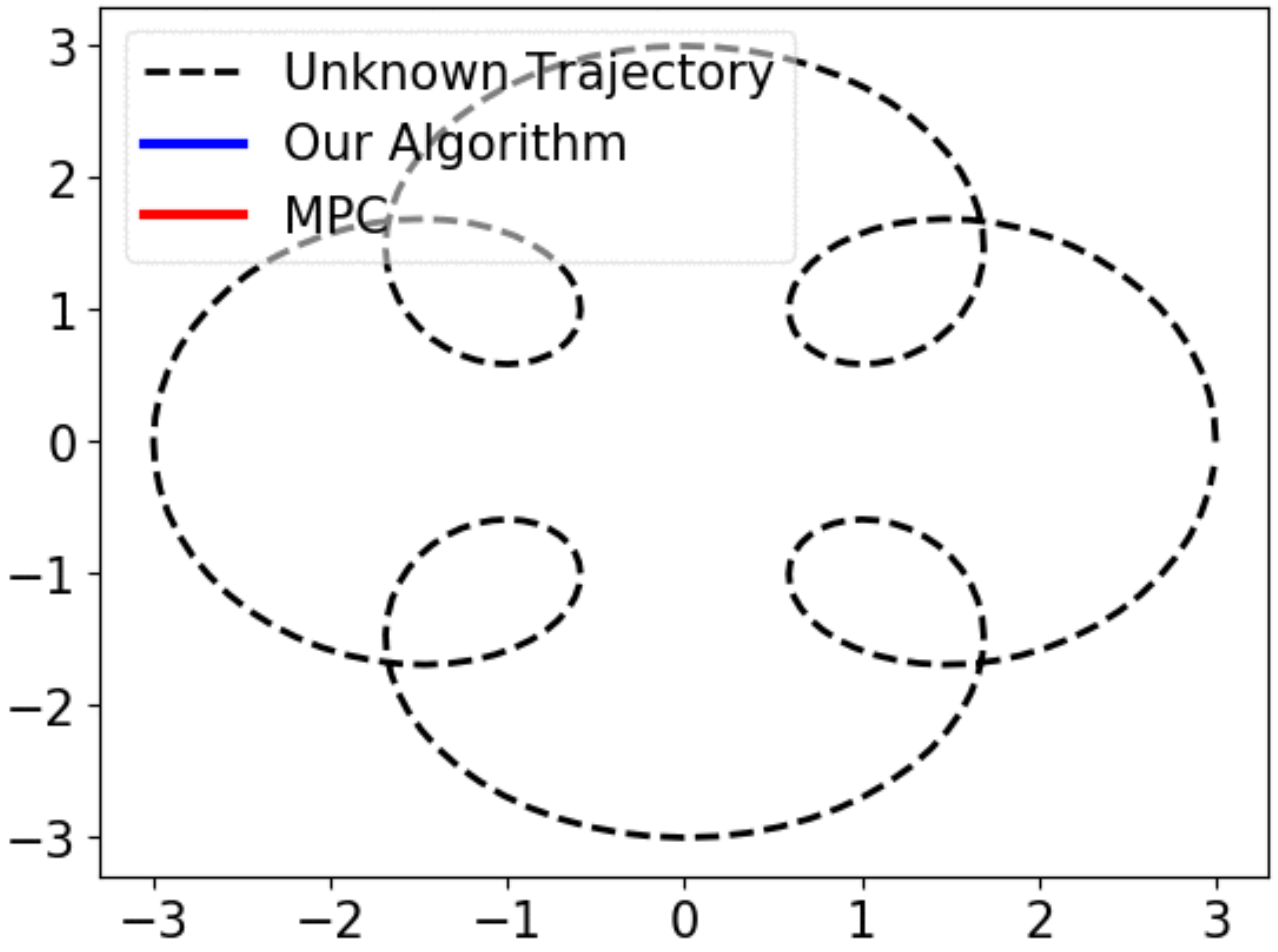


Verify the Convergence of Trust Parameters

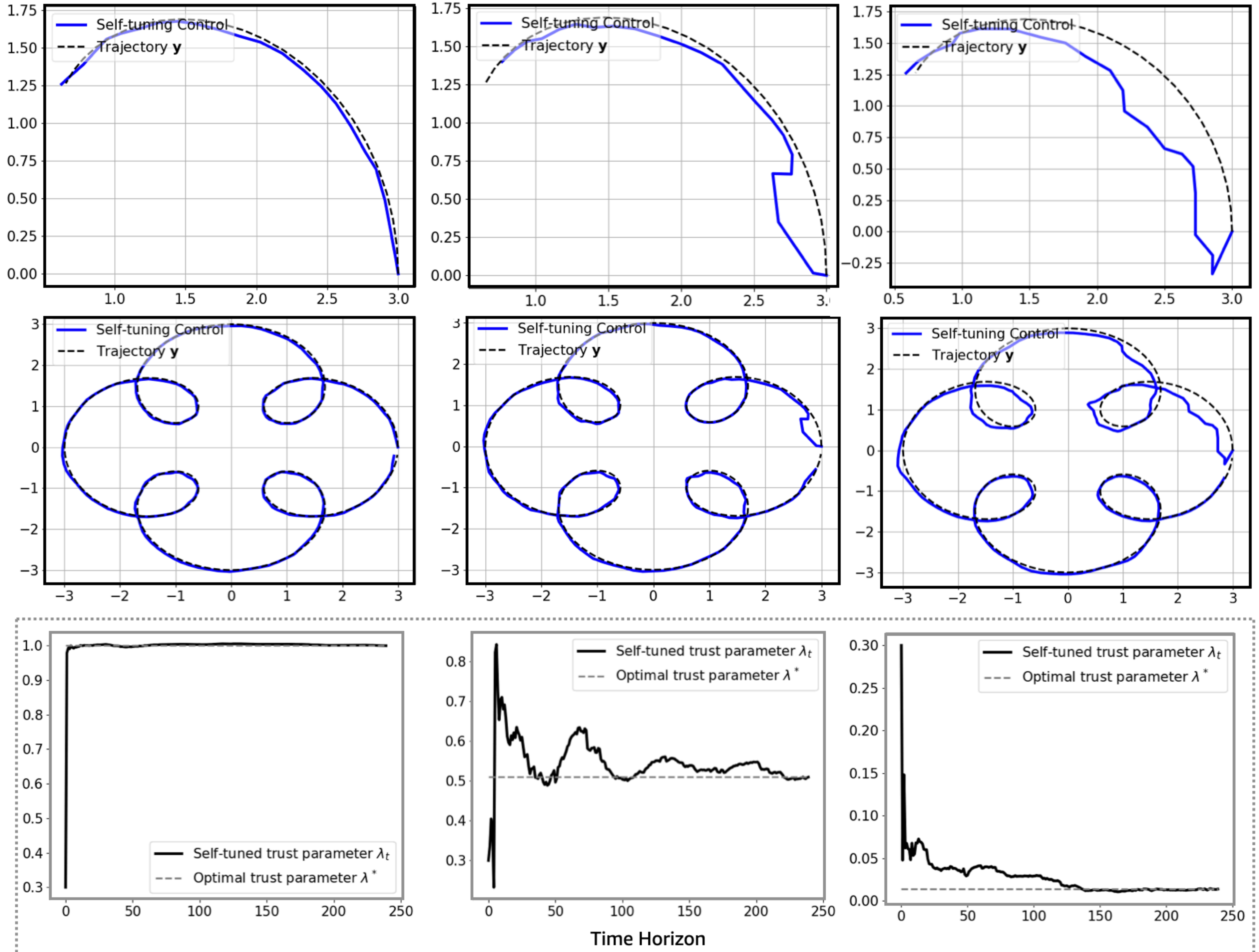
Case Study: Robot Tracking

High Error Case: Optimal $\lambda \approx 0$



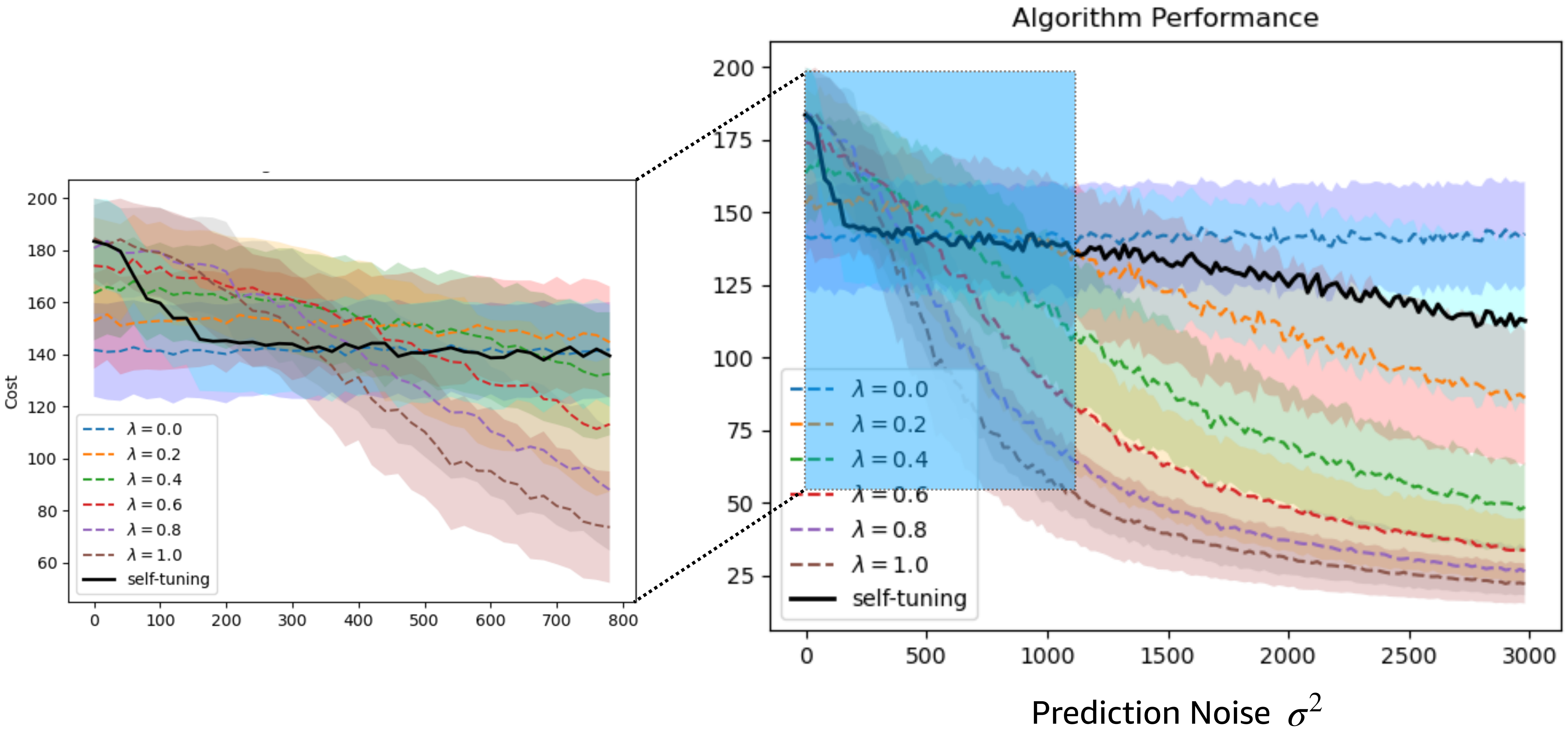


Verify the Convergence of Trust Parameters



Works for Nonlinear Dynamics ...

Empirically works well for the CartPole problem (nonlinear dynamics)!



Future Directions and Extensions

Learning-Augmented Online Control

- ***Linear System with Uncertain Predictions* [SIGMETRICS '22]** ⟵ *We have covered*
- *Nonlinear/Time-Varying Dynamics/MDP*
- *Gradient Descent for Learning the Trust Parameter Online*
- *Decentralized Linear Quadratic Regulator with Untrusted Predictions*
- *Non-convex combination of control policies*

Future Directions and Extensions

Learning-Augmented Online Control

- *Linear System with Uncertain Predictions [SIGMETRICS '22]* ← *We have covered*
- ***Nonlinear/Time-Varying Dynamics/MDP***
- *Gradient Descent for Learning the Trust Parameter Online*
- *Decentralized Linear Quadratic Regulator with Untrusted Predictions*

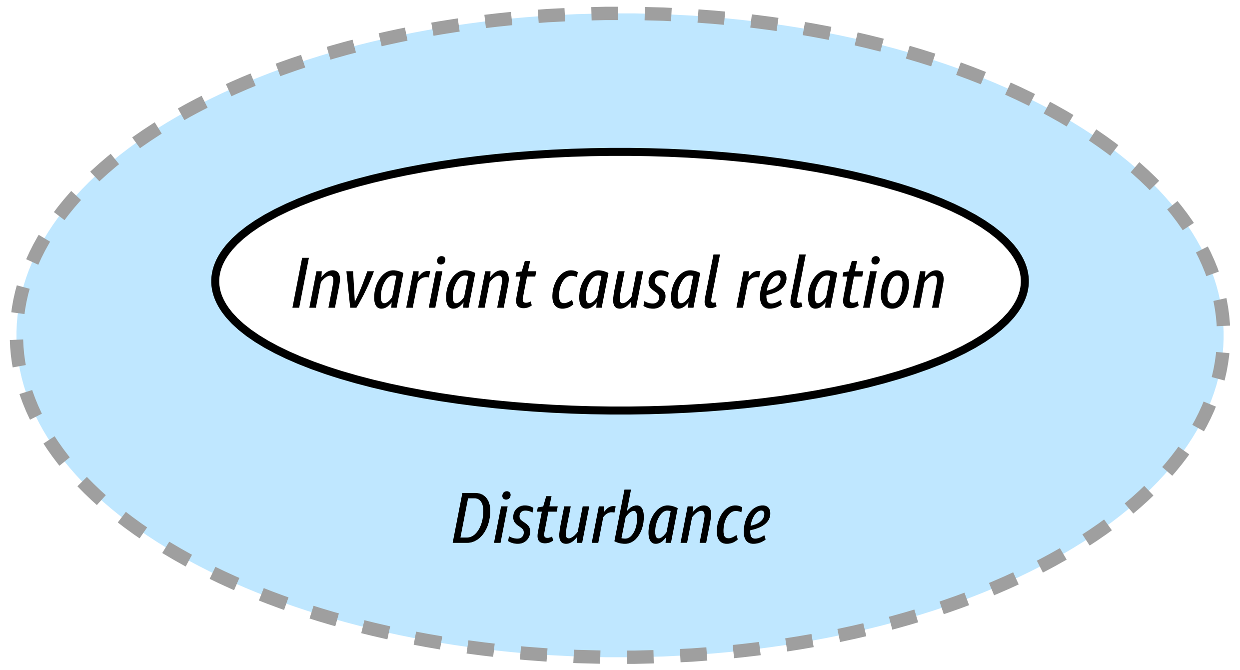
Advice: Pre-trained RL policies that are not robustness to …

- *Dataset Shift* [Bai et. al. NeurIPS 2019]
- *Mode Collapse* [Jabri et. al. NeurIPS 2019]
- *Reward Sparsity* [Riedmiller et. al. ICML 2018]
- *Sample Inefficiency* [Botvinick et. al. TrendsCognSci 2019]
- *Variability of Policy Gradient* [Cheng et. al. ICML 2019, Recht et. al. Annual Reviews 2019]

Future Directions and Extensions

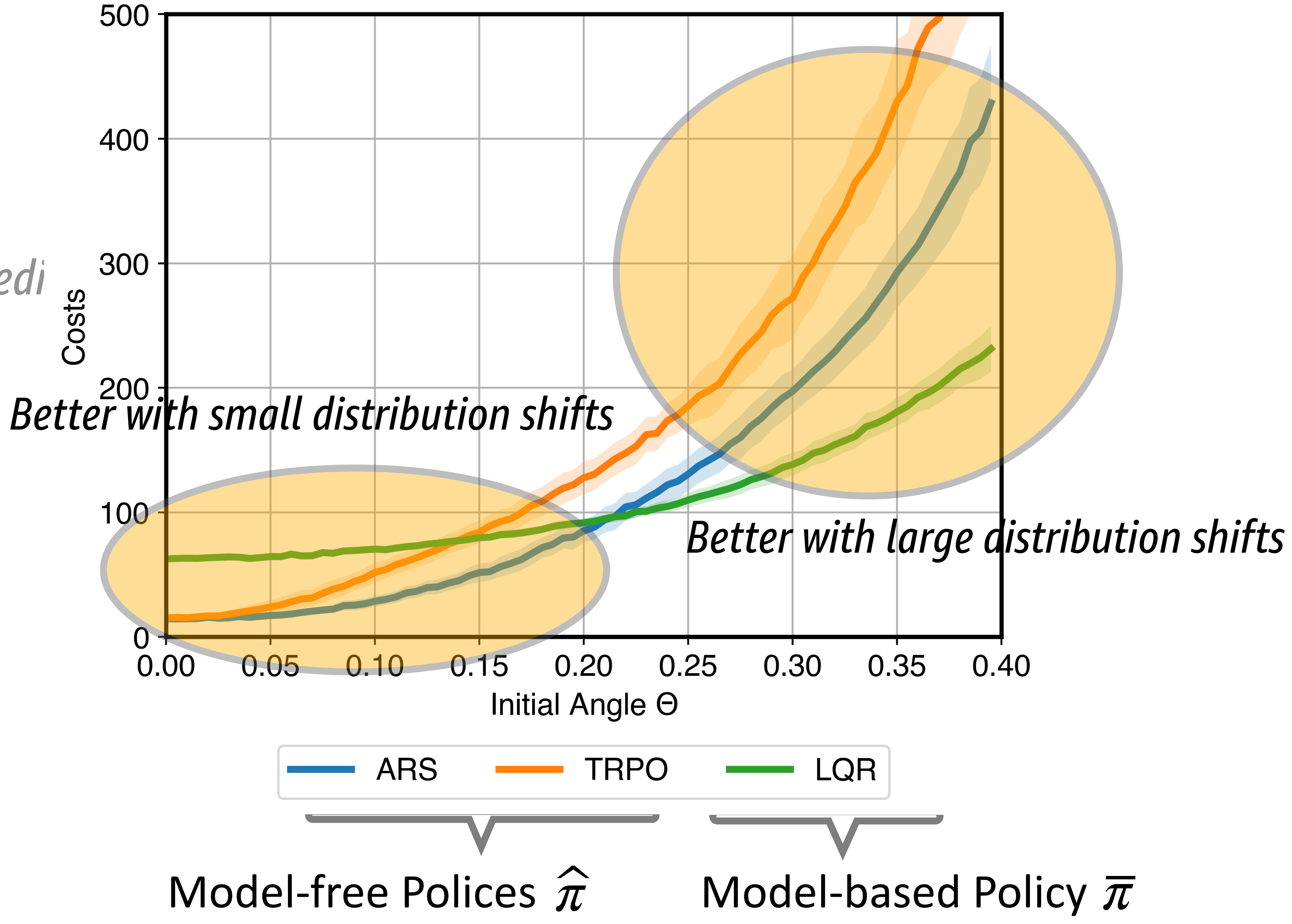
Learning-Augmented Online Control

- *Linear System with Uncertain Predictions [SIGMETRICS '22]*
- ***Nonlinear/Time-Varying Dynamics/MDP***
- *Gradient Descent for Learning the Trust Parameter Online*
- *Decentralized Linear Quadratic Regulator with Untrusted Predi*



Statistics change, but causal relation remains unchanged

CartPole Problem with Varying Initial Angle



More Learning-Augmented Control Models

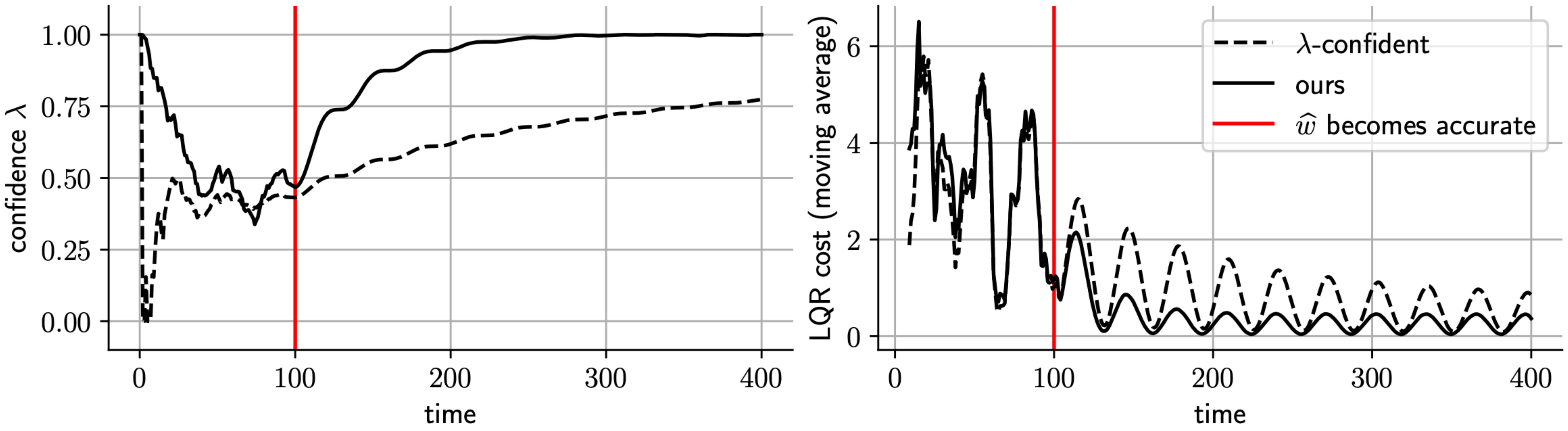
Dynamics	Cost Function	Imperfect Predictions/Black-box AI/ML Advice	
$x_{t+1} = Ax_t + Bu_t + w_t$	$\sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_f x_T$	$\widehat{w}_0, \ldots, \widehat{w}_{T-1}$	***LQR with Uncertain Predictions [SIGMETRICS '22]***
$x_{t+1} = Ax_t + Bu_t + f_t(x_t, u_t)$	$\sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t$	$\widehat{\pi} : \mathbb{R}^n \rightarrow \mathbb{R}^m$	*Stabilize Black-box Policies [Preprint '22]*
$x_{t+1} = f_t(x_t, u_t)$ $x_t \in \mathsf{X}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t}) \subseteq \mathbb{X}$ $u_t \in \mathsf{U}_t(\mathbf{x}_{<t}, \mathbf{u}_{<t}) \subseteq \mathbb{U}$	$\sum_{t=1}^{T} c_t(u_t)$	$p_t^* : \mathbb{U} \rightarrow \mathbb{X}$	*Learning-based MPC for DER Coordination [e-Energy '20a, SIGMETRICS '21, TSG '21]*

- ***Online-learning helps automatically tuning a trust parameter and achieve competitive ratio guarantees***
- *Crude model information can be used to guarantee stability for black-box policies in a single-trajectory*
- *Flexibility/feasibility information can be used to guarantee sub-linear regret in nonlinear control (with assumptions on constraints)*

Future Directions and Extensions

Learning-Augmented Online Control

- *Linear System with Uncertain Predictions [SIGMETRICS '22]*
- *Nonlinear/Time-Varying Dynamics*
- ***Other Online Learning Methods for Tuning the Trust Parameters***
- *Decentralized Linear Quadratic Regulator with Untrusted Predictions*



A recent work by Yiheng Lin et. al.

Gradient-based Adaptive Policy Selection (GAP)

Future Directions and Extensions

Learning-Augmented Online Control

- ***Non-convex Combination of Control Policies***

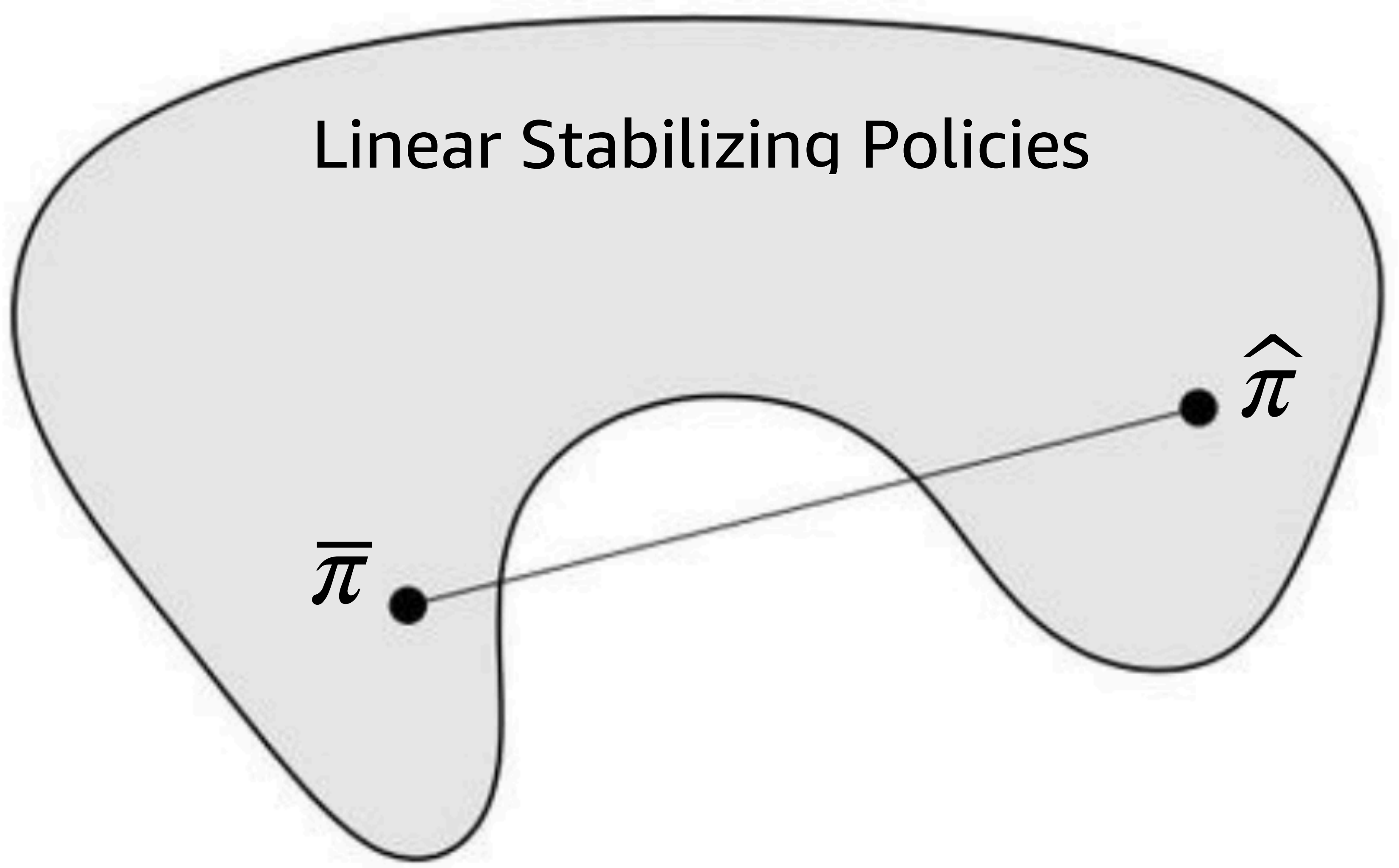
In general ... convex combination of control policies can lead to an unstable policy

Convex combination

$$\lambda \widehat{\pi}(x_t) + (1-\lambda)\overline{\pi}(x_t)$$

↑

Trust Parameter



Need a better way to combine classic algorithms and pre-trained policies!

[2019, Zheng et al.] On the equivalence of Youla, System-level and Input-output parameterizations.

Thank You!

(1) Li, Tongxin, Ruixiao Yang, Guannan Qu, Guanya Shi, Chenkai Yu, Adam Wierman, and Steven Low. "**Robustness and consistency in linear quadratic control with untrusted predictions.**" Proceedings of the ACM on Measurement and Analysis of Computing Systems 6, no. 1 (2022): 1-35.

(2) Li, Tongxin, Ruixiao Yang, Guannan Qu, Yiheng Lin, Steven Low, and Adam Wierman. "**Equipping Black-Box Policies with Model-Based Advice for Stable Nonlinear Control.**" arXiv preprint arXiv:2206.01341 (2022).

(2) Jianyi Yang, Li, Tongxin, Pengfei Li, Adam Wierman, Shaolei Ren. "**Reinforcement Learning for Online Competitive Control with Policy Priors.**" Submitted to SIGMETRICS '23

Tongxin Li litongxin@cuhk.edu.cn

School of Data Science

The Chinese University of Hong Kong (Shenzhen)