DDA 6201 Online Decision-Making Lecture 11

Application 1: Linear Quadratic Control

Tongxin Li

School of Data Science

The Chinese University of Hong Kong (Shenzhen)

Motivation and General Picture

The community has developed many AI/ML tools for making decisions in practical systems, e.g. power systems, transportation ⋯

But it’s hard to see them being widely used ⋯

The community has developed many AI/ML tools for making decisions in practical systems, e.g. power systems, transportation ⋯

But it’s hard to see them being widely used ⋯

How can we better introduce AI in practice to help make critical online decisions?

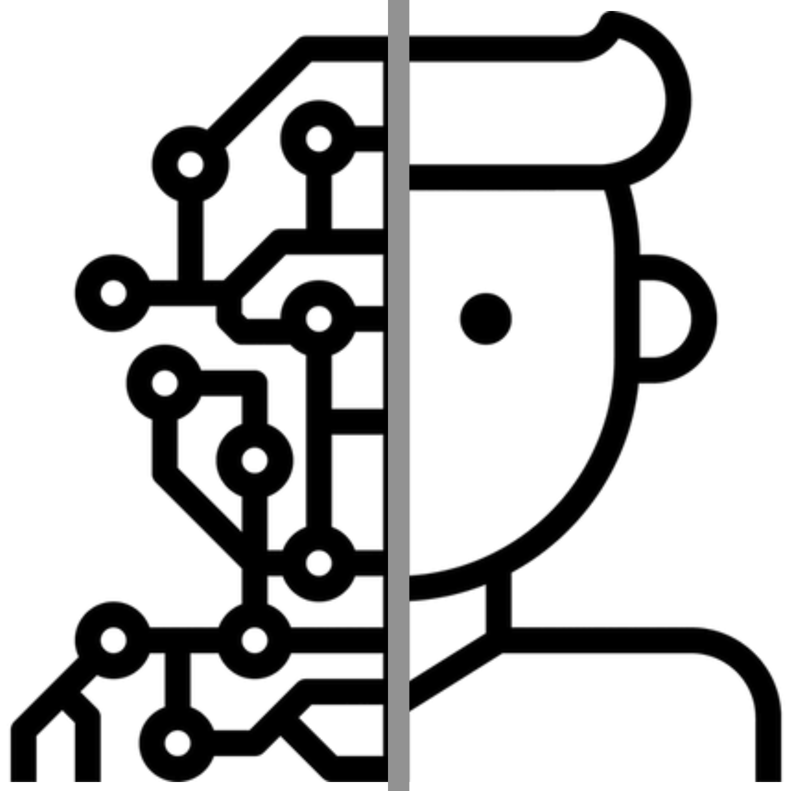
Going From Digital to Physical Worlds …

Digital World

AlphaGo

OpenAI

GPT-4



Physical World

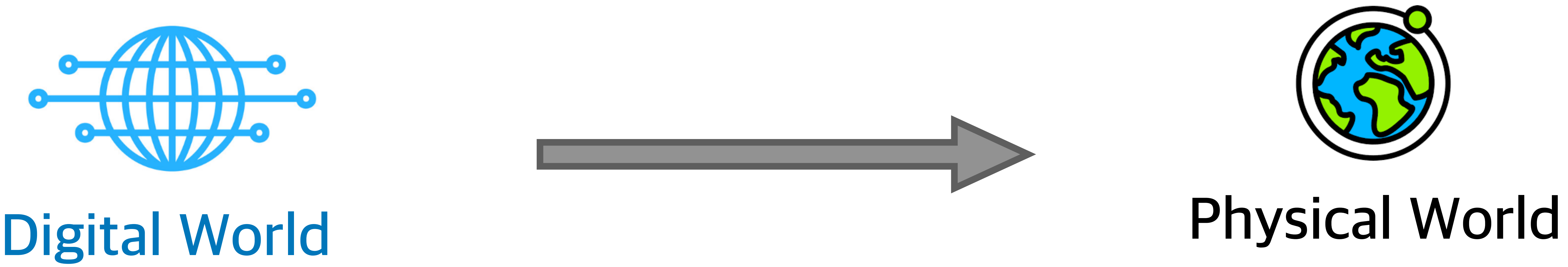
Power Systems

Autonomous Driving

Embodied AI

…

What Makes the AI Methods Less Responsible?



Digital World

Physical World

Key Challenges:

1. Environments are more complicated and more sensitive to mistakes

2. Many existing and well-established industrial methods that are hard to be replaced entirely (more unique in power systems)

e.g. Control Agent: Why should I use RL for scheduling?

Some Quick Thoughts

Idea: Use Classic Methods as Backup Plans!

Black-Box

Classic Problems and Methods

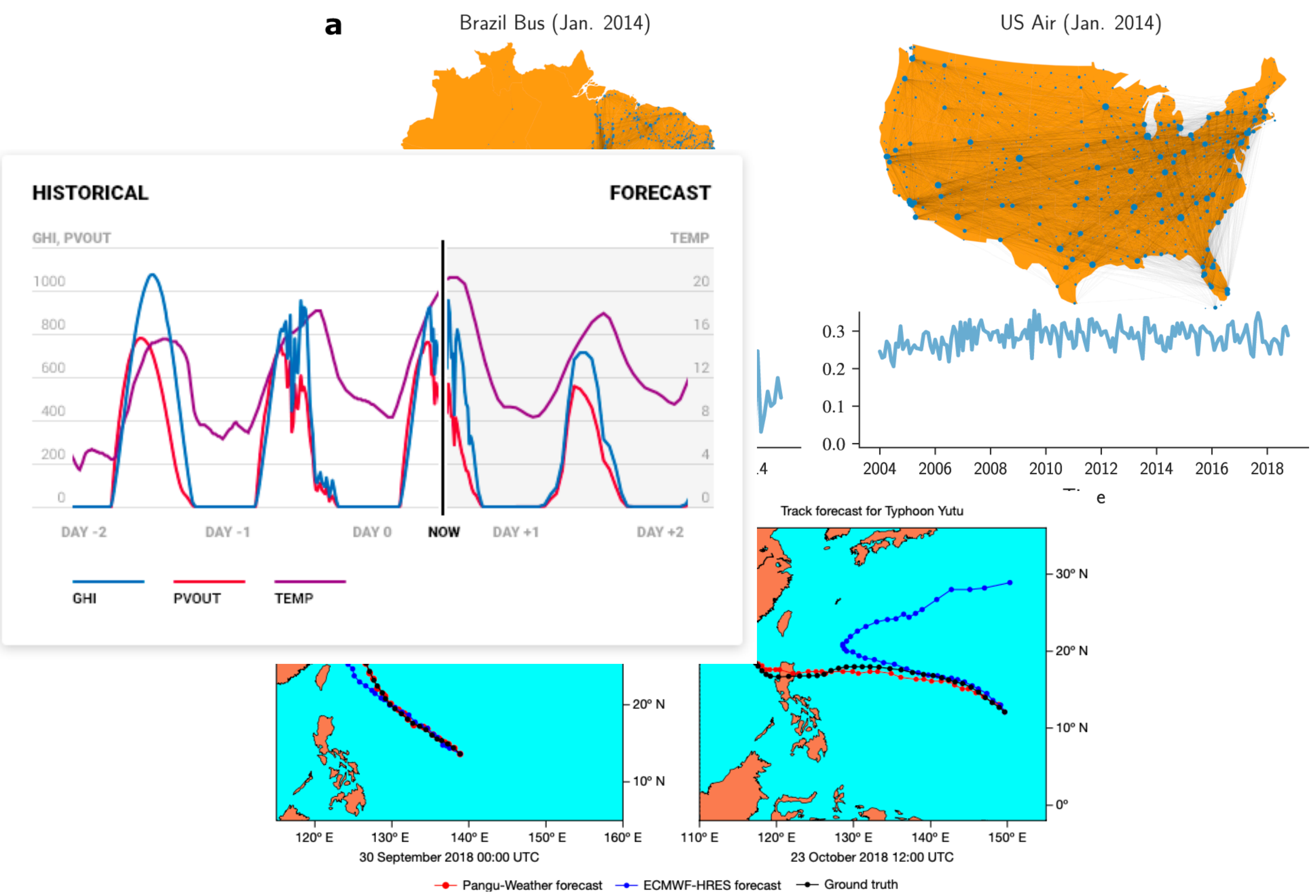
Machine-Learned Predictions

Online Optimization

Bandit Problems

MDP

?



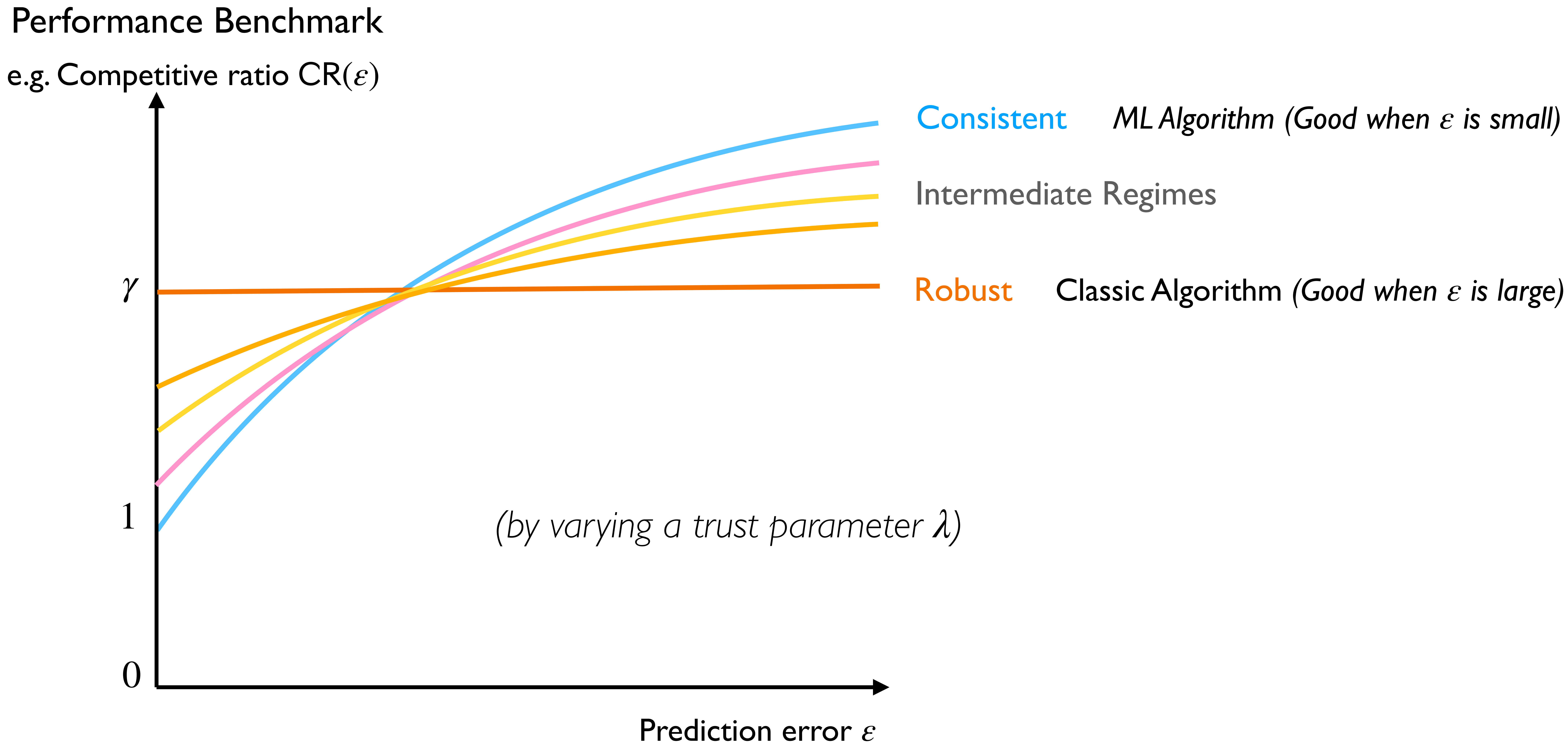
Linear Controller

Online Algorithms

ChatGPT

The Goal of Learning-Augmented Algorithms

Meta-algorithms   Consistency vs Robustness Trade-off



Consistent   *ML Algorithm (Good when ε is small)*

Intermediate Regimes

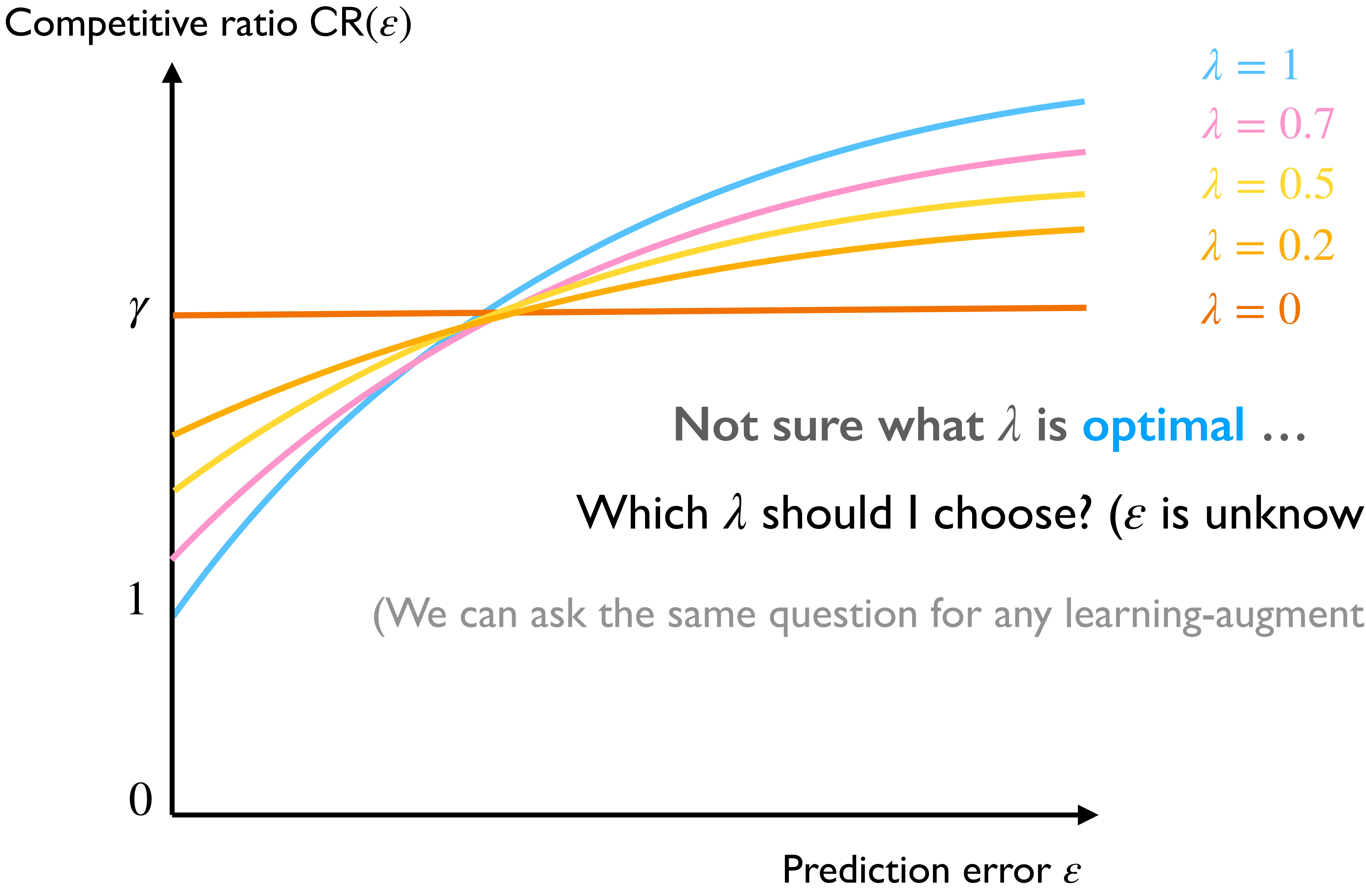
Robust   Classic Algorithm *(Good when ε is large)*

(by varying a trust parameter λ)

First Limitation

General Goal of Learning-Augmented Algorithms

Consistency vs Robustness Trade-off



Not sure what λ is optimal …

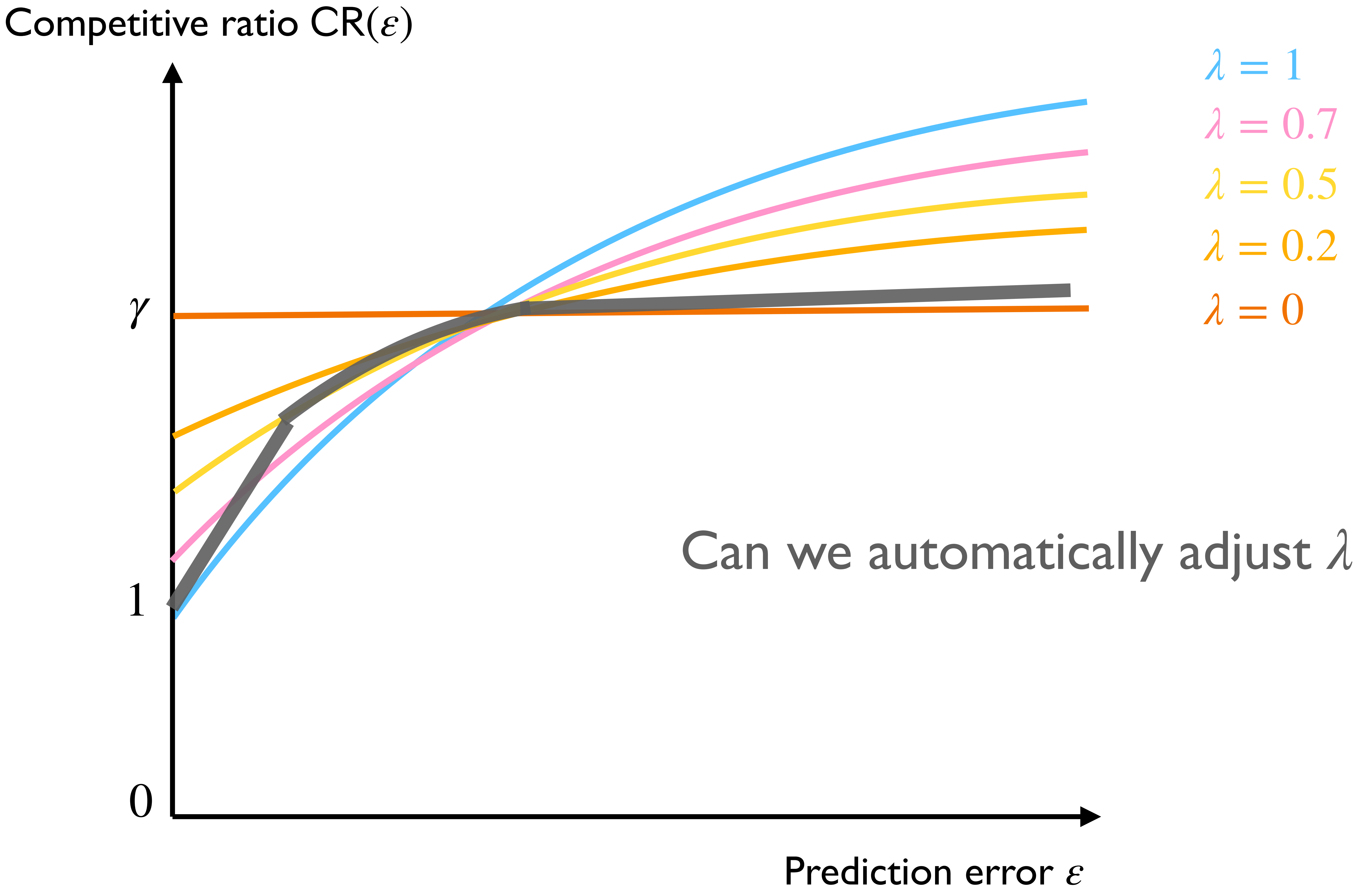
Which λ should I choose? (ε is unknown)

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

First Limitation

Issue: Prediction error ε is not known a priori

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

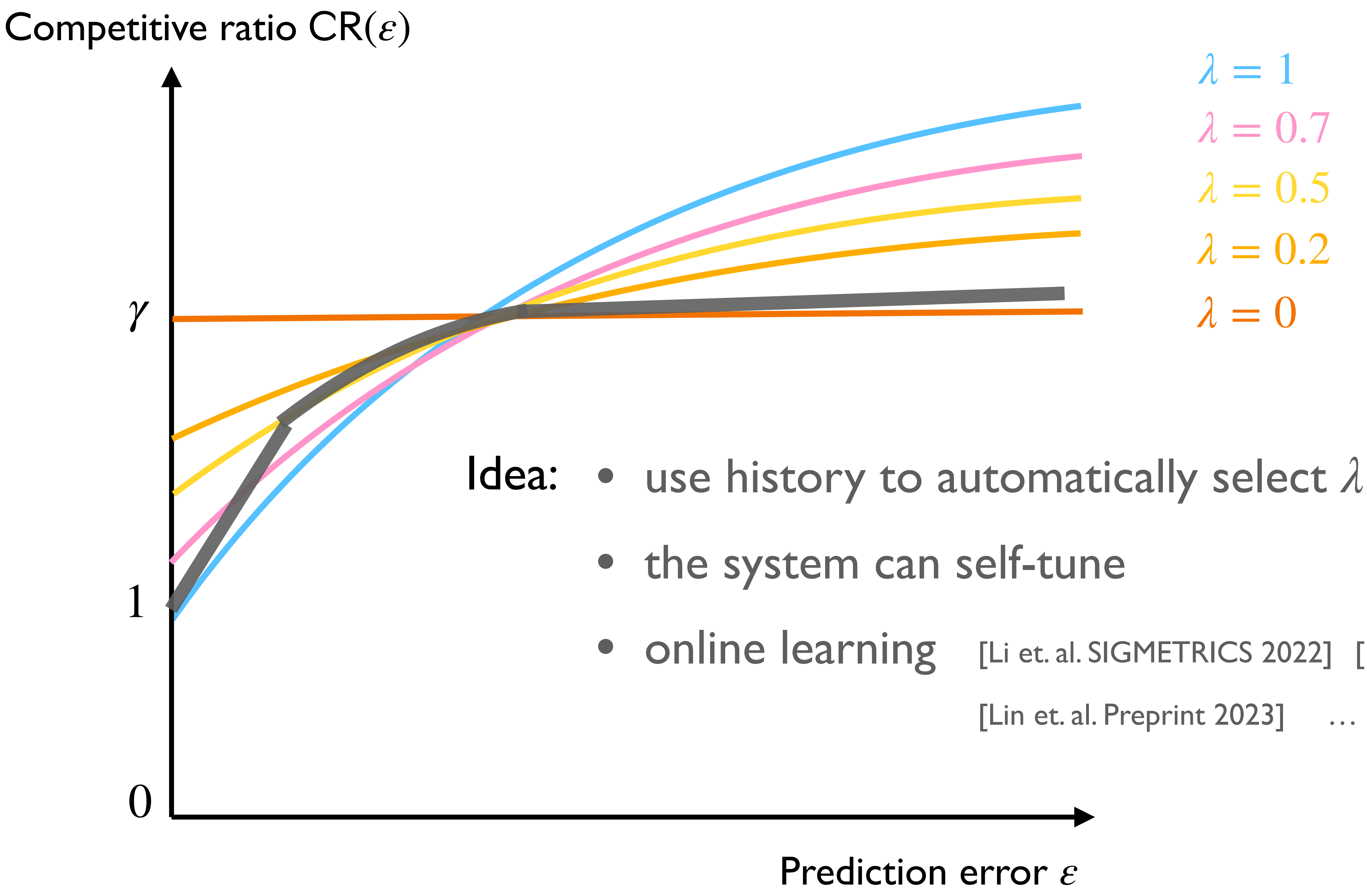


Can we automatically adjust λ ?

One Solution: **Online Learning**

General Goal of Learning-Augmented Algorithms

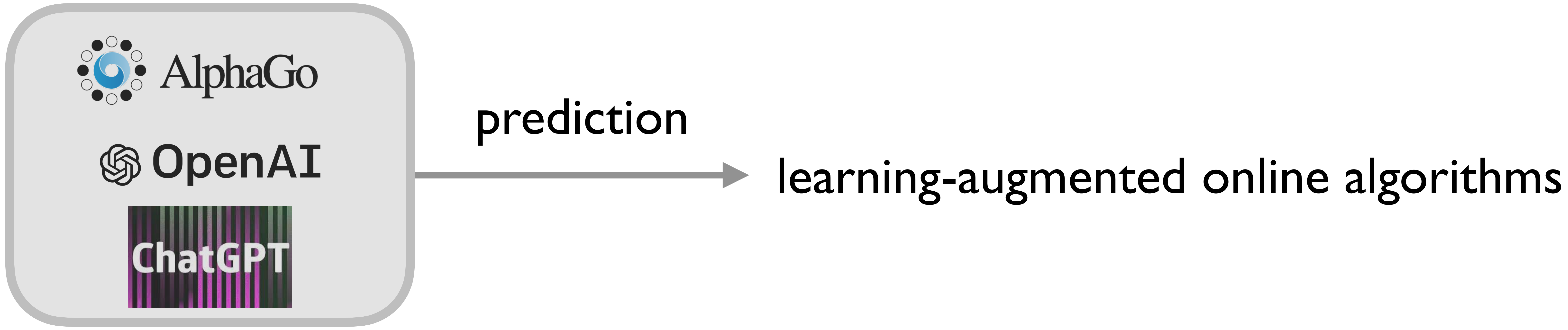
Consistency vs Robustness Trade-off



Idea:

- use history to automatically select λ
- the system can self-tune
- online learning [Li et. al. SIGMETRICS 2022] [Khodak et. al. NeurIPS 2022] [Lin et. al. Preprint 2023] … [Li et. al. NeurIPS 2024]

Second Limitation



prediction

learning-augmented online algorithms

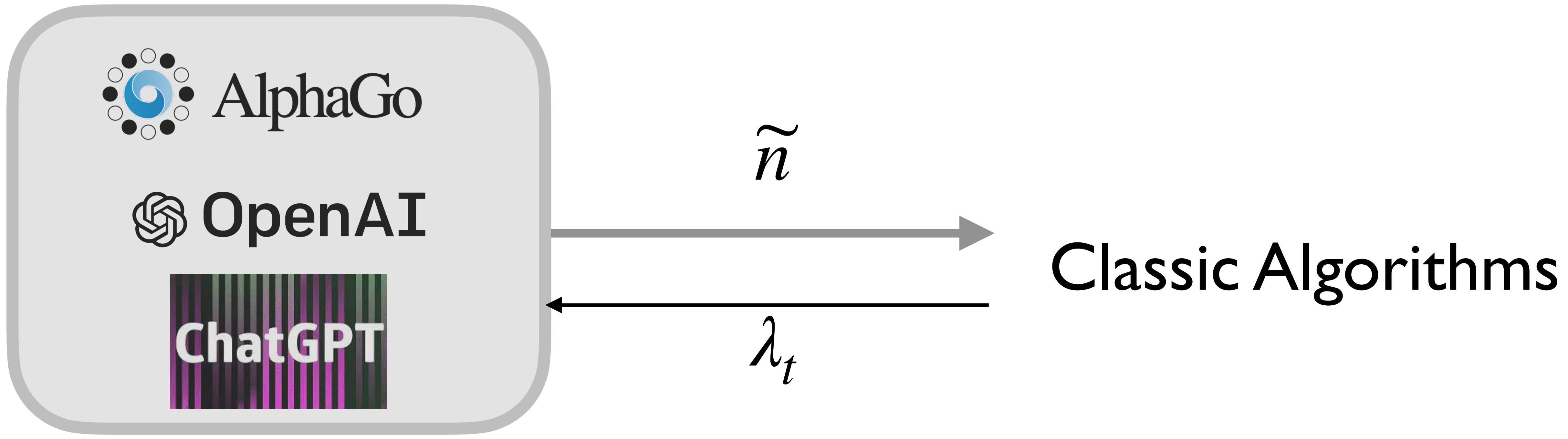
- The machine learning tools are considered as **black-boxes**
- **Structural information** of the model and ML tools can be helpful
  - specific forms of predictions [Li et. al. SIGMETRICS 2022]
  - grey-box ML models (Q-value functions of value-based policies)

    [Li et. al. Preprint 2023]
  - can be used to self-tune λ (second solution)

Second Limitation

- Learning-augmented —> Learning-infused

- Q-learning
- Linear Regression
- Multi-arm bandit



$\tilde{n}$

Classic Algorithms

λ_t

- The machine learning tools are considered as **black-boxes**

- Structural information of the model and ML tools can be helpful
  - specific forms of predictions [Li et. al. SIGMETRICS 2022]
  - grey-box ML models (Q-value functions of value-based policies)

    [Li et. al. Preprint 2023]

  - can be used to self-tune λ (second solution)

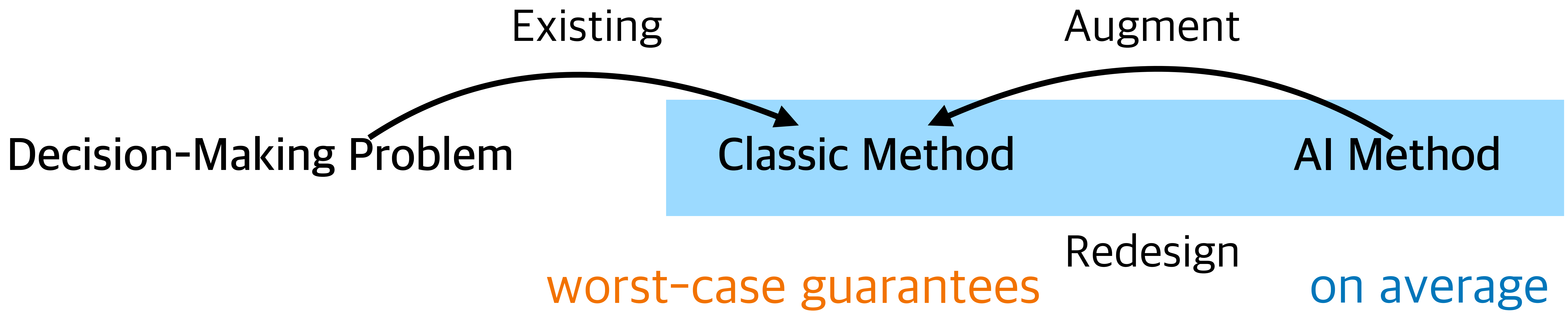
Learning-Augmented Algorithms

Online Problems	**不准确预测 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	
Linear Quadratic Control	**System Perturbations**	**[Li et. al. SIGMETRICS 2022] [Li et. al. NeruIPS 2024]**
	不可信AI建议 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]
Q Learning	Q-Value Functions	[Golowich et. al. NeurIPS 2022]
Value-Based RL	**Q-Value Functions (灰盒)/Actions (黑盒)**	**[Li et. al. NeurIPS 2023]**
Stochastic Game	**Type Beliefs**	**[Li et. al. NeurIPS 2024]**
…	…	

Over 100 topics on this website: https://algorithms-with-predictions.github.io/

Methods and Results

Partial Solution: Combine Classic and AI Algorithms

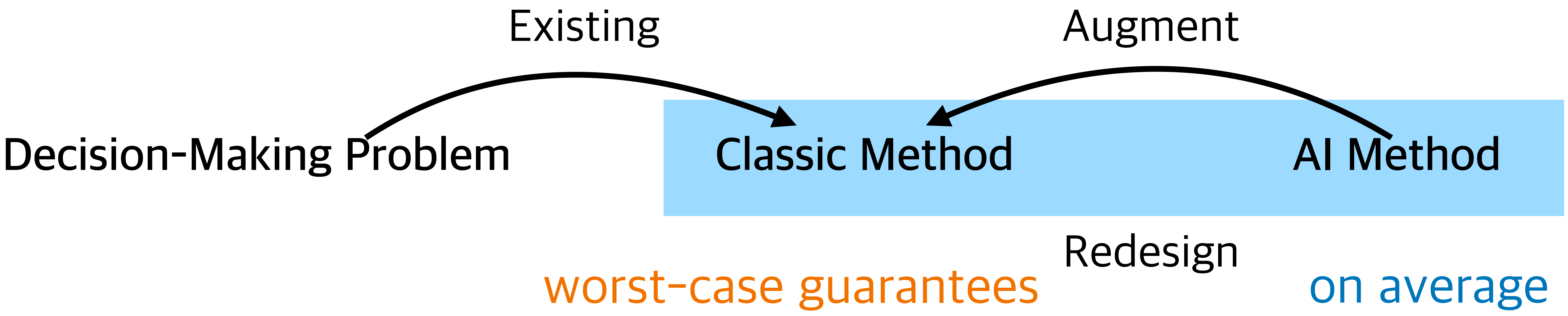


- Goal: **take advantage of both worlds**
- AI tools ~~make decisions alone~~ help classic algorithms make decisions

How to combine them?

- Switching
- Convex combination
- Projection ⋯

Partial Solution: Combine Classic and AI Algorithms



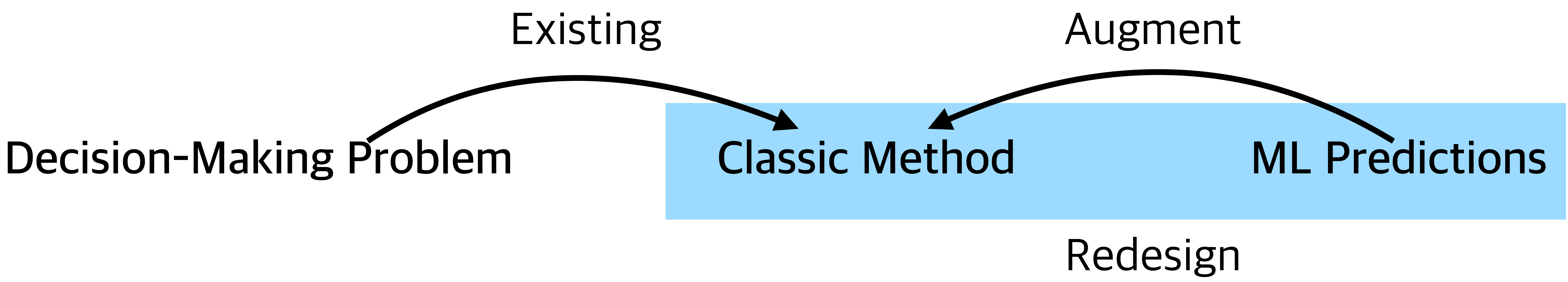
- Goal: **take advantage of both worlds**
- AI tools ~~make decisions alone~~ help classic algorithms make decisions

How to combine them?

- Switching
- Convex combination
- Projection ⋯

Next: Stepping into concrete examples

Partial Solution: Combine Classic and AI Algorithms



- Classic methods that are existing and have worst-case guarantees
- AI methods that are better on average
- Augment ML predictions or advice to the classic method and redesign algorithms

Revisit: Combine Classic and ML Algorithms

Classic Agent

$\bar{\pi} : X \rightarrow U$

State Space: X

Action Space: U

ML Agent

$\tilde{\pi} : X \rightarrow U$

- Goal: **take advantage of both worlds**

How to combine them?

- Switching
- Convex combination
- Projection ⋯

Combining Classic and ML Agents

Classic Agent

$\bar{\pi} : \mathrm{X} \rightarrow \mathrm{U}$

State Space: X

Action Space: U

ML Agent

$\tilde{\pi} : \mathrm{X} \rightarrow \mathrm{U}$

- Goal: **take advantage of both worlds**

How to combine them?

- Switching
- Convex combination
- Projection ⋯

Next: Stepping into concrete examples

Concrete Models

Classic Agent

$\bar{\pi} : \mathrm{X} \to \mathrm{U}$

State Space: X

Action Space: U

ML Agent

$\widetilde{\pi} : \mathrm{X} \to \mathrm{U}$

System Model	Classic Agent	ML Agent	
Linear Dynamics	LQR	MPC+Perturbation Predictions	[SIGMETRICS ’22]

Concrete Models

Classic Agent

$\overline{\pi} : \mathsf{X} \rightarrow \mathsf{U}$

State Space: X

Action Space: U

ML Agent

$\widetilde{\pi} : \mathsf{X} \rightarrow \mathsf{U}$

System Model	**Classic Agent**	**ML Agent**	
Linear Dynamics	LQR	MPC+Perturbation Predictions	[SIGMETRICS ’22]
NonLinear Dynamics	LQR	Black-Box RL	[OJCSYS ’23]

Linear Quadratic Control

Decision-Making Problem	Classic Method	AI Method
Linear Quadratic Control	Linear Quadratic Regulator	MPC with Perturbation Predictions

Dynamics

$$x_{t+1} = f_t(x_t, u_t) = Ax_t + Bu_t + w_t$$

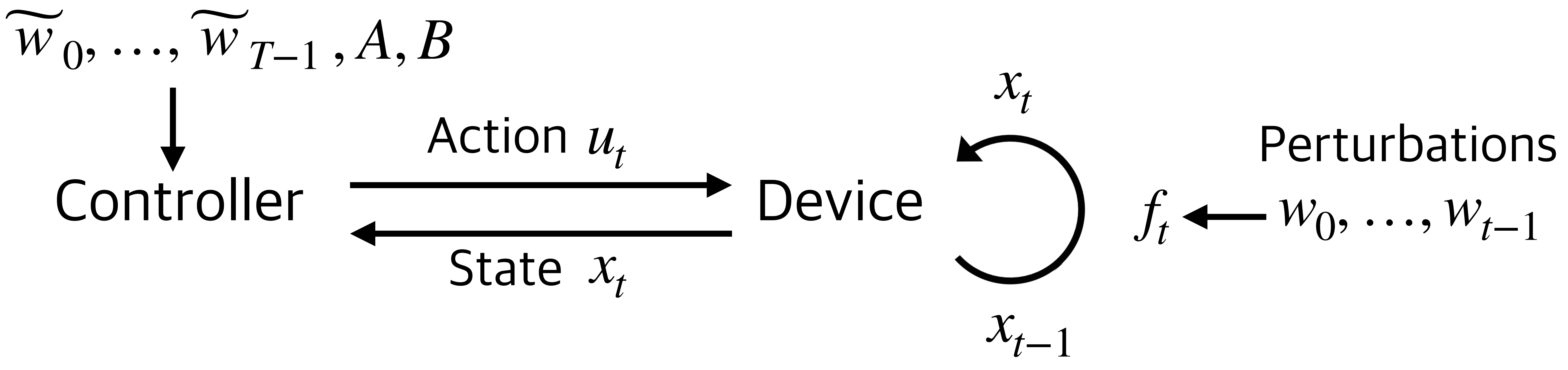
Costs

$$\sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t + x_T^\top Q_f x_T$$

Linear Quadratic Control

Machine-Learned Perturbations

System



$\widetilde{w}_0, \ldots, \widetilde{w}_{T-1}, A, B$

Controller

Action u_t

State x_t

Device

x_t

x_{t-1}

f_t

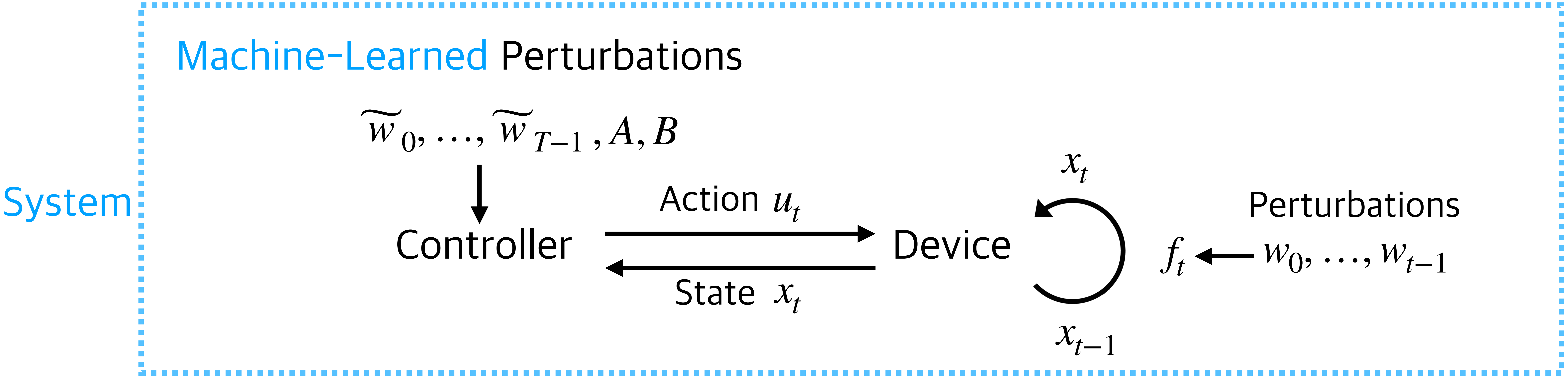
Perturbations

$w_0, \ldots, w_{t-1}$

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

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

Linear Quadratic Control



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

[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

Robust MPC cannot actively adapt based 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

$$\mathrm{CR}(\varepsilon) := \max_{\mathbf{w},\widetilde{\mathbf{w}}:d(\mathbf{w},\widetilde{\mathbf{w}})\leq\varepsilon} \frac{\mathrm{ALG}(\varepsilon)}{\mathrm{OPT}} \qquad \mathrm{CR} := \max_{\varepsilon\geq 0} \mathrm{CR}(\varepsilon)$$

$\mathrm{ALG}(\varepsilon) :=$ 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

$$\varepsilon$$

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

$P :=$ 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”

Prediction Error

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

ε

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

Why is it a "weighted sum"?

Quick Answer:

- Simplify expressions in our analysis

More fundamental Answers:

- Per-step error impact is not uniform in a dynamical system
- Impact decays exponentially
- It is actually the "error in the actions"

Model Predictive Control

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

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

Good when ε is small

$$x_{\tau+1} = A x_\tau + B u_\tau + \widetilde{w}_\tau, \forall \tau = t, \ldots, T-1.$$

(Explicit Expressions [2020 Yu et al.])

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

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

Taking benefit of Two Policies ⋯

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

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

(Optimal linear controller for LQR with Gaussian perturbations)

Taking benefit of Two Policies ⋯

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

Good when ε is small

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

Drop the predictions Good when ε is large

(LQR; optimal with Gaussian perturbations)

MPC Policy + LQR Policy

How about a convex combination?

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

Trust Parameter

λ-Confident Control

“1-confident”

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

“ λ-confident”

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

Trust parameter

“0-confident”

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

λ-Confident Control

“ λ-confident”

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

Trust parameter

(Equivalent to)

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

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

Trust parameter

Revisit Our Paradigm

Decision-Making Problem | Classic Method | AI Method

Linear Quadratic Control | Linear Quadratic Regulator | MPC with Machine Learned Predictions

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

$P :=$ Solution of DARE

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

Revisit Our Paradigm

Decision-Making Problem	Classic Method	AI Method
Linear Quadratic Control	Linear Quadratic Regulator	MPC with Machine Learned Predictions

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

$P :=$ Solution of DARE

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

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

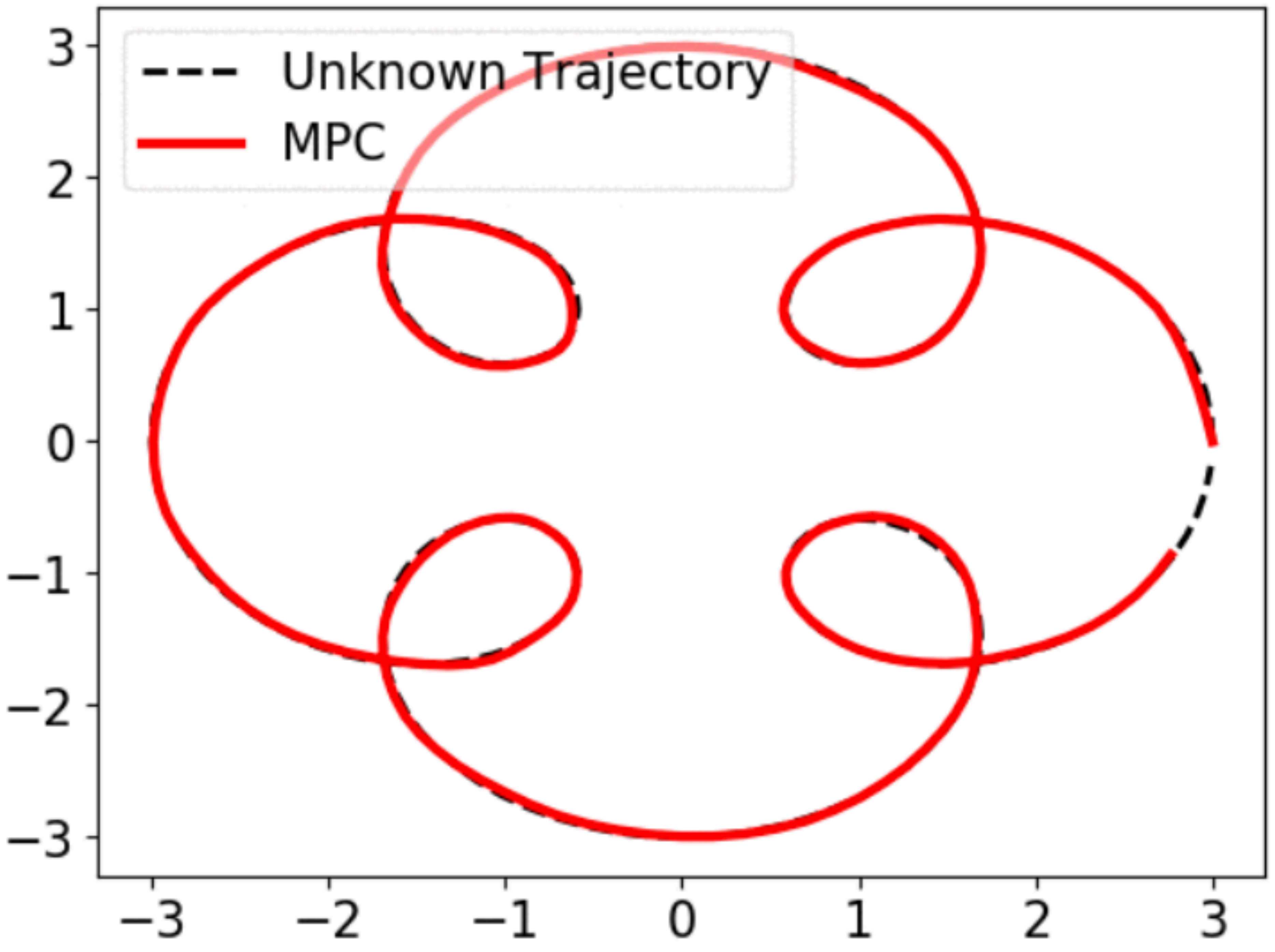
$$x_{\tau+1} = Ax_\tau + Bu_\tau + \widetilde{w}_\tau, \forall \tau = t, \ldots, T-1.$$

Predictions

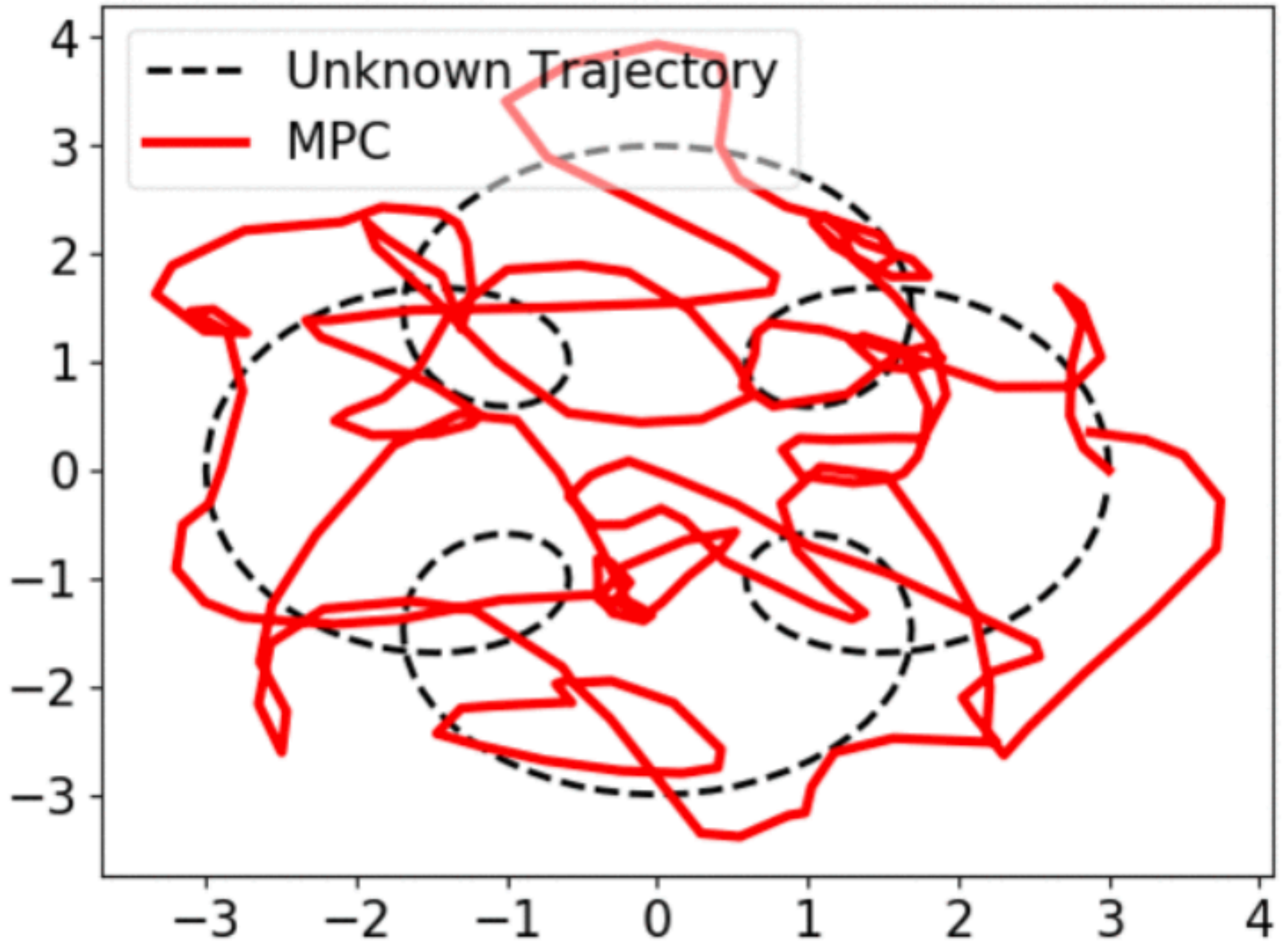
MPC with Untrusted Predictions

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

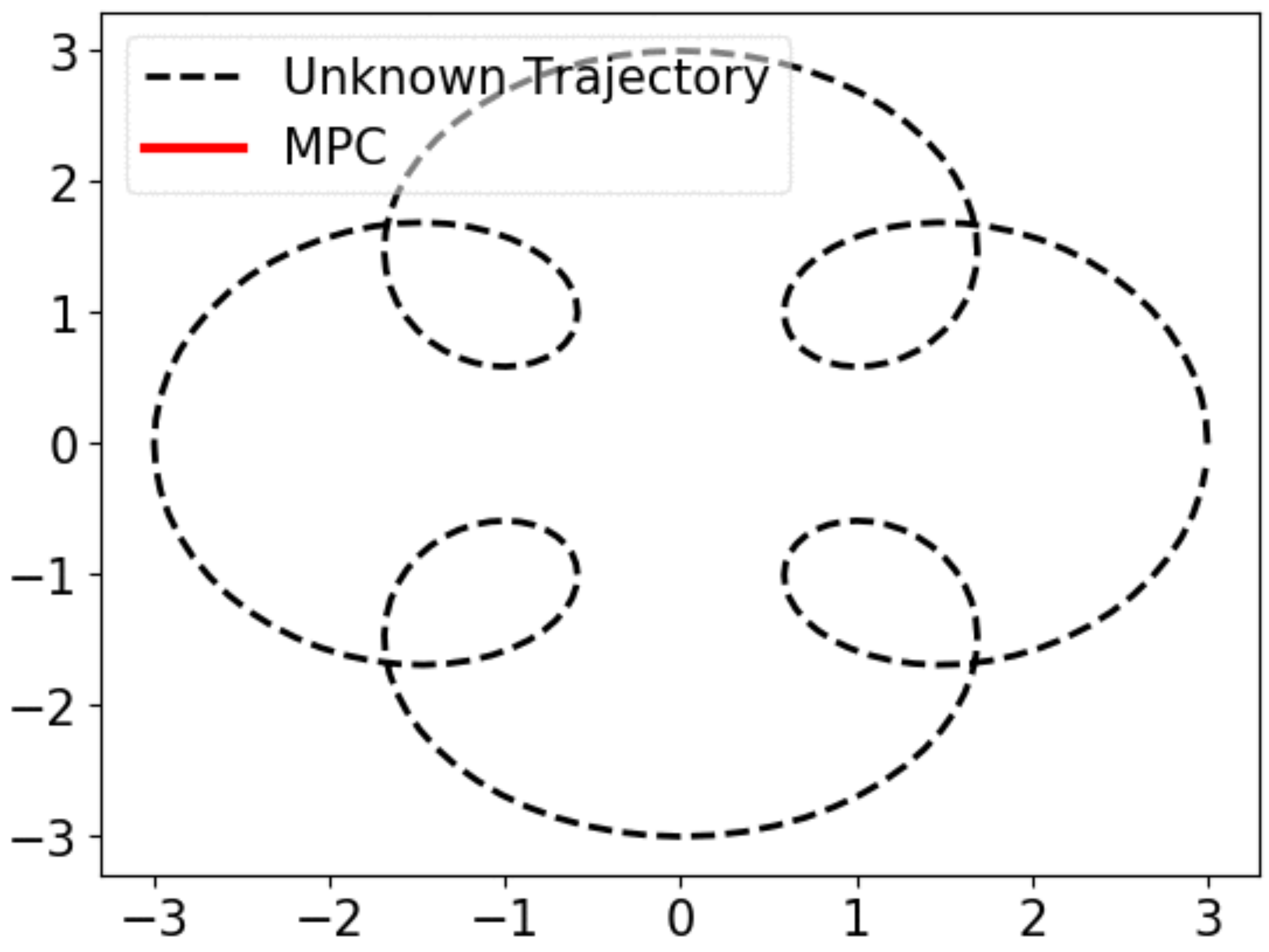
Prefect Predictions



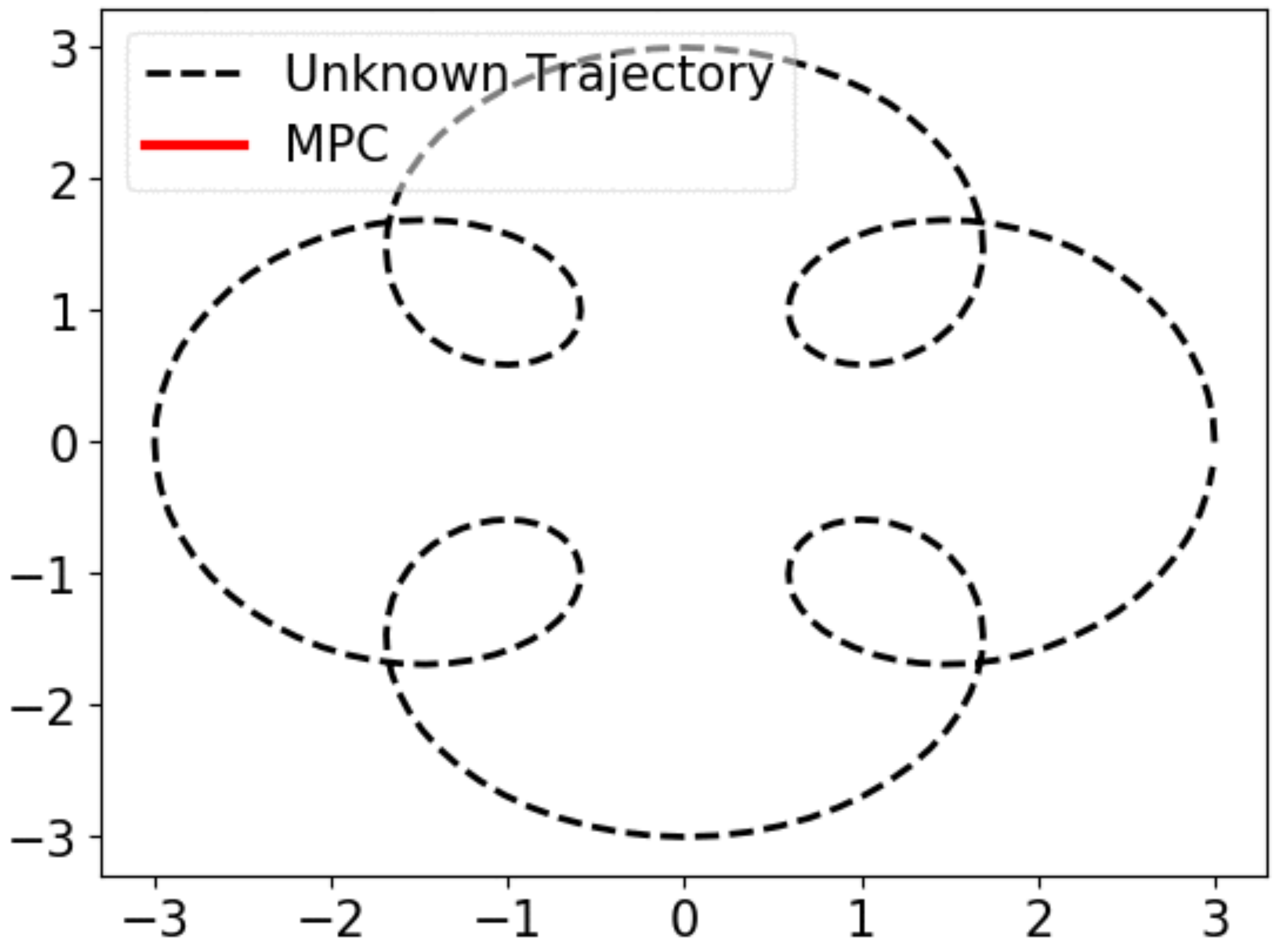
Untrusted ML Predictions



MPC with Untrusted Predictions



Prefect Predictions



Untrusted ML Predictions

Revisit Our Paradigm

Decision-Making Problem	Classic Method	AI Method
Linear Quadratic Control	Linear Quadratic Regulator	MPC with Machine Learned Predictions

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

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

Revisit Our Paradigm

Augment

Decision-Making Problem	**Classic Method**	**AI Method**
	Good when ε is large	Good when ε is small
Linear Quadratic Control	Linear Quadratic Regulator	MPC with Machine Learned Predictions

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

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

How about a convex combination?

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

Trust Parameter $\lambda \in [0,1]$

Revisit Our Paradigm

Augment

Decision-Making Problem	Classic Method	AI Method
	Good when ε is large	Good when ε is small
Linear Quadratic Control	Linear Quadratic Regulator	MPC with Machine Learned Predictions

“ λ-confident”

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

Competitive Ratio Results

Theorem (Informal; SIGMETRICS '22) Meta Theorem

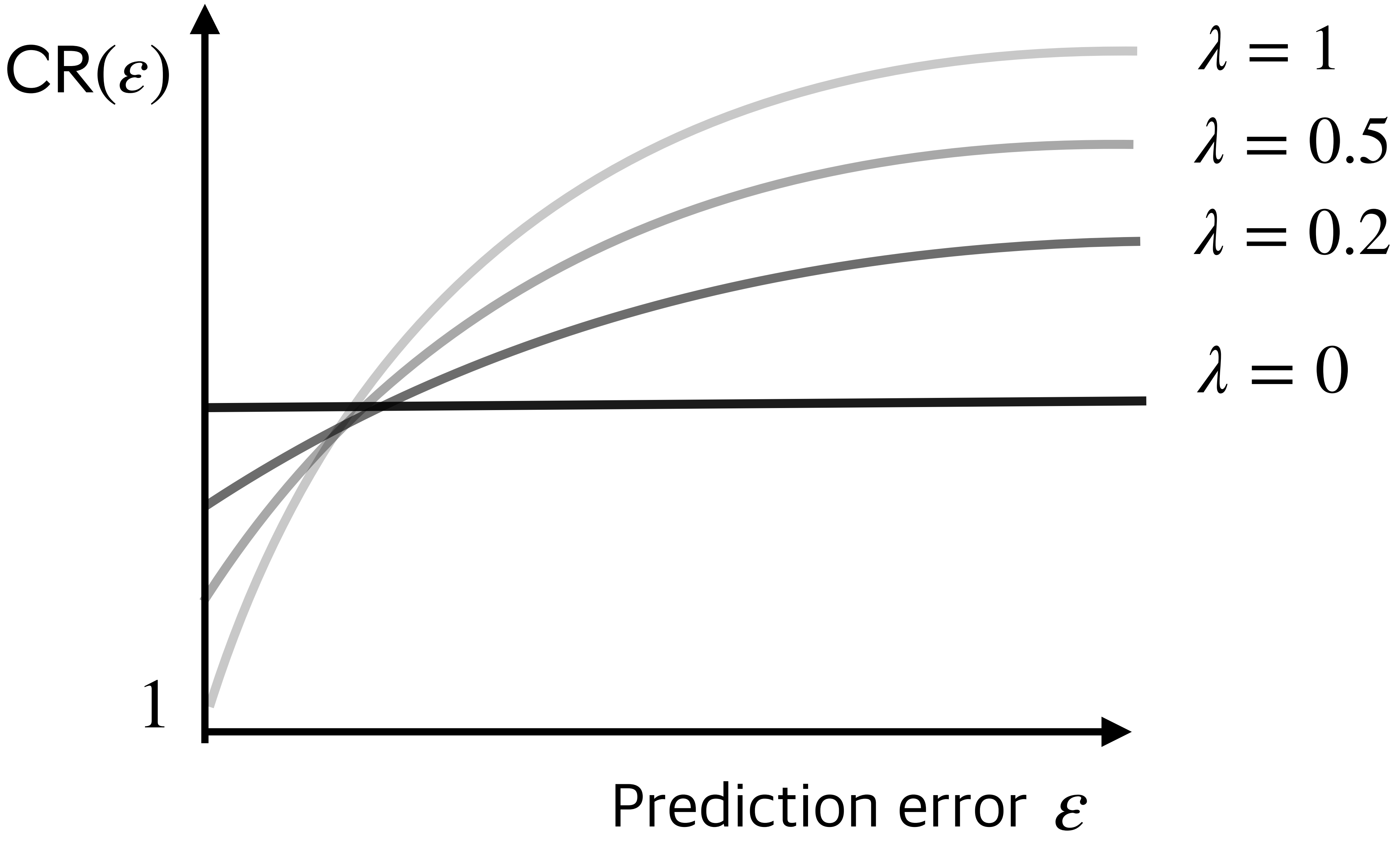
Under model assumptions, with a fixed trust parameter $\lambda > 0$, the λ-confident algorithm has a worst-case competitive ratio of at most

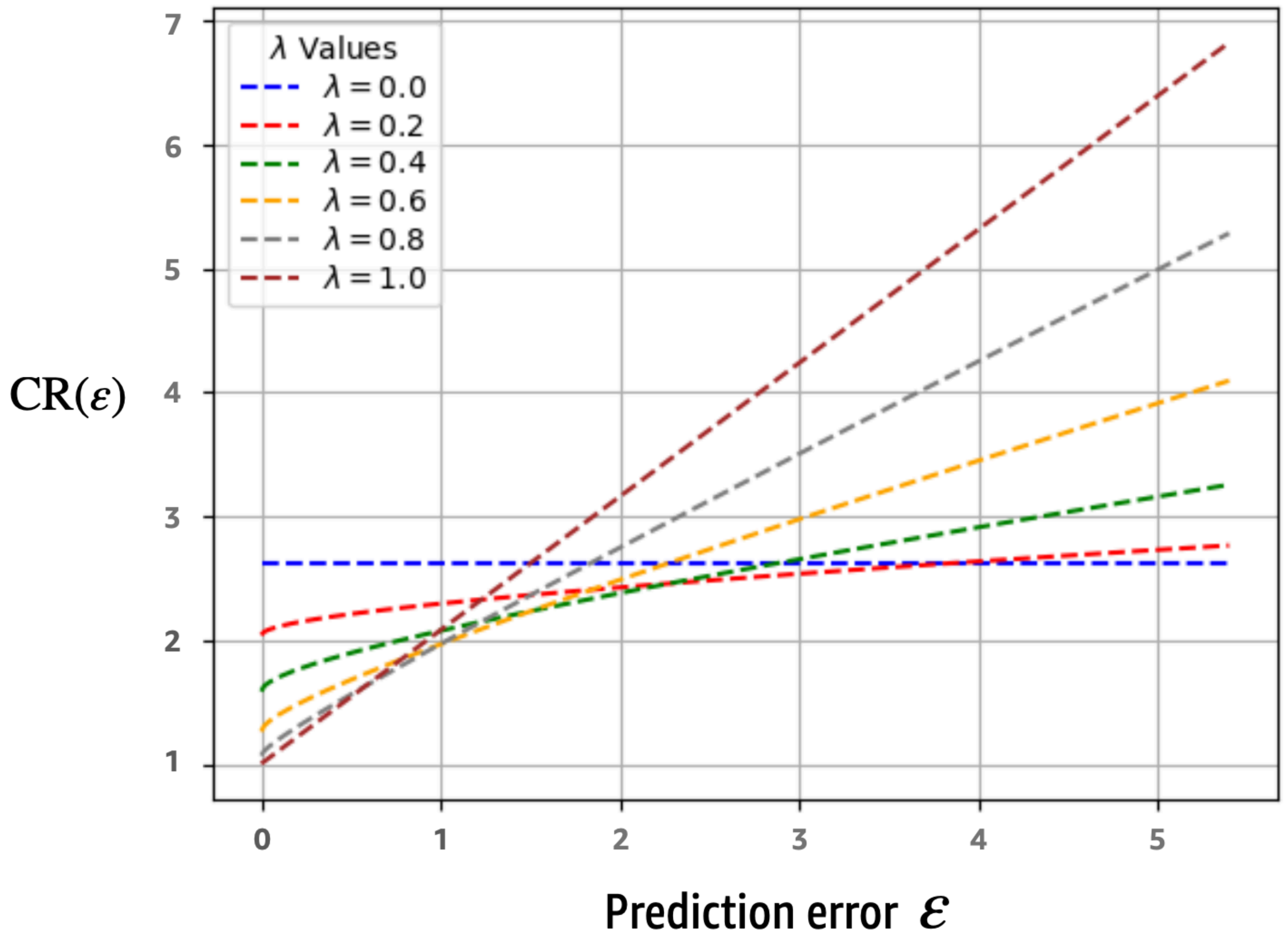
$$\mathsf{CR}(\varepsilon) \leq 1 + 2\|H\| \min\left\{\left(\frac{\lambda^2}{\mathsf{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C}\right), \left(\frac{1}{C} + \frac{\lambda^2}{\mathsf{OPT}}\overline{W}\right)\right\}$$

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

Varying Trust Parameter λ

Consistency vs Robustness Trade-off



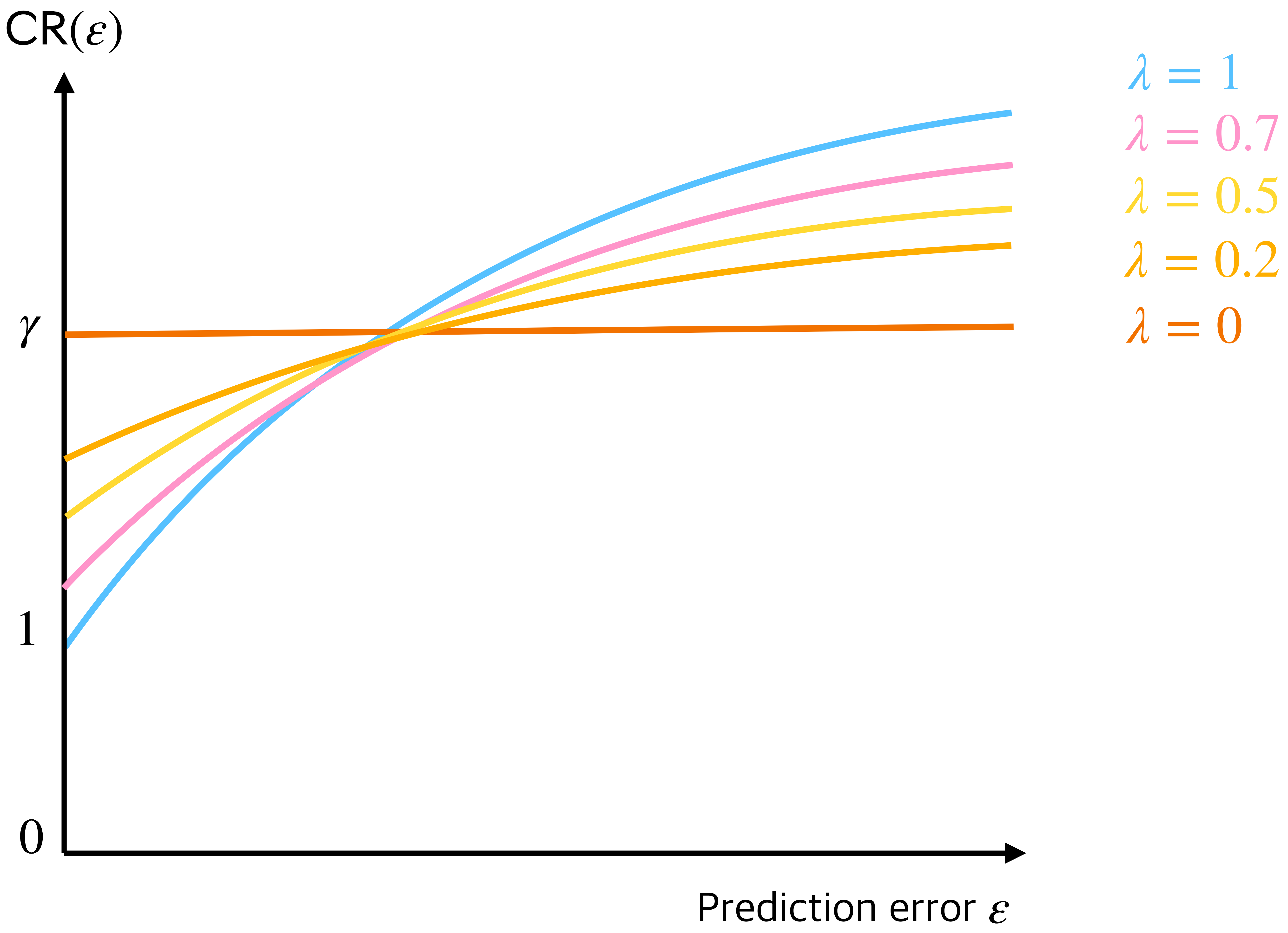


$$\text{CR}(\varepsilon) \le 1 + 2\|H\| \left(\frac{\lambda^2}{\text{OPT}}\varepsilon + \frac{(1-\lambda)^2}{C} \right)$$ **[SIGMETRICS '22]**

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

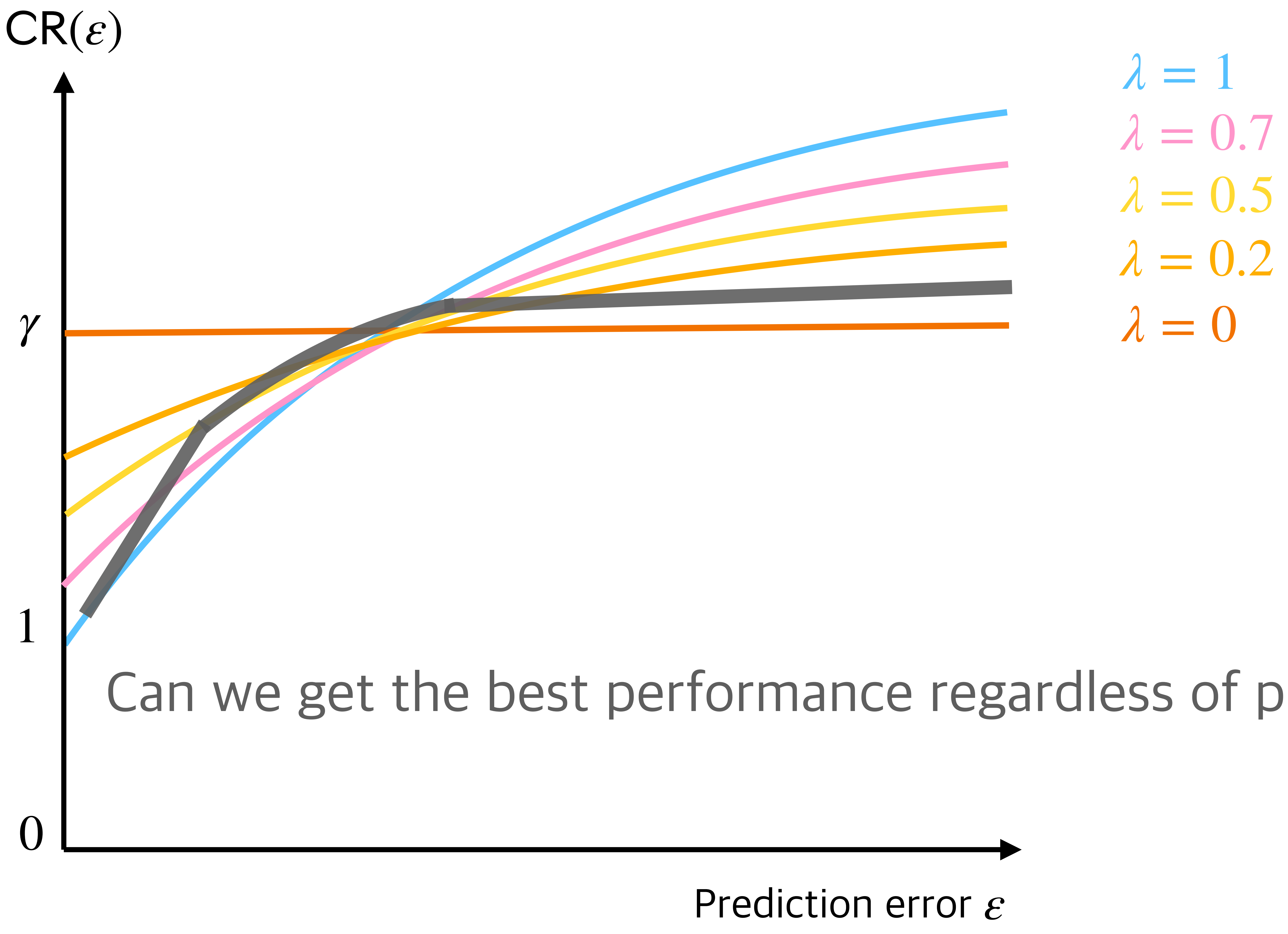
What λ Should I Choose?

(ε is unknown)



What λ Should I Choose?

(ε is unknown)



Can we get the best performance regardless of prediction error?

Our Solution: Online Learning Approach

Quadratic function of λ

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

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

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

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

Self-Tuning Control Algorithm

```
For $t = 0,\ldots,T-1$
    If $t = 0$ Initialize $\lambda_0$
    Else Compute
        $\lambda_t = \frac{\sum_{s=0}^{t-1}\left(\eta(w;s,t-1)\right)^\top H\left(\eta(\widetilde{w};s,t-1)\right)}{\sum_{s=0}^{t-1}\left(\eta(\widetilde{w};s,t-1)\right)^\top H\left(\eta(\widetilde{w};s,t-1)\right)}$
        where $\eta(w;s,t) := \sum_{\tau=s}^{t}\left(F^\top\right)^{\tau-s} P w_\tau$
    Generate an action using the $\lambda_t$-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

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

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

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

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

Competitive Ratio Bound for Self-tuning Control

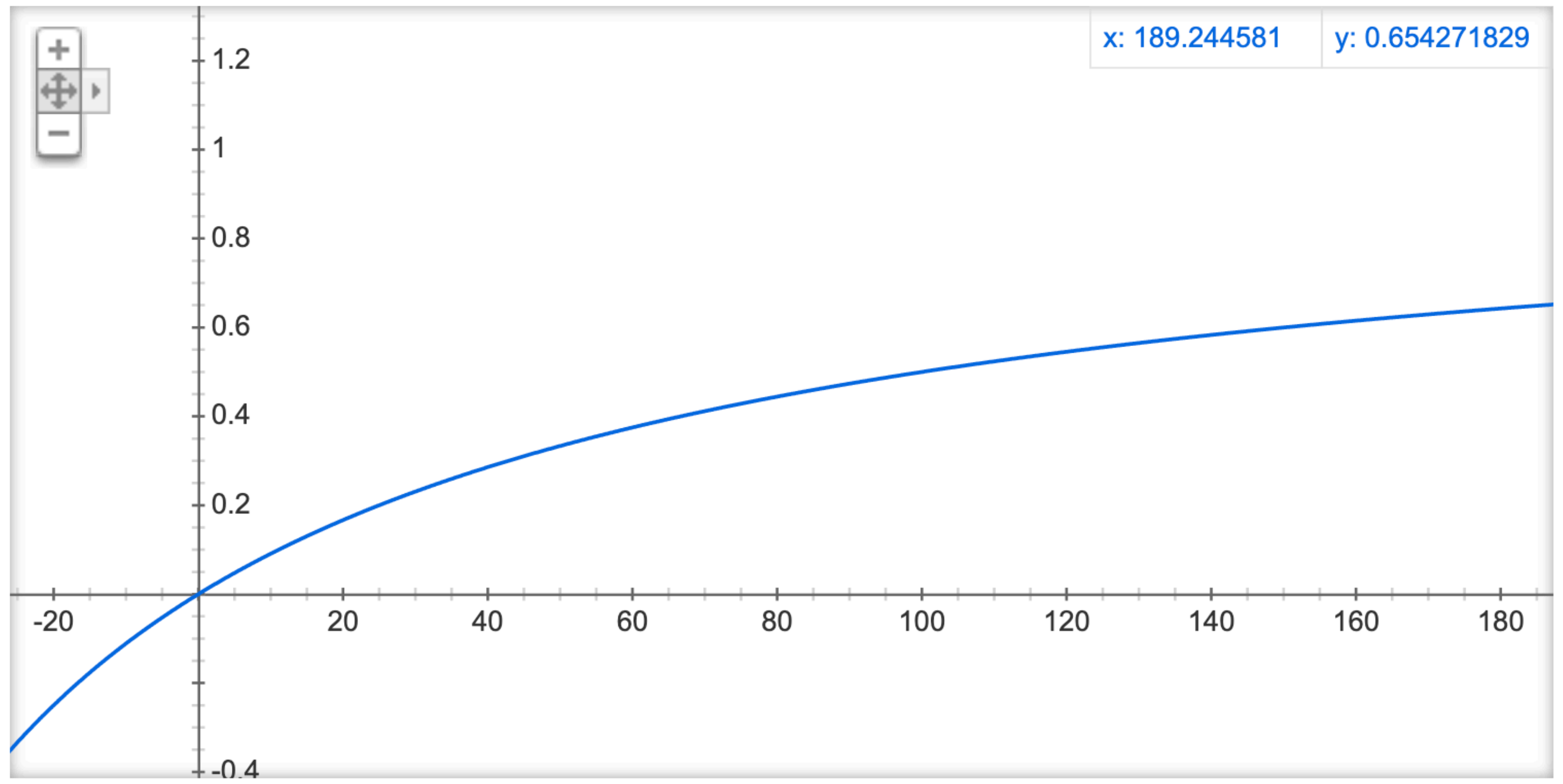
Theorem (Informal; SIGMETRICS ’22) **CR Theorem**

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

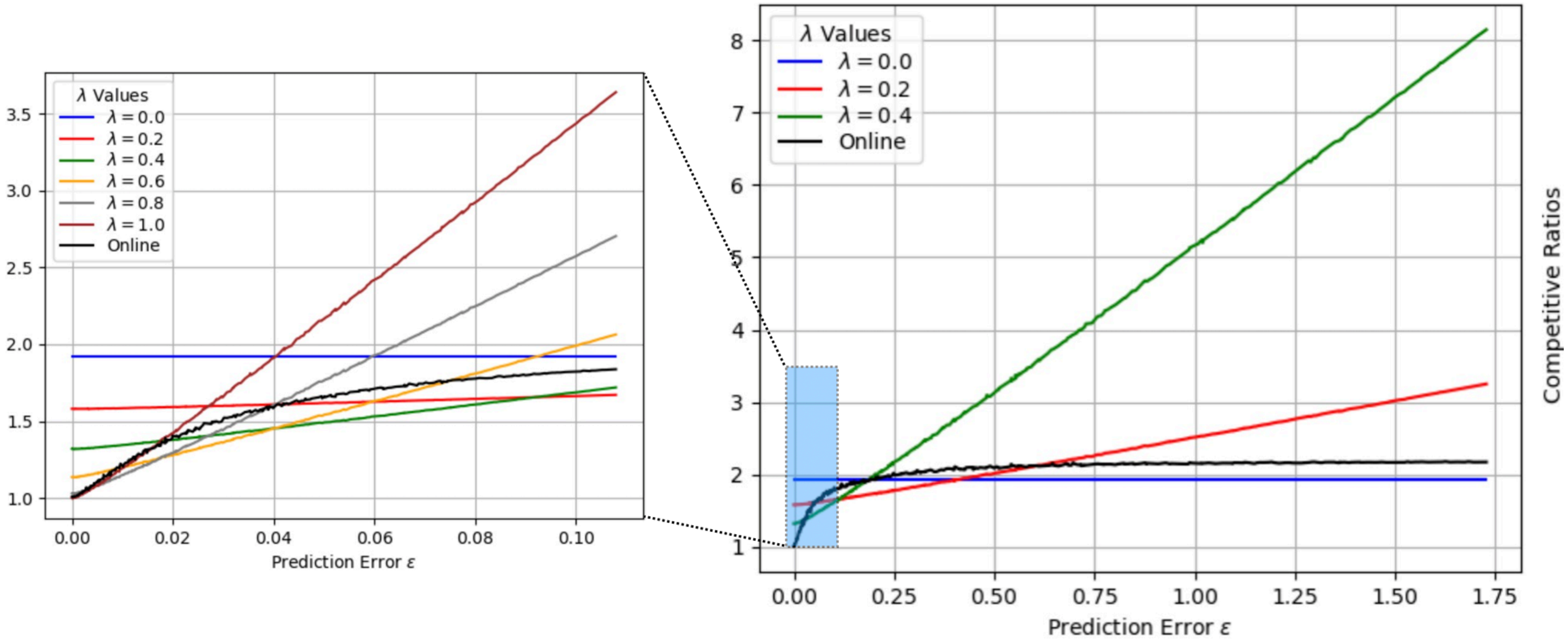
$$\mathsf{CR}(\varepsilon) \leq 1 + 2\|H\|\frac{\varepsilon}{\mathsf{OPT} + C\varepsilon} + O\left(\frac{\left(\mu_{\mathsf{VAR}}(\mathbf{w}) + \mu_{\mathsf{VAR}}(\widetilde{\mathbf{w}})\right)^2}{\mathsf{OPT}}\right).$$

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

Graph for x/(100+x)



Apply Our Algorithm



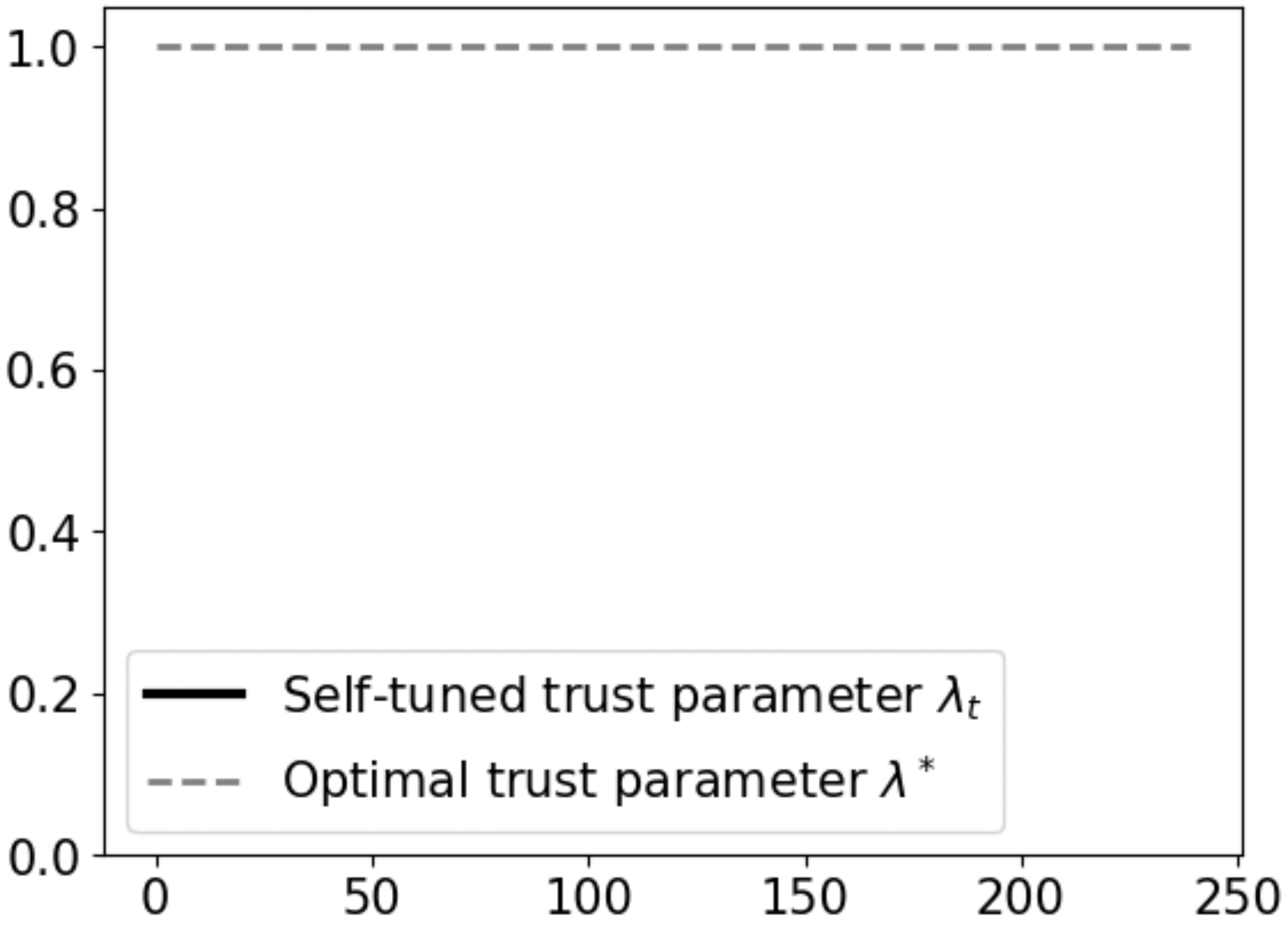
Main Results:

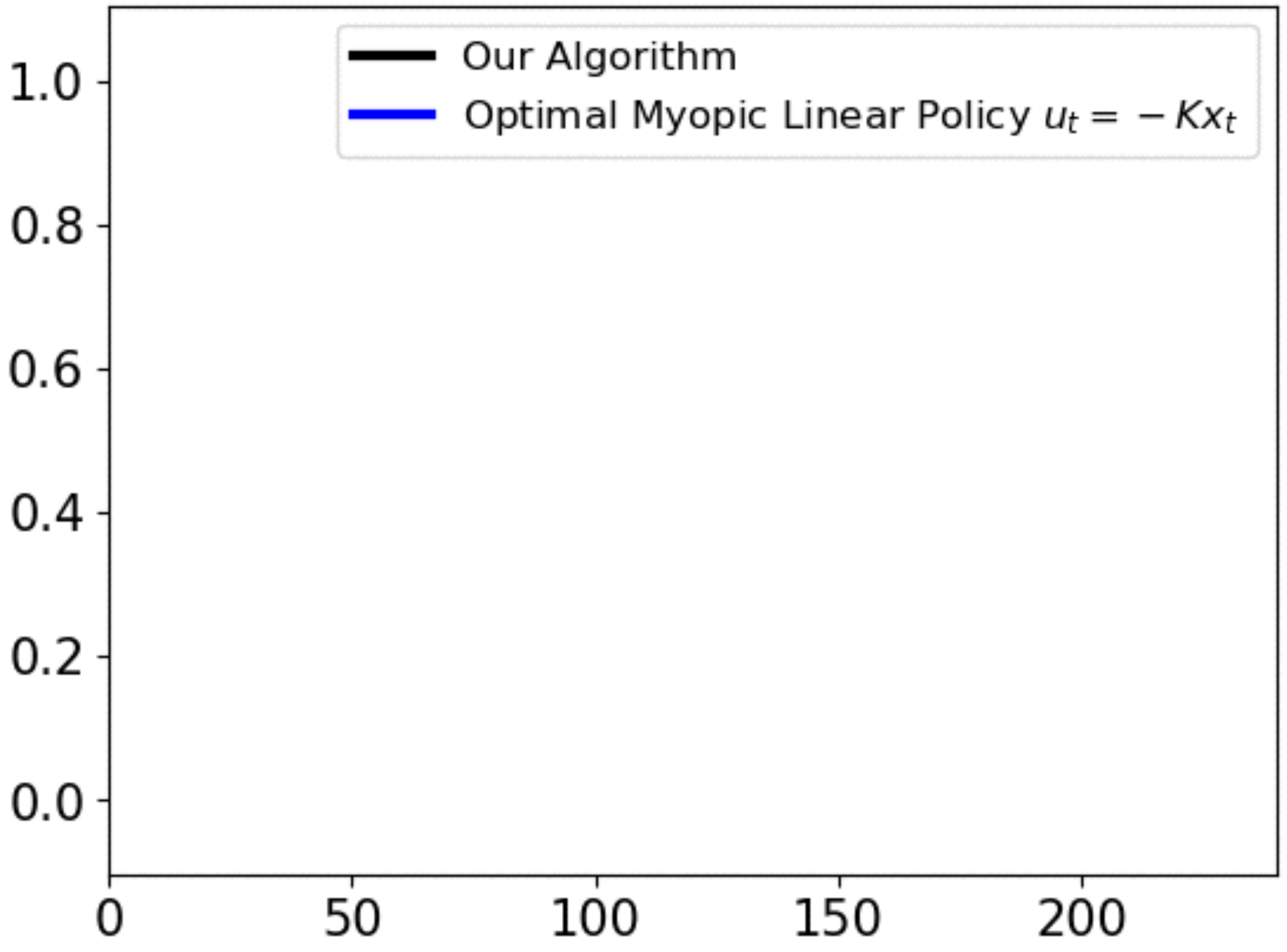
$$\mathrm{CR}(\varepsilon) \leq 1 + O(\lambda^2 \varepsilon)$$

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

Apply Our Algorithm

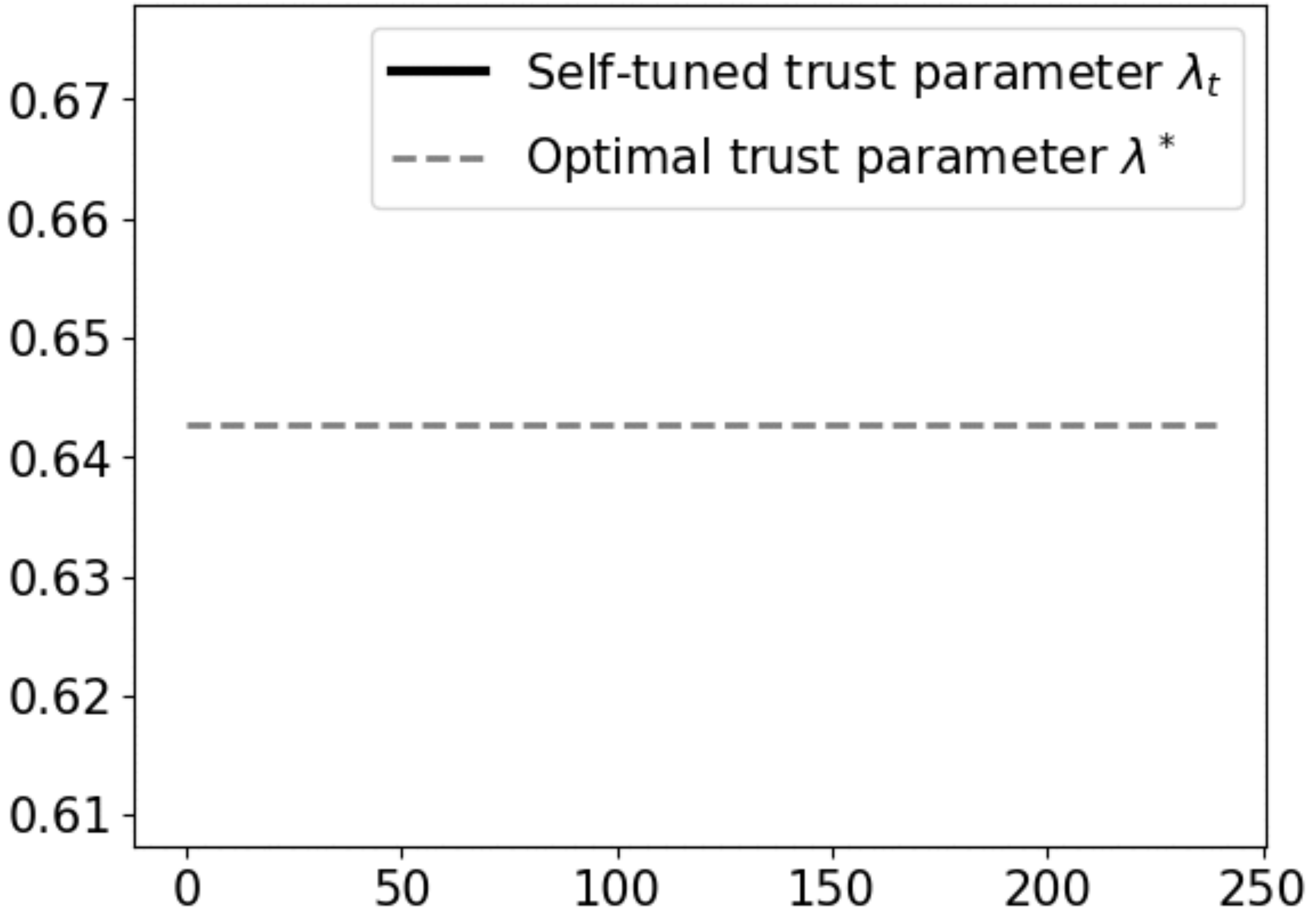
Low Error Case: Optimal $\lambda \approx 1$

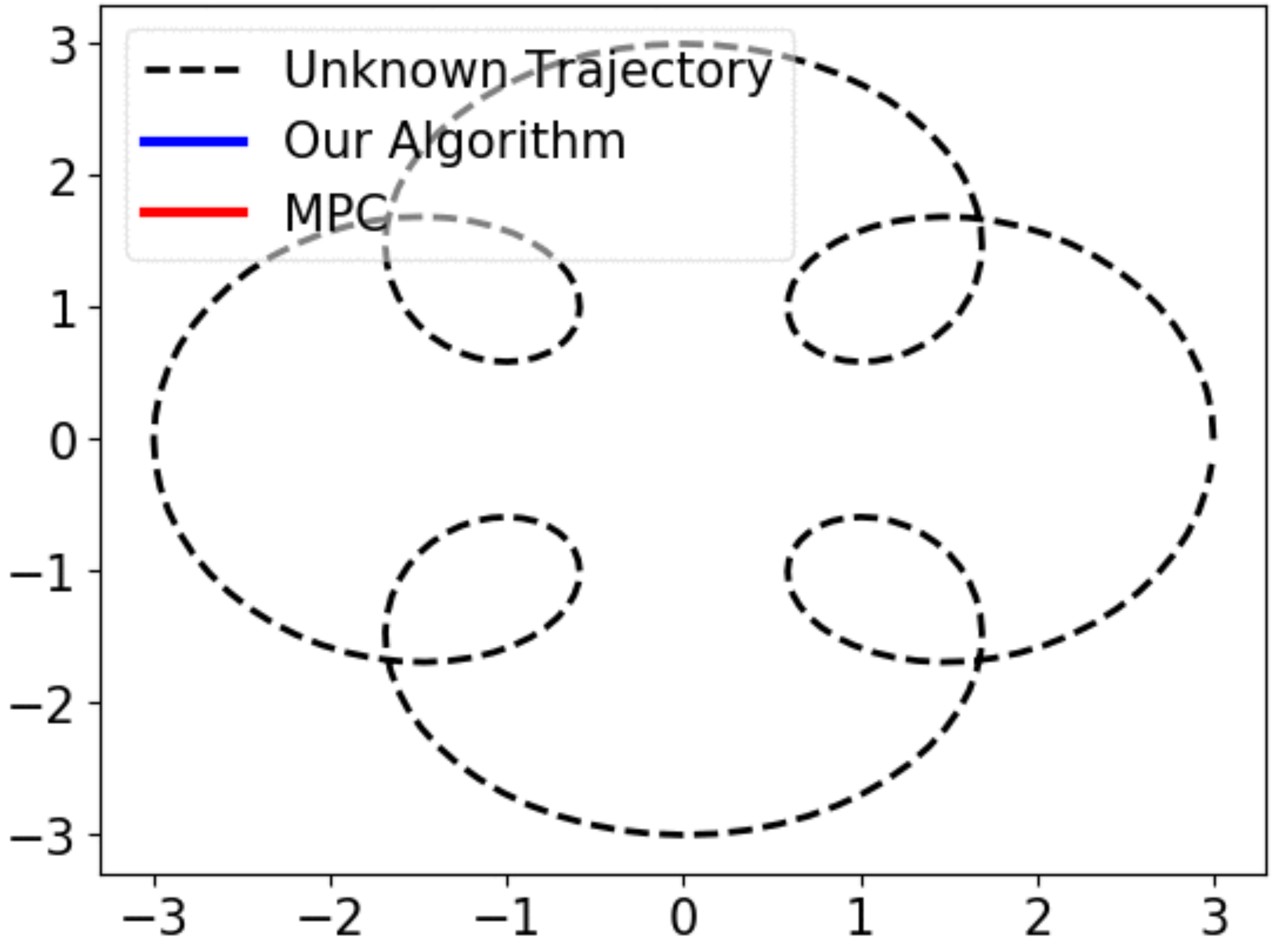




Apply Our Algorithm

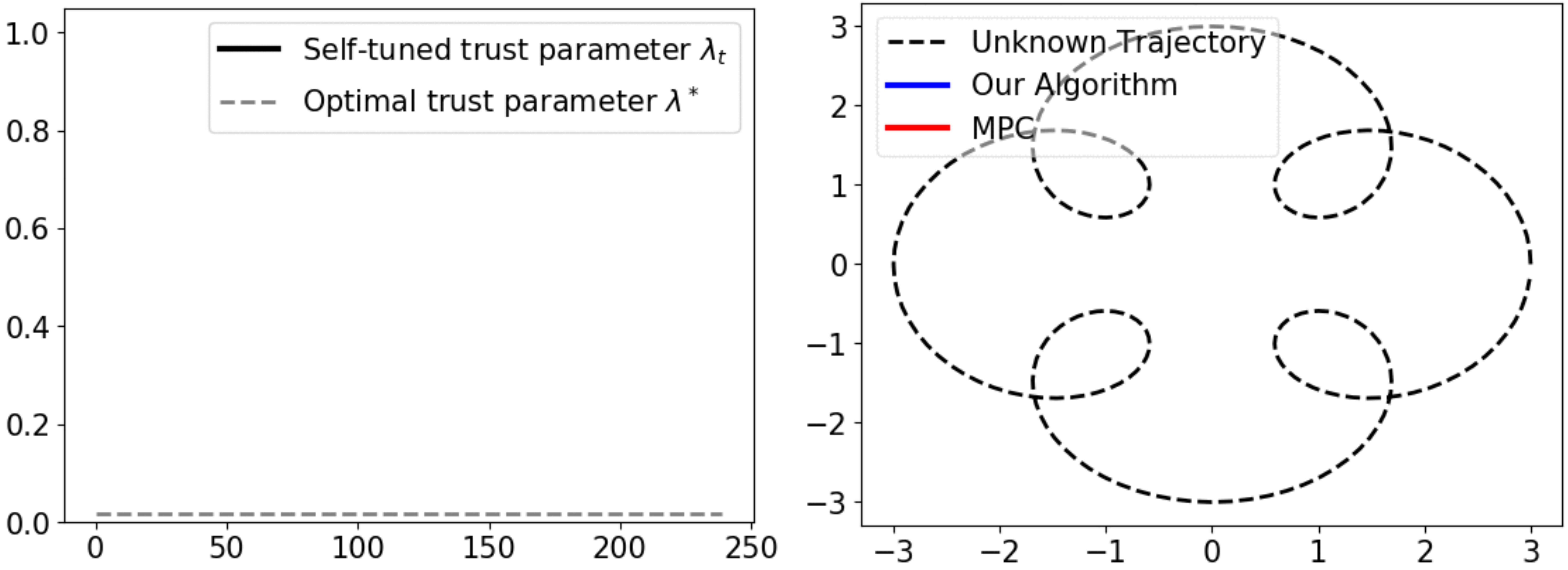
Medium Error Case: Optimal $0 < \lambda < 1$





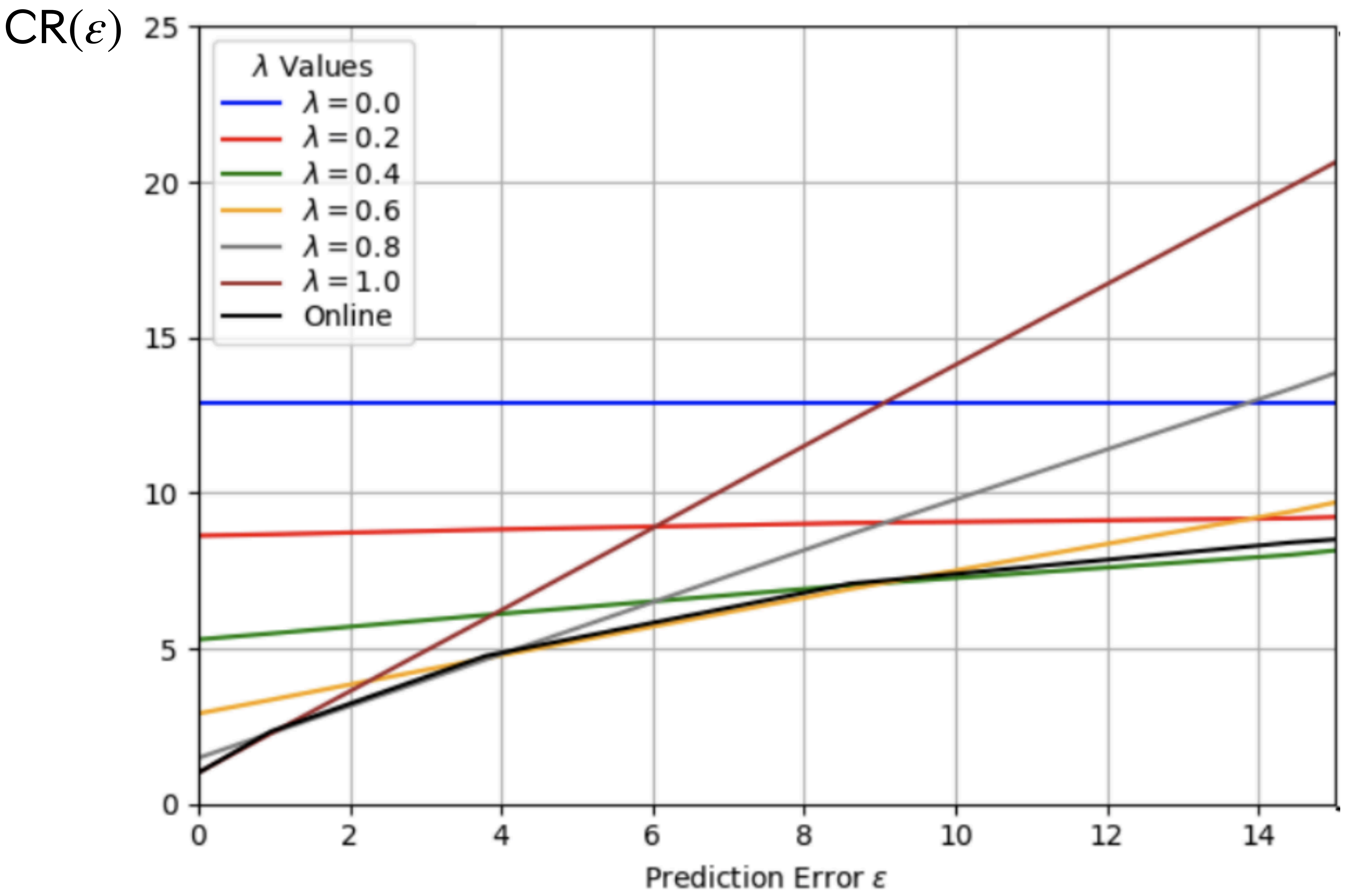
Apply Our Algorithm

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



What λ Should I Choose?

(ε is unknown) Use online learning to tune λ_t **[SIGMETRICS ’22]**



- Without online learning:

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

- With online learning:

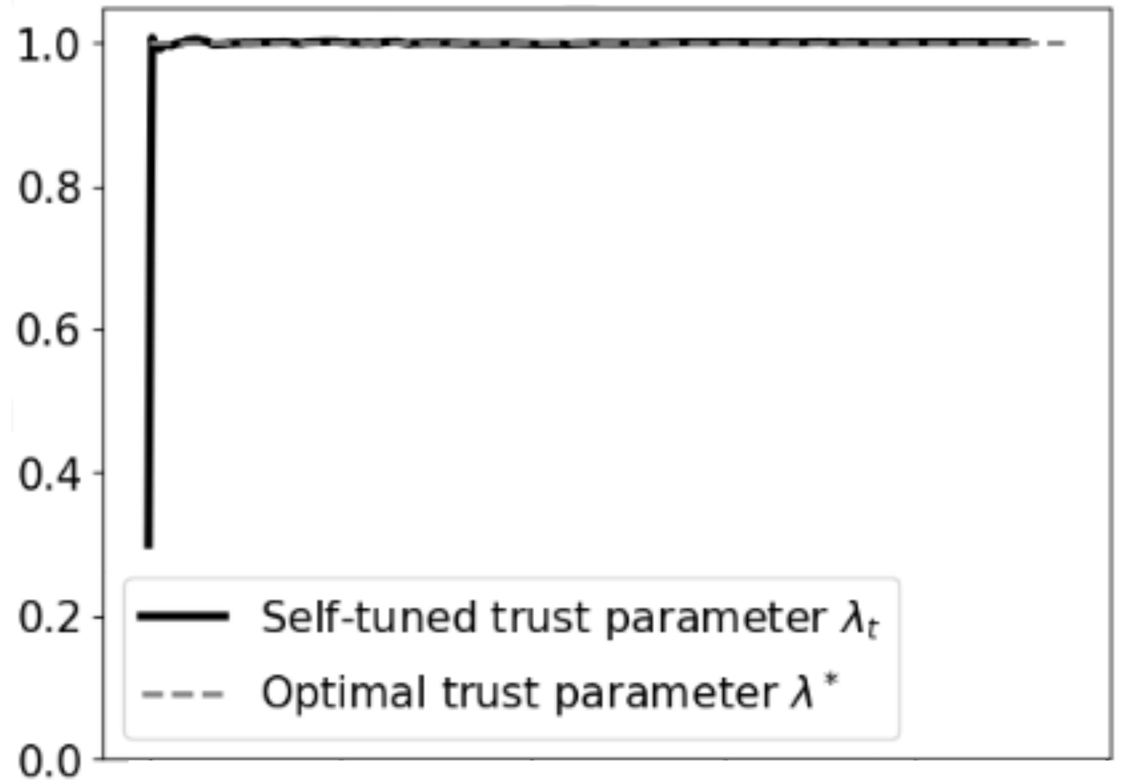
$$\text{CR}(\varepsilon) \le 1 + \frac{O(\varepsilon)}{\Theta(1) + \Theta(\varepsilon)} + \text{Variation}$$

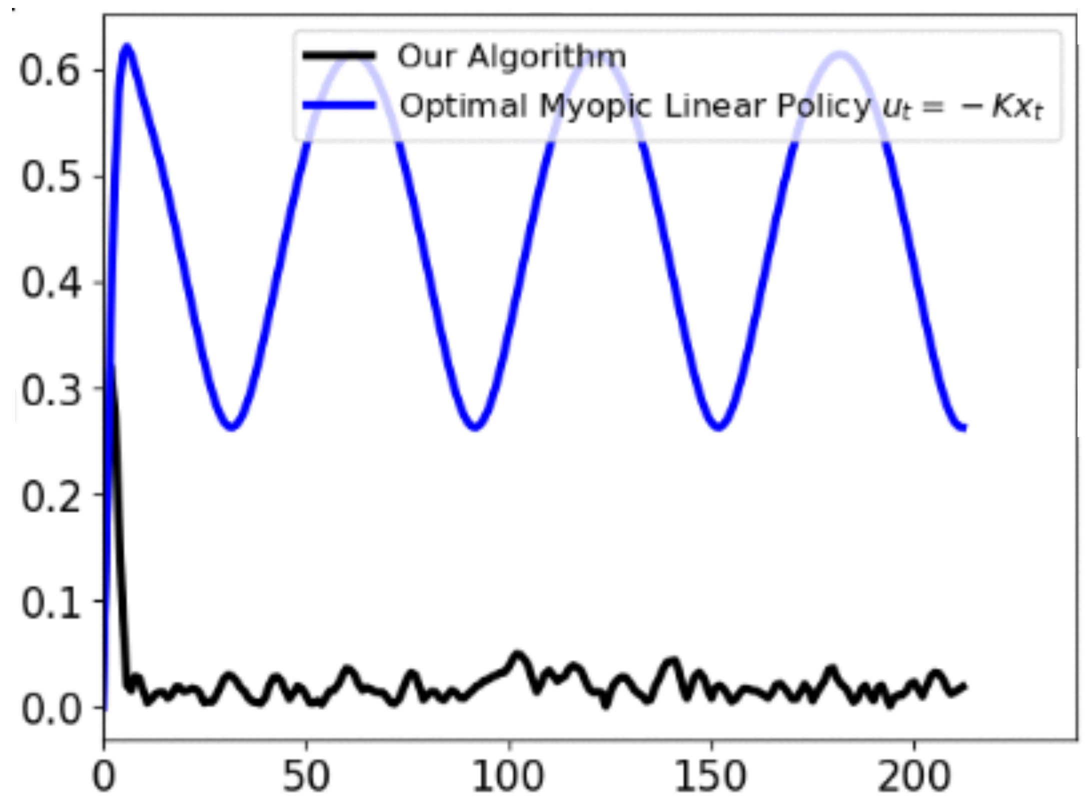
Prediction error ε is small

Prediction error ε is large

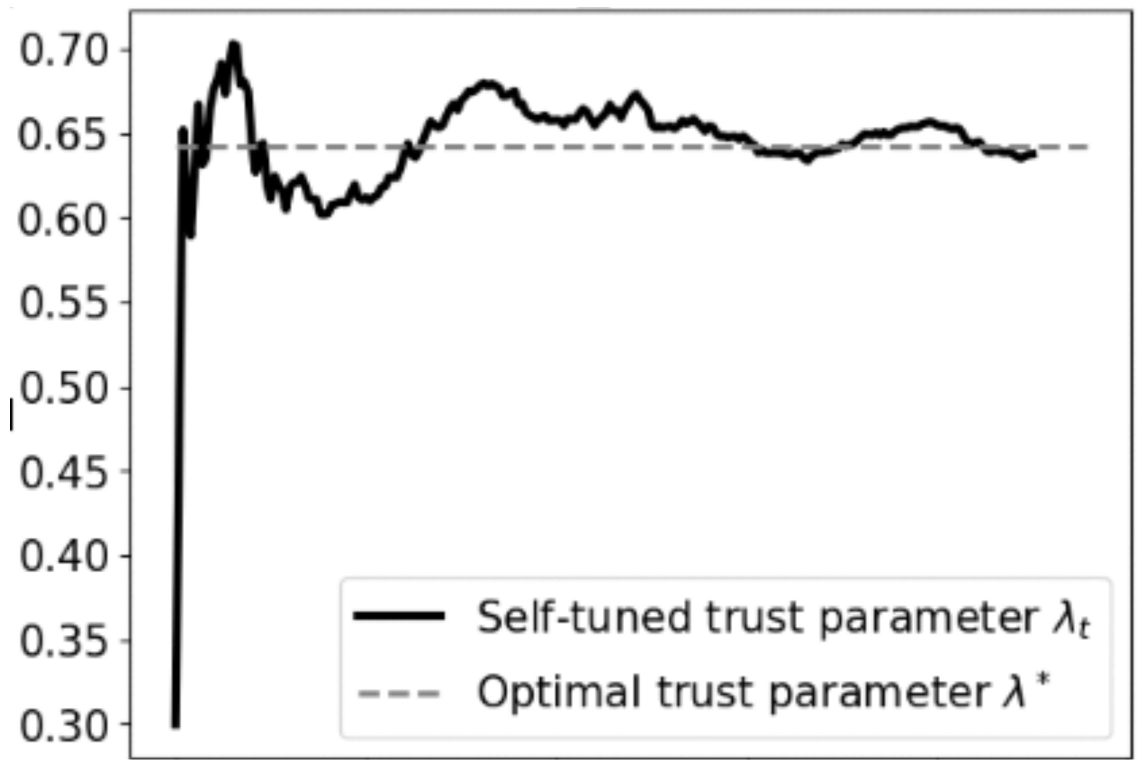
What λ Should I Choose?

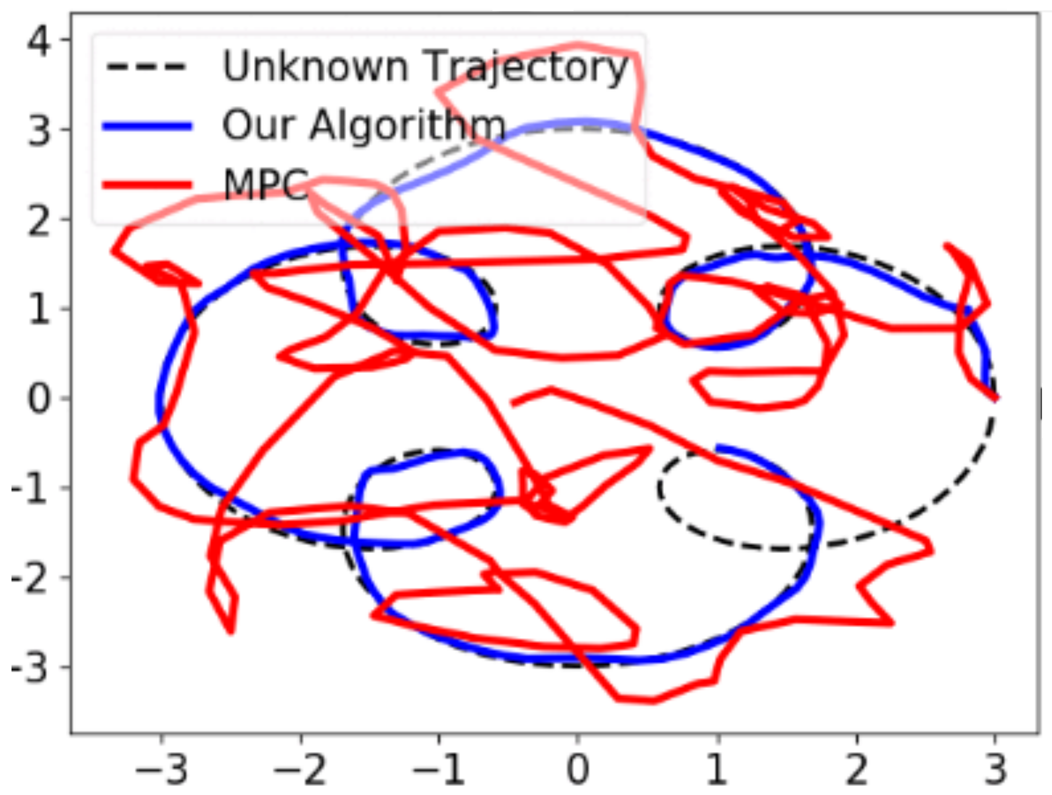
Low Error: Optimal $\lambda \approx 1$



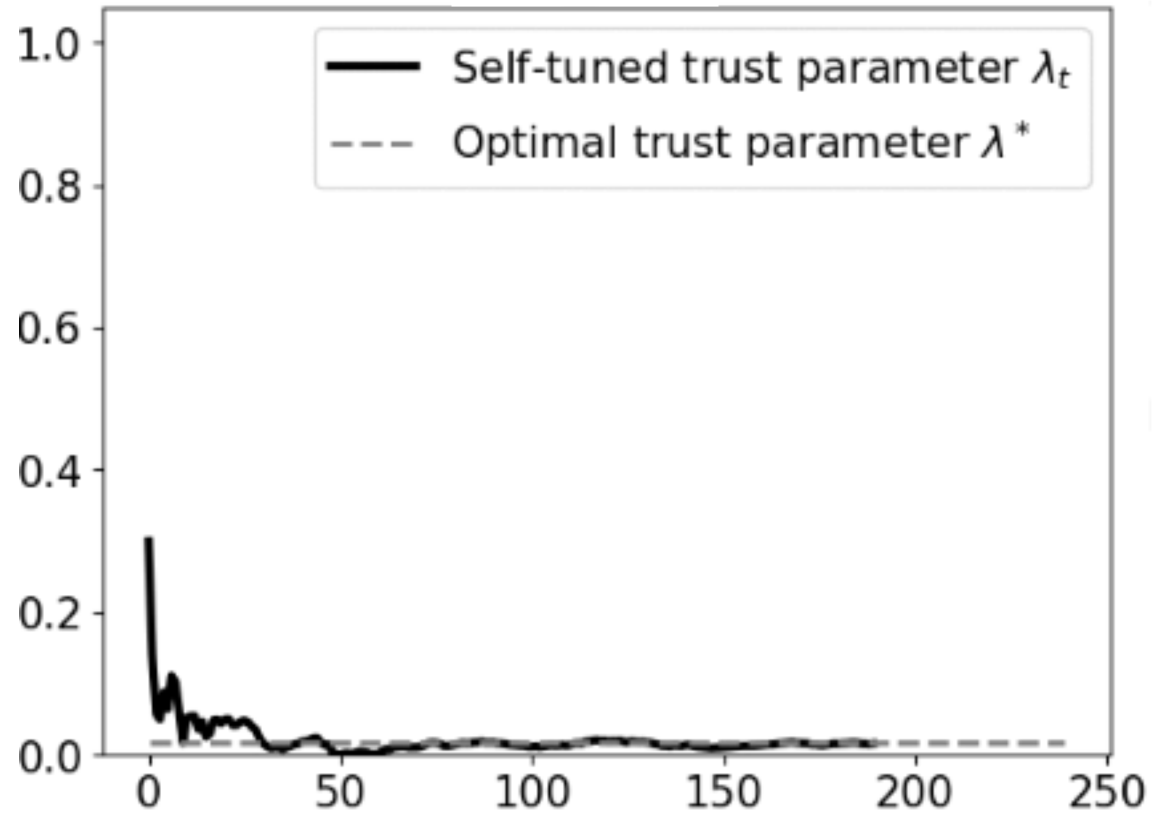


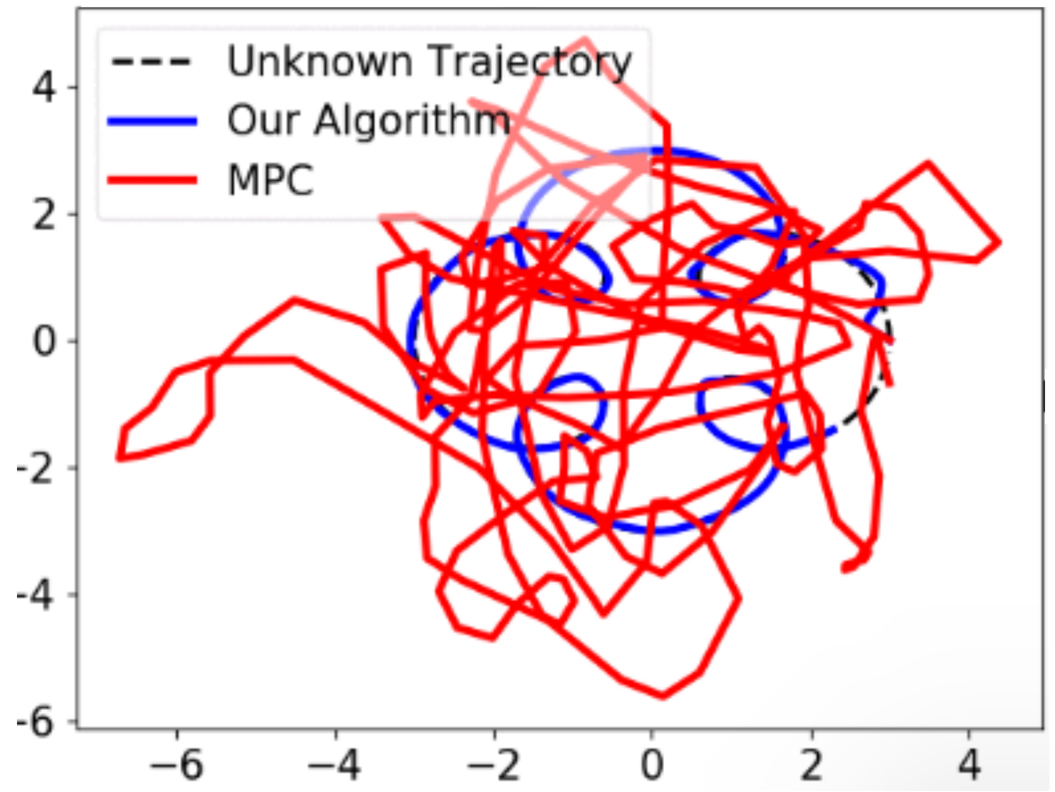
Medium Error: Optimal $0 < \lambda < 1$





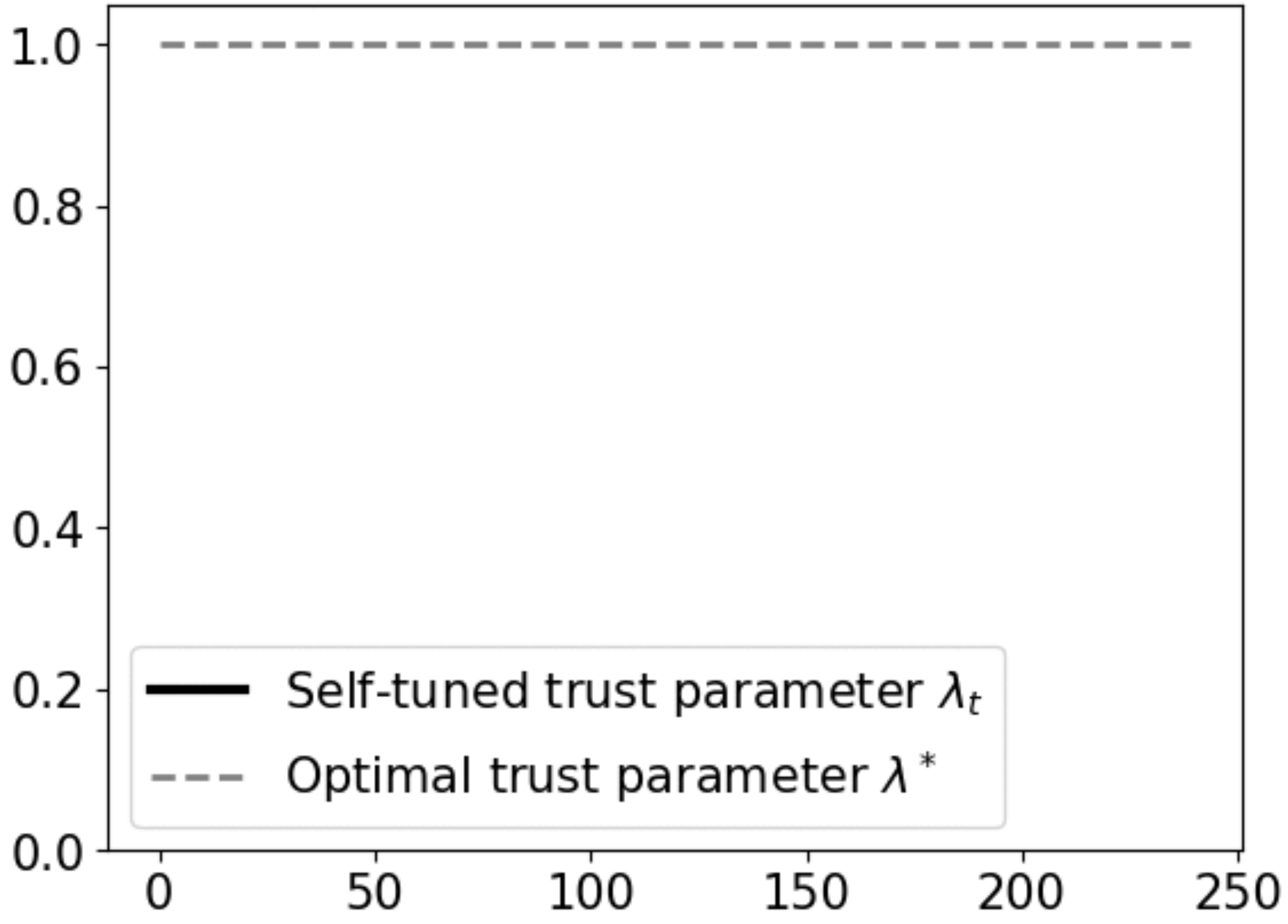
High Error: Optimal $\lambda \approx 0$

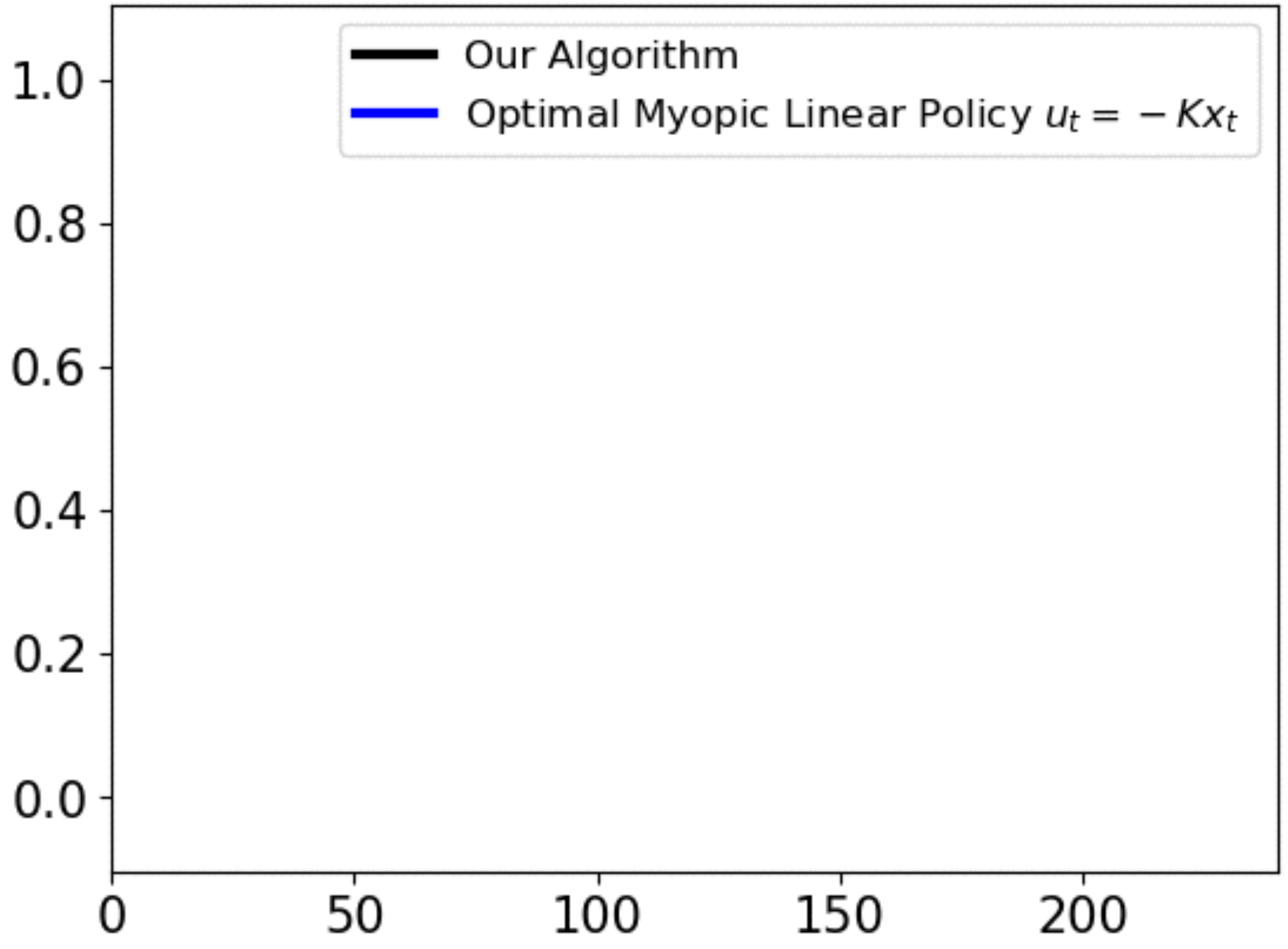




Apply Our Algorithm

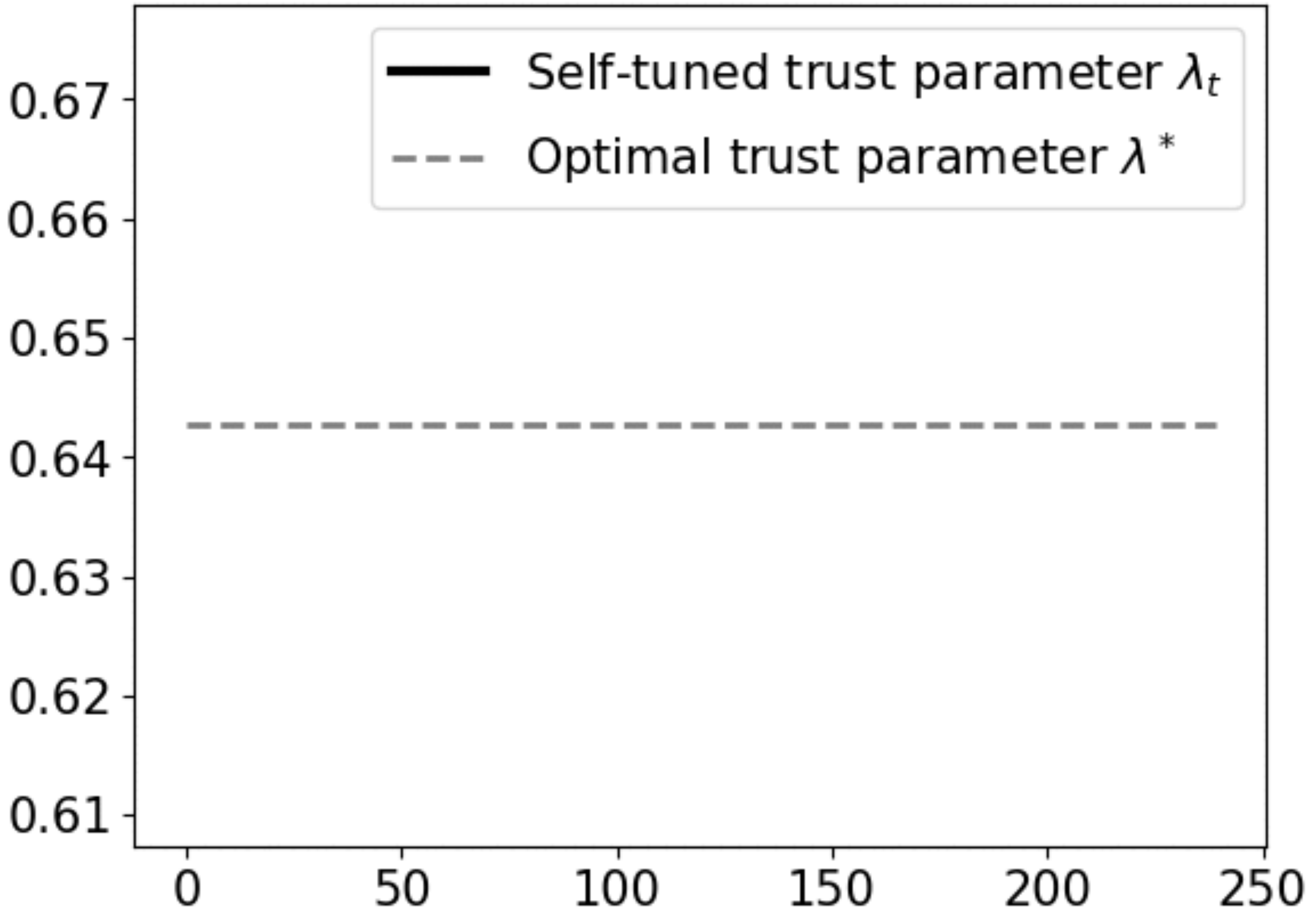
Low Error Case: Optimal $\lambda \approx 1$

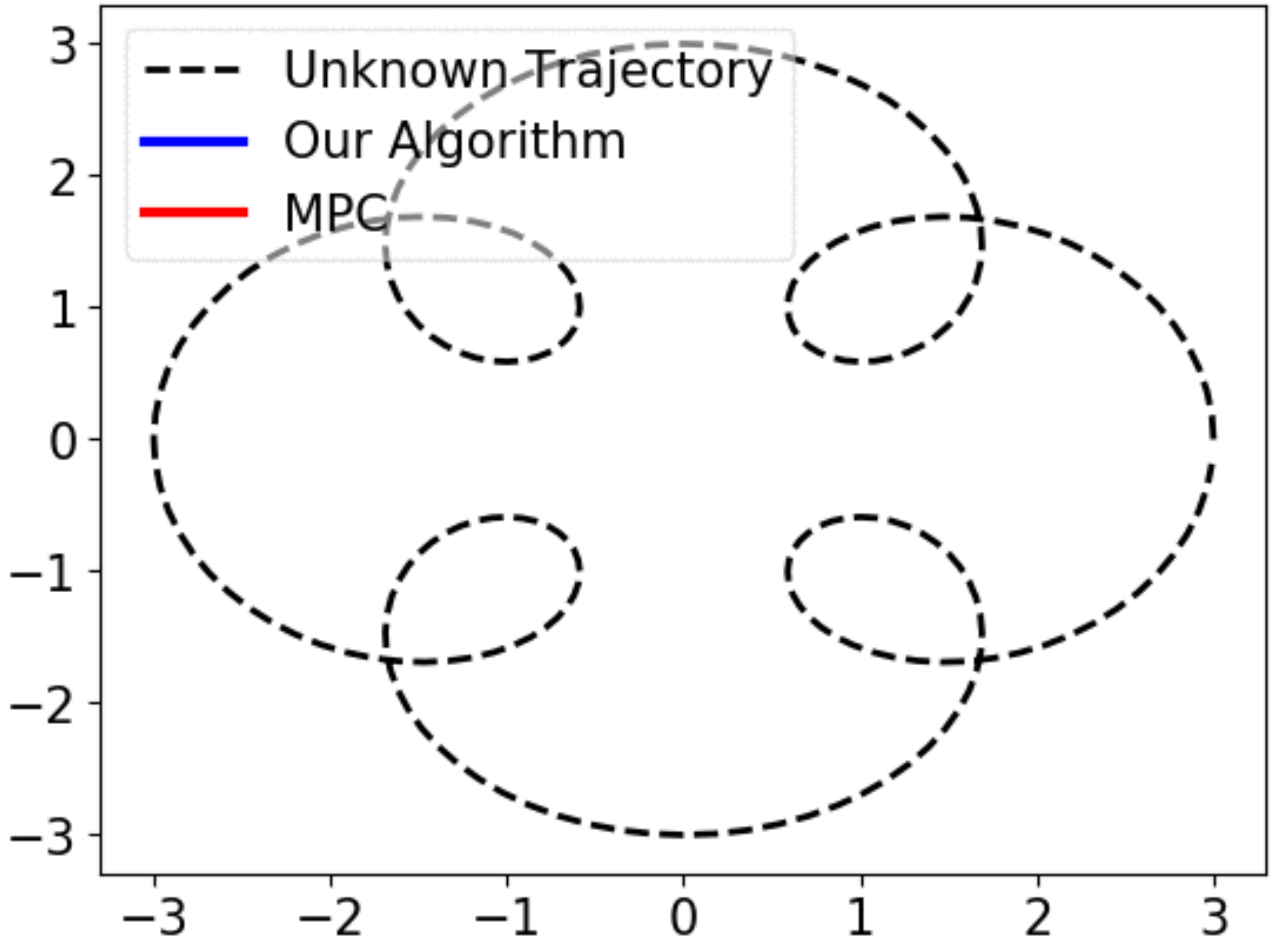




Apply Our Algorithm

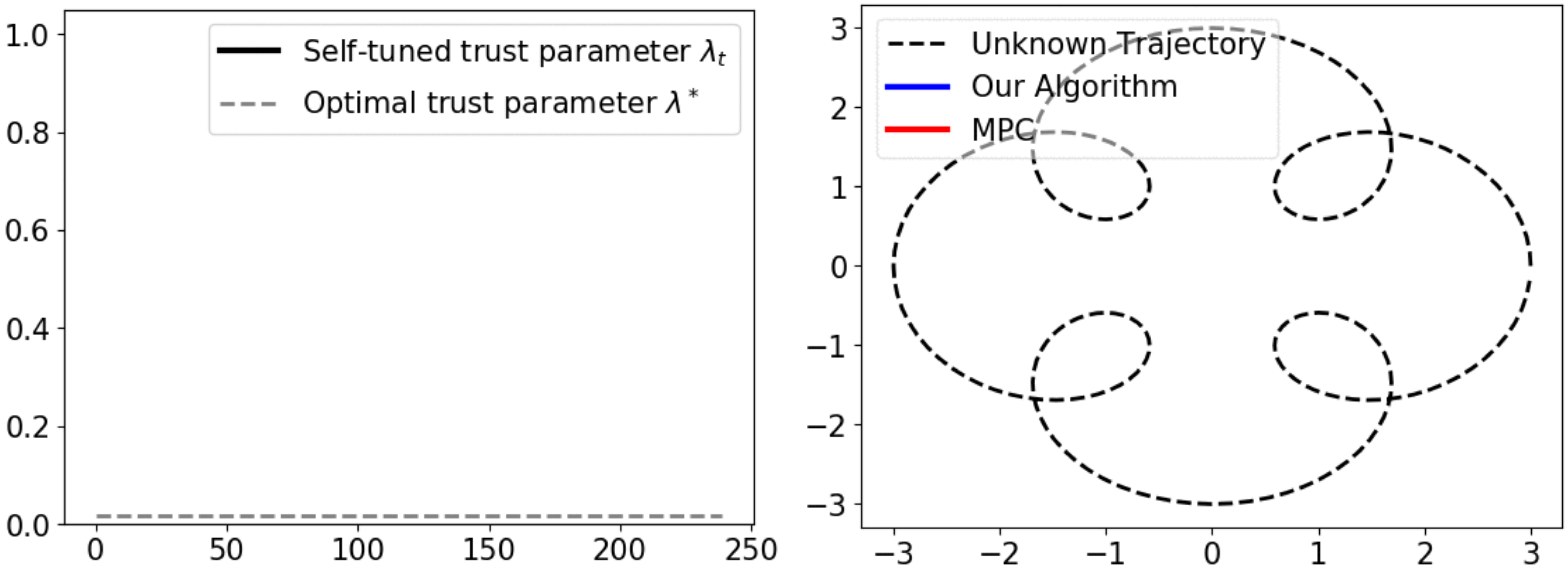
Medium Error Case: Optimal $0 < \lambda < 1$



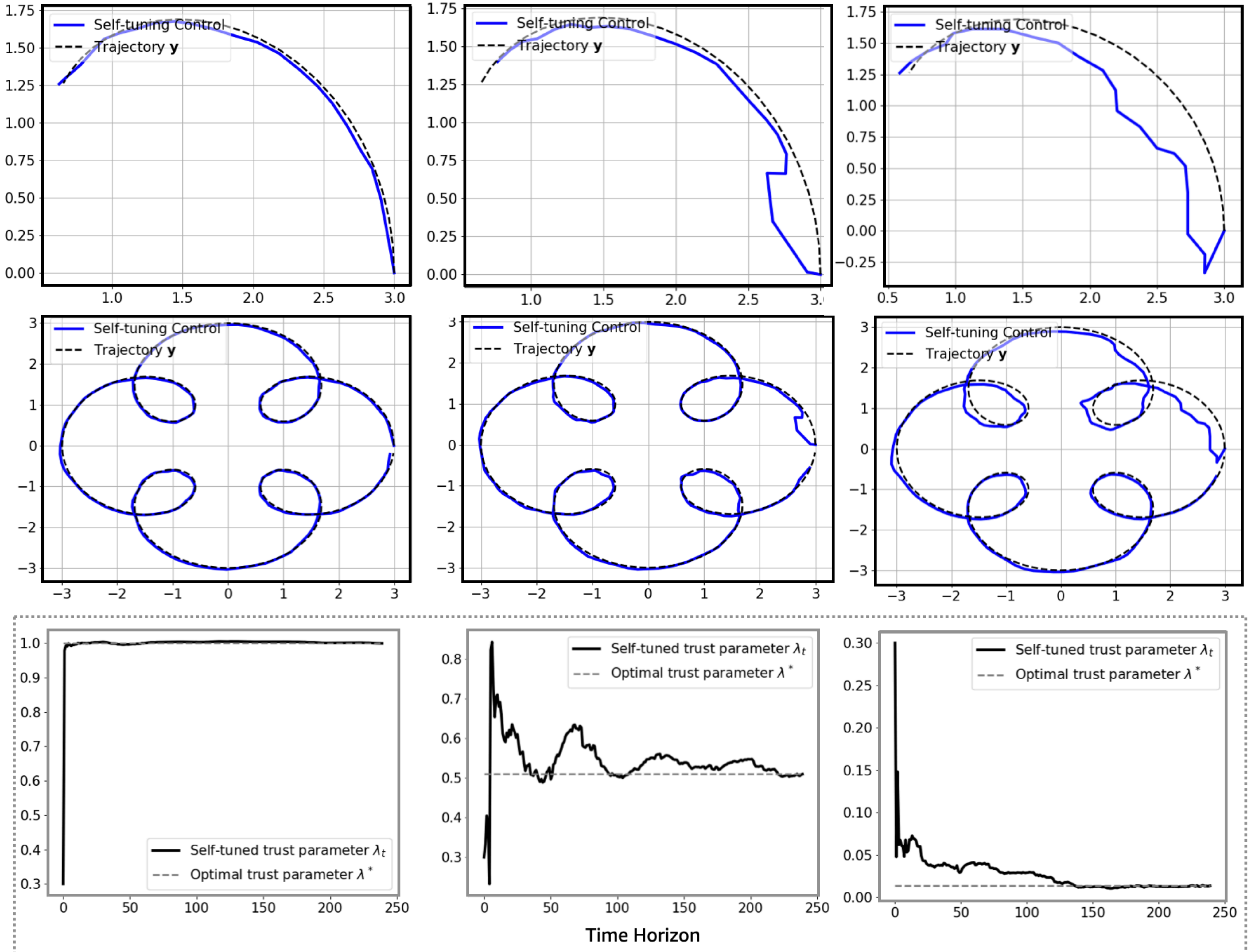


Apply Our Algorithm

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



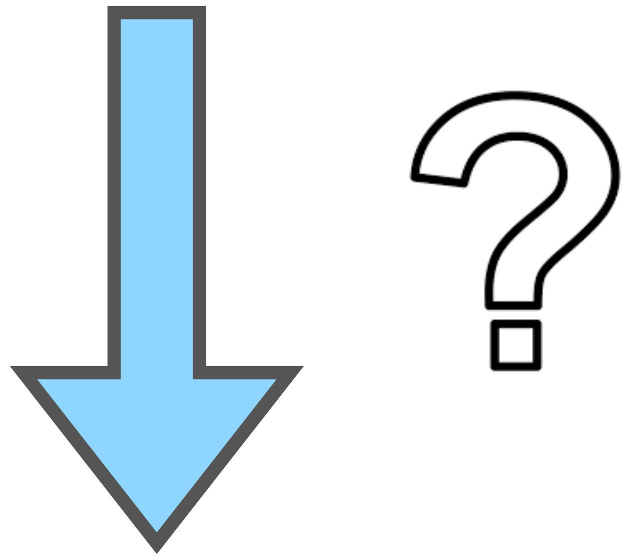
Verify the Convergence of Trust Parameters



Sketched Proof

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

λ-Confident Control



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

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

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

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

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

λ-Confident Control

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

Optimize the upper bound over λ

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

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

λ-Confident Control

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

Optimize the upper bound over λ

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

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

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

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

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

⇓

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

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

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

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

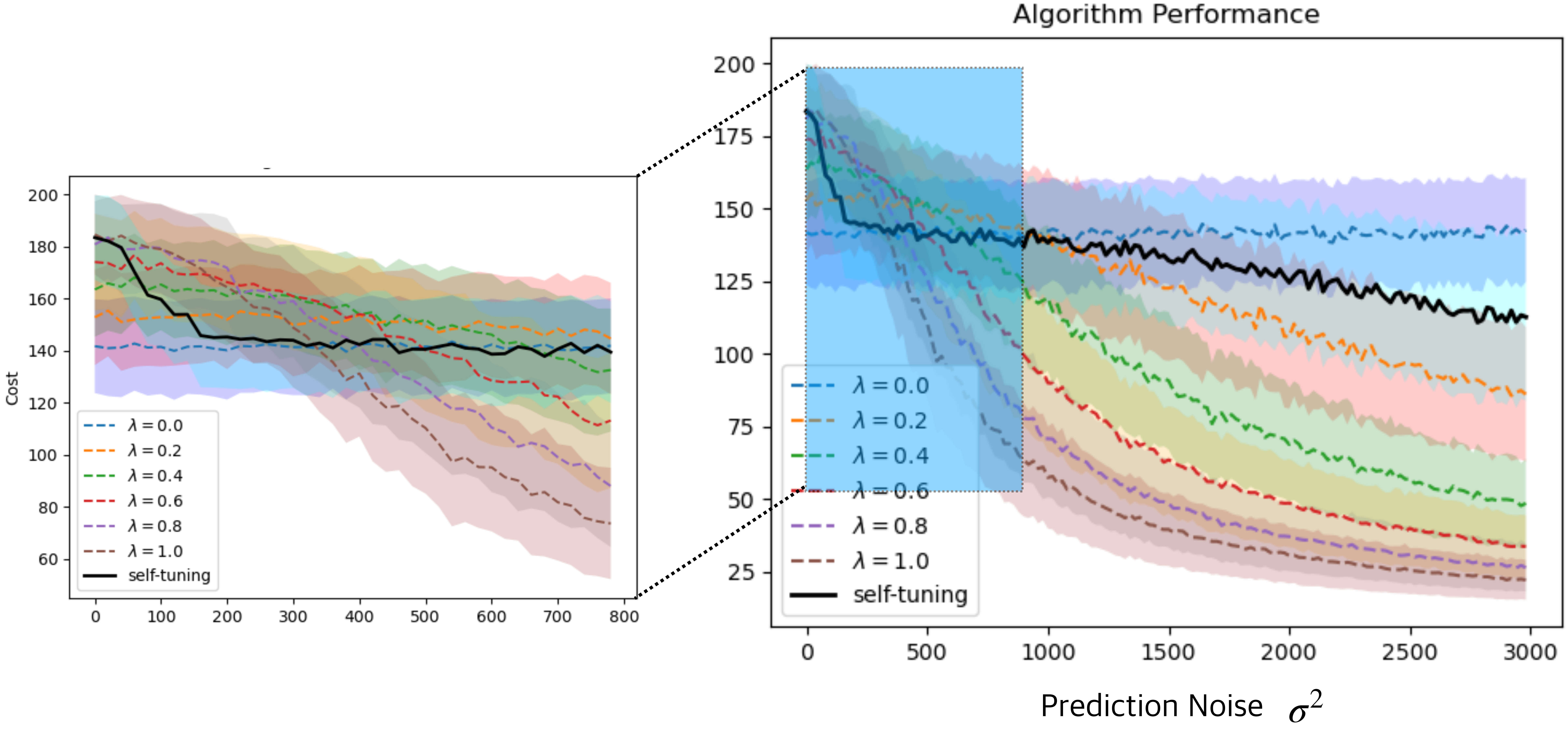
$$\text{CR}_{\text{self}}(\varepsilon) \leq 1 + 2\|H\|\frac{\varepsilon}{\text{OPT} + C\varepsilon} + O\left(\frac{\left(\mu_{\text{VAR}}(\mathbf{w}) + \mu_{\text{VAR}}(\widetilde{\mathbf{w}})\right)^2}{\text{OPT}}\right)$$

CR Theorem

Self-Tuning Control

Generalize to Nonlinear Cases

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



Tradeoff in Linear Models

System Model	Classic Agent	ML Agent	Remarks	Tradeoffs
Linear Dynamics	LQR	MPC+Perturbation Predictions	Convex Combination	Consistency vs Robustness
NonLinear Dynamics	LQR	Black-Box RL	Switching	Consistency vs Stability

Theorem (Informal; SIGMETRICS'22) Consistency vs Robustness

Under model assumptions, there exists an algorithm whose competitive ratio can be bounded by

$$\mathrm{CR}(\varepsilon) \leq 1 + 2\|H\| \frac{\varepsilon}{\mathrm{OPT} + C\varepsilon} + O\left(\text{Variation of } w, \widetilde{w}\right).$$

Nonlinear Model is Harder

System Model	Classic Agent	ML Agent	Remarks	Tradeoffs
Linear Dynamics	LQR	MPC+Perturbation Predictions	Convex Combination	Consistency vs Robustness
NonLinear Dynamics	LQR	Black-Box RL	Switching	Consistency vs Stability

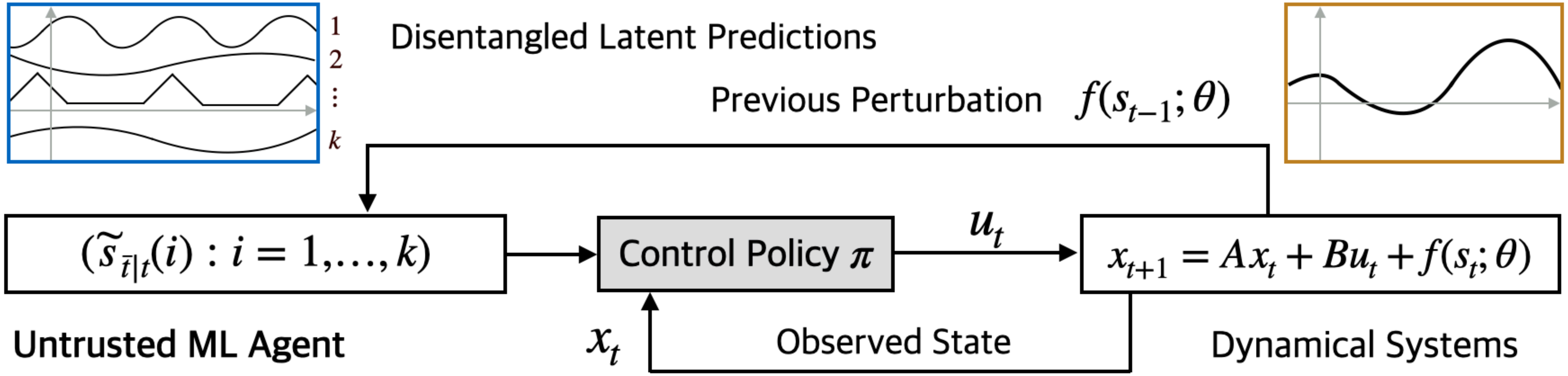
Theorem (Informal; OJCSYS ’23) **Consistency vs Stability**

Under model assumptions, there exists a policy satisfying

(1) If prediction error is smaller than a threshold, then the competitive ratio can be bounded;

(2) If prediction error is larger than that threshold, then the policy is exponentially stabilizing.

Given Disentangled Predictions in LQC ···



- Disentangling time series to obtain higher prediction accuracy (FastICA, nonlinear ICA)
- Learn to trust each independent components

[3] Joint work with Liu H, Yue Y, 2024.

Informally ⋯

Without disentangled predictions [1] ⋯

$$\mathrm{CR}(\varepsilon) \leq 1 + O\left(\frac{\varepsilon}{\underbrace{\Omega(T)}_{} + \varepsilon}\right) + O(\text{variability of } w)$$

time horizon

overall prediction error

With disentangled predictions (this work) ⋯

$$\mathrm{CR}(\varepsilon) \leq 1 + O\left(\sum_{i=1}^{k} \frac{\varepsilon(i)}{\Omega(T/w) + \varepsilon(i)}\right) + O(\rho^{2w})$$

closed-loop system spectral radius

summing over disentangled components

prediction window size

individual component prediction error

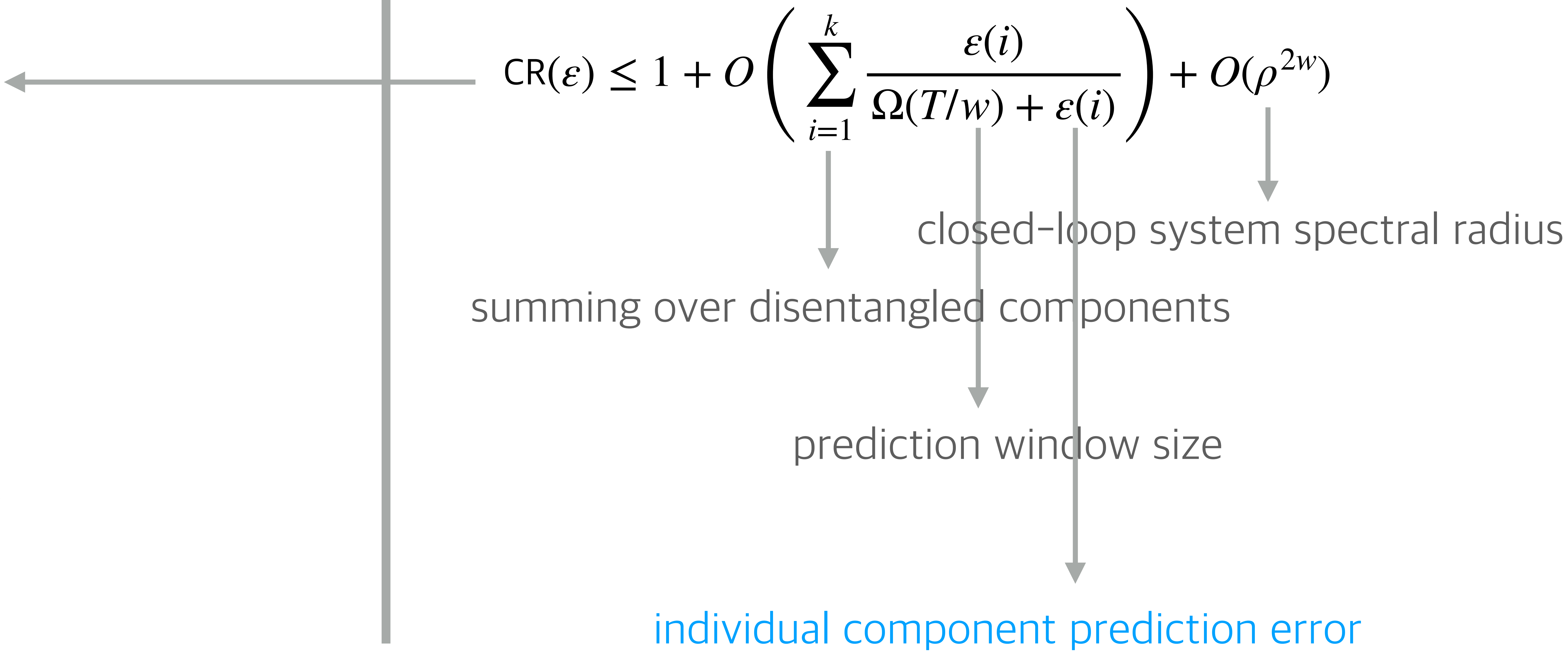
[1] Li T, Yang R, Qu G, Shi G, Yu C, Wierman A, Low S. Robustness and consistency in linear quadratic control with untrusted predictions. ACM SIGMETRICS 2022

Informally ⋯

best-of-both-worlds
utilization of
untrusted ML predictions

- If $\epsilon(i) = 0$, near-optimal
- If $\epsilon(i) = \infty$, bounded CR

With disentangled predictions (this work) ⋯



$$\mathrm{CR}(\varepsilon) \le 1 + O\left(\sum_{i=1}^{k} \frac{\varepsilon(i)}{\Omega(T/w) + \varepsilon(i)}\right) + O(\rho^{2w})$$

closed-loop system spectral radius

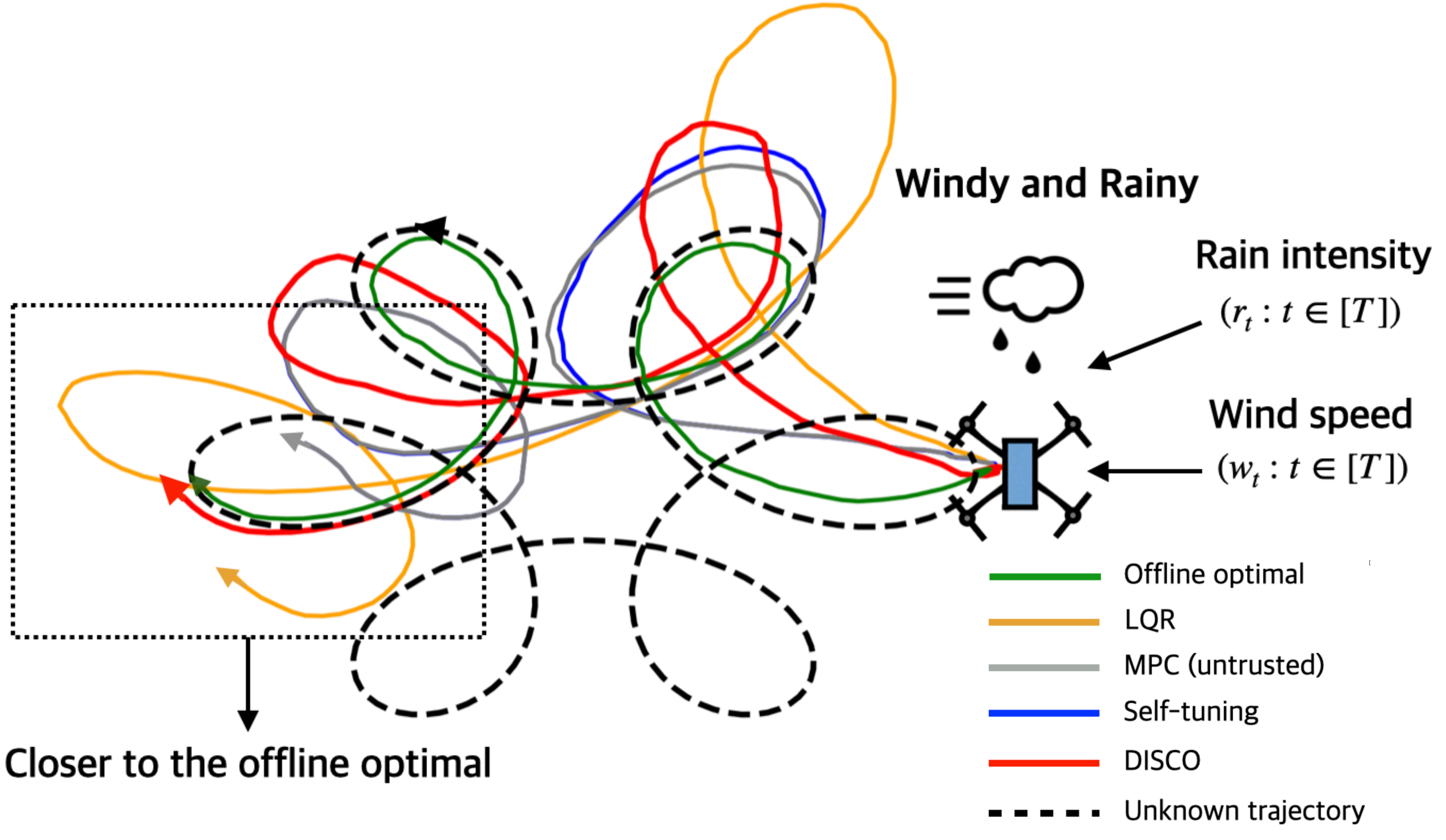
summing over disentangled components

prediction window size

individual component prediction error

Controlling a drone under challenging **windy** and **rainy** weather conditions

disentangled forces



Windy and Rainy

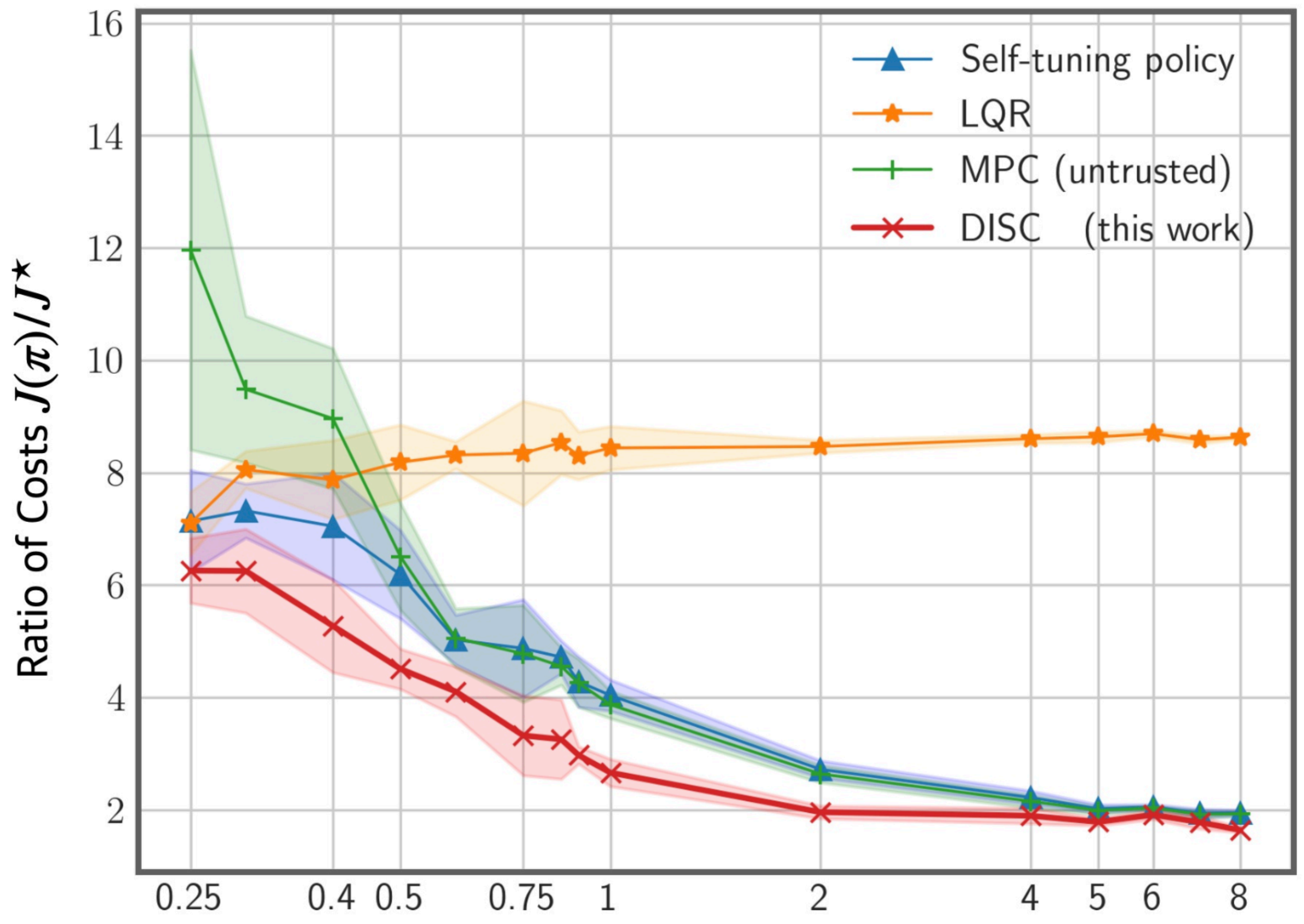
Rain intensity

$(r_t : t \in [T])$

Wind speed

$(w_t : t \in [T])$

Closer to the offline optimal



Untrusted → Trustworthy

References

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

(3) Jianyi Yang, Pengfei Li, Tongxin Li, Adam Wierman, Shaolei Ren.
"Anytime-Constrained Reinforcement Learning with Policy Prior.” (Accepted NeruIPS 2023)