Active Learning for Control-Oriented Identification

Nikolai Matni

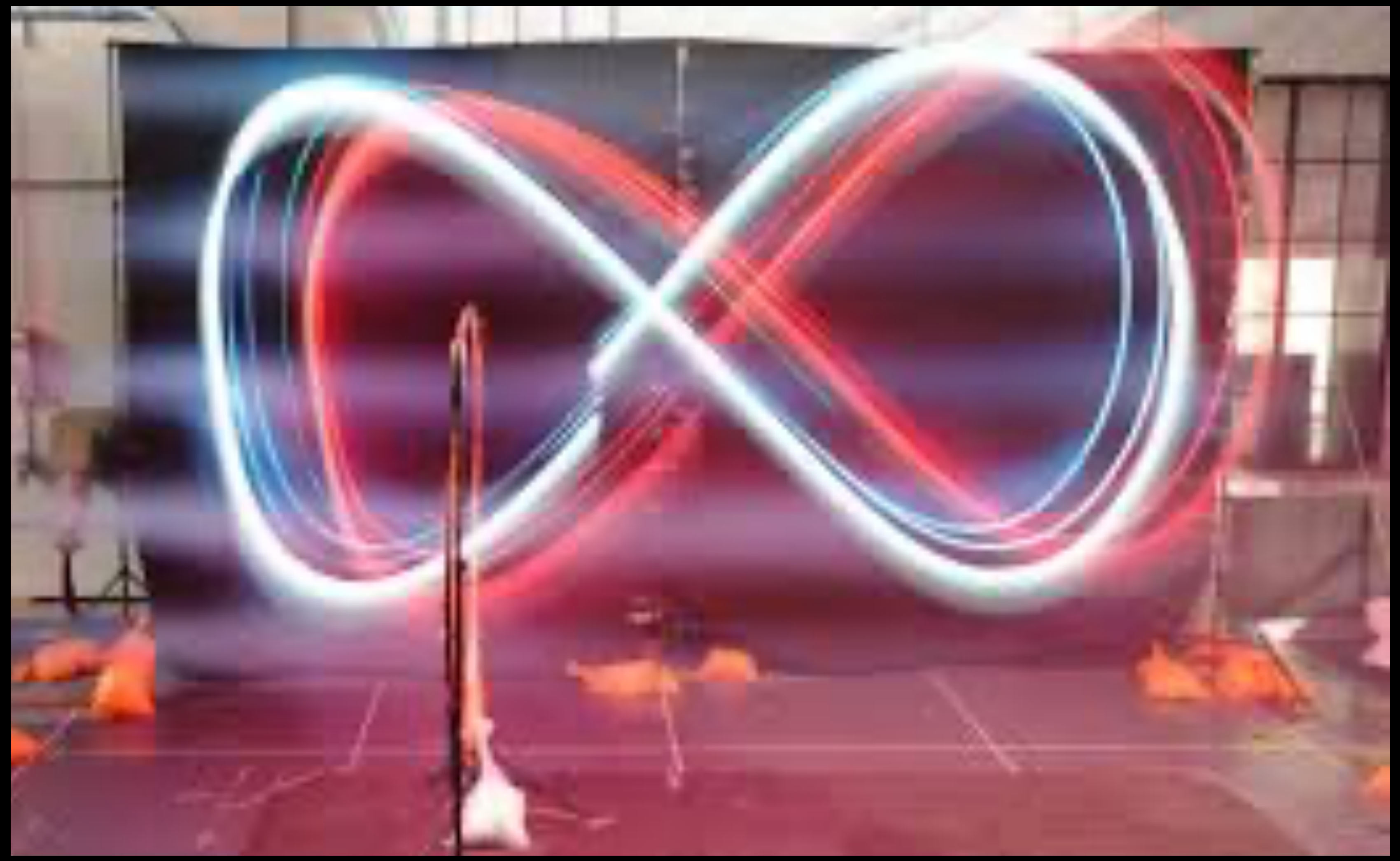
Department of Electrical and System Engineering, University of Pennsylvania

2024 AFOSR Dynamical Systems and Control Theory Review
Work supported by AFOSR YIP FA9550-24-1-0102

Penn
University *of* Pennsylvania

Experiment

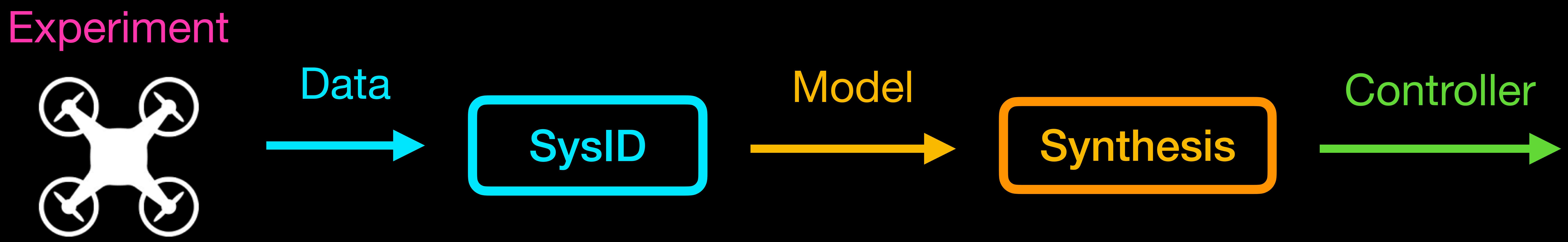
Data → **SysID** → Model → **Synthesis** → Controller →



Agile Flight in Strong Wind

Manipulation

Locomotion on Difficult Terrain



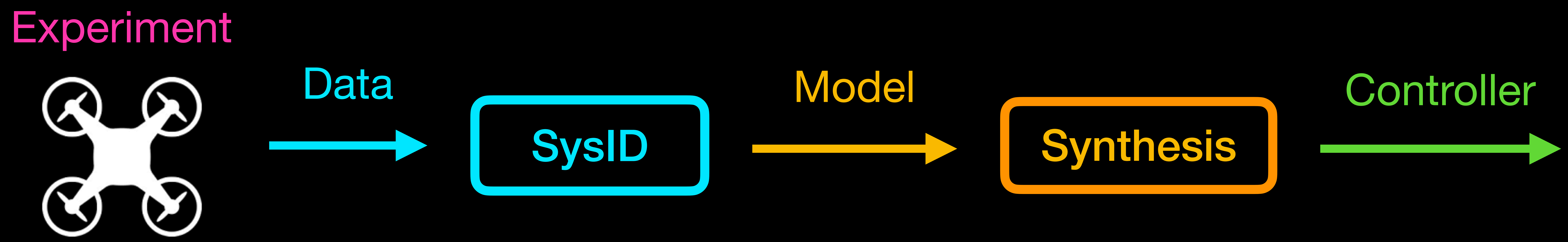
Agile Flight in Strong Wind

“Useful” data from physical systems

Expensive

Time-consuming

Dangerous



UAVs in Combat Scenarios

“Useful” data from military systems

Very Expensive

~~Time-consuming~~ Near Impossible

Very Dangerous

Statistical Learning Theory for Nonlinear Control

Thrust 1: Single trajectory learning for nonlinear control systems

- **Sharp (near mixing-free) rates** even when learning from a single rollout
- Ziemann, Tu, Pappas, **M**, *Sharp Rates in Dependent Learning Theory: Avoiding Sample Size Deflation for the Square Loss*, ICML 2024 Spotlight

Thrust 2: Representation learning for nonlinear control systems

- Use **related cheap & big data to speed up learning in expensive data settings**
- Zhang, Lee, Ziemann, Pappas, **M**, *Guarantees for Nonlinear Representation Learning: Non-identical Covariates, Dependent Data, Fewer Samples,* ICML 2024

Thrust 3: Fundamental limits of learning to control nonlinear systems

- Build systems that are **easy to learn to control**, and design **optimal learning algorithms**
- Lee, Ziemann, Pappas, **M**, *Active Learning for Control-Oriented Identification of Nonlinear Systems,* CDC 2024

Bruce Lee

Ingvar Ziemann

George J. Pappas

Experiment

Data → **SysID** → Model → **Synthesis** → Controller

UAVs in Combat Scenarios

“Useful” data from military systems

Very Expensive

~~Time-consuming~~ Near Impossible

Very Dangerous

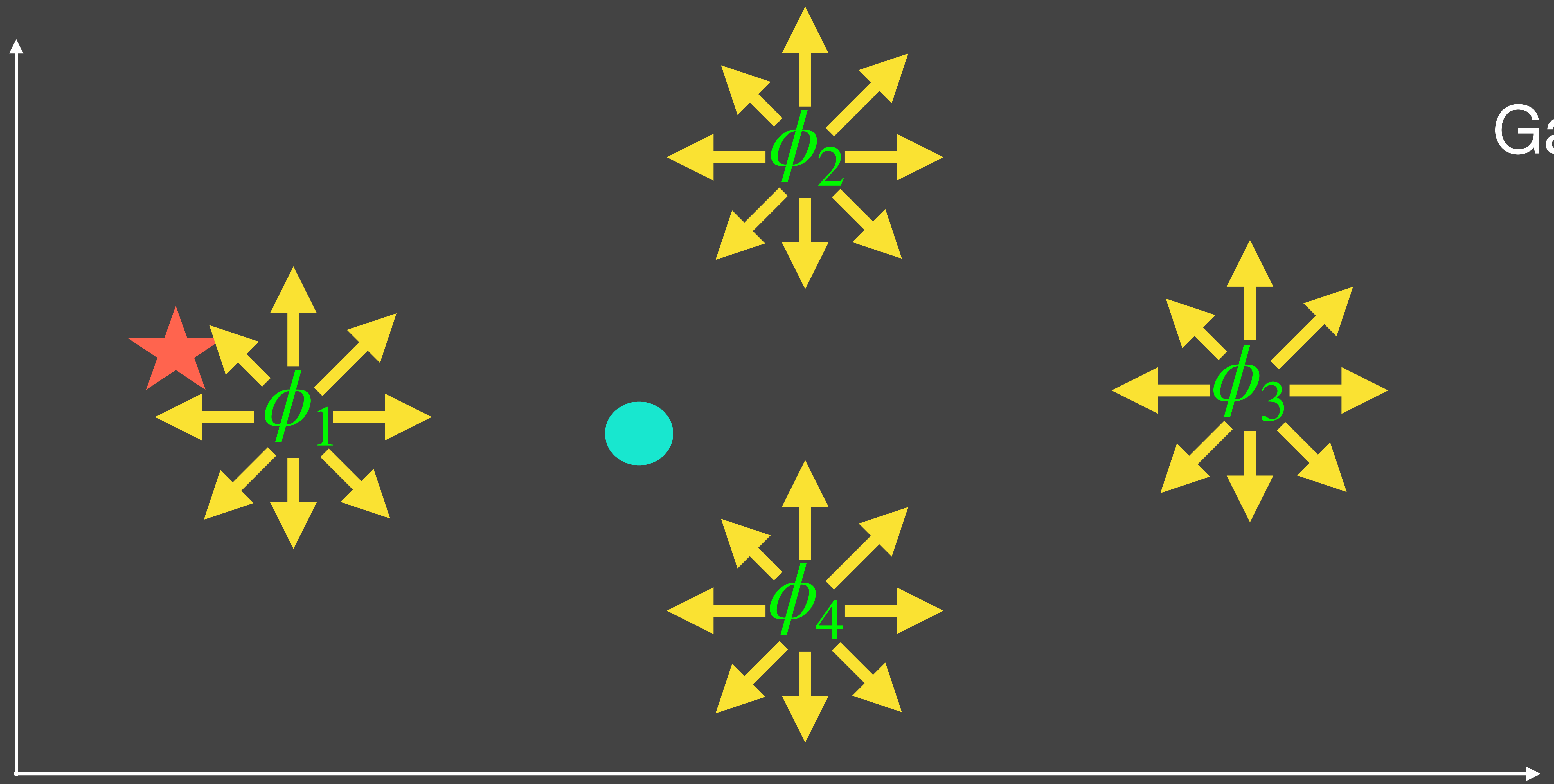
Experiments should maximize relevant information

- Relevant: characterized by the downstream control task

Example: Consider two dimensional system $X_{t+1} = X_t + U_t + W_t + \sum_{i=1}^{4} \boxed{\sigma}(X_t - \boxed{\phi_i})$

unknown centers

Gaussian kernel



Naive exploration policies are inefficient

Existing approaches only rigorously study

1) Tabular/linear/low-rank MDPs
2) Linear/kernelized dynamical systems

Otherwise apply heuristics motivated by asymptotic analysis

Objective 1: propose end-to-end pipeline for learning to control a nonlinear dynamical system

Experiments should provide **maximum information** relevant to **downstream control objective**

Objective 2: provide ***finite sample guarantees*** verifying the effectiveness of the pipeline

Outline

Problem Formulation

Motivate Algorithm

Correctness Guarantees

Numerical Validation

Outline

Problem Formulation

Motivate Algorithm

Correctness Guarantees

Numerical Validation

System dynamics: $X_{t+1} = f(X_t, U_t; \phi_\star) + W_t$

state input unknown parameters noise $\sim N(0, \sigma_W^2 I)$

Want control policy $U_t = \pi_t(X_t)$ to minimize objective:

$$J(\pi, \phi) = \mathbf{E}_\phi^\pi \left[\sum_{t=1}^{T} c_t(X_t, U_t) + c_{T+1}(X_{T+1}) \right]$$

- Expectation as X_t and U_t are RVs due to noise
- U_t from feedback policy π on X_t
- ϕ: dynamics in cost evaluation evolve under $X_{t+1} = f(X_t, U_t; \phi) + W_t$

System dynamics: $X_{t+1} = f(X_t, U_t; \phi_\star) + W_t$

Control Objective: $J(\pi, \phi) = \mathbf{E}_\phi^\pi \left[\sum_{t=1}^{T} c_t(X_t, U_t) + c_{T+1}(X_{T+1}) \right]$

Suppose $\phi_\star$ is known - we would select $\pi_\star = \arg\min_{\pi \in \Pi^\star} J(\pi, \phi_\star)$

With an estimate $\hat{\phi}$, we can use *certainty equivalence* $\pi_\star(\hat{\phi}) = \arg\min_{\pi \in \Pi^\star} J(\pi, \hat{\phi})$

System dynamics: $X_{t+1} = f(X_t, U_t; \phi_\star) + W_t$

Control Objective: $J(\pi, \phi) = \mathbf{E}_\phi^\pi \left[\sum_{t=1}^{T} c_t(X_t, U_t) + c_{T+1}(X_{T+1}) \right]$

Experiment

Data

SysID

$\hat{\phi}$

Synthesis

N experiment episodes, each run closed-loop under policy $\pi_{\text{exp}} \in \Pi_{\text{exp}}$

Episode 1: use π_{exp}^1 to collect dataset $D_1 = \{X_1^{(1)}, U_1^{(1)}, \ldots, X_T^{(1)}, U_T^{(1)}, X_{T+1}^{(1)}\}$

Episode 2: use π_{exp}^2 $\quad D_2$

⋮

Episode N: use π_{exp}^N $\quad D_N$

Use least squares over data: $\hat{\phi} = \arg\min_{\phi} \sum_{(X,U,X^+)\in\text{Data}} \left\| X^+ - f(X, U; \phi) \right\|^2$

Certainty equivalence: $\pi_\star(\hat{\phi}) = \arg\min_{\pi \in \Pi^\star} J(\pi, \hat{\phi})$

Central question: which experiments should we perform to achieve a controller with low cost?

Related Work

- **“Classical” Control Oriented Identification:** Gevers 1993, Hjalmarsson, et. al, 1996, many more. **Importance of closed-loop identification**

- **Active learning + finite sample guarantees:** Mania et al 2020, Wagenmaker et al 2021, Wagenmaker et al 2023. **Linear in the parameters, algorithmically complex**

We provide finite sample guarantees for
general nonlinear models via a simple and interpretable algorithm

Outline

Problem Formulation

Motivate Algorithm

Correctness Guarantees

Numerical Validation

Outline

Problem Formulation

Motivate Algorithm

Correctness Guarantees

Numerical Validation

Excess cost $J(\pi_\star(\hat{\phi}), \phi_\star) - J(\pi_\star(\phi_\star), \phi_\star) \approx (\hat{\phi} - \phi_\star)^\top H(\phi_\star)(\hat{\phi} - \phi_\star)$

$H(\phi_\star)$ is the Hessian of $J(\pi_\star(\hat{\phi}), \phi_\star)$ evaluated at $\phi_\star$

Wagenmaker et. al, 2023

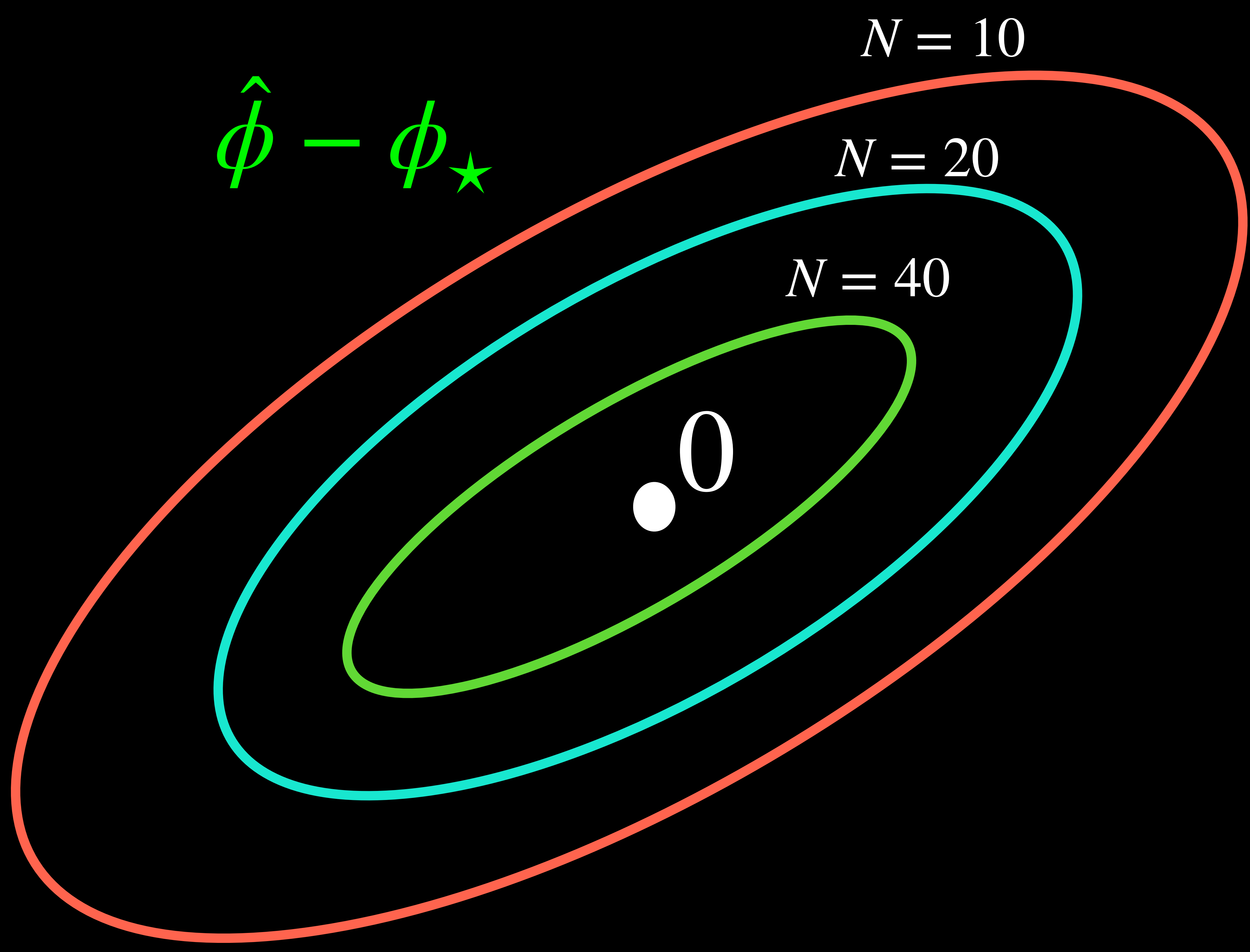
$\Longrightarrow$ Excess cost characterized by weighted SysID error

Let $\hat{\phi}$ be the least squares solution:

$$\hat{\phi} = \arg\min_{\phi} \sum_{X, U, X_+ \in \mathcal{D}_1 \cup \ldots \cup \mathcal{D}_N} \|f(X, U; \phi) - X_+\|^2$$

Then $\lim_{N \to \infty} \sqrt{N}(\hat{\phi} - \phi_\star) \sim \mathcal{N}(0, \mathrm{FI}^{\pi_{\exp}}(\phi_\star)^{-1})$

$$\mathrm{FI}^{\pi}(\phi) = \frac{\mathbf{E}_{\phi}^{\pi}\left[\sum_{t=1}^{T} Df(X_t, U_t; \phi) Df(X_t, U_t; \phi)^\top\right]}{\sigma_W^2}$$



Consider setting $\hat{\phi}$ as LS estimate using datasets $D_1, \ldots, D_N$ collected with π_{exp}

Excess cost $J(\pi_\star(\hat{\phi}), \phi_\star) - J(\pi_\star(\phi_\star), \phi_\star) \approx \dfrac{\mathrm{Tr}(H(\phi_\star)\mathrm{FI}^{\pi_{\text{exp}}}(\phi_\star)^{-1})}{N}$

$H(\phi)$ is the Hessian of $J(\pi_\star(\phi), \phi)$

$$\mathrm{FI}^{\pi}(\phi) = \frac{\mathbf{E}_{\phi}^{\pi}\left[\sum_{t=1}^{T} D_\phi f(X_t, U_t; \phi) D_\phi f(X_t, U_t; \phi)^\top\right]}{\sigma_W^2}$$

Minimize the quantity $\mathrm{Tr}(H(\phi_\star)\mathrm{FI}^{\pi}(\phi_\star)^{-1})$ over π

$H(\phi_\star)\mathrm{FI}^{\pi_{\text{exp}}}(\phi_\star)^{-1}$ depends on unknown $\phi_\star \implies$ replace with coarse est. $\hat{\phi}^{-}$

ALCOI: Given exploration budget of N episodes and an initial exploration policy

Naive Experiments

Play initial exploration policy for $N/2$ episodes to collect data $\mathscr{D}^- = D_1 \cup \ldots D_{N/2}$

SysID 1

Set $\hat{\phi}^-$ as least squares solution using data $\mathscr{D}^-$

Targeted Experiments

$$\pi_{\exp} = \arg \min_{\pi \in \Pi_{\exp}} \mathrm{Tr}(H(\hat{\phi}^-)\mathrm{FI}^{\pi_{\exp}}(\hat{\phi}^-)^{-1})$$

Play $\pi_{\exp}$ for $N/2$ episodes to collect data $\mathscr{D}^+ = D_{N/2+1} \cup \ldots \cup D_N$

SysID 2

Set $\hat{\phi}^+$ as least squares solution using data $\mathscr{D}^+$

Synthesis

Return certainty equivalent policy $\pi_\star(\hat{\phi}^+)$

Outline

Problem Formulation

Motivate Algorithm

Correctness Guarantees

Numerical Validation

Outline

Problem Formulation

Motivate Algorithm

Correctness Guarantees

Numerical Validation

Main Theorem (informal) Suppose N is sufficiently large. Then with probability at least $1-\delta$, the algorithm's output satisfies*

$$J(\hat{\pi}, \phi_\star) - J(\pi_\star, \phi_\star) \leq C \log\frac{1}{\delta} \min_{\pi \in \Pi_{\mathrm{exp}}} \frac{\mathrm{Tr}(H(\phi_\star)\mathrm{FI}^{\pi}(\phi_\star)^{-1})}{N}$$

*Under smoothness assumptions on the dynamics and policy class, and identifiability conditions for the parameters

Proof consists of novel identification error bound arising from the "**delta method**" (clever Taylor expansion) along with "**learning with little mixing**"

Main Theorem (informal) Suppose N is sufficiently large. Then with probability at least $1-\delta$, the algorithm's output satisfies*

$$J(\hat{\pi}, \phi_\star) - J(\pi_\star, \phi_\star) \leq C \log\frac{1}{\delta} \min_{\pi \in \Pi_{\text{exp}}} \frac{\text{Tr}(H(\phi_\star)\text{FI}^\pi(\phi_\star)^{-1})}{N}$$

Remark 1: Tight up to the log term when f is linear in ϕ - probably in general?

Remark 2: Bound captures interplay of hardness of control with hardness of identification divided by amount of data

Outline

Problem Formulation

Motivate Algorithm

Correctness Guarantees

Numerical Validation

Outline

Problem Formulation

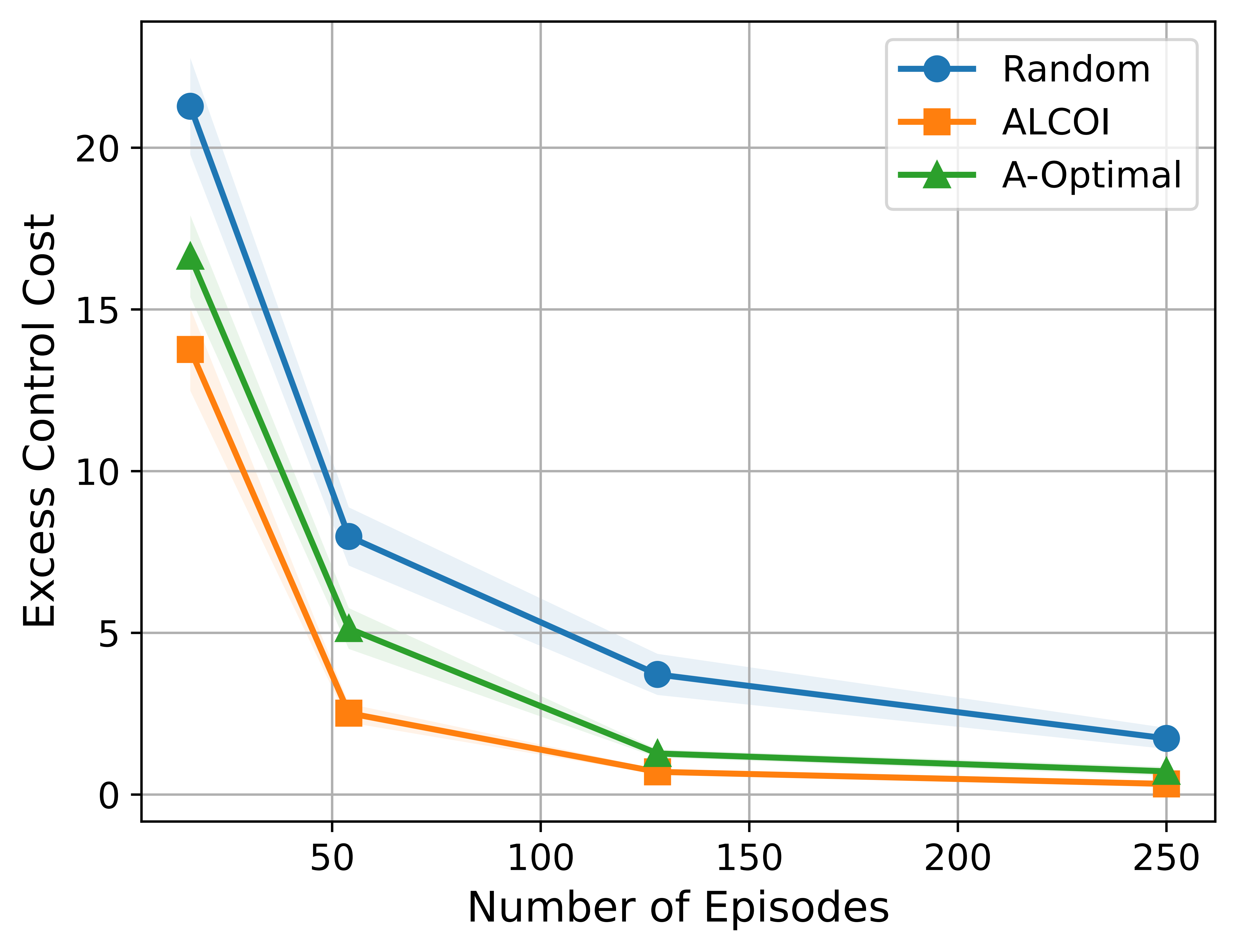
Motivate Algorithm

Correctness Guarantees

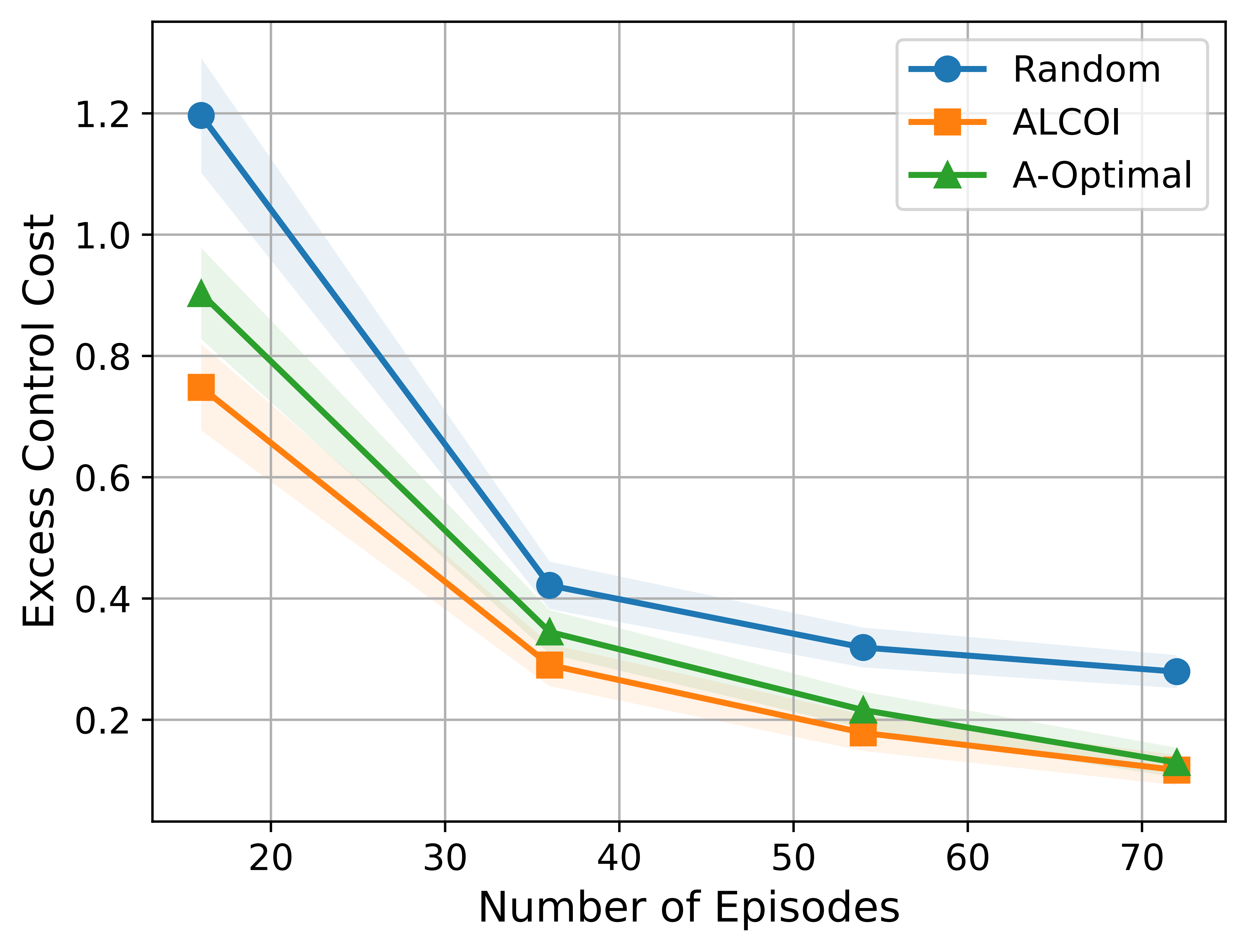
Numerical Validation

Illustrative example

Recall system $X_{t+1} = X_t + U_t + \sum_{i=1}^{4} \sigma(X_t - \phi_i) + W_t$



Cartpole SwingUp Example



Next steps

Ongoing Extensions

Partially Observed

Lower Bounds

Future directions

Active Exploration for Multi-task Representation Learning

Choose both the task to collect data from and the exploration policy

Thank you!

Relevant Papers

- B. Lee, I. Ziemann, G. J. Pappas, N. Matni, *Active Learning for Control-Oriented Identification of Nonlinear Systems,* IEEE Conference on Decision and Control, 2024 (to appear)

- T. TCK Zhang, B. Lee, I. Ziemann, G. J. Pappas, N. Matni, *Guarantees for Nonlinear Representation Learning: Non-identical Covariates, Dependent Data, Fewer Samples,* International Conference on Machine Learning, 2024

- I. Ziemann, S. Tu, G. J. Pappas, N. Matni, *Sharp Rates in Dependent Learning Theory: Avoiding Sample Size Deflation for the Square Loss*, International Conference on Machine Learning, 2024 (Spotlight)