Implicit Biases of Gradient Descent in Offline System Identification and Optimal Control

---

Nadav Cohen

Tel Aviv University & Imubit

AI Week *26 June 2024*

Talk Sources

Implicit Bias of Policy Gradient in Linear Quadratic Control: Extrapolation to Unseen Initial States

Razin* + Alexander* + Cohen-Karlik + Giryes + Globerson + C | *ICML 2024*

Learning Low Dimensional State Spaces with Overparameterized Recurrent Neural Nets

Cohen-Karlik + Menuhin-Gruman + Giryes + C + Globerson | *ICLR 2023*

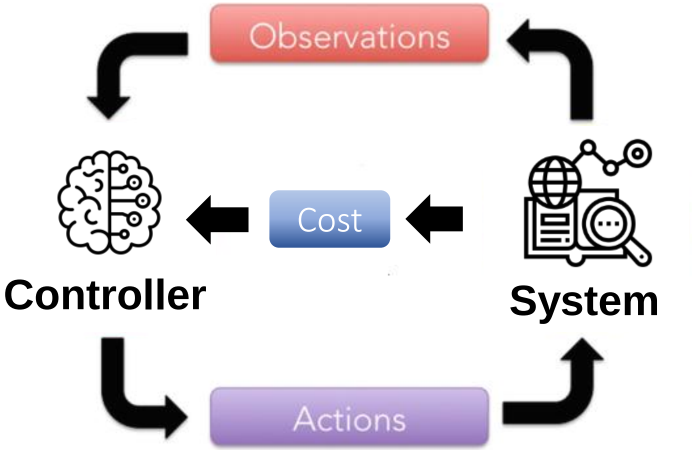
On the Implicit Bias of Gradient Descent for Temporal Extrapolation

Cohen-Karlik + Ben David + C + Globerson | *AISTATS 2022*

Optimal Control (Reinforcement Learning)

Goal

Design controller that steers a dynamical system to minimize a cost



Observations

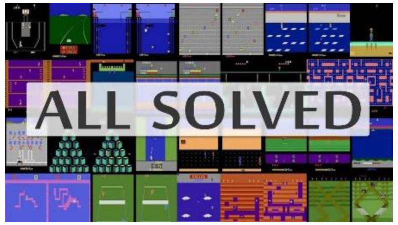
Cost

Controller

System

Actions

Applications



Computer gaming

Playing Go

Autonomous driving

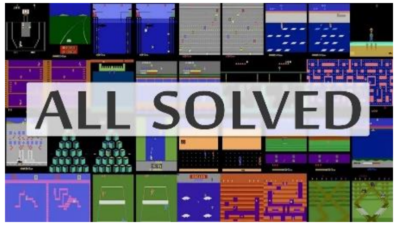
Medical treatment

Manufacturing optimization

Learning via Trial & Error

Learning a controller typically entails **trial & error** over system

Feasible in some applications; **prohibitively costly/dangerous** in others



Computer gaming

✓

Playing Go

✓

Autonomous driving

▬

Medical treatment

⚠

Manufacturing optimization

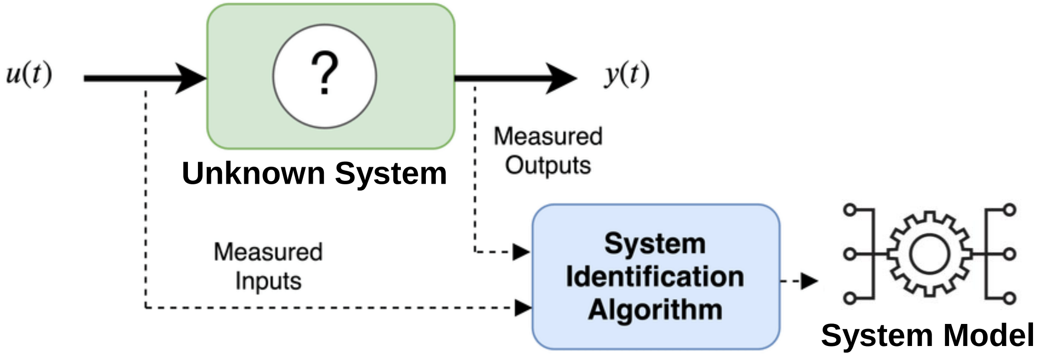
⚠

Offline System Identification and Optimal Control

Natural approach for learning a controller **without trail & error**:

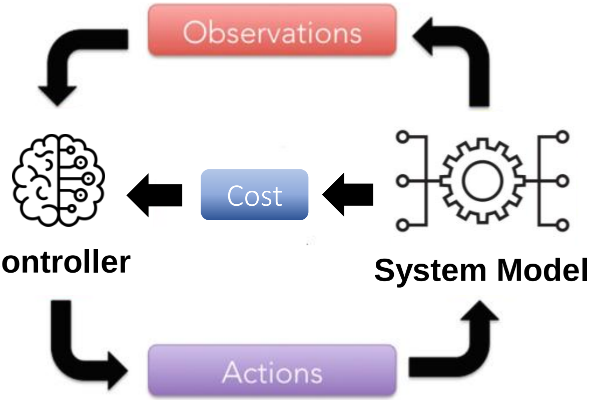
Step 1 Offline System Identification

Use pre-recorded data for offline learning of a **system model**

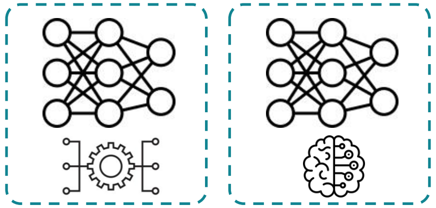


Step 2 Optimal Control

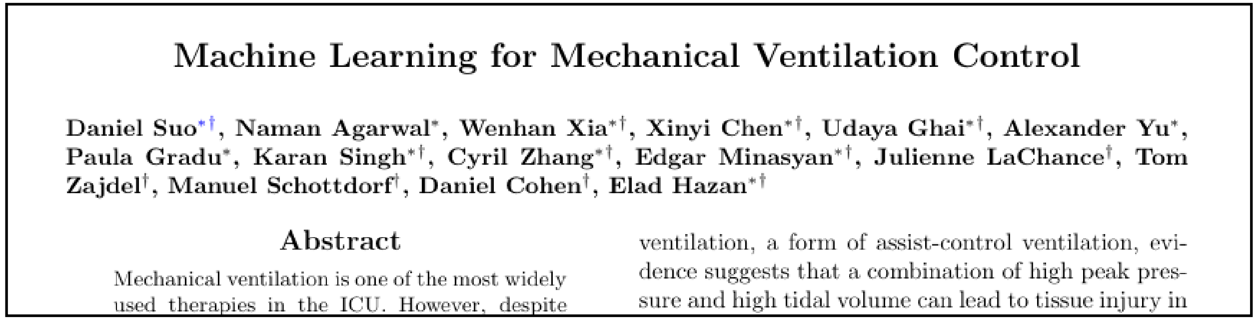
Use system model for offline learning of an **optimal controller**



Realizing approach via **overparameterized models** (e.g. neural networks) trained by **gradient descent** (**GD**) yields breakthrough results



Case Study I: Medical Treatment



Machine Learning for Mechanical Ventilation Control

Daniel Suo[*†], Naman Agarwal[*], Wenhan Xia[*†], Xinyi Chen[*†], Udaya Ghai[*†], Alexander Yu[*], Paula Gradu[*], Karan Singh[*†], Cyril Zhang[*†], Edgar Minasyan[*†], Julienne LaChance[†], Tom Zajdel[†], Manuel Schottdorf[†], Daniel Cohen[†], Elad Hazan[*†]

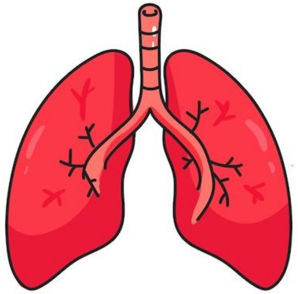
Abstract

Mechanical ventilation is one of the most widely used therapies in the ICU. However, despite

ventilation, a form of assist-control ventilation, evidence suggests that a combination of high peak pressure and high tidal volume can lead to tissue injury in

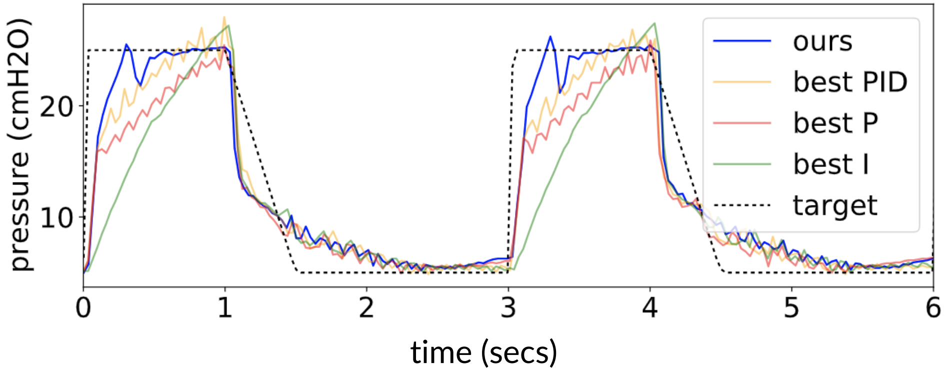
Step 1 Offline System Identification

Use pre-recorded data for offline learning of a lungs model



Step 2 Optimal Control

Use lungs model for offline learning of a mechanical ventilator controller

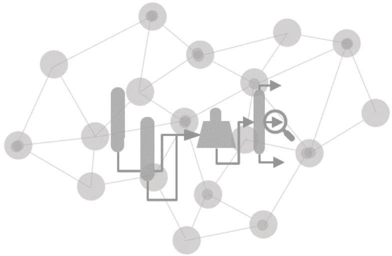


Case Study II: Manufacturing Optimization

Unlock new value

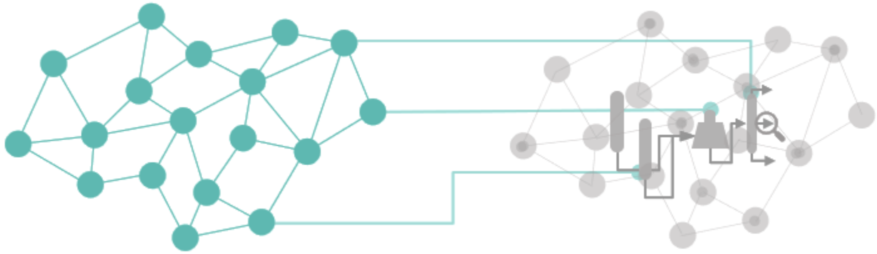
Step 1 Offline System Identification

Use pre-recorded data for offline learning of a neural network plant model



Step 2 Optimal Control

Use plant model for offline learning of a neural network controller



10-20%
Reduction In Energy Consumption

1K+ MT
Reduction In Scope 1 Emissions

1%-3%
Yield Improvement

→

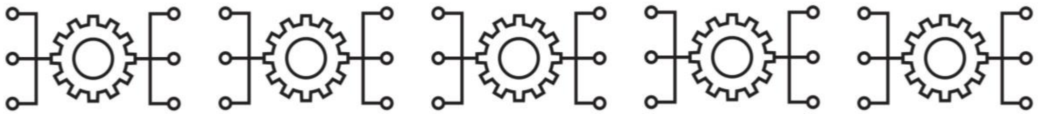
$10M+
Bottom-Line Value

Implicit Biases of Gradient Descent

When using overparameterized models:

Offline System Identification

Multiple system models fit pre-recorded data, **some generalize well while others do not**



Optimal Control

Multiple controllers are optimal for learned system model and given set of initial states, **some generalize well while others do not**

Fact that learned solutions generalize well results from **implicit biases** of GD

Not only in-distribution but also
out-of-distribution (**extrapolation**)

Q: Can we theoretically characterize when implicit bias of GD leads to extrapolation?

Outline

Offline System Identification: Extrapolation to Longer Horizon in Overparameterized Linear Models

Optimal Control: Extrapolation to Unseen Initial States in Overparameterized Linear Controllers

Conclusion

Linear Dynamical System

A cornerstone of control theory is the **linear dynamical system** (**LDS**):

- **Parameters**

$$\mathbf{A} \in \mathbb{R}^{D \times D} \;,\; \mathbf{B} \in \mathbb{R}^{D \times M} \;,\; \mathbf{C} \in \mathbb{R}^{K \times D}$$

State transition matrix   Action matrix   Observation matrix

- **Dynamics**

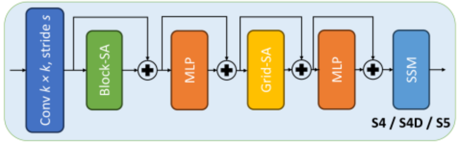
$$\text{For } h = 0,1,2,\ldots : \quad \mathbf{x}_{h+1} = \mathbf{A}\mathbf{x}_h + \mathbf{B}\mathbf{u}_h \quad , \quad y_h = \mathbf{C}\mathbf{x}_h$$

Next state   Current state   Action   Observation

For system identification, we focus on common case of symmetric $\mathbf{A}$ and $M = K = 1$ (SISO)

Covers modern state space models (e.g. S4, Mamba,...)

(Gu et al. 2022, Gu & Dao 2023,...)



Offline Overparameterized Linear System Identification

Consider (unknown) ground truth LDS $\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*$ with state dim D^*

Given

Pre-recorded data of horizon H: $\{(\mathbf{u}^{(1)}, y^{*(1)}), \ldots, (\mathbf{u}^{(N)}, y^{*(N)})\}$

Action sequence $(u_1^{(1)}, \ldots, u_H^{(1)})$ Ground truth observation at time H

Goal

Identify ground truth LDS, thereby **extrapolating to longer horizon**

Method

Train **overparameterized** LDS $\mathbf{A}, \mathbf{B}, \mathbf{C}$ with state dim $D > D^*$ by running **GD over squared loss**:

$$\mathcal{L}_H(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \frac{1}{N} \sum_{n=1}^{N} \left(y_H^{(n)} - y^{*(n)} \right)^2$$

To decouple extrapolation from in-distribution generalization we assume unlimited data ($N \to \infty$)

$$\implies \quad \mathcal{L}_H(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{h=0}^{H-1} \left(\mathbf{C}\mathbf{A}^h\mathbf{B} - \mathbf{C}^*\mathbf{A}^{*h}\mathbf{B}^* \right)^2$$

Quantifying Extrapolation

$$\boxed{\mathcal{L}_H(\mathbf{A},\mathbf{B},\mathbf{C}) = \textstyle\sum_{h=0}^{H-1} \left(\mathbf{C}\mathbf{A}^h\mathbf{B} - \mathbf{C}^*\mathbf{A}^{*h}\mathbf{B}^*\right)^2}$$

Optimality Condition

Loss $\mathcal{L}_H(\cdot)$ is minimized **if and only if**

$$\mathbf{C}\mathbf{A}^h\mathbf{B} = \mathbf{C}^*\mathbf{A}^{*h}\mathbf{B}^* \text{ for all } h = 0, \ldots, H-1$$

Definition: *Extrapolation to Longer Horizon*

For $\epsilon > 0$, we say learned LDS $\mathbf{A},\mathbf{B},\mathbf{C}$ **ϵ-extrapolates to horizon $H' > H$** if:

$$\left|\mathbf{C}\mathbf{A}^h\mathbf{B} - \mathbf{C}^*\mathbf{A}^{*h}\mathbf{B}^*\right| \leq \epsilon \text{ for all } h \in \{0, \ldots, H'-1\}$$

Implicit Bias of Gradient Descent Leads to Extrapolation

Theory

Proposition: *Existence of Non-Extrapolating Models*

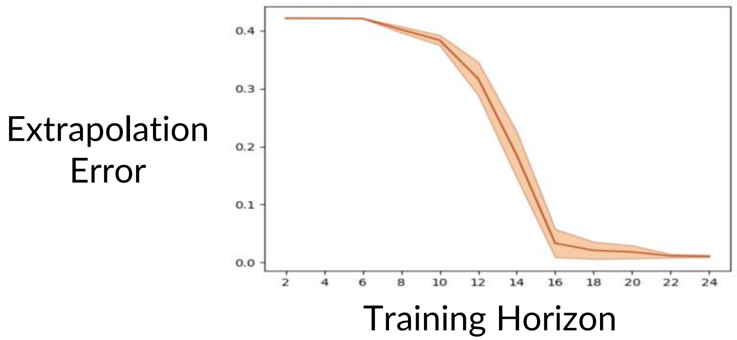
When $D > H$, for any $\epsilon > 0$, $H' > H$, and ground truth LDS $\mathbf{A}^*, \mathbf{B}^*, \mathbf{C}^*$, there exist model parameters $\mathbf{A}, \mathbf{B}, \mathbf{C}$ that minimize $\mathcal{L}_H(\cdot)$ but do not ϵ-extrapolate to horizon H'

Theorem: *GD Leads to Extrapolating Model*

When $D > H > 2D^*$, if GD learns model parameters $\mathbf{A}, \mathbf{B}, \mathbf{C}$ that minimize $\mathcal{L}_H(\cdot)$, then for any $\epsilon > 0$ and $H' > H$, the parameters ϵ-extrapolate to horizon H'

Experiments

- Validate theory (with linear models)
- Demonstrate theoretical results with neural networks



Outline

Offline System Identification: Extrapolation to Longer Horizon in Overparameterized Linear Models

Optimal Control: Extrapolation to Unseen Initial States in Overparameterized Linear Controllers

Conclusion

Linear Quadratic Regulator

A fundamental problem in control theory is the **linear quadratic regulator** (**LQR**):

LDS

$$\mathbf{x}_{h+1} = \mathbf{A}\mathbf{x}_h + \mathbf{B}\mathbf{u}_h$$

Quadratic Cost

$$\sum_{h=0}^{H} \mathbf{x}_h^\top \mathbf{Q} \mathbf{x}_h + \mathbf{u}_h^\top \mathbf{R} \mathbf{u}_h$$

Linear Controller

$$\mathbf{u}_h = \mathbf{K}\mathbf{x}_h$$

State cost matrix (parameter)

Action cost matrix (parameter)

Control matrix (designed to minimize cost)

Set of **seen initial states** $\mathcal{S}$ induces a **training cost**:

$$cost_{\mathcal{S}}(\mathbf{K}) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}_0 \in \mathcal{S}} \sum_{h=0}^{H} \mathbf{x}_h^\top \big(\mathbf{Q} + \mathbf{K}^\top \mathbf{R} \mathbf{K}\big) \mathbf{x}_h$$

We study practically motivated setting where multiple controllers minimize training cost

Consider learning controller $\mathbf{K}_{\text{GD}}$ by running GD over training cost

Q of prime importance: Does $\mathbf{K}_{\text{GD}}$ extrapolate to **unseen initial states**?

Quantifying Extrapolation

Optimality Condition

A controller $\mathbf{K}$ minimizes the training cost **if and only if** $\underbrace{\|(\mathbf{A}+\mathbf{B}\mathbf{K})\mathbf{x}_0\|^2 = 0}_{\mathbf{K} \text{ sends } \mathbf{x}_0 \text{ to zero}}$ for all $\mathbf{x}_0 \in \mathcal{S}$

Definition: *Error in Extrapolation to Unseen Initial States*

$\mathcal{E}(\mathbf{K}) := \frac{1}{|\mathcal{U}|} \sum_{\mathbf{x}_0 \in \mathcal{U}} \|(\mathbf{A}+\mathbf{B}\mathbf{K})\mathbf{x}_0\|^2$, where $\mathcal{U}$ is a basis of $\mathcal{S}^\perp$ (unseen subspace)

Baseline Controllers

Perfectly Extrapolating $\mathbf{K}_{\text{ext}}$

Satisfies $(\mathbf{A}+\mathbf{B}\mathbf{K}_{\text{ext}})\mathbf{x}_0 = \mathbf{0}$ for all $\mathbf{x}_0$

Minimizes training cost and $\mathcal{E}(\mathbf{K}_{\text{ext}}) = 0$

Non-Extrapolating $\mathbf{K}_{\text{no-ext}}$

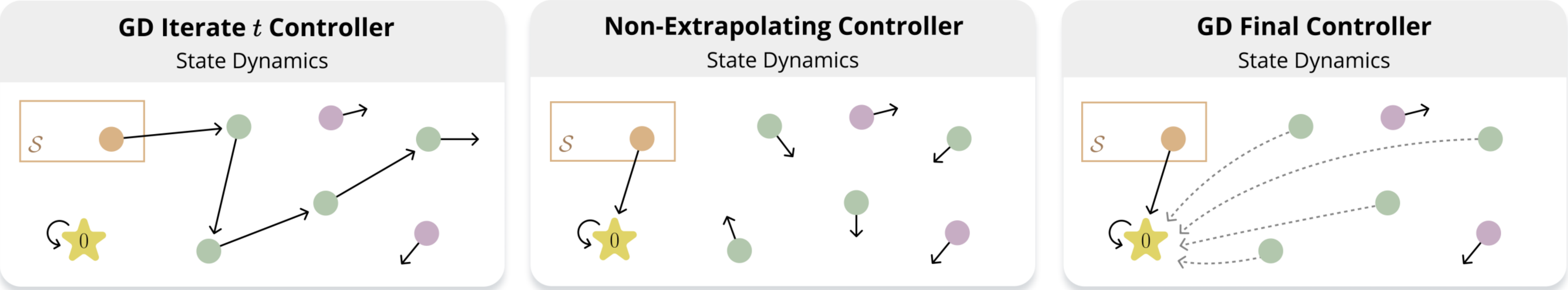
Satisfies $(\mathbf{A}+\mathbf{B}\mathbf{K}_{\text{no-ext}})\mathbf{x}_0 = \begin{cases} \mathbf{0} & , \mathbf{x}_0 \in \mathcal{S} \\ \mathbf{A}\mathbf{x}_0 & , \mathbf{x}_0 \in \mathcal{U} \end{cases}$

Minimizes training cost but $\mathcal{E}(\mathbf{K}_{\text{no-ext}})$ is **typically high**

Intuition: Extrapolation is Determined by Exploration

Intuition Behind Our Results

Extrapolation of $\mathbf{K}_{\mathrm{GD}}$ is determined by **exploration induced by system from seen initial states**



GD Iterate t Controller

State Dynamics

Non-Extrapolating Controller

State Dynamics

GD Final Controller

State Dynamics

- ● initial state seen in training
- ● explored state
- ● unexplored state

Theoretical Analysis

Proposition: *Extrapolation Requires Exploration*

- For states orthogonal to those reached during GD, $\mathbf{K}_{\text{GD}}$ and $\mathbf{K}_{\text{no-ext}}$ produce identical controls
- There exist non-exploratory systems in which: $\mathcal{E}(\mathbf{K}_{\text{GD}}) = \mathcal{E}(\mathbf{K}_{\text{no-ext}})$

Proposition: *Extrapolation in Exploration-Inducing Setting*

There exist exploration-inducing settings in which: $\mathcal{E}(\mathbf{K}_{\text{GD}}) << \mathcal{E}(\mathbf{K}_{\text{no-ext}})$

Theorem: *Extrapolation in Typical Setting*

When $\mathbf{A}$ is random Gaussian:

Learning rate

Horizon

$$\mathcal{E}(\mathbf{K}_{\text{GD}}) \leq \mathcal{E}(\mathbf{K}_{\text{no-ext}}) - \Omega\left(\eta \cdot \frac{H^2}{D}\right)$$

State dim

in expectation, and with high probability if D is large

Intuition: random system generically induces exploration (formalized via random matrix theory + topology)

Experiments: Non-Linear Systems and Neural Network Controllers

Our Theory

Linear system induces exploration from seen initial states ➡ linear controller typically extrapolates

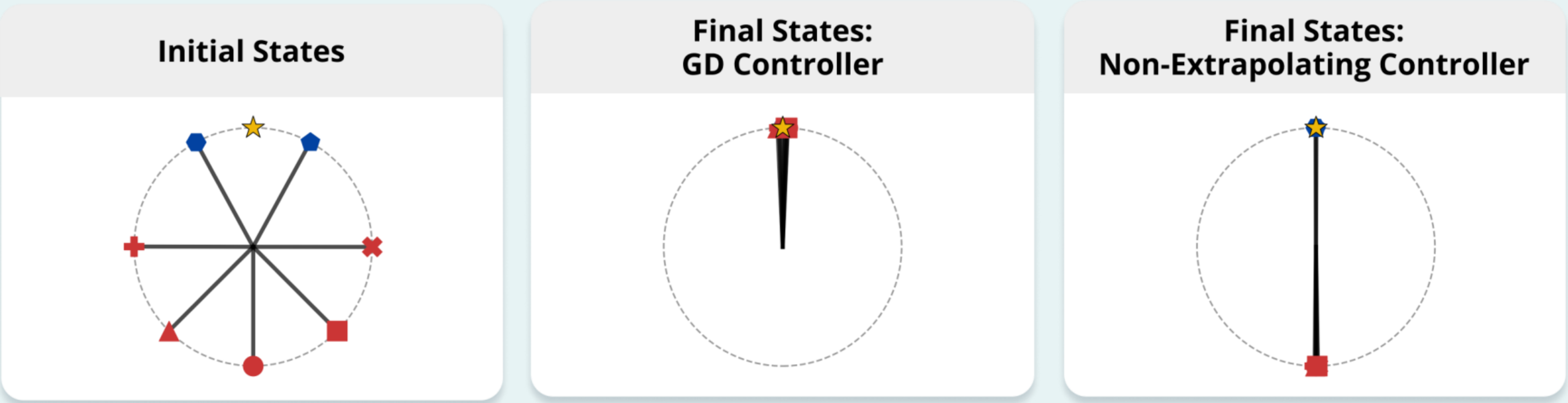
Experiments

Phenomenon extends to **non-linear systems** and **neural network controllers**

Pendulum Control Problem

(analogous experiments for a quadcopter control problem)

- ★ target state
- ● initial state seen in training
- ● initial state unseen in training



GD controller extrapolates despite existence of non-extrapolating controllers!

Outline

Offline System Identification: Extrapolation to Longer Horizon in Overparameterized Linear Models

Optimal Control: Extrapolation to Unseen Initial States in Overparameterized Linear Controllers

Conclusion

Recap

Learning to control critical systems via **trial & error is prohibitively costly/dangerous**

Natural approach: **offline system identification and optimal control**

➡ Breakthrough results with **overparameterized models** trained by **gradient descent** (**GD**)

↗

Fueled by **implicit biases of GD**, which lead to out-of-distribution generalization (**extrapolation**)

Theory: for overparameterized linear models:

Offline System Identification

If training horizon is sufficiently long, **GD extrapolates to longer horizon**

Optimal Control

If exploration induced by system from initial states seen in training is sufficient, **GD extrapolates to unseen initial states**

Experiments: phenomena extend to **neural networks**

Practical Implications

Our results suggest avenues for **improving extrapolation in practical settings**:

Offline System Identification

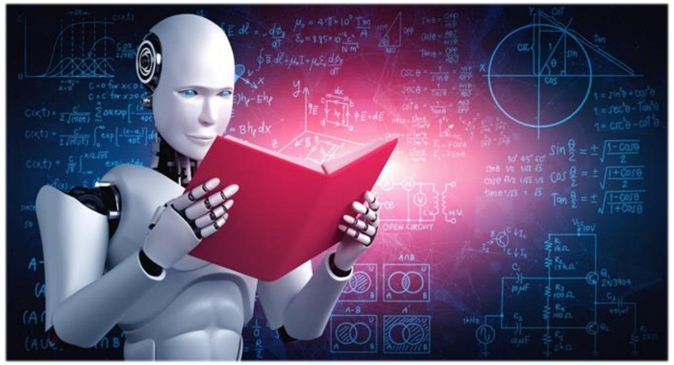
Developing algorithms for inferring required training horizon

Optimal Control

Developing algorithms for selecting exploration promoting initial states to train on

For practical progress in control of critical systems:

Trial & error ✗



Theory may be necessary

Thank You!

Work supported by:
Apple scholars in AI/ML PhD fellowship, Google Research Scholar Award, Google Research Gift, the Yandex Initiative in Machine Learning, the Israel Science Foundation (grant 1780/21), Len Blavatnik and the Blavatnik Family Foundation, Tel Aviv University Center for AI and Data Science, and Amnon and Anat Shashua