Fundamental Limits to Learning-Based Control

Ingvar Ziemann (UPenn), Henrik Sandberg (KTH)

Regret Lower Bounds for Linear Quadratic Gaussian Control, To appear, IEEE Transactions on Automatic Control

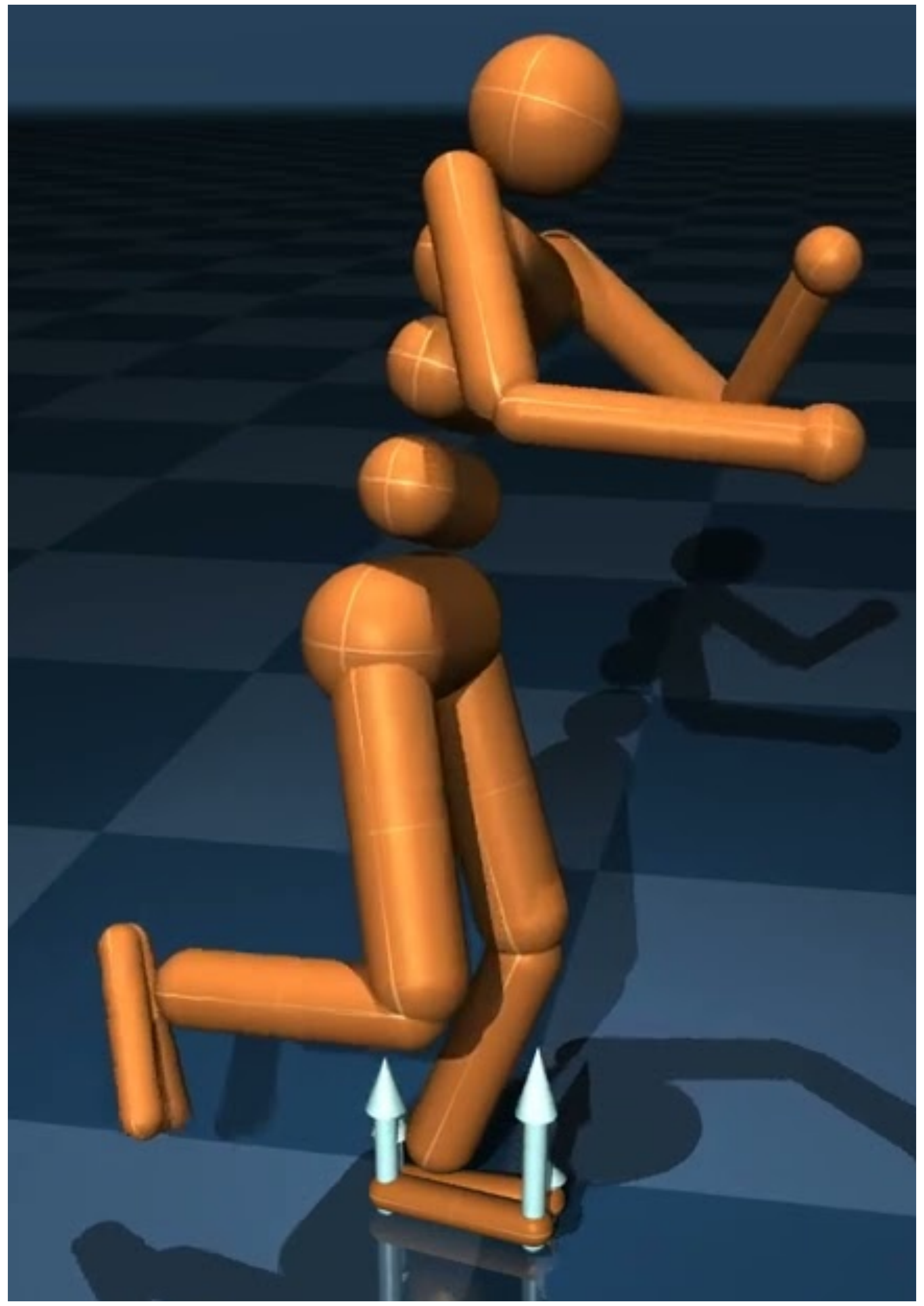
Why Study Fundamental Limits?

Why Study Fundamental Limits?

Ambition

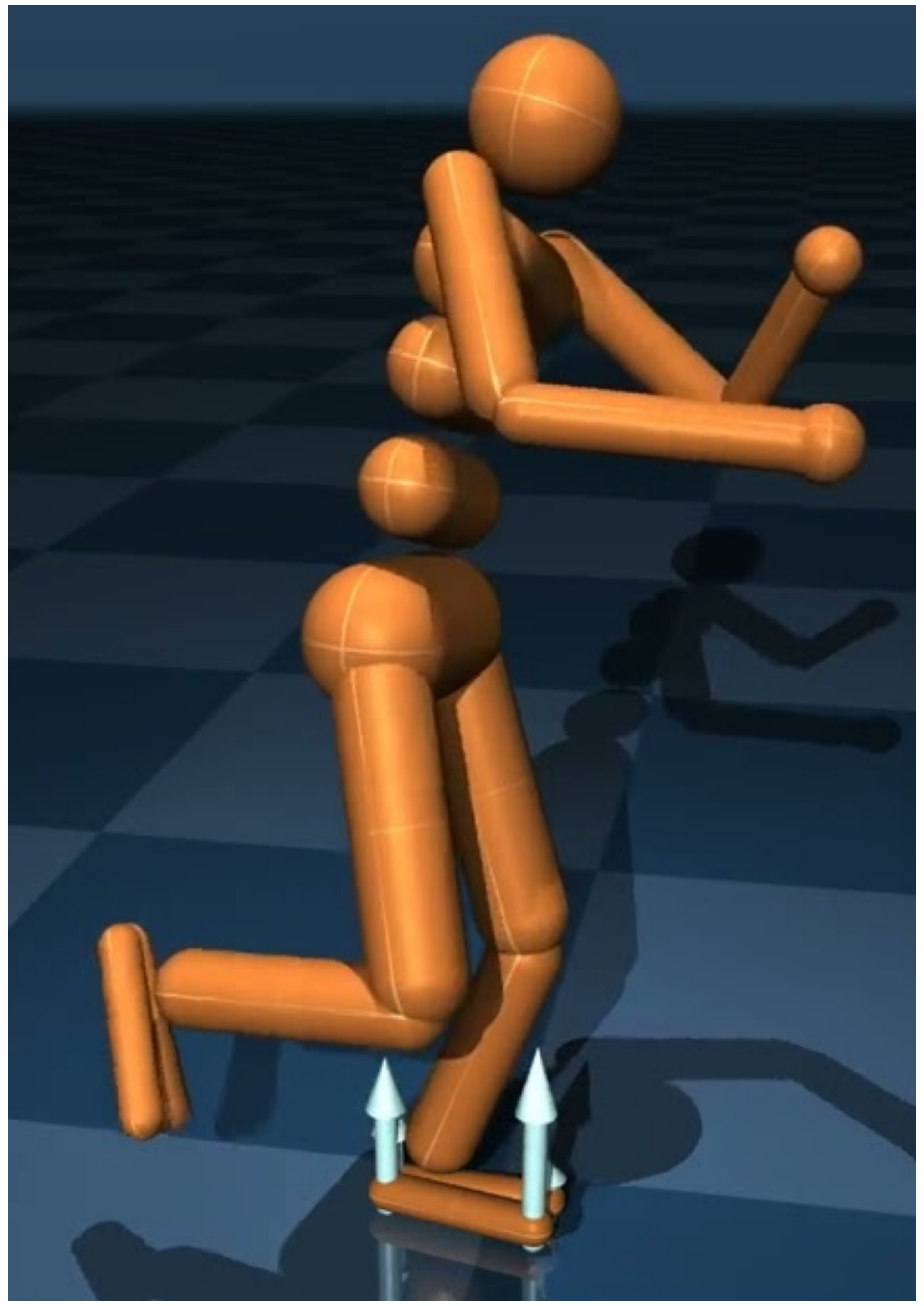
Why Study Fundamental Limits?

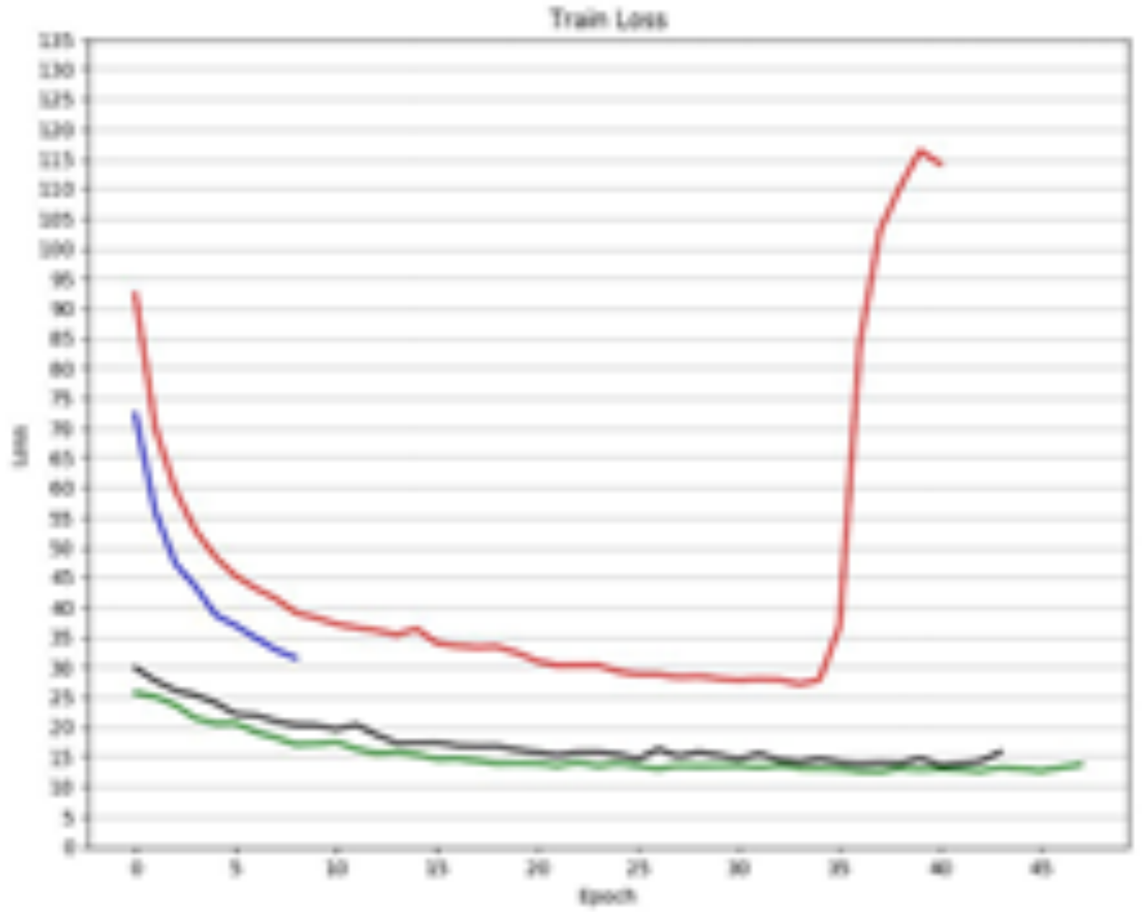
Ambition



Why Study Fundamental Limits?

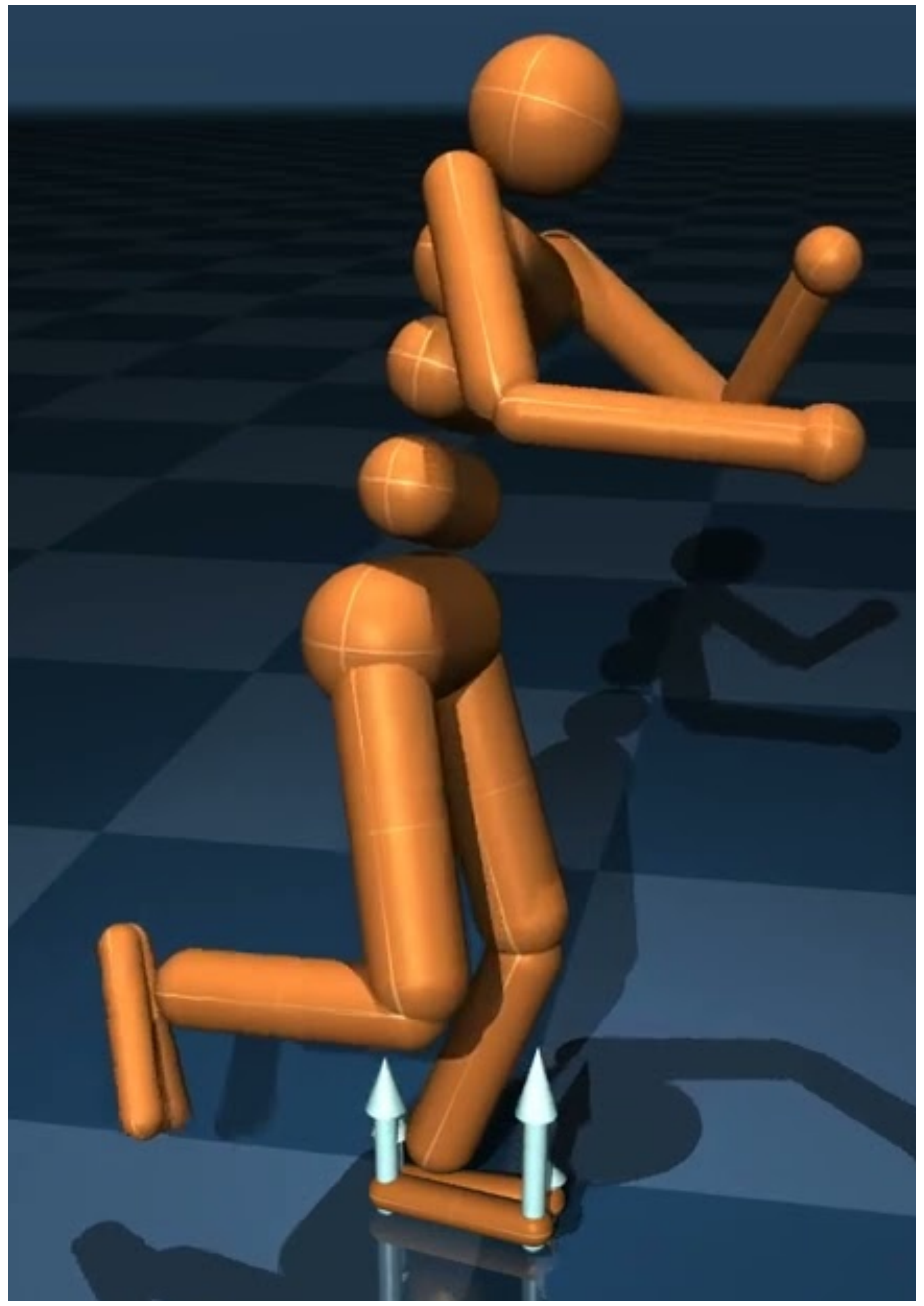
Ambition

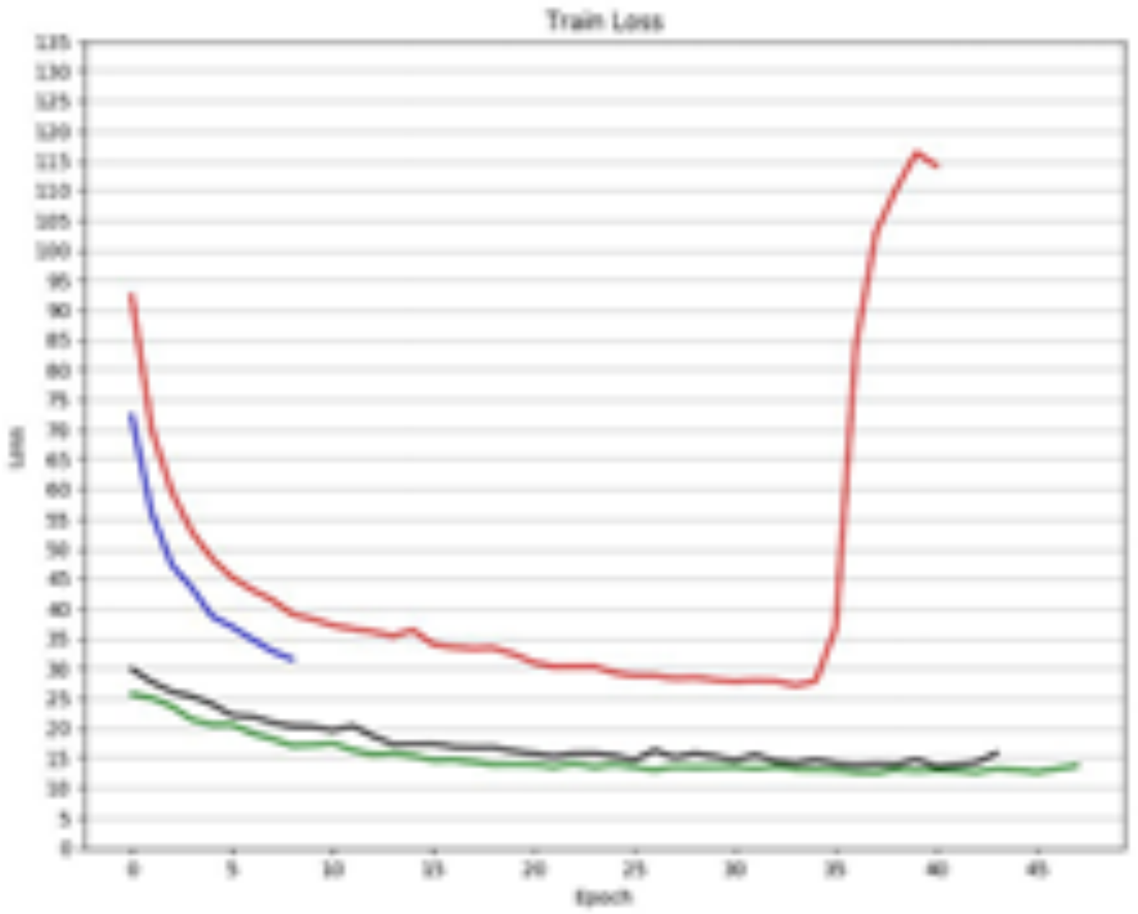




Why Study Fundamental Limits?

Ambition

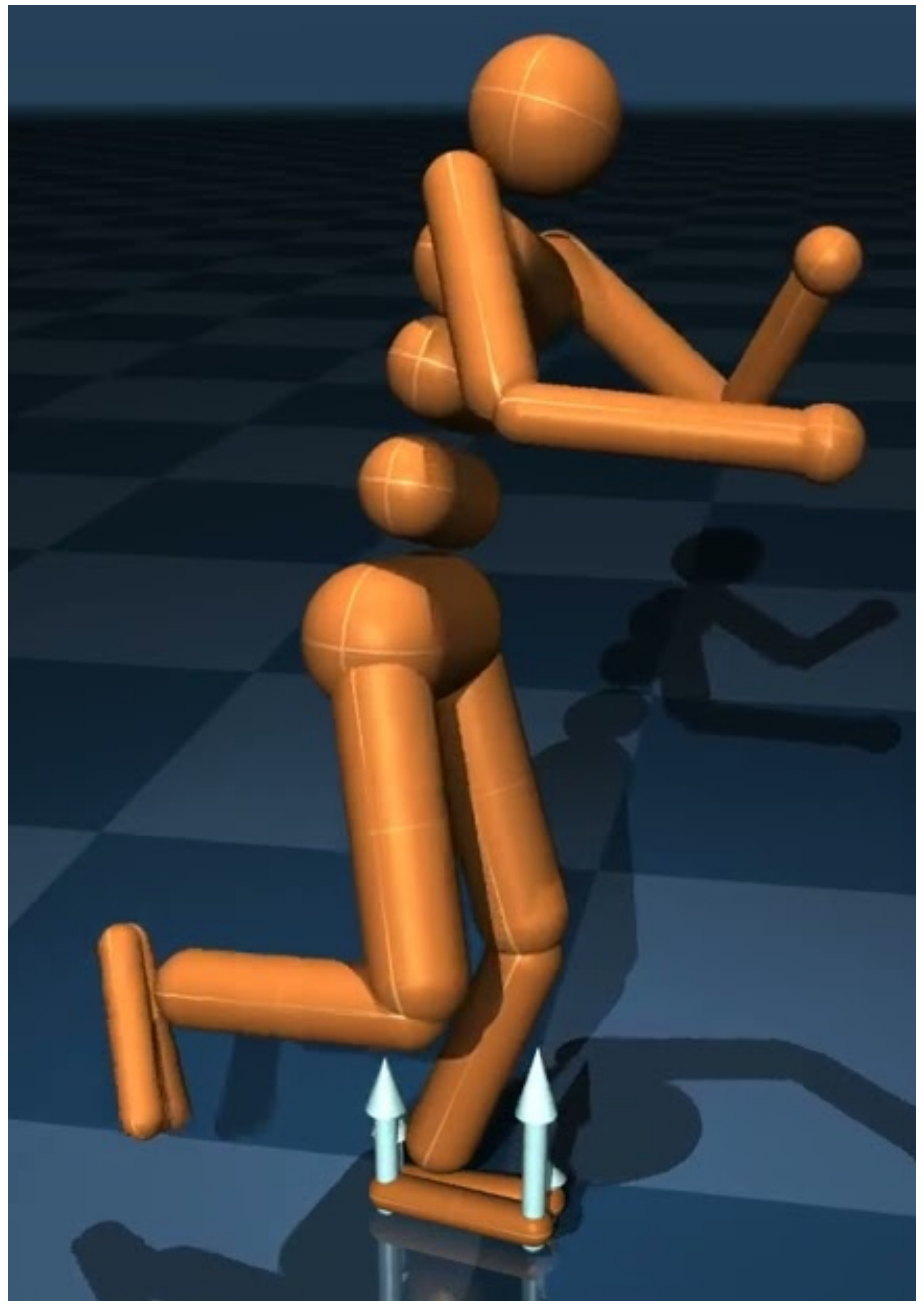


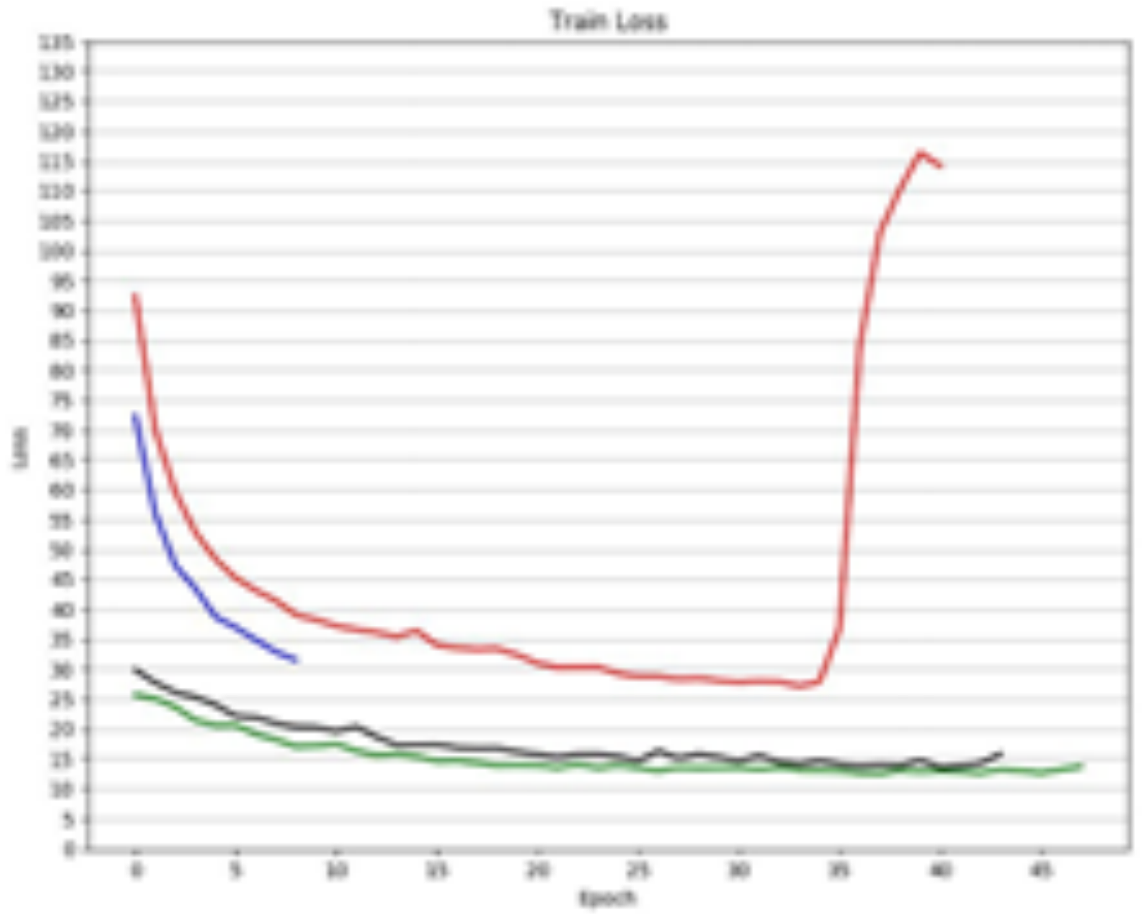


Reality

Why Study Fundamental Limits?

Ambition

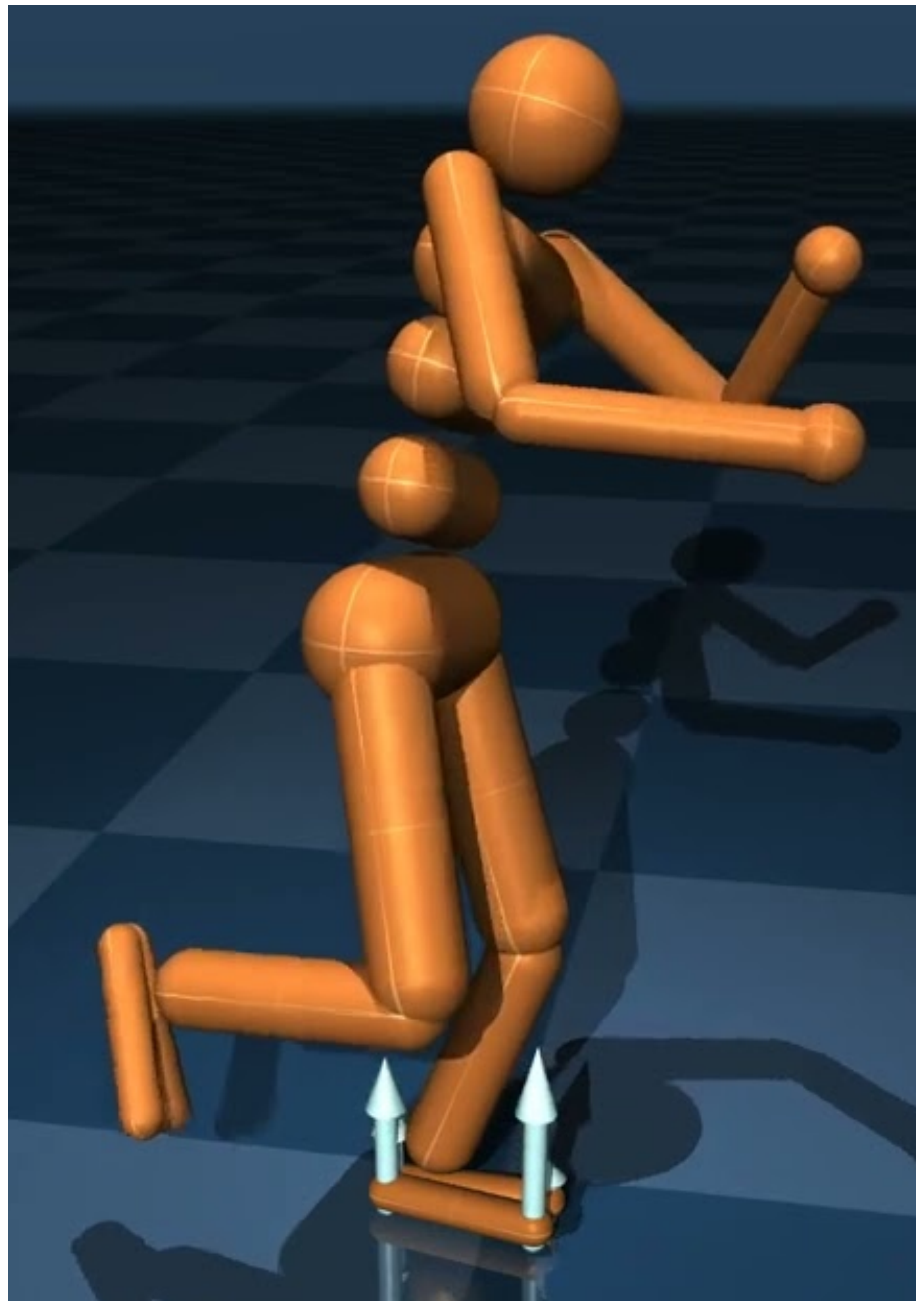


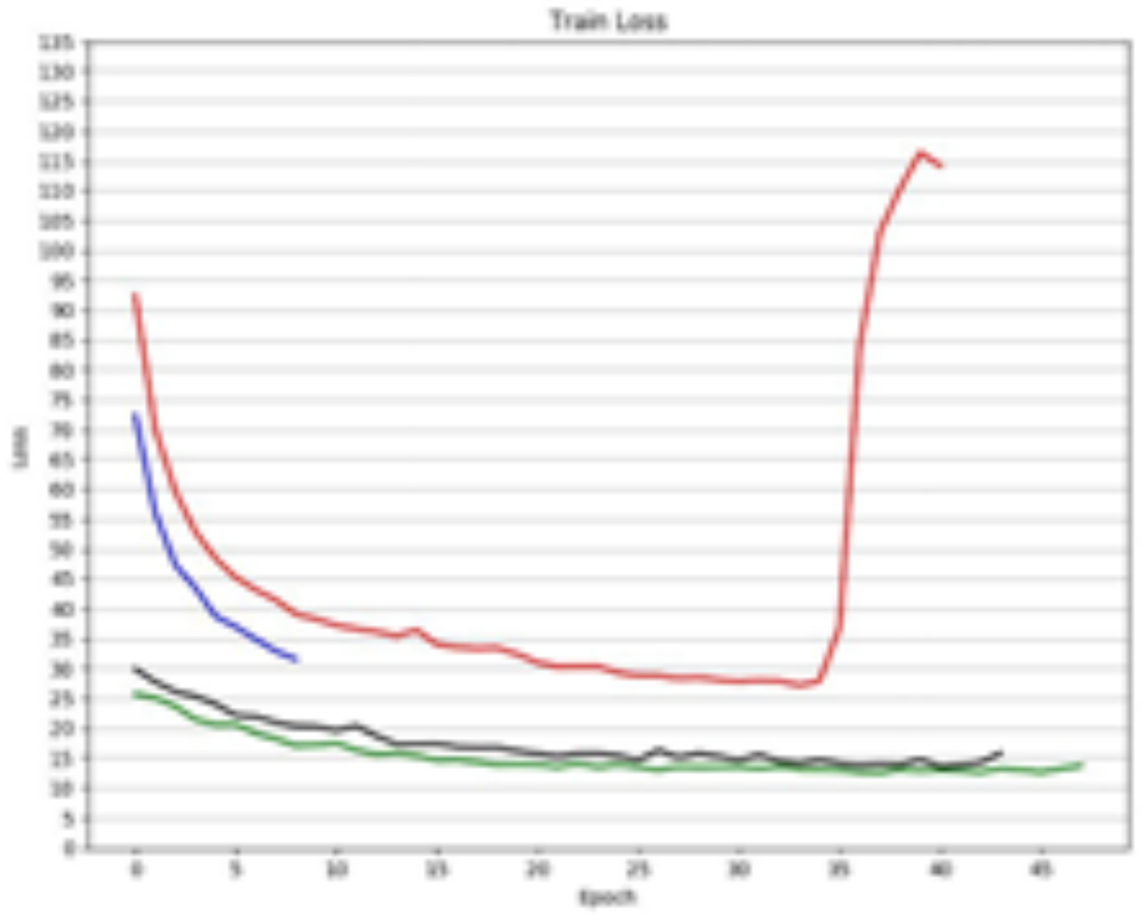


Reality

Why Study Fundamental Limits?

Ambition



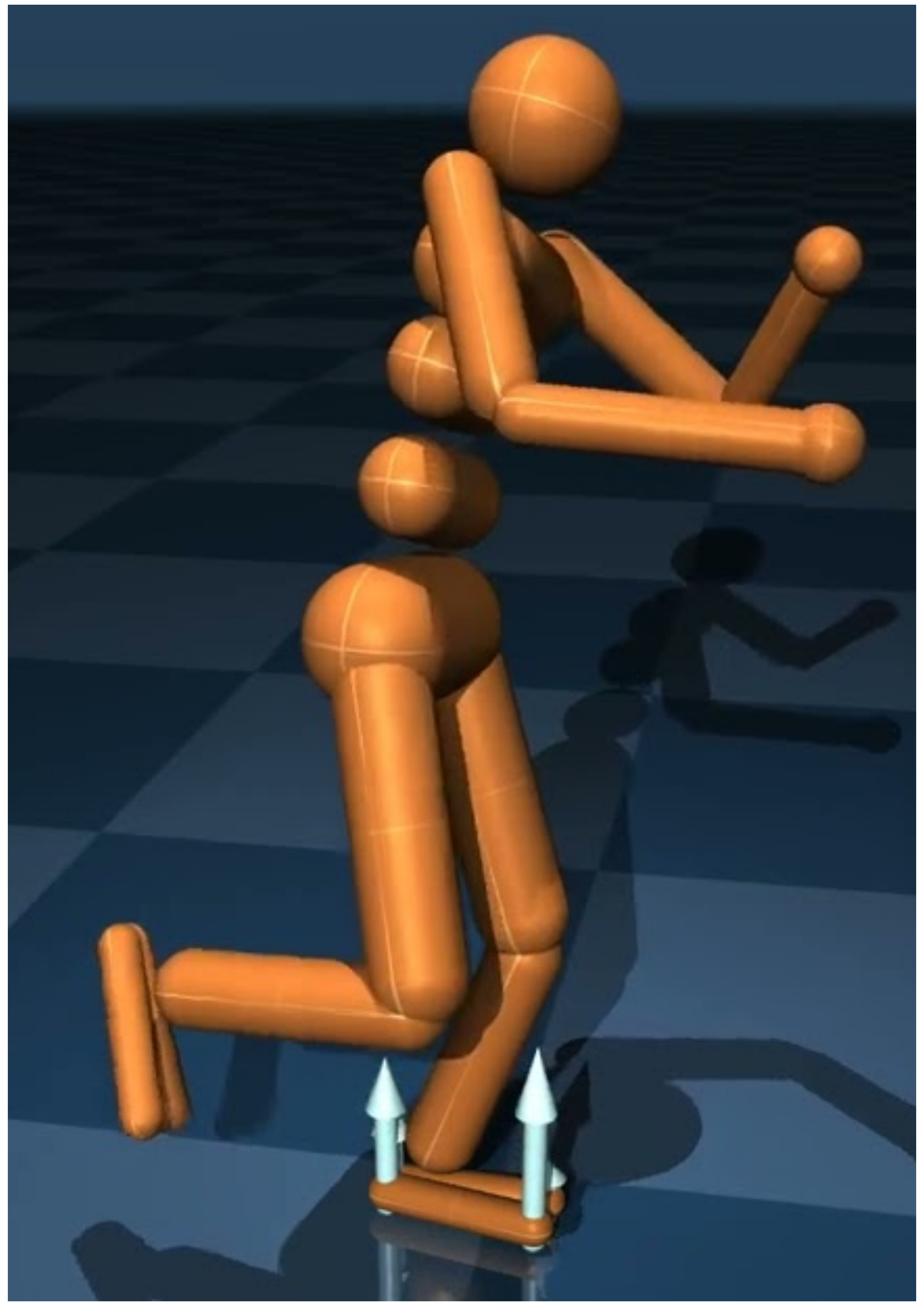


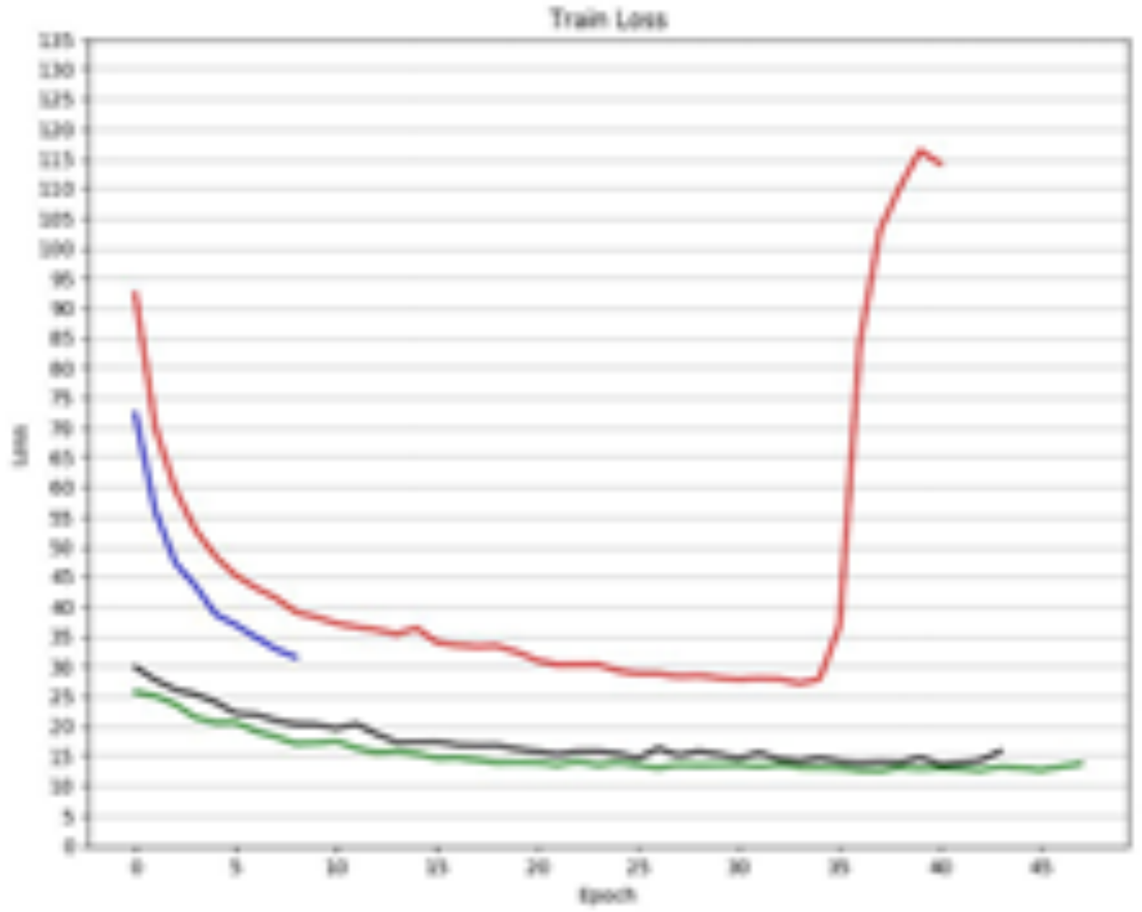
Reality

Not just in sim

Why Study Fundamental Limits?

Ambition





Reality

Not just in sim

Guaranteed Margins for LQG Regulators

JOHN C. DOYLE

Abstract—There are none.

Introduction

Considerable attention has been given lately to the issue of robustness of linear–quadratic (LQ) regulators. The recent work by Safonov and

Problem Formulation

Adaptive LQG

Problem Formulation

Adaptive LQG

Dynamics

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^{\star}(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^\star(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^\star(\theta)$ cannot be attained in practice

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^\star(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^\star(\theta)$ cannot be attained in practice

Regret of π under θ is:

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^\star(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^\star(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^{\pi}(\theta) \triangleq V_T^{\pi}(\theta) - V_T^\star(\theta)$$

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^{\star}(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^{\star}(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^{\pi}(\theta) \triangleq V_T^{\pi}(\theta) - V_T^{\star}(\theta)$$

We prove lower bounds of the form

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^{\top} Q X_t + U_T^{\top} R U_t + X_T^{\top} Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^{\star}(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^{\star}(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^{\pi}(\theta) \triangleq V_T^{\pi}(\theta) - V_T^{\star}(\theta)$$

We prove lower bounds of the form

$$\inf_{\pi} \sup_{\theta' \in B(\theta, \epsilon)} R_T^{\pi}(\theta') \gtrsim f(\theta) \times \sqrt{T}$$

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^{\top} Q X_t + U_T^{\top} R U_t + X_T^{\top} Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^{\star}(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^{\star}(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^{\pi}(\theta) \triangleq V_T^{\pi}(\theta) - V_T^{\star}(\theta)$$

We prove lower bounds of the form

$$\inf_{\pi} \sup_{\theta' \in B(\theta, \epsilon)} R_T^{\pi}(\theta') \gtrsim f(\theta) \times \sqrt{T}$$

Where $f(\theta)$ depends on system properties

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^{\star}(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^{\star}(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^{\pi}(\theta) \triangleq V_T^{\pi}(\theta) - V_T^{\star}(\theta)$$

We prove lower bounds of the form

$$\inf_{\pi} \sup_{\theta' \in B(\theta, \epsilon)} R_T^{\pi}(\theta') \gtrsim f(\theta) \times \sqrt{T}$$

Where $f(\theta)$ depends on system properties

Stability, Observability, Dimension etc.

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^\pi(\theta) \triangleq \mathbf{E}_\theta^\pi \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0,\ Q_T = P,\ \Sigma_{X_0} = S$

Let also $V_T^\star(\theta) \triangleq \inf_\pi V_T^\pi(\theta)$

Since θ is unknown, $V_T^\star(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^\pi(\theta) \triangleq V_T^\pi(\theta) - V_T^\star(\theta)$$

We prove lower bounds of the form

$$\inf_\pi \sup_{\theta' \in B(\theta, \epsilon)} R_T^\pi(\theta') \gtrsim f(\theta) \times \sqrt{T}$$

Where $f(\theta)$ depends on system properties

Stability, Observability, Dimension etc.

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^\pi(\theta) \triangleq \mathbf{E}_\theta^\pi \sum_{t=1}^{T-1} X_t^\top Q X_t + U_T^\top R U_t + X_T^\top Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^\star(\theta) \triangleq \inf_\pi V_T^\pi(\theta)$

Since θ is unknown, $V_T^\star(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^\pi(\theta) \triangleq V_T^\pi(\theta) - V_T^\star(\theta)$$

We prove lower bounds of the form

$$\inf_\pi \sup_{\theta' \in B(\theta, \epsilon)} R_T^\pi(\theta') \gtrsim f(\theta) \times \sqrt{T}$$

These are called

Local minimax lower bounds

Where $f(\theta)$ depends on system properties

Stability, Observability, Dimension etc.

Problem Formulation

Adaptive LQG

Dynamics

$$X_{t+1} = AX_t + BU_t + W_t$$

$$Y_t = CX_t + V_t$$

W_t, V_t are iid mean zero Gaussian

$(A, B, C) \triangleq \theta$ unknown to the controller/learner

Goal: choose $\pi = (\pi)_{1:T}$, with π_t depending on $(X, Y)_{1:t}$ to min:

$$V_T^{\pi}(\theta) \triangleq \mathbf{E}_{\theta}^{\pi} \sum_{t=1}^{T-1} X_t^{\top} Q X_t + U_T^{\top} R U_t + X_T^{\top} Q_T X_T$$

$(A, B, C) \in$ stab $\wedge$ det, $Q, R \succ 0$, $Q_T = P$, $\Sigma_{X_0} = S$

Let also $V_T^{\star}(\theta) \triangleq \inf_{\pi} V_T^{\pi}(\theta)$

Since θ is unknown, $V_T^{\star}(\theta)$ cannot be attained in practice

Regret of π under θ is:

$$R_T^{\pi}(\theta) \triangleq V_T^{\pi}(\theta) - V_T^{\star}(\theta)$$

We prove lower bounds of the form

$$\inf_{\pi} \sup_{\theta' \in B(\theta, \epsilon)} R_T^{\pi}(\theta') \gtrsim f(\theta) \times \sqrt{T}$$

These are called

Local minimax lower bounds

Where $f(\theta)$ depends on system properties

Stability, Observability, Dimension etc.

Regret Lower Bound

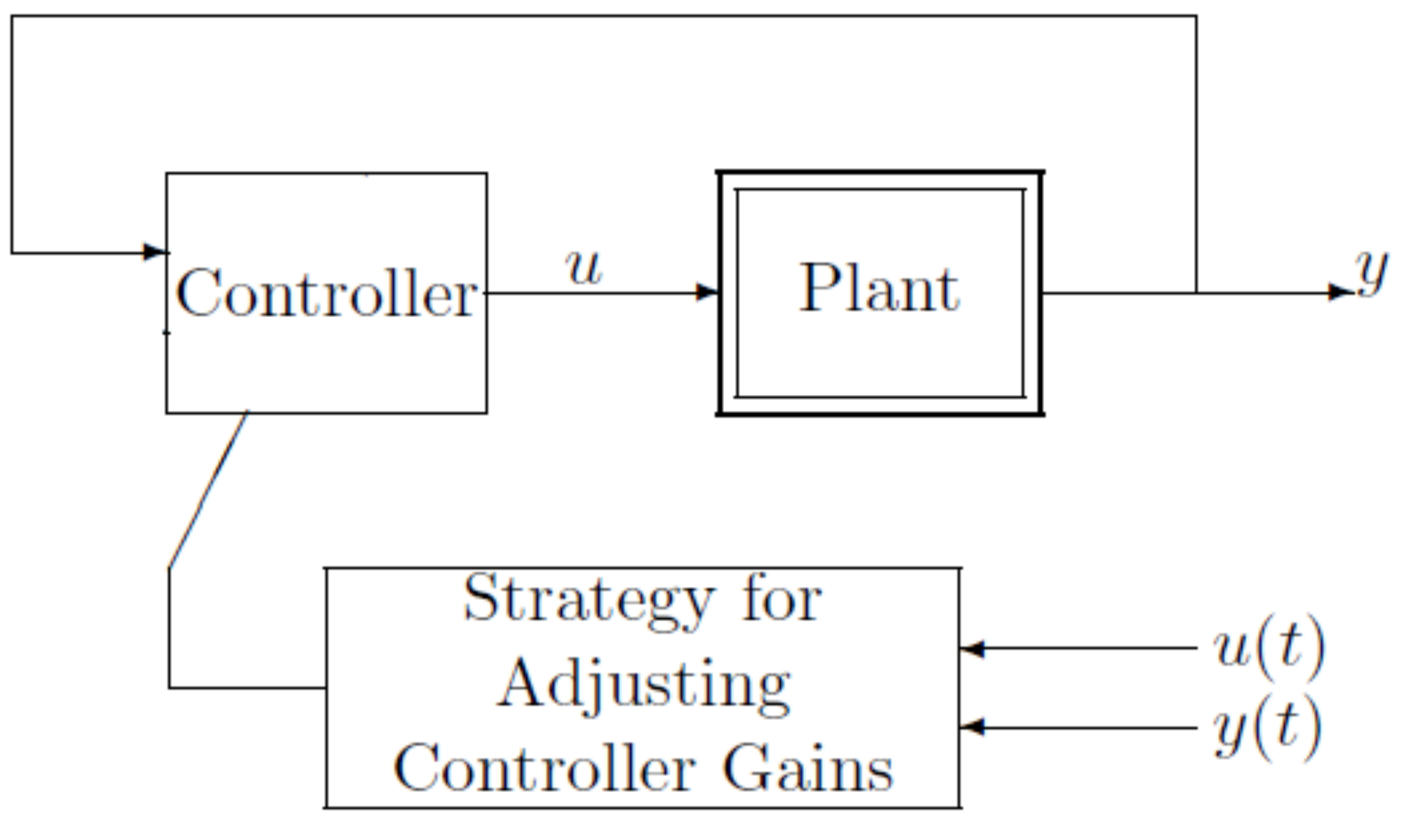
=

Fundamental Limit

to Adaptive Control

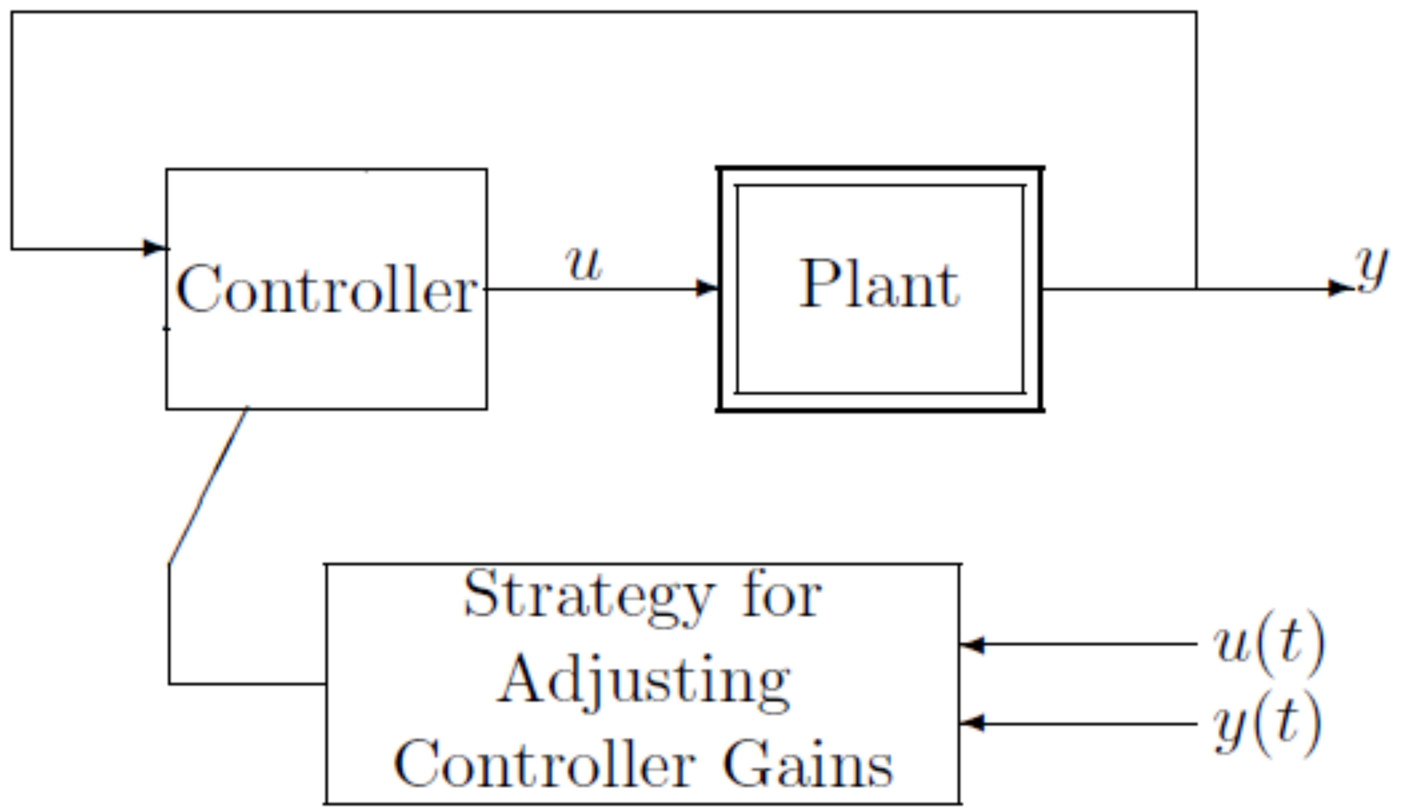
Performance

Background on Certainty Equivalence



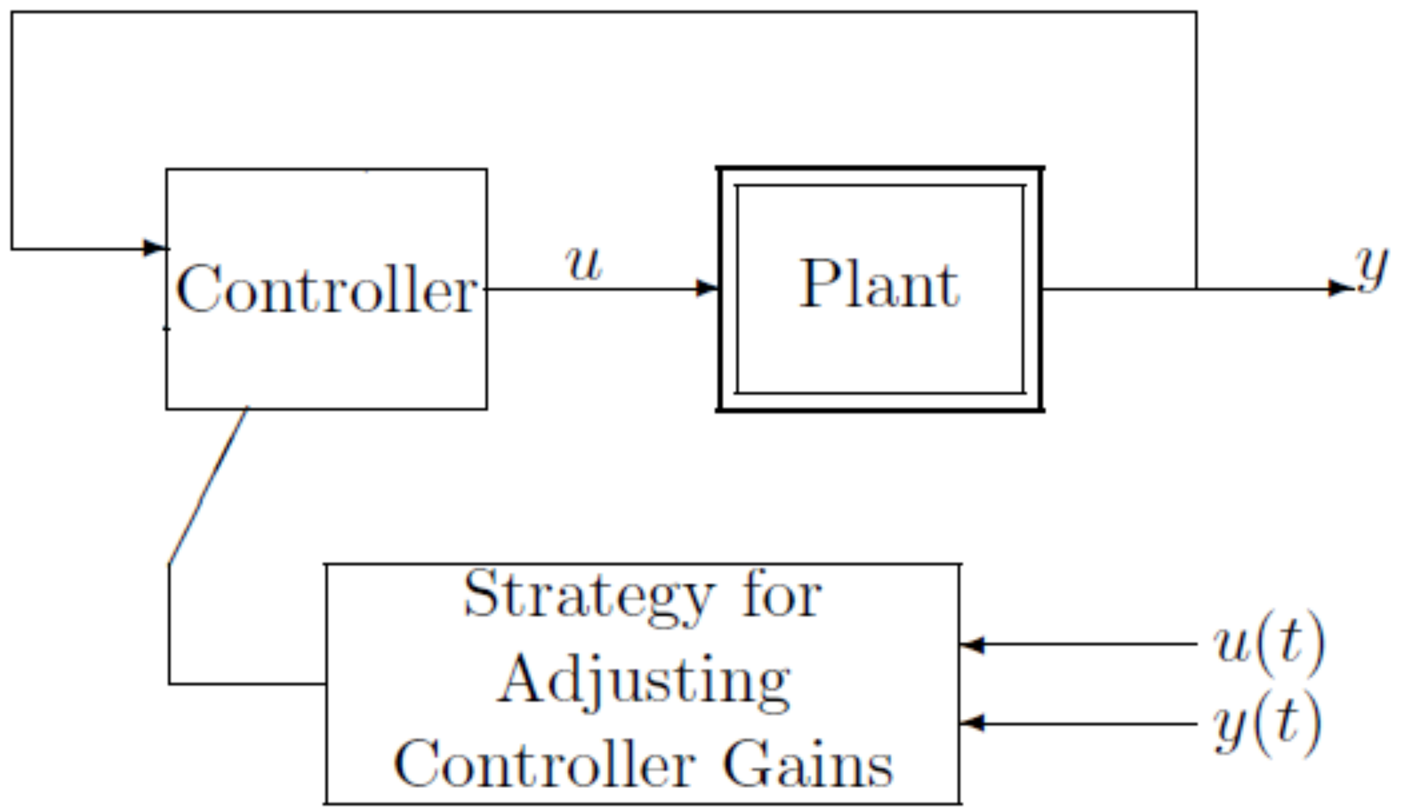
Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



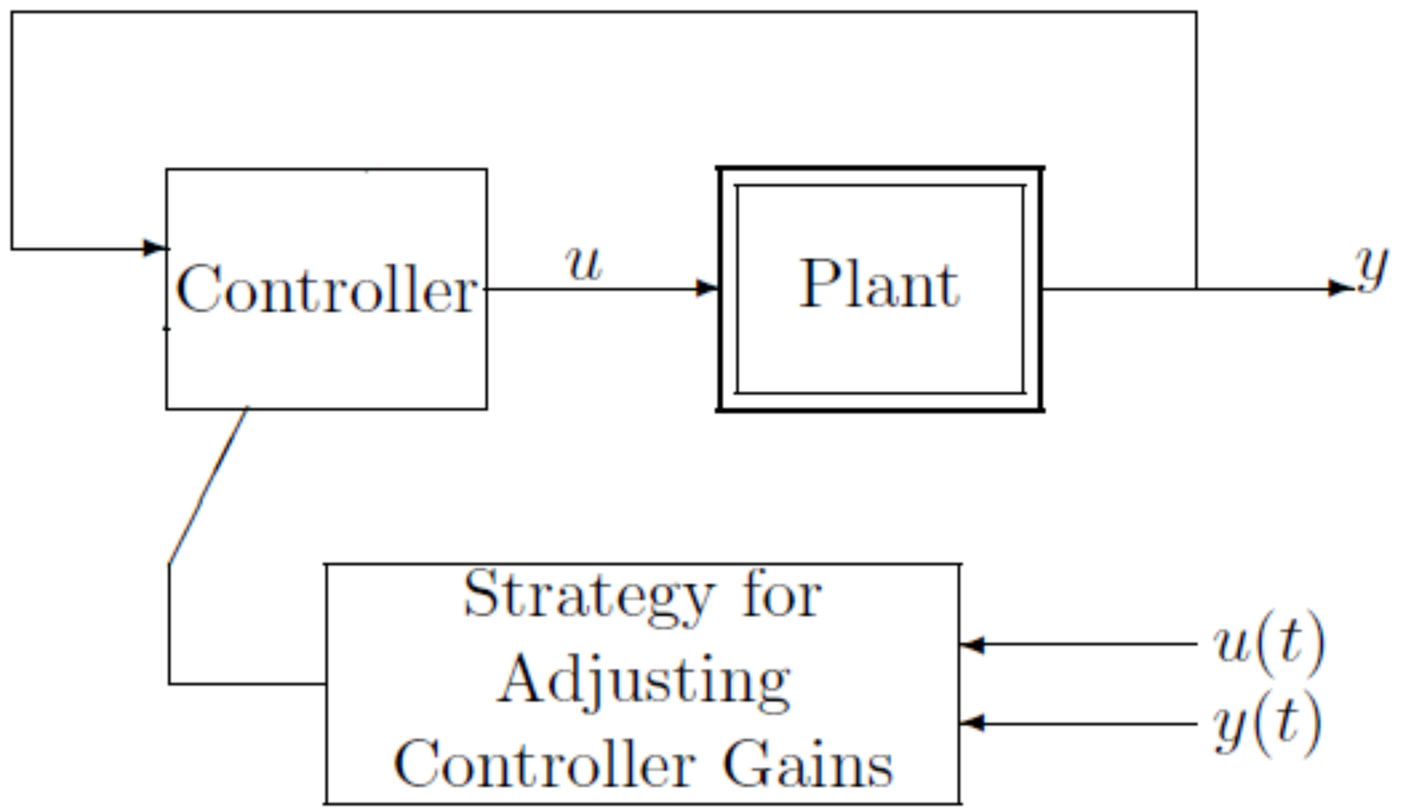
Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



Background on Certainty Equivalence

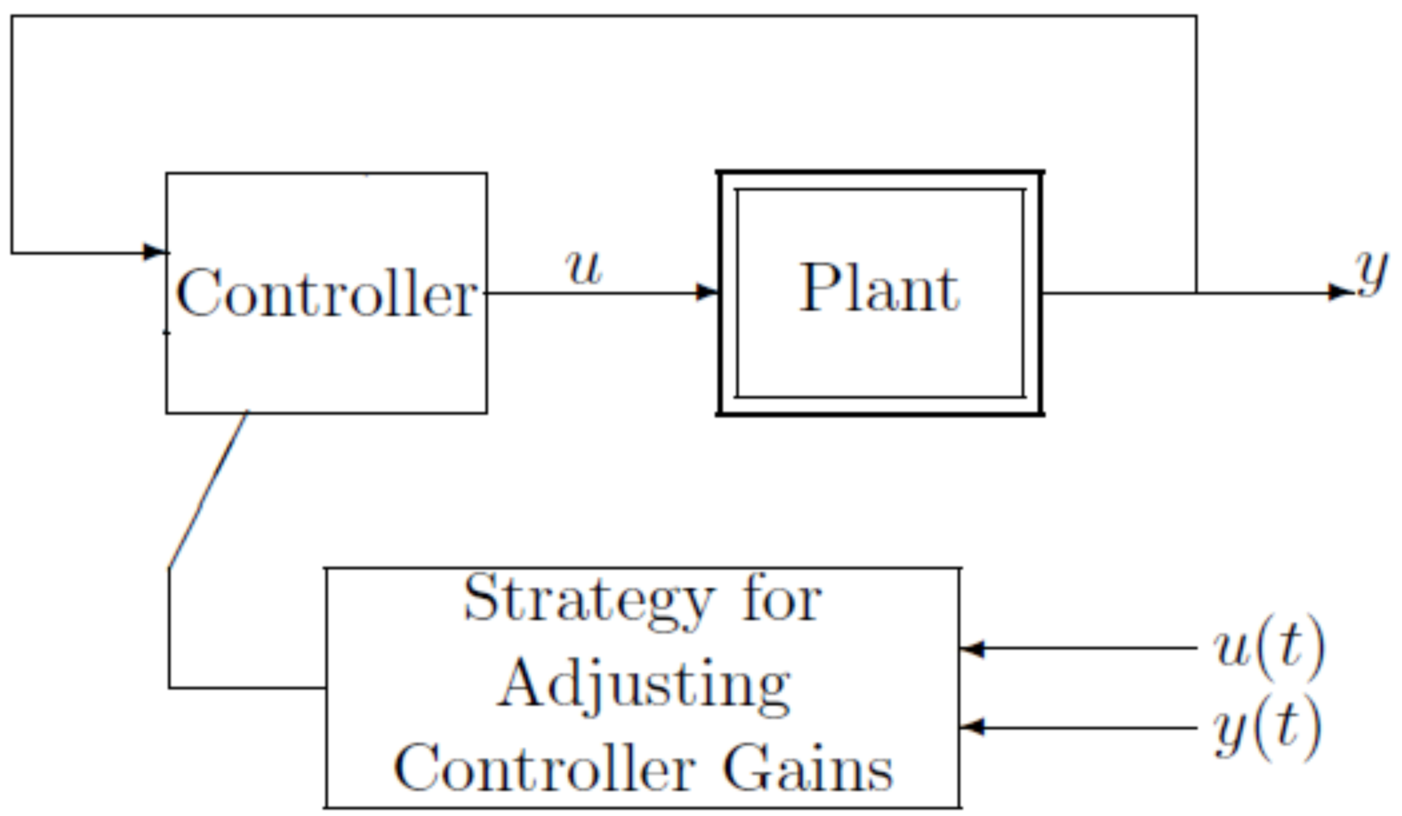
Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):

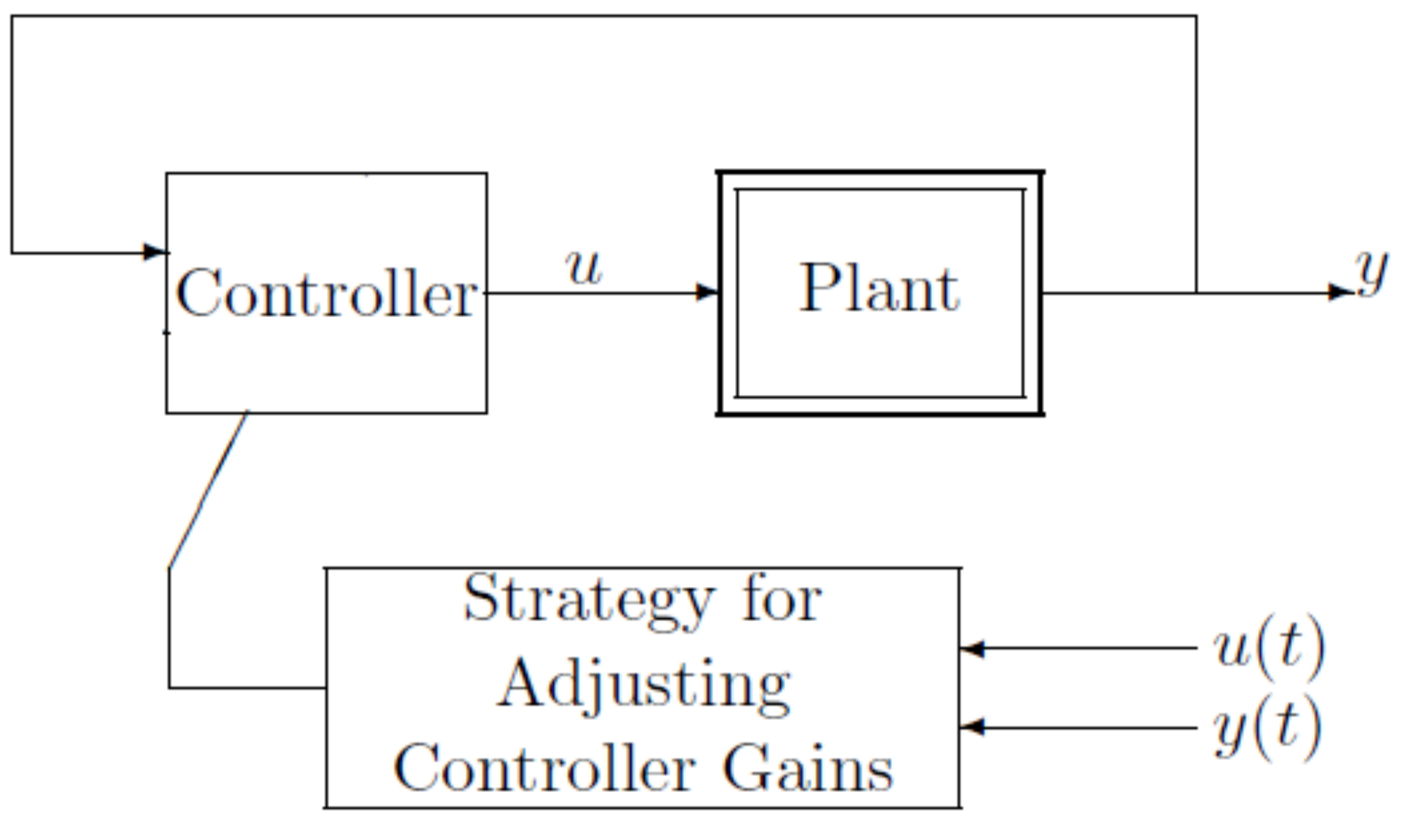


What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



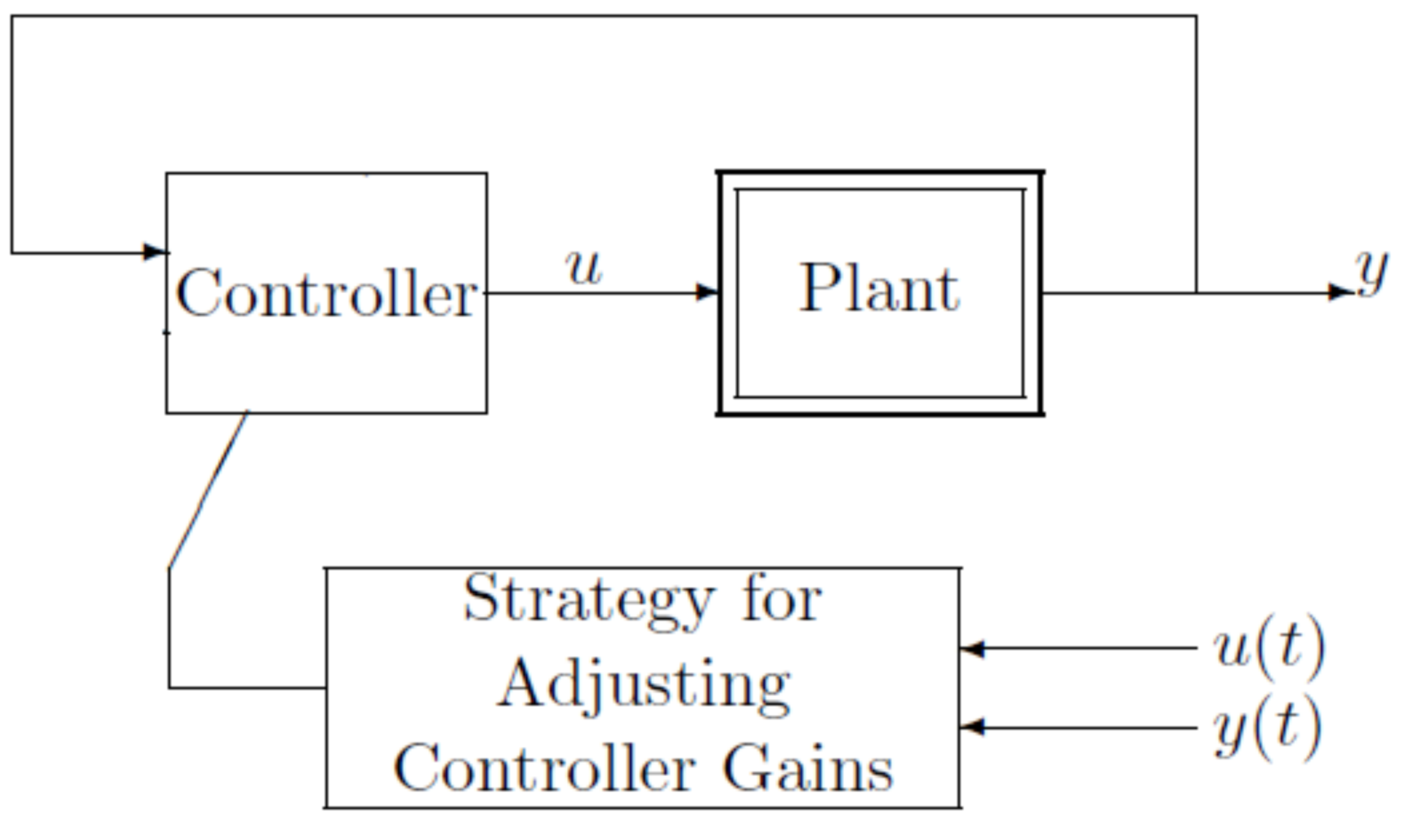
What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a “reasonable” strategy:

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

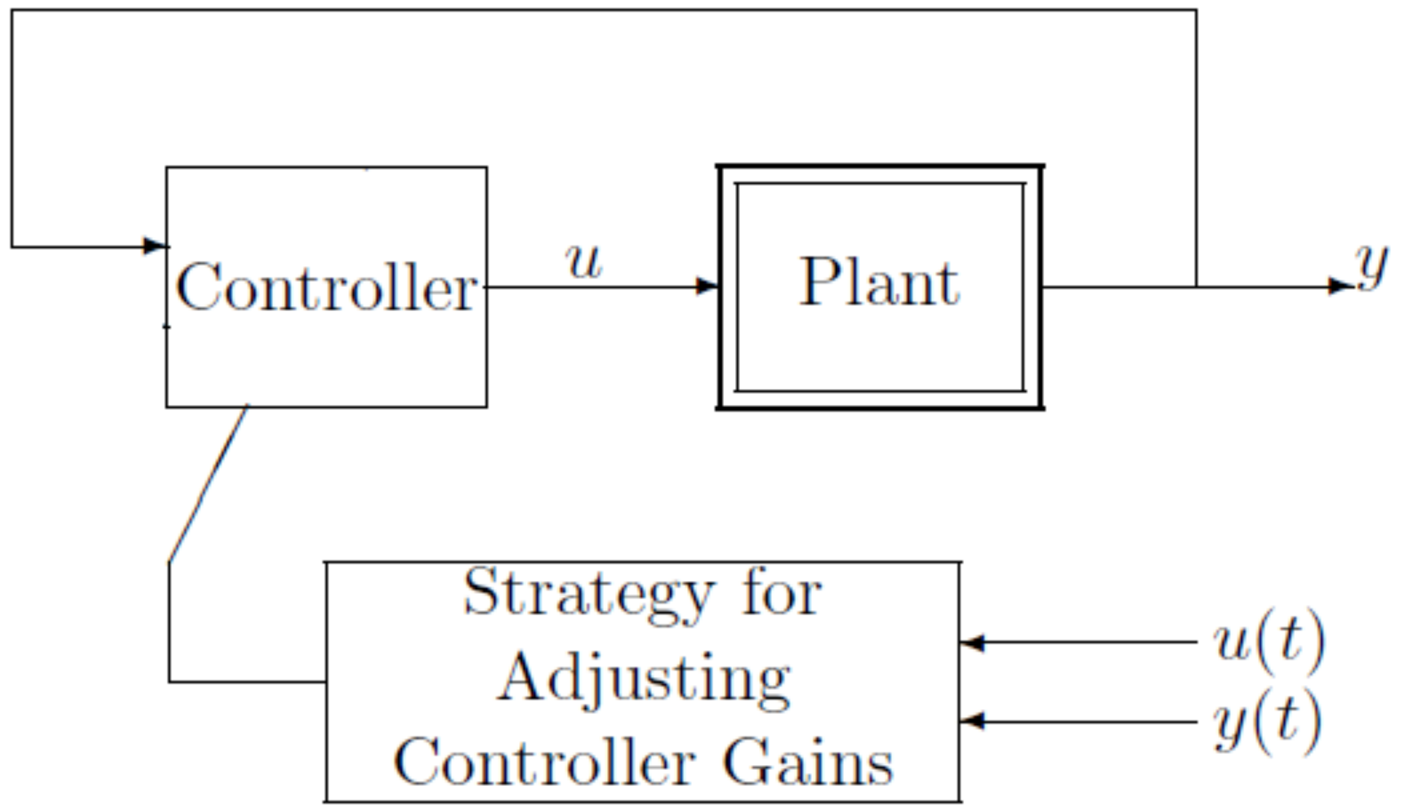
Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

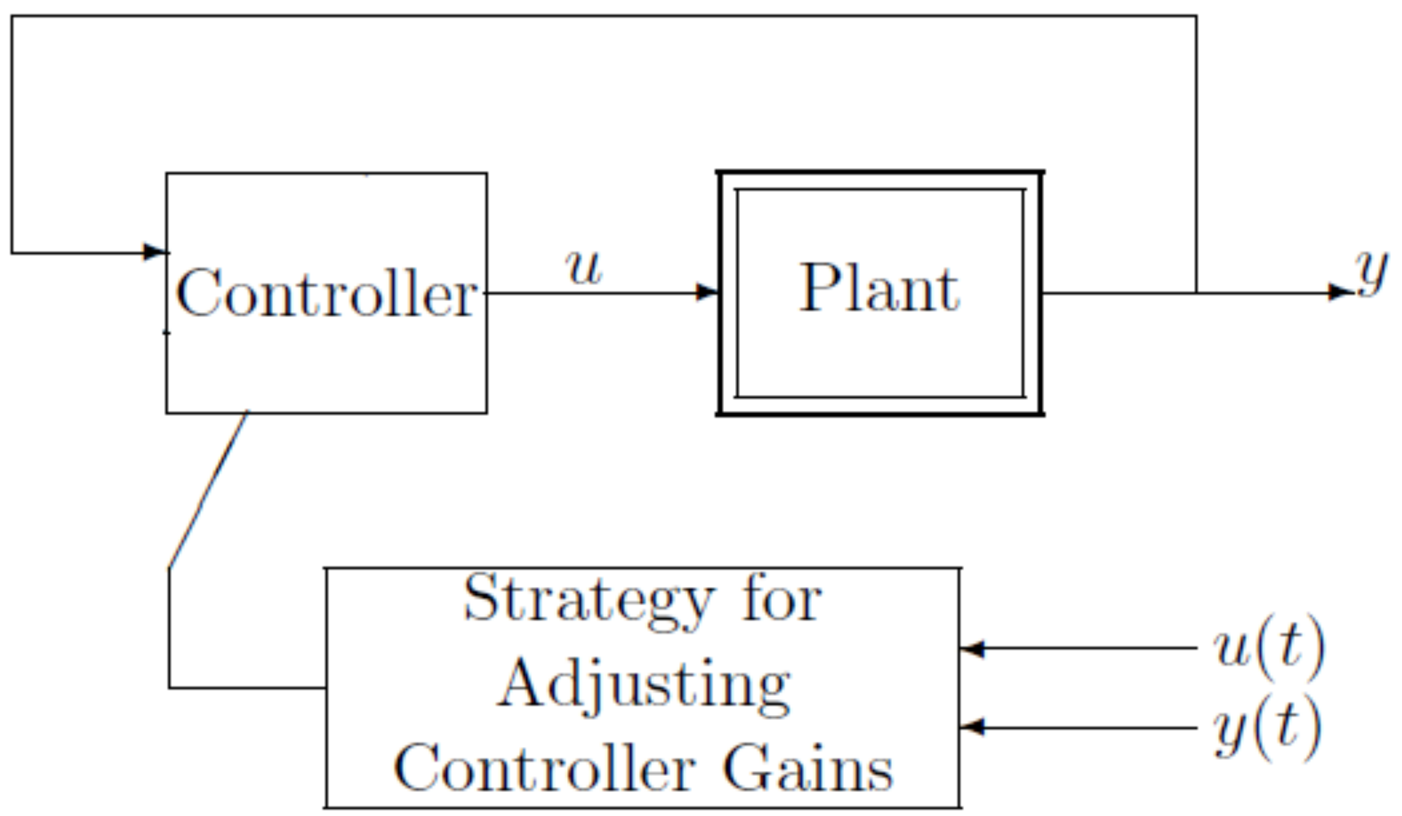
Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$X_{t+1} = AX_t + BU_t + W_t$

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

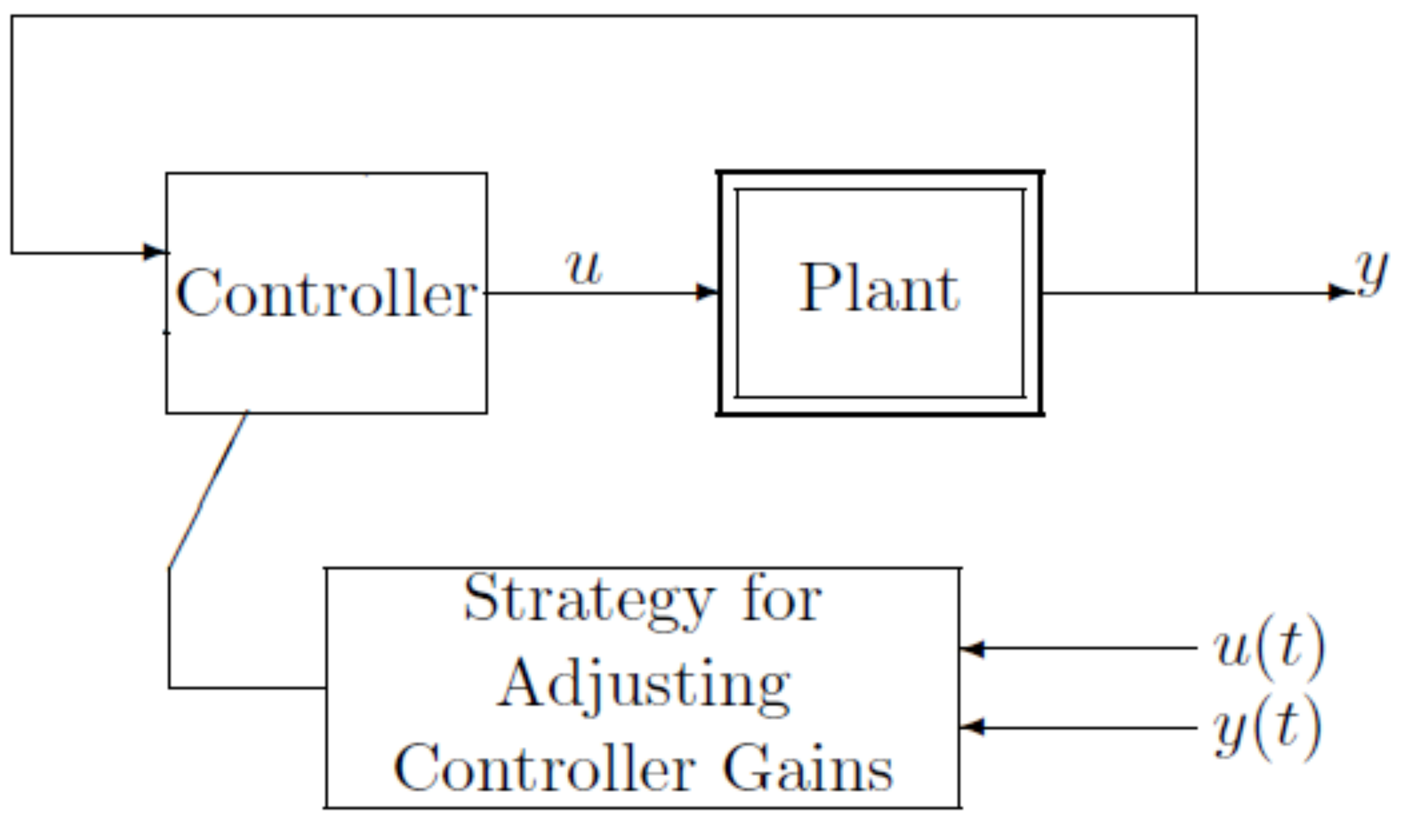
Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$X_{t+1} = AX_t + BU_t + W_t$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

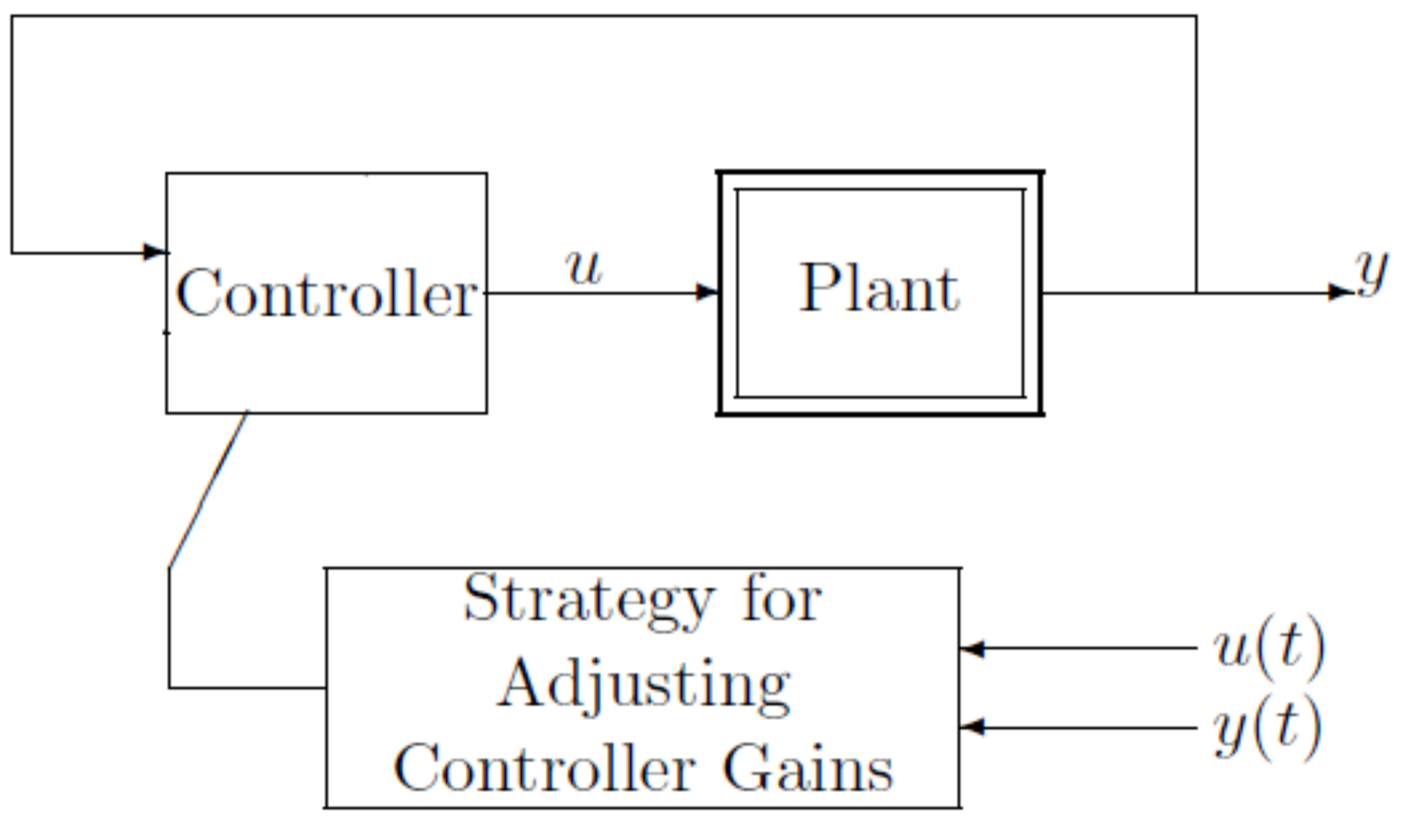
$X_{t+1} = AX_t + BU_t + W_t$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$X_{t+1} = AX_t + BU_t + W_t$

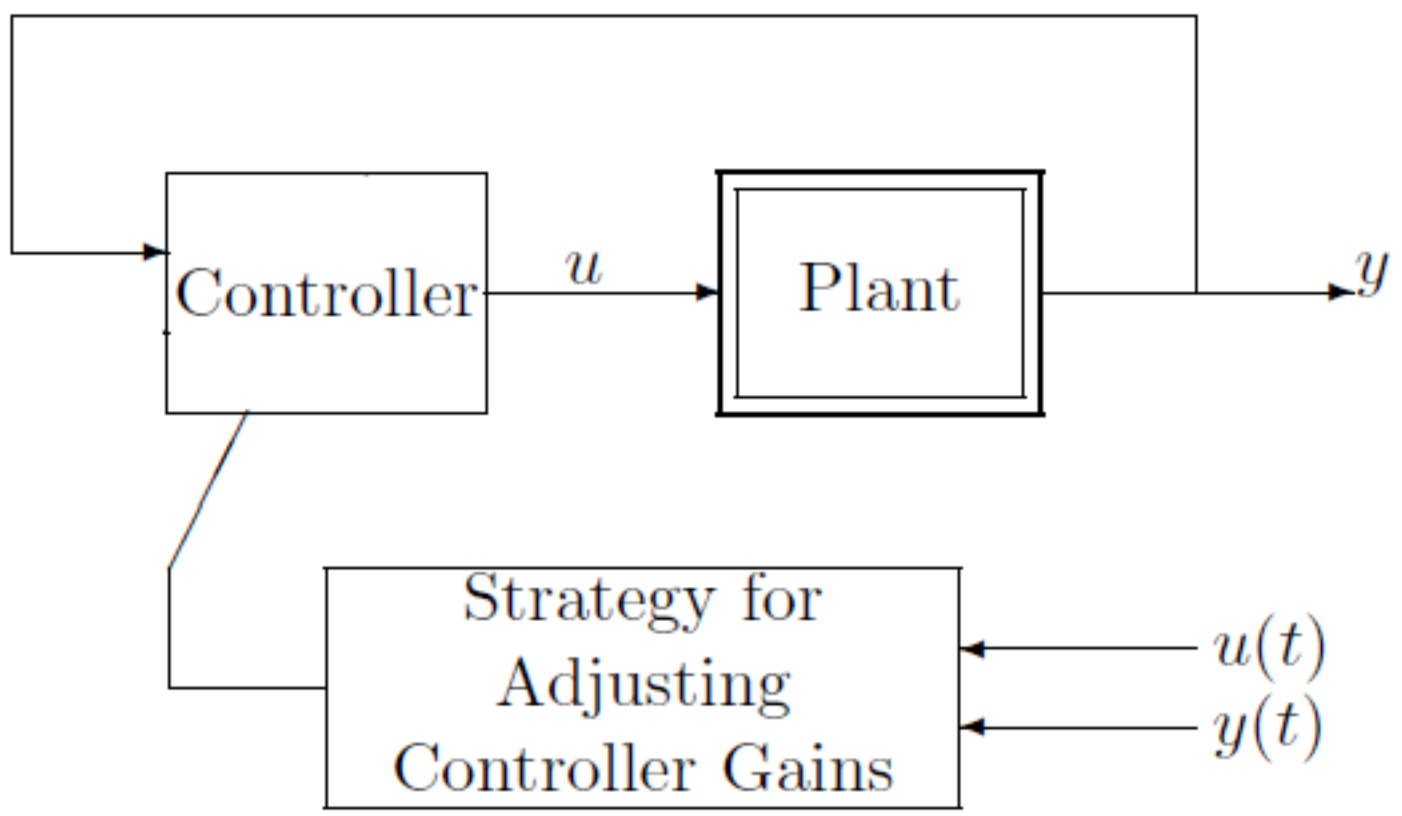
Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}, V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t, \widehat{B}_t$ via regression

$$X_{t+1} = AX_t + BU_t + W_t$$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

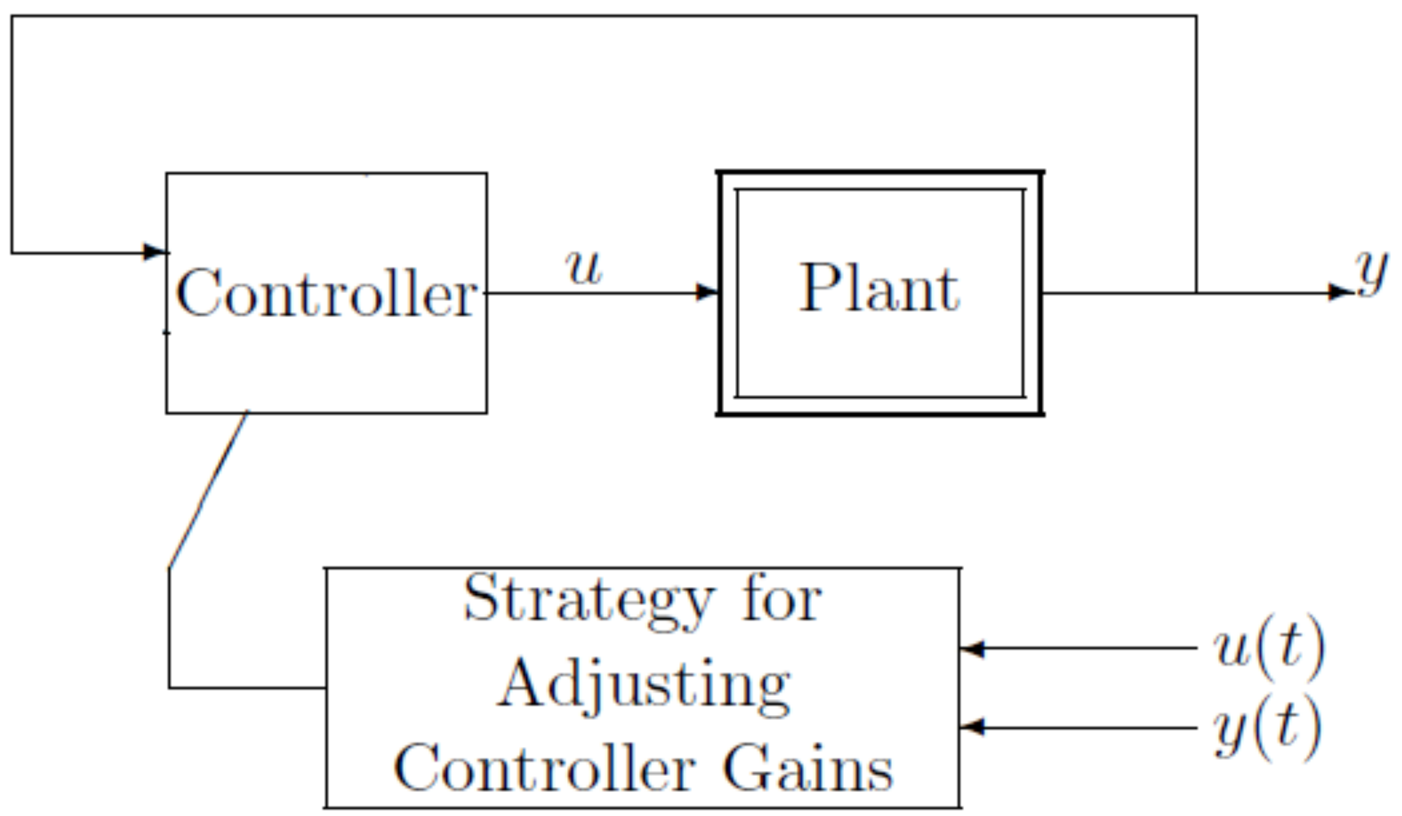
Plug these estimates, $\widehat{A}_t, \widehat{B}_t$ into

$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

$$K(A, B) = -(B^\top P_K B + R)^{-1} B^\top P_K A$$

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$X_{t+1} = AX_t + BU_t + W_t$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

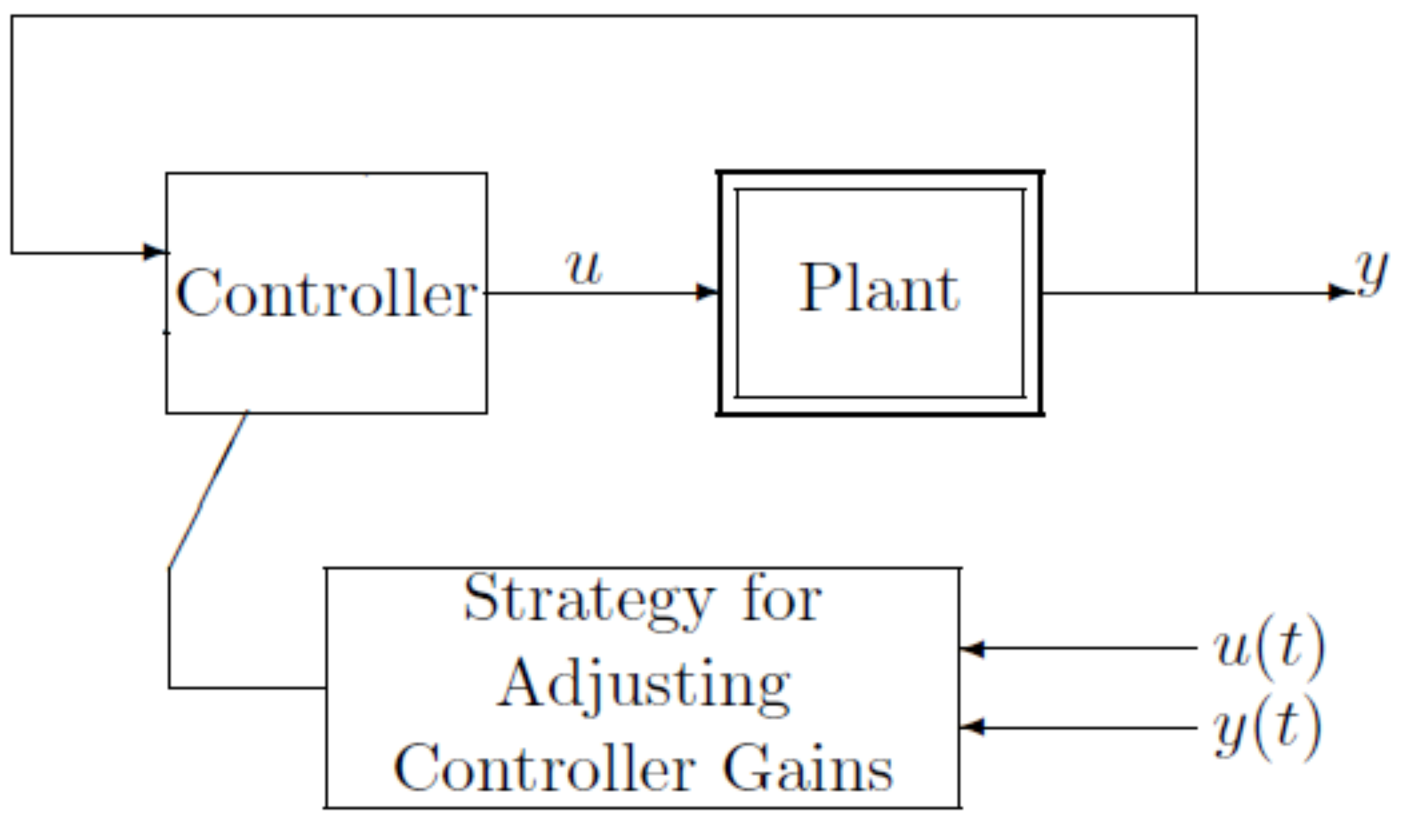
Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

$$K(A, B) = -(B^\top P_K B + R)^{-1} B^\top P_K A$$

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$X_{t+1} = AX_t + BU_t + W_t$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

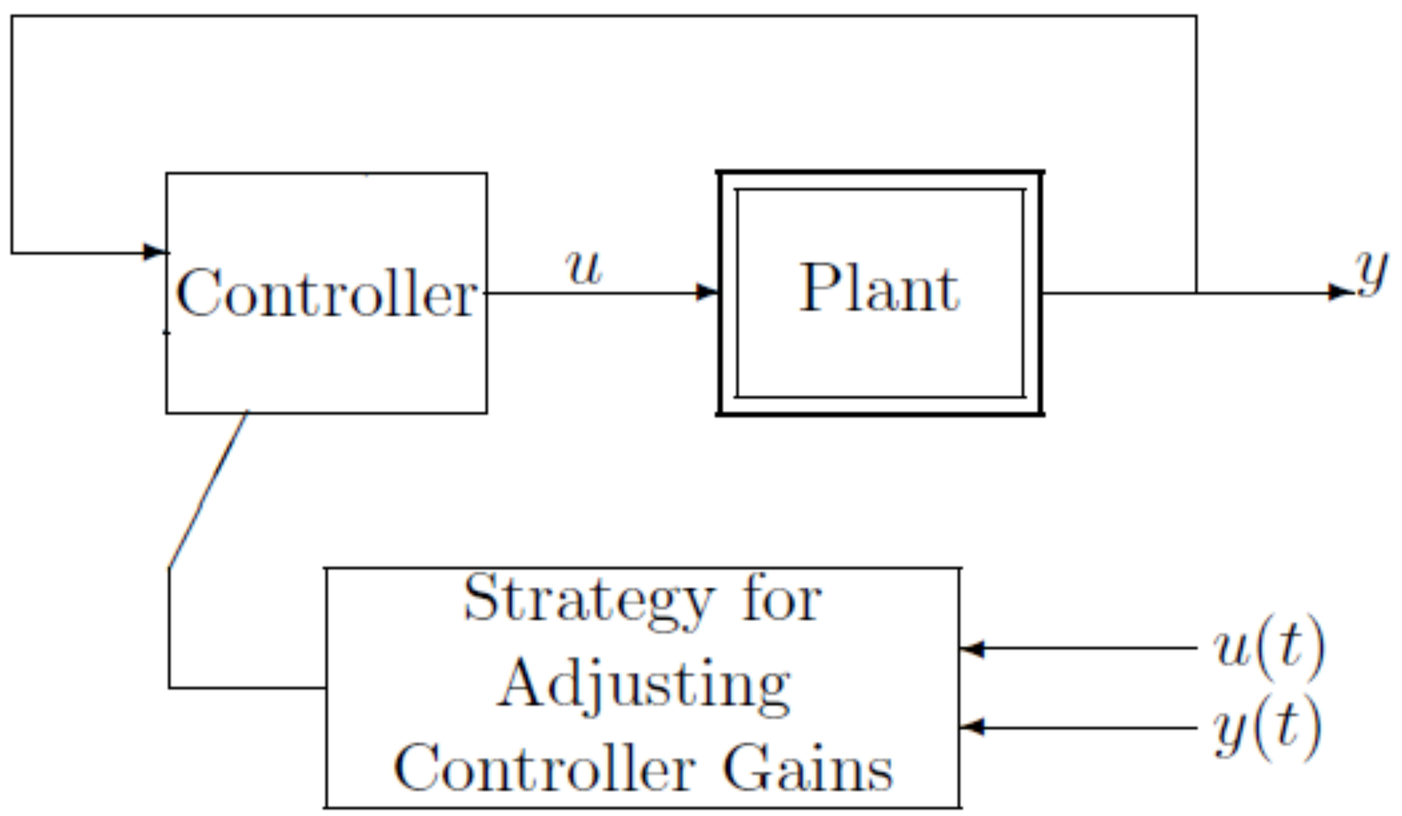
$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

$$K(A, B) = -(B^\top P_K B + R)^{-1} B^\top P_K A$$

So that $\widehat{K}_t = K(\widehat{A}_t, \widehat{B}_t)$

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$X_{t+1} = AX_t + BU_t + W_t$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

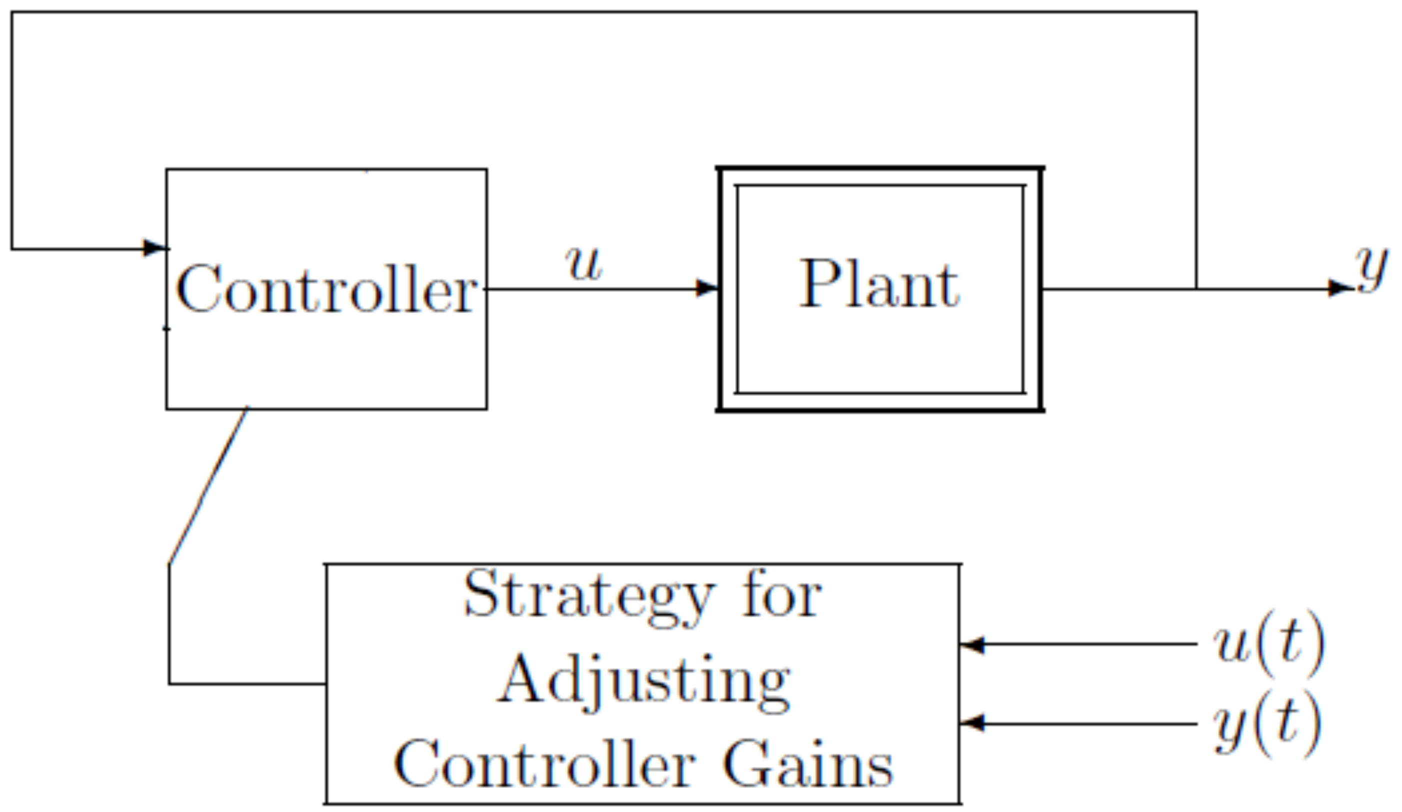
$$K(A, B) = -(B^\top P_K B + R)^{-1} B^\top P_K A$$

So that $\widehat{K}_t = K(\widehat{A}_t, \widehat{B}_t)$

Often use policies of the form:

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$$X_{t+1} = AX_t + BU_t + W_t$$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

$$K(A, B) = -(B^\top P_K B + R)^{-1} B^\top P_K A$$

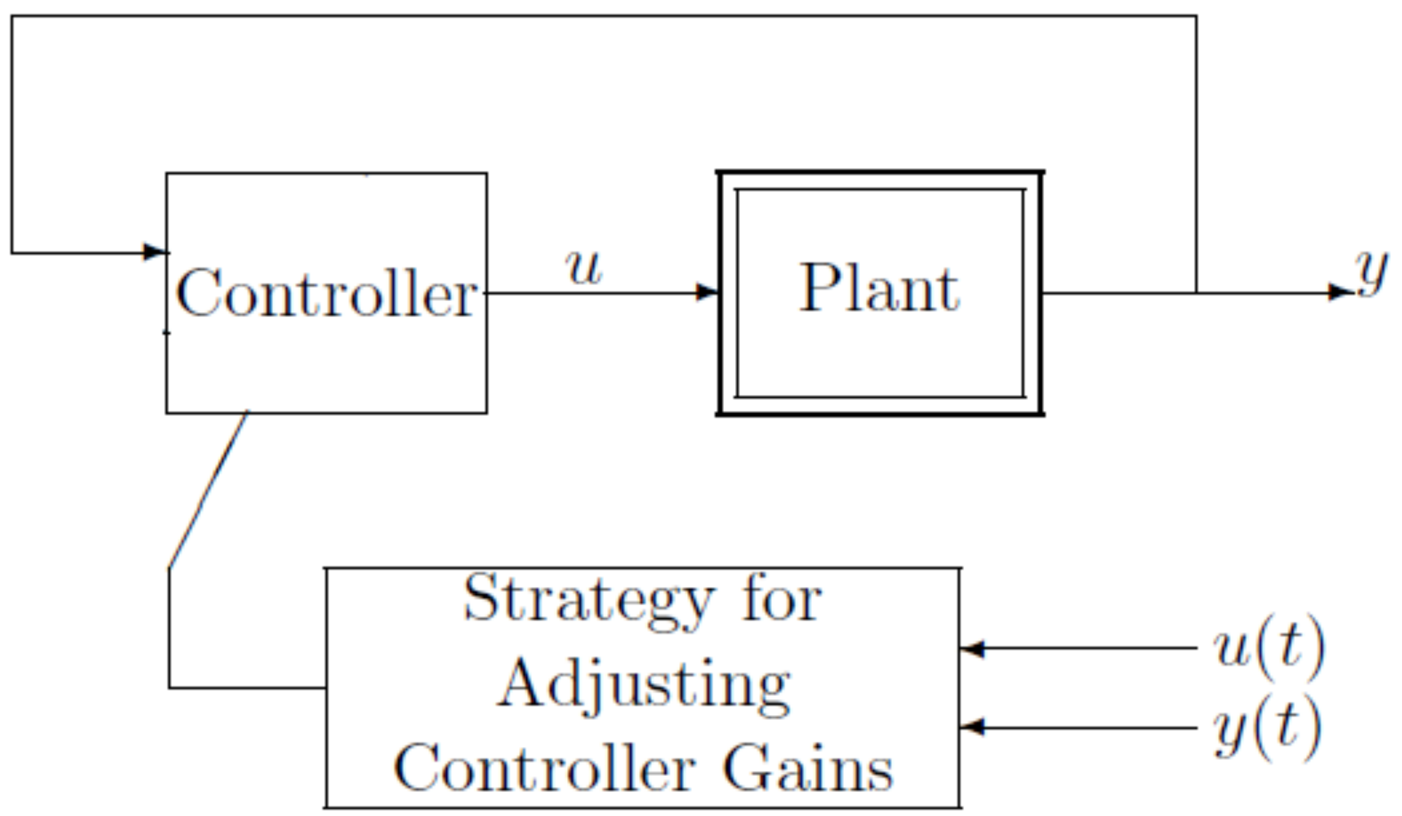
So that $\widehat{K}_t = K(\widehat{A}_t, \widehat{B}_t)$

Often use policies of the form:

$U_t = \widehat{K}_t X_t + \eta_t$ where η_t is some sequence added for "exploration"

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a “reasonable” strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$X_{t+1} = AX_t + BU_t + W_t$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

$$K(A, B) = -(B^\top P_K B + R)^{-1} B^\top P_K A$$

So that $\widehat{K}_t = K(\widehat{A}_t, \widehat{B}_t)$

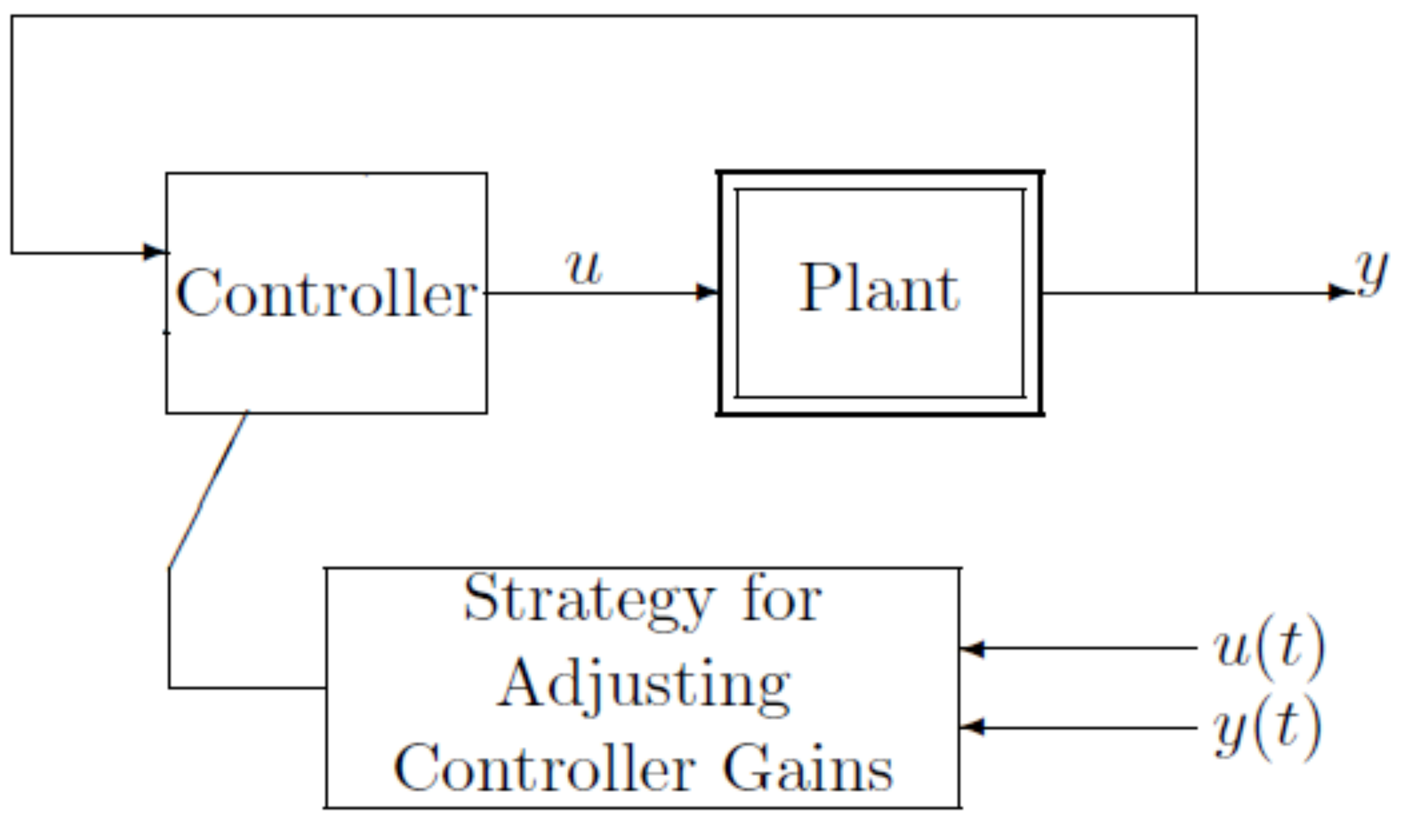
Often use policies of the form:

$U_t = \widehat{K}_t X_t + \eta_t$ where η_t is some sequence added for “exploration”

Such algorithms can achieve $O(\sqrt{T})$ regret

Background on Certainty Equivalence

Unknown System/Plant, (A, B, C):



What strategy for adjusting gains?

Suppose $C = I_{d_x}$, $V_t \equiv 0$

Certainty Equivalence, a "reasonable" strategy:

Form estimates $\widehat{A}_t$, $\widehat{B}_t$ via regression

$$X_{t+1} = AX_t + BU_t + W_t$$

Soooo…. How do we now adjust the gain, $\widehat{K}_t$?

Plug these estimates, $\widehat{A}_t$, $\widehat{B}_t$ into

$$P_K = A + A^\top P_K A - A^\top P_K B(B^\top P_K B + R)^{-1} B^\top P_K A$$

$$K(A, B) = -(B^\top P_K B + R)^{-1} B^\top P_K A$$

So that $\widehat{K}_t = K(\widehat{A}_t, \widehat{B}_t)$

Often use policies of the form:

$U_t = \widehat{K}_t X_t + \eta_t$ where η_t is some sequence added for "exploration"

Such algorithms can achieve $O(\sqrt{T})$ regret

Similar ideas can be used in the partially observed setting but one also has to estimate the Kalman gain / Youla parametrization

Overview of convergence rates, adaptive LQR

Paper	Setting	Upper Bound	Lower Bound	
[AYS11]	SF: (A,B) unknown	$\tilde{O}(\sqrt{T})$ Intractable		
[DMM+18]	SF: (A,B) unknown	$\tilde{O}(T^{2/3})$		
[FTM20/MTR19/ CKM19]	SF: (A,B) unknown	$\tilde{O}(\sqrt{T})$		
[SSH20**]**	PO: (A,B,C) unknown	$\tilde{O}(\sqrt{T})$		
[SF20]	SF: (A,B) unknown	$O\left(\sqrt{d_x d_u^2 T}\right)$	$\Omega\left(\sqrt{d_x d_u^2 T}\right)$	
[**Z**S22]	SF/PO		$\Omega\left(c_{\mathsf{sys}}\sqrt{d_x d_u^2 T}\right)$	Instance dep.
[T**Z**MMP22]	SF: (A,B) unknown	$O\left(\sqrt{c^{d_x} T}\right)$	$\Omega\left(\sqrt{c^{d_x} T}\right)$	Worst Case

NB: big-Oh is potentially hiding system/dimension factors

Proof Outline

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^{\pi}(\theta') - V_T^{\star}(\theta')$.

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^{\pi}(\theta') - V_T^{\star}(\theta').$

First step:

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^{\pi}(\theta') - V_T^{\star}(\theta')$.

First step:

$$\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta) = \sup_{\theta' \in B(\theta,\epsilon)} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t)$$

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^{\pi}(\theta') - V_T^{\star}(\theta')$.

First step:

$$\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta) = \sup_{\theta' \in B(\theta,\epsilon)} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t)$$

$$\geq \mathbf{E}_{\theta' \sim \mu} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t) \qquad \text{for all “priors” } \mu \text{ on } B(\theta, \epsilon)$$

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^{\pi}(\theta') - V_T^{\star}(\theta')$.

First step:

$$\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta) = \sup_{\theta' \in B(\theta,\epsilon)} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t)$$

$$\geq \mathbf{E}_{\theta' \sim \mu} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t) \qquad \text{for all "priors" } \mu \text{ on } B(\theta, \epsilon)$$

Perspective: regard the control input U_t as an estimate of $K(\theta)$ in the "direction" X_t

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^{\pi}(\theta') - V_T^{\star}(\theta')$.

First step:

$$\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta) = \sup_{\theta' \in B(\theta,\epsilon)} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t)$$

$$\geq \mathbf{E}_{\theta' \sim \mu} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t) \qquad \text{for all "priors" } \mu \text{ on } B(\theta, \epsilon)$$

Perspective: regard the control input U_t as an estimate of $K(\theta)$ in the "direction" X_t

Lower bound proceeds by lower bounding the "estimation errors" $(U_t - K(\theta)X_t)(U_t - K(\theta)X_t)^\top \succeq ?$

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^{\pi}(\theta') - V_T^{\star}(\theta').$

First step:

$$\sup_{\theta' \in B(\theta,\epsilon)} R_T^{\pi}(\theta) = \sup_{\theta' \in B(\theta,\epsilon)} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t)$$

$$\geq \mathbf{E}_{\theta' \sim \mu} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^{\pi}(U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t) \qquad \text{for all "priors" } \mu \text{ on } B(\theta, \epsilon)$$

Perspective: regard the control input U_t as an estimate of $K(\theta)$ in the "direction" X_t

Lower bound proceeds by lower bounding the "estimation errors" $(U_t - K(\theta)X_t)(U_t - K(\theta)X_t)^\top \succeq ?$

Next: detour to information-theoretic lower bounds for (Bayesian) estimation problems

Proof Outline

Want to lower bound $\sup_{\theta' \in B(\theta,\epsilon)} R_T^\pi(\theta') = \sup_{\theta' \in B(\theta,\epsilon)} V_T^\pi(\theta') - V_T^\star(\theta')$.

First step:

$$\sup_{\theta' \in B(\theta,\epsilon)} R_T^\pi(\theta) = \sup_{\theta' \in B(\theta,\epsilon)} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^\pi (U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t)$$

$$\geq \mathbf{E}_{\theta' \sim \mu} \sum_{t=1}^{T-1} \mathbf{E}_{\theta'}^\pi (U_t - K(\theta')X_t)^\top (B^\top P(\theta')B + R)(U_t - K(\theta')X_t) \qquad \text{for all "priors" } \mu \text{ on } B(\theta, \epsilon)$$

Perspective: regard the control input U_t as an estimate of $K(\theta)$ in the "direction" X_t

Lower bound proceeds by lower bounding the "estimation errors" $(U_t - K(\theta)X_t)(U_t - K(\theta)X_t)^\top \succeq ?$

Next: detour to information-theoretic lower bounds for (Bayesian) estimation problems

Question for the audience: why can't we lower bound $R_T^\pi(\theta)$ without the sup?

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Heuristically this becomes:

$$R_T^{\pi}(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Heuristically this becomes:

$$R_T^{\pi}(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Van Trees Inequality states that:

Heuristically this becomes:

$$R_T^{\pi}(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Van Trees Inequality states that:

$$\mathrm{vec}(\widehat{K}_t - K(\theta'))\mathrm{vec}(\widehat{K}_t - K(\theta'))^\top \succeq$$

Heuristically this becomes:

$$R_T^{\pi}(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Van Trees Inequality states that:

$$\mathrm{vec}(\widehat{K}_t - K(\theta'))\mathrm{vec}(\widehat{K}_t - K(\theta'))^\top \succeq$$

$$\mathsf{D}_\theta \mathrm{vec} K(\theta) \left(\mathbf{E}_{\theta' \sim \mu} \mathtt{I}(\theta, \pi) + \mathsf{J}(\mu) \right)^{-1} \mathsf{D}_\theta \mathrm{vec} K(\theta)$$

Heuristically this becomes:

$$R_T^{\pi}(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi)) \right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Van Trees Inequality states that:

$$\mathrm{vec}(\widehat{K}_t - K(\theta'))\mathrm{vec}(\widehat{K}_t - K(\theta'))^\top \succeq$$

$$\mathsf{D}_\theta \mathrm{vec} K(\theta)\left(\mathbf{E}_{\theta' \sim \mu}\mathtt{I}(\theta, \pi) + \mathsf{J}(\mu)\right)^{-1}\mathsf{D}_\theta \mathrm{vec} K(\theta)$$

Where $\mathsf{J}(\mu)$ can be thought of as a constant depending only on prior radius ϵ and:

Heuristically this becomes:

$$R_T^\pi(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Van Trees Inequality states that:

$$\mathrm{vec}(\widehat{K}_t - K(\theta'))\mathrm{vec}(\widehat{K}_t - K(\theta'))^\top \succeq$$

$$\mathsf{D}_\theta \mathrm{vec} K(\theta)\left(\mathbf{E}_{\theta' \sim \mu} \mathtt{I}(\theta, \pi) + \mathsf{J}(\mu)\right)^{-1} \mathsf{D}_\theta \mathrm{vec} K(\theta)$$

Where $\mathsf{J}(\mu)$ can be thought of as a constant depending only on prior radius ϵ and:

for the special case $\mathrm{vec}(A, B) = \theta \quad \Rightarrow \quad \mathtt{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^\pi Z_t Z_t^\top \otimes \Sigma_W^{-1}$ with $Z_t^\top = \begin{bmatrix} X_t & U_t \end{bmatrix}$

Heuristically this becomes:

$$R_T^\pi(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Van Trees Inequality states that:

$$\mathrm{vec}(\widehat{K}_t - K(\theta'))\mathrm{vec}(\widehat{K}_t - K(\theta'))^\top \succeq$$

$$\mathsf{D}_\theta \mathrm{vec} K(\theta)\left(\mathbf{E}_{\theta' \sim \mu} \mathtt{I}(\theta, \pi) + \mathsf{J}(\mu)\right)^{-1} \mathsf{D}_\theta \mathrm{vec} K(\theta)$$

Where $\mathsf{J}(\mu)$ can be thought of as a constant depending only on prior radius ϵ and:

for the special case $\mathrm{vec}(A, B) = \theta \quad \Rightarrow \quad \mathtt{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^\pi Z_t Z_t^\top \otimes \Sigma_W^{-1}$ with $Z_t^\top = \begin{bmatrix} X_t & U_t \end{bmatrix}$

Remark: connection to least squares estimate

Heuristically this becomes:

$$R_T^\pi(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Detour to Estimation Theory

Van Trees Inequality (AKA Bayesian Cramer-Rao)

Want to estimate $K(\theta')X_t$ given information $X_0, U_0, X_1, U_1, \ldots, U_{t-1}, X_t$ and prior μ

Van Trees Inequality states that:

$$\mathrm{vec}(\widehat{K}_t - K(\theta'))\mathrm{vec}(\widehat{K}_t - K(\theta'))^\top \succeq$$

$$\mathsf{D}_\theta \mathrm{vec} K(\theta)\left(\mathbf{E}_{\theta' \sim \mu} \mathtt{I}(\theta, \pi) + \mathsf{J}(\mu)\right)^{-1} \mathsf{D}_\theta \mathrm{vec} K(\theta)$$

Where $\mathsf{J}(\mu)$ can be thought of as a constant depending only on prior radius ϵ and:

for the special case $\mathrm{vec}(A, B) = \theta \quad \Rightarrow \quad \mathtt{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^\pi Z_t Z_t^\top \otimes \Sigma_W^{-1}$ with $Z_t^\top = \begin{bmatrix} X_t & U_t \end{bmatrix}$

Remark: connection to least squares estimate

$$\begin{bmatrix} \widehat{A}_t & \widehat{B}_t \end{bmatrix} = \left(\sum_{t=0}^{t-1} Z_t Z_t\right)^{-1} \sum_{t=0} Z_t X_{t+1}^\top$$

Heuristically this becomes:

$$R_T^\pi(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathtt{I}(\theta, \pi))\right)^{-1}$$

Regret and closed loop identifiability

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathtt{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_{\theta}^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_{\theta}^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathbf{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Singular!

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathrm{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_{\theta}^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_{\theta}^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Singular!

So what if $\mathrm{I}(\theta, \pi_\star)$ is singular? what about $\mathrm{I}(\theta, \pi)$ for arbitrary π?

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathrm{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Singular!

So what if $\mathrm{I}(\theta, \pi_\star)$ is singular? what about $\mathrm{I}(\theta, \pi)$ for arbitrary π?

*Fact [**Z**S22]:* $\lambda_{\min}(\mathrm{I}(\theta, \pi)) \lesssim R_T^\pi(\theta)$ whenever $\mathrm{I}(\theta, \pi_\star)$ is singular

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathrm{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Singular!

So what if $\mathrm{I}(\theta, \pi_\star)$ is singular? what about $\mathrm{I}(\theta, \pi)$ for arbitrary π?

*Fact [**Z**S22]:* $\lambda_{\min}(\mathrm{I}(\theta, \pi)) \lesssim R_T^\pi(\theta)$ whenever $\mathrm{I}(\theta, \pi_\star)$ is singular

Controllers with lower regret lead to poorer excitation ⇒ LQR has issues with closed loop identifiability

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathrm{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Singular!

So what if $\mathrm{I}(\theta, \pi_\star)$ is singular? what about $\mathrm{I}(\theta, \pi)$ for arbitrary π?

*Fact [**Z**S22]:* $\lambda_{\min}(\mathrm{I}(\theta, \pi)) \lesssim R_T^\pi(\theta)$ whenever $\mathrm{I}(\theta, \pi_\star)$ is singular

Controllers with lower regret lead to poorer excitation ⇒ LQR has issues with closed loop identifiability

Exploration vs Exploitation tradeoff is nontrivial

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathrm{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Singular!

So what if $\mathrm{I}(\theta, \pi_\star)$ is singular? what about $\mathrm{I}(\theta, \pi)$ for arbitrary π?

*Fact [**Z**S22]:* $\lambda_{\min}(\mathrm{I}(\theta, \pi)) \lesssim R_T^\pi(\theta)$ whenever $\mathrm{I}(\theta, \pi_\star)$ is singular

Controllers with lower regret lead to poorer excitation ⇒ LQR has issues with closed loop identifiability

Exploration vs Exploitation tradeoff is nontrivial

This gives us the proof sketch: $R_T^\pi(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathrm{I}(\theta, \pi))\right)^{-1}$ and $\lambda_{\min}(\mathrm{I}(\theta, \pi)) \lesssim R_T^\pi(\theta)$

Regret and closed loop identifiability

Recall that we hope that under our policy π that $U_t \approx K_\star X_t$

Now for $\pi_\star$ we find that:

$$\mathrm{I}(\theta, \pi) = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} Z_t Z_t^\top \otimes \Sigma_W^{-1} = \sum_{t=1}^{T-1} \mathbf{E}_\theta^{\pi_\star} \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix} X_t X_t^\top \begin{bmatrix} I_{d_x} \\ K_\star \end{bmatrix}^\top \otimes \Sigma_W^{-1}$$

Singular!

So what if $\mathrm{I}(\theta, \pi_\star)$ is singular? what about $\mathrm{I}(\theta, \pi)$ for arbitrary π?

*Fact [**Z**S22]:* $\lambda_{\min}(\mathrm{I}(\theta, \pi)) \lesssim R_T^\pi(\theta)$ whenever $\mathrm{I}(\theta, \pi_\star)$ is singular

Controllers with lower regret lead to poorer excitation ⇒ LQR has issues with closed loop identifiability

Exploration vs Exploitation tradeoff is nontrivial

This gives us the proof sketch: $R_T^\pi(\theta) \gtrsim T \times \left(\lambda_{\min}(\mathrm{I}(\theta, \pi))\right)^{-1}$ and $\lambda_{\min}(\mathrm{I}(\theta, \pi)) \lesssim R_T^\pi(\theta)$

Necessary condition for both inequalities to hold: $R_T^\pi(\theta) \gtrsim \sqrt{T}$

Result for the State Feedback Setting

Result for the State Feedback Setting

Theorem (informal) [**Z**S22]

Result for the State Feedback Setting

Theorem (informal) [**Z**S22]

Let $\theta = \mathrm{vec}[A\ \ B]$. For every policy π, $\alpha \in (0,1/4)$ we have that:

Result for the State Feedback Setting

Theorem (informal) [**Z**S22]

Let $\theta = \text{vec}[A \;\; B]$. For every policy π, $\alpha \in (0,1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x d_u} \times \sqrt{T} \times \sigma_{\min}\left(P(\theta)\right)$$

Result for the State Feedback Setting

Theorem (informal) [**Z**S22]

Let $\theta = \mathrm{vec}[A \;\; B]$. For every policy π, $\alpha \in (0,1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x d_u} \times \sqrt{T} \times \sigma_{\min}\left(P(\theta)\right)$$

Result for the State Feedback Setting

Theorem (informal) [**Z**S22]

Let $\theta = \mathrm{vec}[A \;\; B]$. For every policy π, $\alpha \in (0,1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x d_u} \times \sqrt{T} \times \sigma_{\min}\left(P(\theta)\right)$$

Closed loop close to marginal stability, RHS above diverges

Result for the State Feedback Setting

Theorem (informal) [**Z**S22]

Let $\theta = \text{vec}[A \;\; B]$. For every policy π, $\alpha \in (0,1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x d_u} \times \sqrt{T} \times \sigma_{\min}\left(P(\theta)\right)$$

Closed loop close to marginal stability, RHS above diverges

Systems that are "hard/costly to control" are also *hard to learn to control*

Scalar Systems Illustration

Scalar Systems Illustration

Consider a scalar system

Scalar Systems Illustration

Consider a scalar system

$$X_{t+1} = X_t + bU_t + W_t$$

Scalar Systems Illustration

Consider a scalar system

$$X_{t+1} = X_t + bU_t + W_t$$

Our lower bound (blue) diverges

Scalar Systems Illustration

Consider a scalar system

$$X_{t+1} = X_t + bU_t + W_t$$

Our lower bound (blue) diverges

when control authority is lost

Scalar Systems Illustration

Consider a scalar system

$$X_{t+1} = X_t + bU_t + W_t$$

Our lower bound (blue) diverges

when control authority is lost

Scalar Systems Illustration

Consider a scalar system

$$X_{t+1} = X_t + bU_t + W_t$$

Our lower bound (blue) diverges

when control authority is lost

Contrast with previous result

Scalar Systems Illustration

Consider a scalar system

$$X_{t+1} = X_t + bU_t + W_t$$

Our lower bound (blue) diverges

when control authority is lost

Contrast with previous result

[SF20]

Scalar Systems Illustration

Consider a scalar system

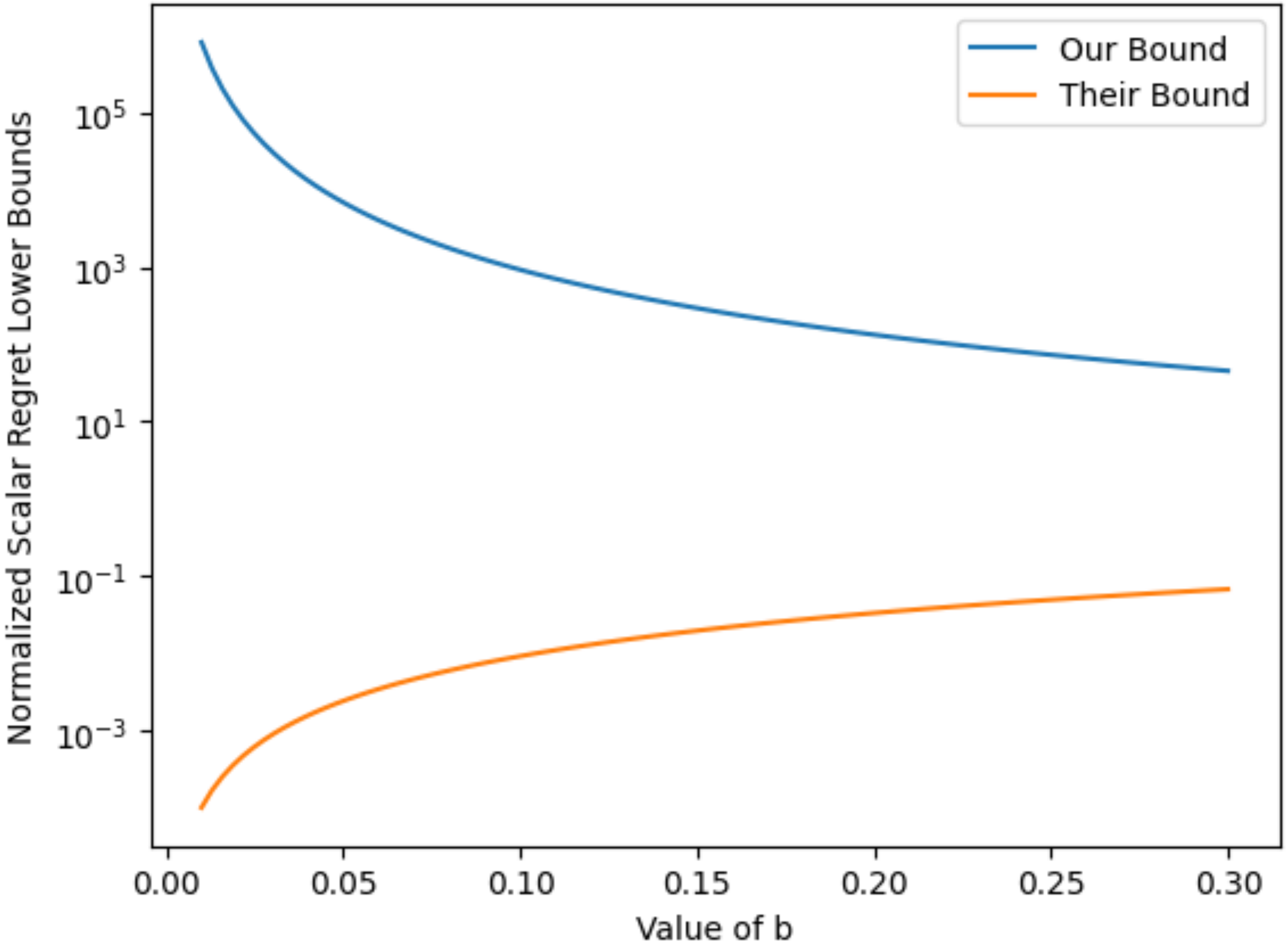
$$X_{t+1} = X_t + bU_t + W_t$$

Our lower bound (blue) diverges

when control authority is lost

Contrast with previous result

[SF20]



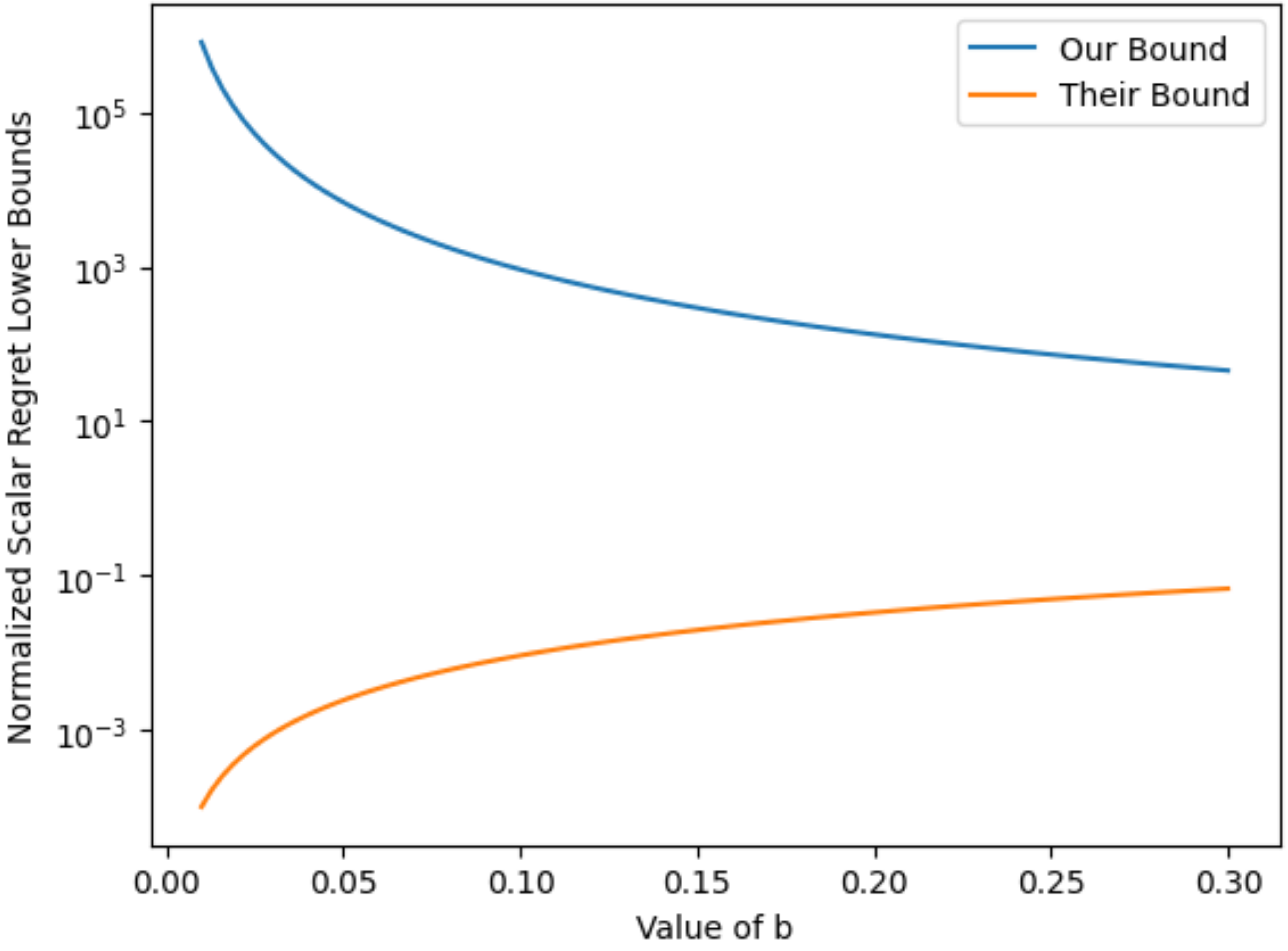
Scalar Systems Illustration

Consider a scalar system

$$X_{t+1} = X_t + bU_t + W_t$$

Our lower bound (blue) diverges when control authority is lost

Contrast with previous result [SF20]



NB: Time horizon T is fixed

State Feedback Systems

Curse of Dimensionality

State Feedback Systems

Curse of Dimensionality

Consider the following “integrator” ($\theta = \text{vec}[A \;\; B]$):

State Feedback Systems

Curse of Dimensionality

Consider the following “integrator” ($\theta = \mathrm{vec}[A\ \ B]$):

$$A = \begin{bmatrix} 0 & 0 & 0 & & 0 & 0 \\ 0 & 1 & 1 & & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \\ 0 & 1 \end{bmatrix}, \quad Q = I_{dx}, \quad R = I_2$$

State Feedback Systems

Curse of Dimensionality

Consider the following "integrator" ($\theta = \mathrm{vec}[A \;\; B]$):

$$A = \begin{bmatrix} 0 & 0 & 0 & & 0 & 0 \\ 0 & 1 & 1 & & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \\ 0 & 1 \end{bmatrix}, \quad Q = I_{d_x}, \quad R = I_2$$

For this system, we show in [T**Z**M+22] that: $\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim 2^{d_x}\sqrt{T}$

State Feedback Systems

Curse of Dimensionality

Consider the following "integrator" ($\theta = \mathrm{vec}[A \;\; B]$):

$$A = \begin{bmatrix} 0 & 0 & 0 & & 0 & 0 \\ 0 & 1 & 1 & & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \\ 0 & 1 \end{bmatrix}, \quad Q = I_{dx}, \quad R = I_2$$

For this system, we show in [T**Z**M+22] that: $\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim 2^{d_x}\sqrt{T}$

More generally, worst case regret is exponential in the so-called *controllability index*

State Feedback Systems

Curse of Dimensionality

Consider the following "integrator" ($\theta = \mathrm{vec}[A \;\; B]$):

$$A = \begin{bmatrix} 0 & 0 & 0 & & 0 & 0 \\ 0 & 1 & 1 & & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & & 1 & 1 \\ 0 & 0 & 0 & & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \\ 0 & 1 \end{bmatrix}, \quad Q = I_{dx}, \quad R = I_2$$

For this system, we show in [T**Z**M+22] that: $\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim 2^{d_x}\sqrt{T}$

More generally, worst case regret is exponential in the so-called *controllability index*

Hard to Control ⇒ Hard to Learn to Control

Result for the Partially Observed Setting

Result for the Partially Observed Setting

Theorem (informal) [**Z**S22]

Result for the Partially Observed Setting

Theorem (informal) [**Z**S22]

There exists a $\theta = \mathrm{vec}[A \ \ B \ \ C]$ such that for every policy π, $\alpha \in (0, 1/4)$ we have that:

Result for the Partially Observed Setting

Theorem (informal) [**Z**S22]

There exists a $\theta = \mathrm{vec}[A \;\; B \;\; C]$ such that for every policy π, $\alpha \in (0, 1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x} d_u \times \sqrt{T} \times \sigma_{\min}\left(\Sigma_\nu(\theta)\right)$$

Result for the Partially Observed Setting

Theorem (informal) [**Z**S22]

There exists a $\theta = \text{vec}[A \ B \ C]$ such that for every policy π, $\alpha \in (0,1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x} d_u \times \sqrt{T} \times \sigma_{\min}\left(\Sigma_\nu(\theta)\right)$$

$\Sigma_\nu(\theta)$ covariance matrix of the innovations process

Result for the Partially Observed Setting

Theorem (informal) [**Z**S22]

There exists a $\theta = \mathrm{vec}[A \ B \ C]$ such that for every policy π, $\alpha \in (0,1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x d_u} \times \sqrt{T} \times \sigma_{\min}\left(\Sigma_\nu(\theta)\right)$$

$\Sigma_\nu(\theta)$ covariance matrix of the innovations process

Result for the Partially Observed Setting

Theorem (informal) [**Z**S22]

There exists a $\theta = \mathrm{vec}[A \;\; B \;\; C]$ such that for every policy π, $\alpha \in (0, 1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x} d_u \times \sqrt{T} \times \sigma_{\min}\left(\Sigma_\nu(\theta)\right)$$

$\Sigma_\nu(\theta)$ covariance matrix of the innovations process

Poor detectability of unstable modes, RHS above diverges

Result for the Partially Observed Setting

Theorem (informal) [**Z**S22]

There exists a $\theta = \mathrm{vec}[A \;\; B \;\; C]$ such that for every policy π, $\alpha \in (0,1/4)$ we have that:

$$\sup_{\theta' \in B(\theta, T^{-\alpha})} R_T^{\pi}(\theta') \gtrsim \sqrt{d_x} d_u \times \sqrt{T} \times \sigma_{\min}\left(\Sigma_\nu(\theta)\right)$$

$\Sigma_\nu(\theta)$ covariance matrix of the innovations process

Poor detectability of unstable modes, RHS above diverges

Systems that are “poorly observable” are also *hard to learn to control*

Takeaways

Takeaways

L4DC problems can often be cast as problems in (statistical) decision theory

Takeaways

L4DC problems can often be cast as problems in (statistical) decision theory

Well established theory of lower bounds (fundamental performance limits) for these by reduction to estimation/hypotheses testing

Takeaways

L4DC problems can often be cast as problems in (statistical) decision theory

Well established theory of lower bounds (fundamental performance limits) for these by reduction to estimation/hypotheses testing

Applied this observation to stochastic adaptive control

Takeaways

L4DC problems can often be cast as problems in (statistical) decision theory

Well established theory of lower bounds (fundamental performance limits) for these by reduction to estimation/hypotheses testing

Applied this observation to stochastic adaptive control

Characterized exploration-exploitation trade-off through a balance of regret and Fisher Information

Takeaways

L4DC problems can often be cast as problems in (statistical) decision theory

Well established theory of lower bounds (fundamental performance limits) for these by reduction to estimation/hypotheses testing

Applied this observation to stochastic adaptive control

Characterized exploration-exploitation trade-off through a balance of regret and Fisher Information

Regret typically scales as $\sqrt{d_x d_u^2 T}$ and increases further with poor controllability

References

Regret Lower Bounds for Learning Linear Quadratic Gaussian Systems, **Ingvar Ziemann** and Henrik Sandberg, to appear, IEEE Transactions on Automatic Control // [**Z**S22]

Simchowitz, Max, and Dylan Foster. "Naive exploration is optimal for online lqr." *International Conference on Machine Learning*. PMLR, 2020. // [SF20]

Ziemann, Ingvar, Anastasios Tsiamis, Henrik Sandberg, and Nikolai Matni. "How are policy gradient methods affected by the limits of control?." In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 5992-5999. IEEE, 2022.

Lee, Bruce D., **Ingvar Ziemann**, Anastasios Tsiamis, Henrik Sandberg, and Nikolai Matni. "The fundamental limitations of learning linear-quadratic regulators." In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 4053-4060. IEEE, 2023.

Statistical Learning Theory for Control, Anastasios Tsiamis, **Ingvar Ziemann**, Nikolai Matni and George J. Pappas, IEEE Control Systems Magazine 2023

Ziemann, Ingvar, Anastasios Tsiamis, Bruce Lee, Yassir Jedra, Nikolai Matni, and George J. Pappas. "A Tutorial on the Non-Asymptotic Theory of System Identification." In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pp. 8921-8939. IEEE, 2023.

Further Reading:

Wainwright, Martin J. *High-dimensional statistics: A non-asymptotic viewpoint*. Vol. 48. Cambridge university press, 2019.

Polyanskiy, Yury, and Yihong Wu. "Lecture notes on information theory." *Lecture Notes for ECE563 (UIUC) and* 6.2012-2016 (2014): 7.