How are policy gradient methods affected by the limits of control?

I. Ziemann
KTH

Anastasios Tsiamis
ETH

Henrik Sandberg
KTH

Nikolai Matni
UPenn

Related Work

Large literature in RL:

Related Work

Large literature in RL:

Related Work

Large literature in RL:

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

In controls:

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

In controls:

PG converges in LDS [Fazel et al., 2018]

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

In controls:

PG converges in LDS [Fazel et al., 2018]

Variants: Zhang et al. [2020a], Gravell et al. [2019, 2020], Zhang et al. [2020b], Yaghmaie et al. [2022]

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

In controls:

PG converges in LDS [Fazel et al., 2018]

Variants: Zhang et al. [2020a], Gravell et al. [2019, 2020], Zhang et al. [2020b], Yaghmaie et al. [2022]

Extensions to partially observed systems: Tang et al. [2021], Mohammadi et al. [2021], Zheng et al. [2021]

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

In controls:

PG converges in LDS [Fazel et al., 2018]

Variants: Zhang et al. [2020a], Gravell et al. [2019, 2020], Zhang et al. [2020b], Yaghmaie et al. [2022]

Extensions to partially observed systems: Tang et al. [2021], Mohammadi et al. [2021], Zheng et al. [2021]

Comparison between PG and CE: Tu and Recht [2019]

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

In controls:

PG converges in LDS [Fazel et al., 2018]

Variants: Zhang et al. [2020a], Gravell et al. [2019, 2020], Zhang et al. [2020b], Yaghmaie et al. [2022]

Extensions to partially observed systems: Tang et al. [2021], Mohammadi et al. [2021], Zheng et al. [2021]

Comparison between PG and CE: Tu and Recht [2019]

Most of these focus on upper bounds

Related Work

Large literature in RL:

DeepMind's alphaGo Silver et al. [2017]

In controls:

PG converges in LDS [Fazel et al., 2018]

Variants: Zhang et al. [2020a], Gravell et al. [2019, 2020], Zhang et al. [2020b], Yaghmaie et al. [2022]

Extensions to partially observed systems: Tang et al. [2021], Mohammadi et al. [2021], Zheng et al. [2021]

Comparison between PG and CE: Tu and Recht [2019]

Most of these focus on upper bounds

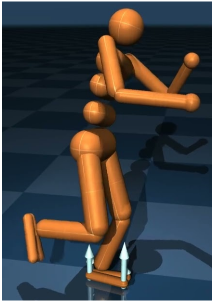
What about fundamental limits? ☕

Why Fundamental Limits?

Ambition:

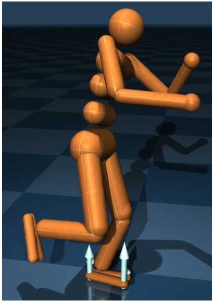
Why Fundamental Limits?

Ambition:



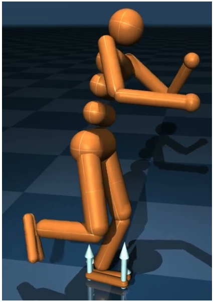
Why Fundamental Limits?

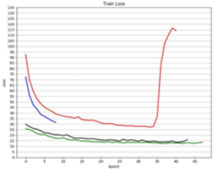
Ambition:



Why Fundamental Limits?

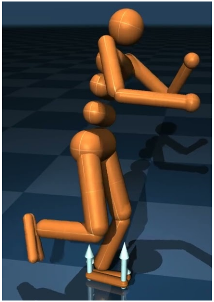
Ambition:

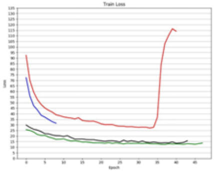




Why Fundamental Limits?

Ambition:

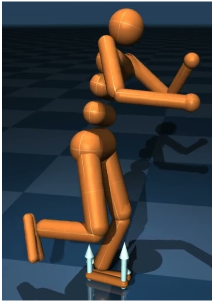


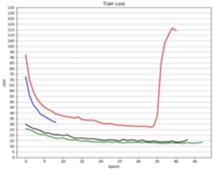


Reality:

Why Fundamental Limits?

Ambition:

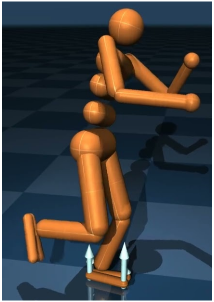


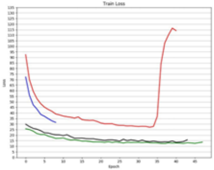


Reality:

Why Fundamental Limits?

Ambition:

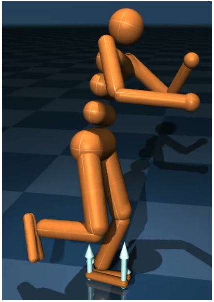


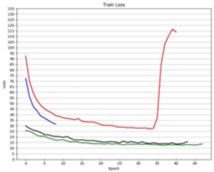


Reality:

Why Fundamental Limits?

Ambition:



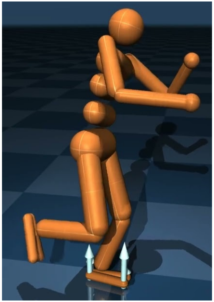


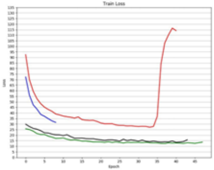
Reality:

Not just in sim:

Why Fundamental Limits?

Ambition:





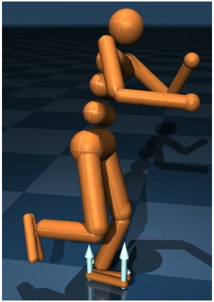
Reality:

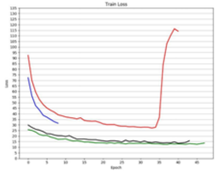
Not just in sim:

Tragic Uber accident

Why Fundamental Limits?

Ambition:





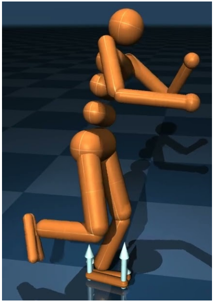
Reality:

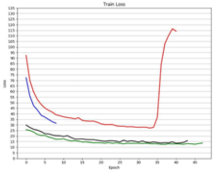
Not just in sim:

Tragic Uber accident

Why Fundamental Limits?

Ambition:





Reality:

Not just in sim:

Tragic Uber accident

Safe use of learning in controls $\Rightarrow$ need to understand fundamental limits

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 Athans [1] has extended to the multivariable case the now well-known guarantee of 60° phase and 6 dB gain margin for such controllers. However, for even the single-input, single-output case there has remained the question of whether there exist any guaranteed margins for the full LQG (Kalman filter in the loop) regulator. By counterexample, this note answers that question; there are none.

A standard two-state single-input single-output LQG control problem is posed for which the resulting closed-loop regulator has arbitrarily small gain margin.

Introduction

Unknown linear dynamics

$$S = (A, B): \qquad x_{t+1} = Ax_t + Bu_t + w_t, \qquad x_0 = 0 \qquad t = 0, 1, \ldots \tag{1}$$

Introduction

Unknown linear dynamics

$$S = (A, B): \qquad x_{t+1} = Ax_t + Bu_t + w_t, \qquad x_0 = 0 \qquad t = 0, 1, \ldots \tag{1}$$

Cost function (LQR):

$$J_S(K) \triangleq \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{E}_{K,S}\left[x_t^\top Q x_t + u_t^\top R u_t\right] \qquad u_t = Kx_t \tag{2}$$

Introduction

Unknown linear dynamics

$$S = (A, B): \qquad x_{t+1} = Ax_t + Bu_t + w_t, \qquad x_0 = 0 \qquad t = 0, 1, \ldots \tag{1}$$

Cost function (LQR):

$$J_S(K) \triangleq \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{E}_{K,S} \left[x_t^\top Q x_t + u_t^\top R u_t \right] \qquad u_t = Kx_t \tag{2}$$

Interested in analyzing algorithms of the form (stochastic policy gradient methods):

$$\widehat{K} \leftarrow \widehat{K} - \alpha \widehat{\nabla_K J(K; S)}$$

Introduction

Unknown linear dynamics

$$S = (A, B) : \qquad x_{t+1} = Ax_t + Bu_t + w_t, \qquad x_0 = 0 \qquad t = 0, 1, \ldots \tag{1}$$

Cost function (LQR):

$$J_S(K) \triangleq \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{E}_{K,S} \left[x_t^\top Q x_t + u_t^\top R u_t \right] \qquad u_t = Kx_t \tag{2}$$

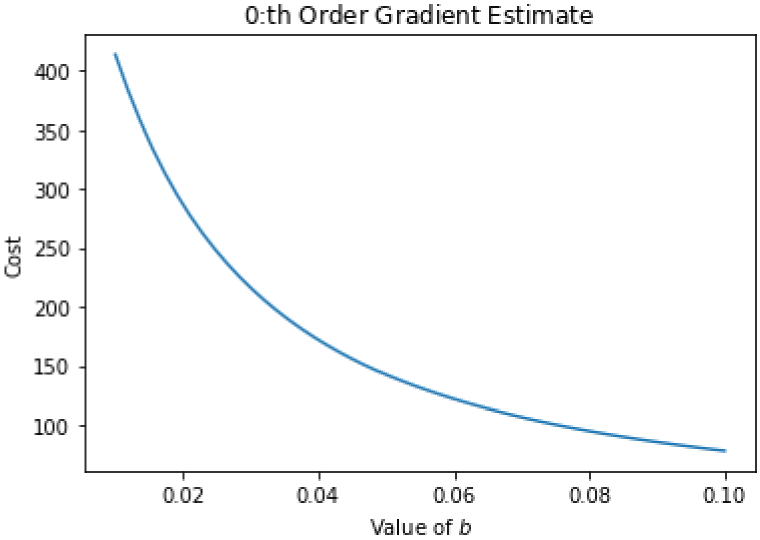
Interested in analyzing algorithms of the form (stochastic policy gradient methods):

$$\widehat{K} \leftarrow \widehat{K} - \alpha \widehat{\nabla_K J(K; S)}$$

Q: How are policy gradient methods affected by the limits of control?

Introduction

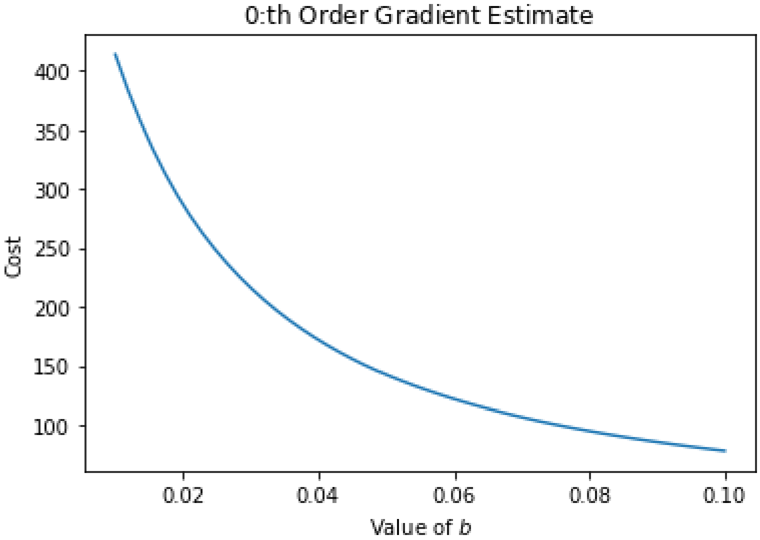
scalar system: $x_t = 1.01x_t + bu_t + w_t$ $\qquad u_t = kx_t$



Introduction

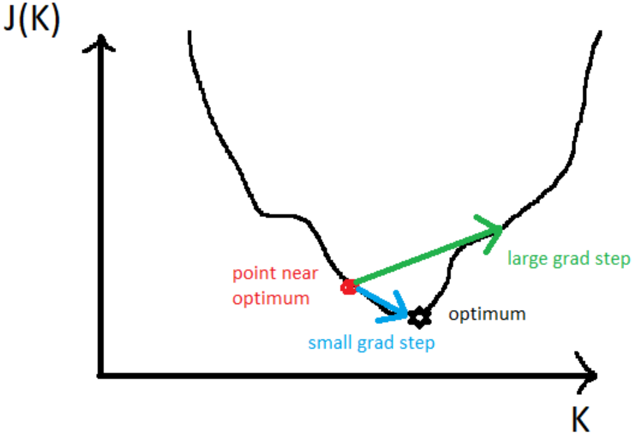
scalar system: $x_t = 1.01x_t + bu_t + w_t$ $\qquad u_t = kx_t$

Do stochastic policy gradient methods work well?



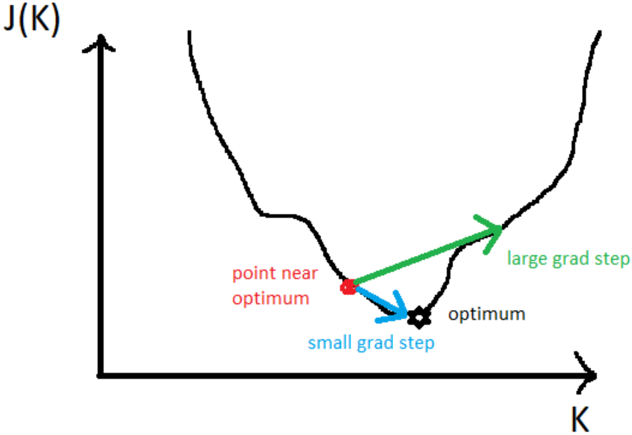
So what goes wrong?

$\widehat{K} \leftarrow \widehat{K} - \alpha \widehat{\nabla_K J(K;S)}$



So what goes wrong?

$\widehat{K} \leftarrow \widehat{K} - \alpha \widehat{\nabla_K J(K;S)}$



large variance in $\widehat{\nabla_K J(K;S)}$ $\Rightarrow$ too large gradient step more likely

Problem Formulation

Q: How noisy is the best possible gradient estimate $\widehat{\nabla_K J(K;S)}$ as a function of system properties?

Stability, Controllability, Observability

Problem Formulation

Q: How noisy is the best possible gradient estimate $\widehat{\nabla_K J(K; S)}$ as a function of system properties?
Stability, Controllability, Observability

Given N trajectories of length T from $S = (A, B)$:

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

Problem Formulation

Q: How noisy is the best possible gradient estimate $\widehat{\nabla_K J(K;S)}$ as a function of system properties?

Stability, Controllability, Observability

Given N trajectories of length T from $S = (A, B)$:

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

Budget ($\beta \in \mathbb{R}_+$):

$$\sum_{n=1}^{N} \sum_{t=0}^{T-1} \mathbf{E}_S u_{t,n}^\top u_{t,n} \leq \beta NT$$

Contribution

Let $K_\star(S)$ be the optimal gain. We prove lower bounds on:

$$\mathfrak{M}_d(\varepsilon; S, K_\star) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S,S') \le \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S); S') - \widehat{\nabla J} \right\|_{\mathrm{op}} \tag{3}$$

Contribution

Let $K_\star(S)$ be the optimal gain. We prove lower bounds on:

$$\mathfrak{M}_d(\varepsilon; S, K_\star) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S,S') \le \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S); S') - \widehat{\nabla J} \right\|_{\mathrm{op}} \qquad (3)$$

In particular, we show that:

Contribution

Let $K_\star(S)$ be the optimal gain. We prove lower bounds on:

$$\mathfrak{M}_d(\varepsilon; S, K_\star) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S,S') \le \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S); S') - \widehat{\nabla J} \right\|_{\mathrm{op}} \qquad (3)$$

In particular, we show that:

Ill-conditioned systems lead to noisy gradients (poor controllability of unstable modes / closed loop marginally stable)

Contribution

Let $K_\star(S)$ be the optimal gain. We prove lower bounds on:

$$\mathfrak{M}_d(\varepsilon; S, K_\star) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S,S') \le \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S); S') - \widehat{\nabla J} \right\|_{\mathrm{op}} \quad (3)$$

In particular, we show that:

Ill-conditioned systems lead to noisy gradients (poor controllability of unstable modes / closed loop marginally stable)

(3) can be exponentially large in the system dimension
integrator ⇒ curse of dimensionality

Contribution

Let $K_\star(S)$ be the optimal gain. We prove lower bounds on:

$$\mathfrak{M}_d(\varepsilon; S, K_\star) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S,S') \leq \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S); S') - \widehat{\nabla J} \right\|_{\mathrm{op}} \qquad (3)$$

In particular, we show that:

Ill-conditioned systems lead to noisy gradients (poor controllability of unstable modes / closed loop marginally stable)

(3) can be exponentially large in the system dimension
integrator ⇒ curse of dimensionality

In the paper we also sketch an extension to partially observed systems

Contribution

Let $K_\star(S)$ be the optimal gain. We prove lower bounds on:

$$\mathfrak{M}_d(\varepsilon; S, K_\star) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S,S') \le \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S); S') - \widehat{\nabla J} \right\|_{\mathrm{op}} \qquad (3)$$

In particular, we show that:

Ill-conditioned systems lead to noisy gradients (poor controllability of unstable modes / closed loop marginally stable)

(3) can be exponentially large in the system dimension
integrator ⇒ curse of dimensionality

In the paper we also sketch an extension to partially observed systems

⇒ Classical control-theoretic limitations can make policy gradient methods suffer arbitrarily noisy gradient estimates

Main Result

$\mathfrak{M}_d(\varepsilon; S_1, K_\star(S_1)) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S_1, S') \leq \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S_1); S') - \widehat{\nabla J} \right\|_{\text{op}}$

df: $P_{K_\star, S_1} = Q + K_\star{}^\top R K_\star + (A + BK_\star)^\top P_{K_\star, S_1} (A + BK_\star)$

Main Result

$\mathfrak{M}_d(\varepsilon; S_1, K_\star(S_1)) \triangleq \inf_{\widehat{\nabla J}} \sup_{S': d(S_1, S') \leq \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S_1); S') - \widehat{\nabla J} \right\|_{\text{op}}$

df: $P_{K_\star, S_1} = Q + K_\star^\top R K_\star + (A + BK_\star)^\top P_{K_\star, S_1} (A + BK_\star)$

df: $\Gamma_{K_\star, S_1} = \sum_{t=0}^{\infty} (A + BK_\star)^t (A + BK_\star)^{t,\top}$

Main Result

$\mathfrak{M}_d(\varepsilon; S_1, K_\star(S_1)) \triangleq \inf_{\widehat{\nabla J}} \sup_{S':d(S_1,S')\leq\varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S_1); S') - \widehat{\nabla J} \right\|_{\mathrm{op}}$

df: $P_{K_\star, S_1} = Q + {K_\star}^\top R K_\star + (A + BK_\star)^\top P_{K_\star, S_1}(A + BK_\star)$

df: $\Gamma_{K_\star, S_1} = \sum_{t=0}^{\infty}(A + BK_\star)^t(A + BK_\star)^{t,\top}$

df: $d_{\mathrm{KL}}(S_1, S_2(\Delta))$ the KL of obs from S_1 vs obs from S_2

Main Result

$\mathfrak{M}_d(\varepsilon; S_1, K_\star(S_1)) \triangleq \inf_{\widehat{\nabla J}} \sup_{S': d(S_1, S') \leq \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S_1); S') - \widehat{\nabla J} \right\|_{\mathrm{op}}$

df: $P_{K_\star, S_1} = Q + {K_\star}^\top R K_\star + (A + BK_\star)^\top P_{K_\star, S_1} (A + BK_\star)$

df: $\Gamma_{K_\star, S_1} = \sum_{t=0}^{\infty} (A + BK_\star)^t (A + BK_\star)^{t, \top}$

df: $d_{\mathrm{KL}}(S_1, S_2(\Delta))$ the KL of obs from S_1 vs obs from S_2

Main Result

$\mathfrak{M}_d(\varepsilon; S_1, K_\star(S_1)) \triangleq \inf_{\widehat{\nabla J}} \sup_{S' : d(S_1, S') \le \varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S_1); S') - \widehat{\nabla J} \right\|_{\text{op}}$

df: $P_{K_\star, S_1} = Q + K_\star^\top R K_\star + (A + BK_\star)^\top P_{K_\star, S_1}(A + BK_\star)$

df: $\Gamma_{K_\star, S_1} = \sum_{t=0}^{\infty} (A + BK_\star)^t (A + BK_\star)^{t,\top}$

df: $d_{\text{KL}}(S_1, S_2(\Delta))$ the KL of obs from S_1 vs obs from S_2

Theorem

Fix $\varepsilon > 0$ and Δ and a metric $d(\cdot,\cdot)$. Let $S_1 = (A, B)$ and $S_2(\Delta) = (A', B')$ with $A' = A - \Delta K_\star$ and $B' = B + \Delta$. We have:

$$\mathfrak{M}_d(\varepsilon; S_1, K_\star(S_1))$$
$$\ge \sup_{d(S_1, S_2(\Delta)) \le \varepsilon} \left\| \Delta^\top P_{K_\star, S_1}(A + BK_\star)\Gamma_{K_\star, S_1} \right\|_{\text{op}} \times \left(1 - \sqrt{\frac{1}{2} d_{\text{KL}}(S_1, S_2(\Delta))} \right)$$

Corollaries (Scalar Systems)

$\mathfrak{M}_d(\varepsilon; S_1, K_\star(S_1)) \triangleq \inf_{\widehat{\nabla J}} \sup_{S':d(S_1,S')\leq\varepsilon} \mathbf{E}_{S'} \left\| \nabla_K J(K_\star(S_1); S') - \widehat{\nabla J} \right\|_{\mathrm{op}}$

Consider the scalar system (with $|a| > 1$):

$$\begin{aligned} s_1 &: x_{t+1} = ax_t + bu_t + w_t \\ s_2 &: x_{t+1} = [a - (1/\sqrt{NT})k_\star(S_1)]x_t + [b + (1/\sqrt{NT})]u_t + w_t \end{aligned} \tag{4}$$

Corollaries (Scalar Systems)

Consider the scalar system (with $|a| > 1$):

$$\begin{aligned} s_1 &: x_{t+1} = a x_t + b u_t + w_t \\ s_2 &: x_{t+1} = [a - (1/\sqrt{NT}) k_\star(S_1)] x_t + [b + (1/\sqrt{NT})] u_t + w_t \end{aligned} \tag{4}$$

We obtain:

$$\mathfrak{M}_{d_\infty}\left(1/\sqrt{NT}, s_1\right) \gtrsim \frac{1}{\sqrt{NT(\beta + k_\star^2 \Gamma_{k_\star, s_1})}} \left| P_{k_\star, s_1}(a + b k_\star) \Gamma_{k_\star, s_1} \right| \tag{5}$$

Corollaries (Scalar Systems)

Consider the scalar system (with $|a| > 1$):

$$\begin{aligned} s_1 &: x_{t+1} = a x_t + b u_t + w_t \\ s_2 &: x_{t+1} = [a - (1/\sqrt{NT}) k_\star(S_1)] x_t + [b + (1/\sqrt{NT})] u_t + w_t \end{aligned} \tag{4}$$

We obtain:

$$\mathfrak{M}_{d_\infty}\left(1/\sqrt{NT}, s_1\right) \gtrsim \frac{1}{\sqrt{NT(\beta + k_\star^2 \Gamma_{k_\star, s_1})}} \left| P_{k_\star, s_1}(a + b k_\star) \Gamma_{k_\star, s_1} \right| \tag{5}$$

Crucially: $b \to 0 \Rightarrow \mathfrak{M}_{d_\infty}(\varepsilon_{NT}, s_1) \gtrsim \sqrt{\dfrac{|P_{k_\star, s_1} \Gamma_{k_\star, s_1}|}{NT}} \to \infty$

Corollaries (Scalar Systems)

Consider the scalar system (with $|a| > 1$):

$$\begin{aligned} s_1 &: x_{t+1} = ax_t + bu_t + w_t \\ s_2 &: x_{t+1} = [a - (1/\sqrt{NT})k_\star(S_1)]x_t + [b + (1/\sqrt{NT})]u_t + w_t \end{aligned} \tag{4}$$

We obtain:

$$\mathfrak{M}_{d_\infty}\left(1/\sqrt{NT}, s_1\right) \gtrsim \frac{1}{\sqrt{NT(\beta + k_\star^2 \Gamma_{k_\star,s_1})}} |P_{k_\star,s_1}(a + bk_\star)\Gamma_{k_\star,s_1}| \tag{5}$$

Crucially: $b \to 0 \Rightarrow \mathfrak{M}_{d_\infty}(\varepsilon_{NT}, s_1) \gtrsim \sqrt{\frac{|P_{k_\star,s_1}\Gamma_{k_\star,s_1}|}{NT}} \to \infty$

$\Rightarrow$ Bad controllability / marginally stable closed loop $\Rightarrow$ noisy gradients! ☹

Corollaries (Scalar Systems)

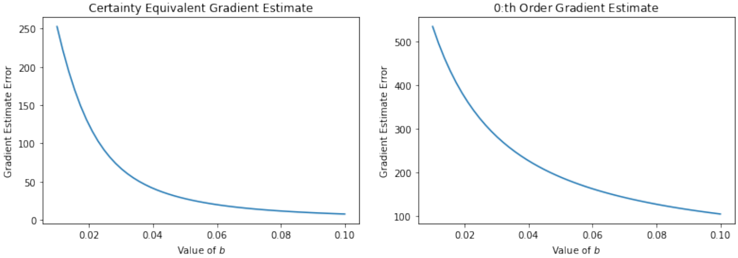


Figure: Gradient estimate spread as a function of b for the scalar system (4). Notice that poor controllability (small b), leads to noisy gradients. The vertical axes show the standard deviation of $\left\| \nabla_K J(K; S) - \widehat{\nabla_K J} \right\|_{\text{op}}$ across multiple trajectories.

Corollaries (Curse of Dimensionality)

Consider (with $0 < \rho < 1$):

Corollaries (Curse of Dimensionality)

Consider (with $0 < \rho < 1$):

$$x_{t+1} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \rho & 2 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ & & & \ddots & & 0 \\ 0 & 0 & 0 & & \rho & 2 \\ 0 & 0 & 0 & \dots & 0 & \rho \end{bmatrix}}_{=A} x_t + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix}}_{=B} u_t + w_t \tag{6}$$

Corollaries (Curse of Dimensionality)

Consider (with $0 < \rho < 1$):

$$x_{t+1} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \rho & 2 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ & & & \ddots & & 0 \\ 0 & 0 & 0 & & \rho & 2 \\ 0 & 0 & 0 & \dots & 0 & \rho \end{bmatrix}}_{=A} x_t + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix}}_{=B} u_t + w_t \tag{6}$$

Proposition

For the system S given in equation (6) we have that

$$\mathfrak{M}_{d_\infty}(\varepsilon_{NT}, S) \gtrsim \frac{4^{d_x}}{\sqrt{\beta NT}} \tag{7}$$

for d_x and NT sufficiently large for any $\varepsilon_{NT} \gtrsim 1/\sqrt{NT}$

Corollaries (Curse of Dimensionality)

Consider (with $0 < \rho < 1$):

$$x_{t+1} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \rho & 2 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ & & & \ddots & & 0 \\ 0 & 0 & 0 & & \rho & 2 \\ 0 & 0 & 0 & \dots & 0 & \rho \end{bmatrix}}_{=A} x_t + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix}}_{=B} u_t + w_t \tag{6}$$

Proposition

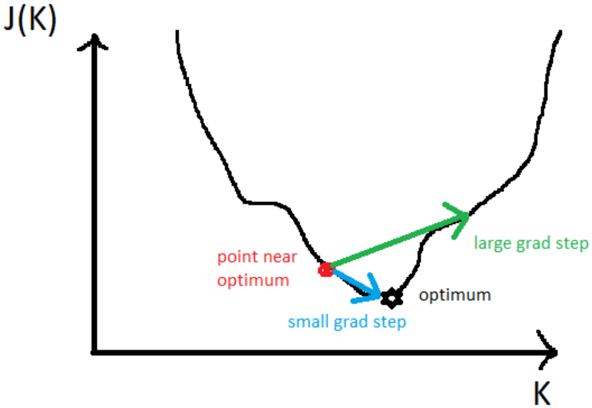
For the system S given in equation (6) we have that

$$\mathfrak{M}_{d_\infty}(\varepsilon_{NT}, S) \gtrsim \frac{4^{d_x}}{\sqrt{\beta NT}} \tag{7}$$

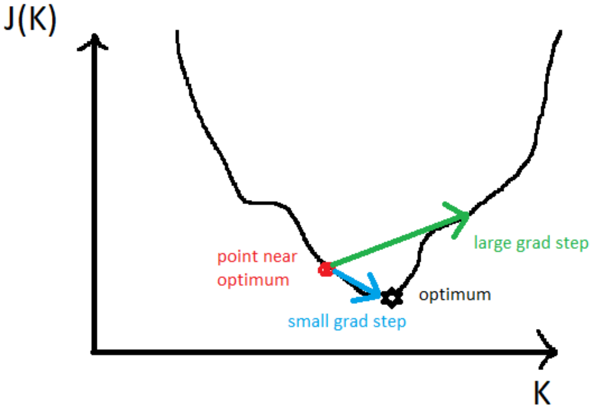
for d_x and NT sufficiently large for any $\varepsilon_{NT} \gtrsim 1/\sqrt{NT}$

$\Rightarrow$ Curse of dimensionality can affect gradient estimates!

What went wrong?



What went wrong?



large variance in $\widehat{\nabla_K J(K;S)}$ happens if:

system is ill-conditioned

has integrator-like structure

$\Rightarrow$ too large gradient step more likely

Conclusion

We showed that:

Conclusion

We showed that:

Ill-conditioned systems lead to noisy gradients (poor controllability / closed loop marginally stable) ☹

Conclusion

We showed that:

- Ill-conditioned systems lead to noisy gradients (poor controllability / closed loop marginally stable)
- gradient estimates can be exponentially bad in the system dimension (integrator $\Rightarrow$ curse of dimensionality)

Conclusion

We showed that:

- Ill-conditioned systems lead to noisy gradients (poor controllability / closed loop marginally stable) ☹
- gradient estimates can be exponentially bad in the system dimension (integrator $\Rightarrow$ curse of dimensionality) ☹

In the paper we also show that:

Conclusion

We showed that:

- Ill-conditioned systems lead to noisy gradients (poor controllability / closed loop marginally stable) ☹
- gradient estimates can be exponentially bad in the system dimension (integrator $\Rightarrow$ curse of dimensionality) ☹

In the paper we also show that:

- "bad markov parameters" $\Rightarrow$ noisy gradients ☹

Conclusion

We showed that:

- Ill-conditioned systems lead to noisy gradients (poor controllability / closed loop marginally stable) ☹
- gradient estimates can be exponentially bad in the system dimension (integrator $\Rightarrow$ curse of dimensionality) ☹

In the paper we also show that:

- "bad markov parameters" $\Rightarrow$ noisy gradients ☹

Future directions

Conclusion

We showed that:

- Ill-conditioned systems lead to noisy gradients (poor controllability / closed loop marginally stable) ☹
- gradient estimates can be exponentially bad in the system dimension (integrator $\Rightarrow$ curse of dimensionality) ☹

In the paper we also show that:

- "bad markov parameters" $\Rightarrow$ noisy gradients ☹

Future directions

- Lower bounds for arbitrary offline methods in LQR/LQG

Conclusion

We showed that:

- Ill-conditioned systems lead to noisy gradients (poor controllability / closed loop marginally stable) ☹
- gradient estimates can be exponentially bad in the system dimension (integrator ⇒ curse of dimensionality) ☹

In the paper we also show that:

- "bad markov parameters" ⇒ noisy gradients ☹

Future directions

- Lower bounds for arbitrary offline methods in LQR/LQG
- See also our concurrent work on the fundamental limits to adaptive control [Tsiamis et al., 2022]

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1467–1476. PMLR, 2018.

Kaiqing Zhang, Alec Koppel, Hao Zhu, and Tamer Basar. Global convergence of policy gradient methods to (almost) locally optimal policies. *SIAM Journal on Control and Optimization*, 58(6):3586–3612, 2020a.

Benjamin Gravell, Peyman Mohajerin Esfahani, and Tyler Summers. Learning robust control for lqr systems with multiplicative noise via policy gradient. *arXiv preprint arXiv:1905.13547*, 2019.

Benjamin Gravell, Peyman Mohajerin Esfahani, and Tyler Summers. Learning optimal controllers for linear systems with multiplicative noise via policy gradient. *IEEE Transactions on Automatic Control*, 66(11):5283–5298, 2020.

Kaiqing Zhang, Bin Hu, and Tamer Basar. Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence. In *Learning for Dynamics and Control*, pages 179–190. PMLR, 2020b.

Farnaz Adib Yaghmaie, Fredrik Gustafsson, and Lennart Ljung. Linear quadratic control using model-free reinforcement learning. *IEEE Transactions on Automatic Control*, 2022.

Yujie Tang, Yang Zheng, and Na Li. Analysis of the optimization landscape of linear quadratic gaussian (lqg) control. In *Learning for Dynamics and Control*, pages 599–610. PMLR, 2021.

Hesameddin Mohammadi, Mahdi Soltanolkotabi, and Mihailo R Jovanović. On the lack of gradient domination for linear quadratic gaussian problems with incomplete state information. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 1120–1124. IEEE, 2021.

Yang Zheng, Luca Furieri, Maryam Kamgarpour, and Na Li. Sample complexity of linear quadratic gaussian (lqg) control for output feedback systems. In *Learning for Dynamics and Control*, pages 559–570. PMLR, 2021.

Stephen Tu and Benjamin Recht. The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint. In *Conference on Learning Theory*, pages 3036–3083. PMLR, 2019.

Anastasios Tsiamis, Ingvar M Ziemann, Manfred Morari, Nikolai Matni, and George J Pappas. Learning to control linear systems can be hard. In *Conference on Learning Theory*, pages 3820–3857. PMLR, 2022.