Asynchronous Heterogeneous Linear Quadratic Regulator Design

Leonardo F. Toso*, Han Wang*, James Anderson

Department of Electrical Engineering, Columbia University

February 28, 2025

Conference on Decision and Control (CDC), 2024

*equal contribution

Motivation

Single-agent model-free optimal control

$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$

Objective: Design a control sequence $\{u_t = -Kx_t\}_t$ that minimizes a cumulative cost $\mathcal{J}(K) = \mathbf{E}_{x_0} \sum_{t=0}^{\infty} c_t(x_t, u_t)$.

Difficulty: Do not have access to the ground-truth system dynamics, i.e., $(A_\star, B_\star)$.

Motivation

Single-agent model-free optimal control

$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$

Objective: Design a control sequence $\{u_t = -Kx_t\}_t$ that minimizes a cumulative cost $J(K) = \mathbf{E}_{x_0} \sum_{t=0}^{\infty} c_t(x_t, u_t)$.

Difficulty: Do not have access to the ground-truth system dynamics, i.e., $(A_\star, B_\star)$.

Policy Gradient (PG)

Collect data: $D_K := \{x_t, u_t\}_t \sim \mathcal{D}_K$

Estimation: $\hat{\nabla}_{D_K} = ZO(D_K)$

Control: PG updates
$K \leftarrow K - \eta \hat{\nabla}_{D_K}$

Sample Complexity:
$J(K) - J(K^\star) \leq \epsilon$,
data points $\approx \mathcal{O}\left(1/\epsilon^2\right)$

Motivation

Single-agent model-free optimal control

$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$

Objective: Design a control sequence $\{u_t = -Kx_t\}_t$ that minimizes a cumulative cost $J(K) = \mathbf{E}_{x_0} \sum_{t=0}^{\infty} c_t(x_t, u_t)$.

Difficulty: Do not have access to the ground-truth system dynamics, i.e., $(A_\star, B_\star)$.

Policy Gradient (PG)

Collect data: $D_K := \{x_t, u_t\}_t \sim \mathcal{D}_K$

Estimation: $\hat{\nabla}_{D_K} = ZO(D_K)$

Control: PG updates $K \leftarrow K - \eta \hat{\nabla}_{D_K}$

Sample Complexity: $J(K) - J(K^\star) \leq \epsilon$, # data points $\approx \mathcal{O}\left(1/\epsilon^2\right)$

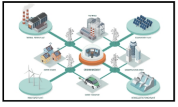
Necessity for a large amount of data samples

Rely on data samples

Autonomous Vehicles

≈ $$$

Smart Grids



≈ $$$$

Industrial Robotics

≈ $$$

- Expensive ($)
- Time consuming
- High risk for human being interaction

Motivation

Single-agent model-free optimal control

$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$

Objective: Design a control sequence $\{u_t = -Kx_t\}_t$ that minimizes a cumulative cost $J(K) = \mathbf{E}_{x_0} \sum_{t=0}^{\infty} c_t(x_t, u_t)$.

Difficulty: Do not have access to the ground-truth system dynamics, i.e., $(A_\star, B_\star)$.

Policy Gradient (PG)

Collect data: $D_K := \{x_t, u_t\}_t \sim \mathcal{D}_K$

Estimation: $\hat{\nabla}_{D_K} = ZO(D_K)$

Control: PG updates $K \leftarrow K - \eta \hat{\nabla}_{D_K}$

Sample Complexity: $J(K) - J(K^\star) \leq \epsilon$, # data points $\approx \mathcal{O}\left(1/\epsilon^2\right)$

Necessity for a large amount of data samples

Rely on data samples

Multi-agent control

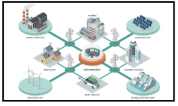
$D_K^{(1)}$ $D_K^{(2)}$ $D_K^{(M)}$

$$D_K^{(i)} \overset{iid}{\sim} \mathcal{D}_K, \forall i \in [M]$$

Autonomous Vehicles

$\approx$ \$\$\$

Smart Grids



$\approx$ \$\$\$\$

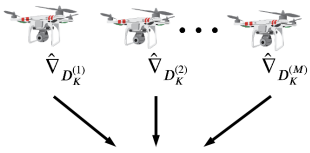
Industrial Robotics

$\approx$ \$\$\$

- Expensive (\$)
- Time consuming
- High risk for human being interaction

Homogeneous

$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$



$$\frac{1}{M}\sum_{i=1}^{M} \hat{\nabla}_{D_K^{(i)}}$$

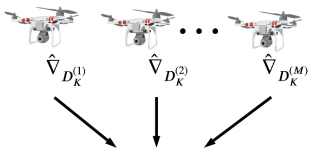
Law of Large Numbers (LLN)

$$\mathcal{J}(K) - \mathcal{J}(K^\star) \leq \epsilon$$

$$\#\text{ data points} \approx \mathcal{O}\left(1/M\epsilon^2\right)$$

Homogeneous

$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$



$$\frac{1}{M}\sum_{i=1}^{M} \hat{\nabla}_{D_K^{(i)}}$$

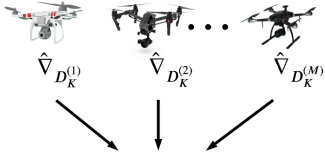
Law of Large Numbers (LLN)

$$J(K) - J(K^\star) \leq \epsilon$$

$$\#\text{ data points} \approx \mathcal{O}\left(1/M\epsilon^2\right)$$

Heterogeneous

$$x_{t+1}^{(i)} = A_\star^{(i)} x_t^{(i)} + B_\star^{(i)} u_t^{(i)} \quad t \geq 0$$



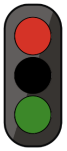
$$\frac{1}{M}\sum_{i=1}^{M} \hat{\nabla}_{D_K^{(i)}}$$

Bias-Variance Tradeoff :

$$J(K) - J(K^\star) \leq \epsilon + \mathcal{O}(\text{het})$$

$$\#\text{ data points} \approx \mathcal{O}\left(1/M\epsilon^2\right)$$

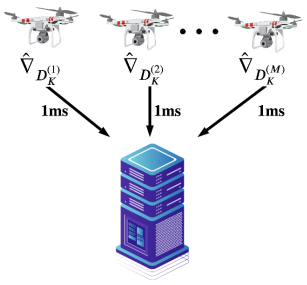
FedLQR **[WTMA, 2023]**



☹ **drawback**
😐 **neutral**
☺ **benefit**

Homogeneous

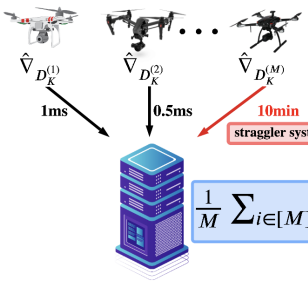
$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$



$$\frac{1}{M}\sum_{i\in[M]} \hat{\nabla}_{D_K^{(i)}}$$

Heterogeneous

$$x_{t+1}^{(i)} = A_\star^{(i)} x_t^{(i)} + B_\star^{(i)} u_t^{(i)} \quad t \geq 0$$



$$\frac{1}{M}\sum_{i\in[M]} \hat{\nabla}_{D_K^{(i)}}$$

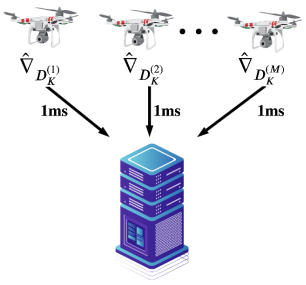
Don't fully exploit the parallelism in the distributed computation

Source of staleness:

- Network latency
- Bandwidth constraints
- Memory overhead
- Packet loss
- Local computation

Homogeneous

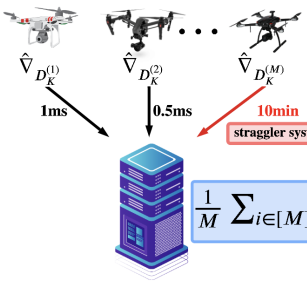
$$x_{t+1} = A_\star x_t + B_\star u_t \quad t \geq 0$$



$$\frac{1}{M}\sum_{i\in[M]} \hat{\nabla}_{D_K^{(i)}}$$

Heterogeneous

$$x_{t+1}^{(i)} = A_\star^{(i)} x_t^{(i)} + B_\star^{(i)} u_t^{(i)} \quad t \geq 0$$

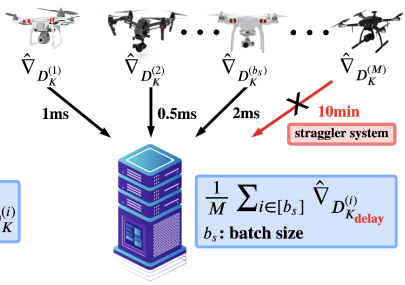


$$\frac{1}{M}\sum_{i\in[M]} \hat{\nabla}_{D_K^{(i)}}$$

Don't fully exploit the parallelism in the distributed computation

Asynchronous

$$x_{t+1}^{(i)} = A_\star^{(i)} x_t^{(i)} + B_\star^{(i)} u_t^{(i)} \quad t \geq 0$$



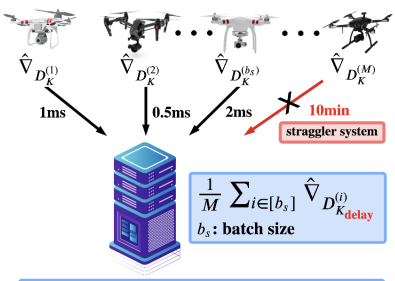
$$\frac{1}{M}\sum_{i\in[b_s]} \hat{\nabla}_{D_{K_{\text{delay}}}^{(i)}}$$

b_s: batch size

Fully exploiting the parallelism

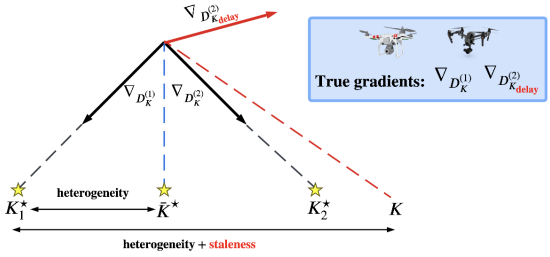
Asynchronous

$$x_{t+1}^{(i)} = A_\star^{(i)} x_t^{(i)} + B_\star^{(i)} u_t^{(i)} \quad t \geq 0$$



Fully exploiting the parallelism

Example: (heterogeneity and staleness)

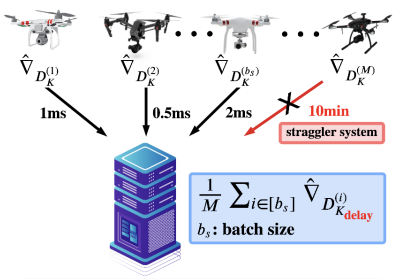


$$\mathbb{E}\| \tfrac{1}{2} \sum_{i \in [2]} \nabla D_{K_{\text{delay}}}^{(i)} \| \leq \mathbb{E}\| \tfrac{1}{2} \sum_{i \in [2]} \nabla D_{K_{\text{delay}}}^{(i)} - \nabla D_K^{(i)} \| + \| \nabla D_K^{(j)} \|$$

heterogeneity: $\nabla D_{K_{\text{delay}}}^{(i)} - \nabla D_K^{(i)}$; staleness: $\nabla D_{K_{\text{delay}}}^{(i)}$

Asynchronous

$$x_{t+1}^{(i)} = A_\star^{(i)} x_t^{(i)} + B_\star^{(i)} u_t^{(i)} \quad t \geq 0$$

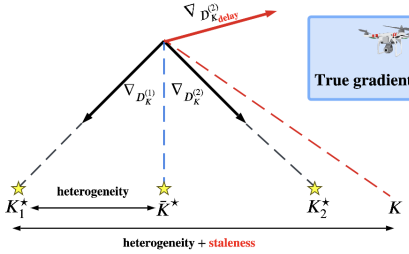


$$\frac{1}{M} \sum_{i \in [b_s]} \hat{\nabla} D_{K_{\text{delay}}}^{(i)}$$

b_s: batch size

Fully exploiting the parallelism

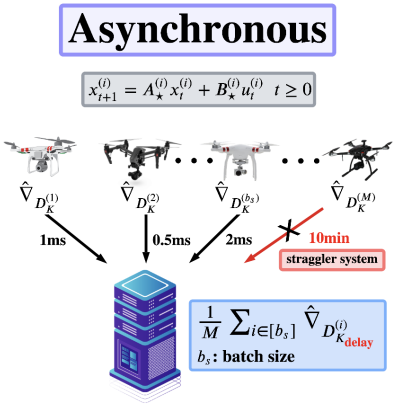
Example: (heterogeneity and staleness)

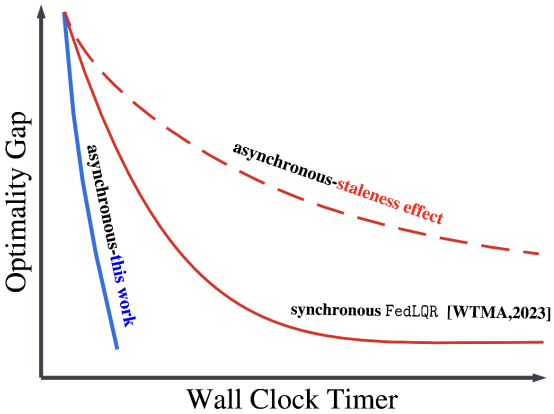


$$\mathbb{E}\| \tfrac{1}{2} \sum_{i \in [2]} \nabla D_{K_{\text{delay}}}^{(i)} \| \leq \mathbb{E}\| \tfrac{1}{2} \sum_{i \in [2]} \nabla D_{K_{\text{delay}}}^{(i)} \overbrace{- \nabla D_K^{(j)}}^{\text{heterogeneity}} \| + \| \nabla D_K^{(j)} \|$$

staleness

Question: Can an asynchronous algorithm produce a controller that is near-optimal, even in the presence of staleness and heterogeneous system dynamics?





Question: Can an asynchronous algorithm produce a controller that is near-optimal, even in the presence of staleness and heterogeneous system dynamics?

Consider a discrete LTI dynamical system

$$x_{t+1}^{(i)} = A^{(i)} x_t^{(i)} + B^{(i)} u_t^{(i)}, \quad t = 0, 1, 2, \ldots \quad \text{(sys-dyn)}$$

where $x_t^{(i)} \in \mathbb{R}^{n_x}$ and $u_t^{(i)} \in \mathbb{R}^{n_u}$.

Consider a discrete LTI dynamical system

$$x_{t+1}^{(i)} = A^{(i)} x_t^{(i)} + B^{(i)} u_t^{(i)}, \quad t = 0, 1, 2, \ldots \quad \text{(sys-dyn)}$$

where $x_t^{(i)} \in \mathbb{R}^{n_x}$ and $u_t^{(i)} \in \mathbb{R}^{n_u}$.

LQR Objective: Design a controller $K_i^\star$ ($u_t^{(i)} = -K_i^\star x_t^{(i)}$) that solves

$$K_i^\star = \text{argmin}_{K \in \mathcal{K}^{(i)}} \mathcal{J}^{(i)}(K) = \mathbb{E}\left[\sum_{t=0}^{\infty} x_t^{(i)\top} \left(Q^{(i)} + K^\top R^{(i)} K\right) x_t^{(i)}\right],$$

subject to (sys-dyn).

Stabilizing set: $\mathcal{K}^{(i)} = \{K \mid \rho(A^{(i)} - B^{(i)} K) < 1\}$

Global convergence despite of the non-convexity* of $\mathcal{J}^{(i)}(K)$ [FGKM,ICML 2018],

Initial Stabilizing Controller: $K_0 \in \mathcal{K}^{(i)}$

Controllability: $(A^{(i)}, B^{(i)})$ is controllable

* Non-convex for $n_x \geq 3$.

Global convergence despite of the non-convexity* of $\mathcal{J}^{(i)}(K)$ [FGKM,ICML 2018],

Initial Stabilizing Controller: $K_0 \in \mathcal{K}^{(i)}$

Controllability: $(A^{(i)}, B^{(i)})$ is controllable

$$K_{n+1} = K_n - \eta \widehat{\nabla} \mathcal{J}^{(i)}(K_n), \text{ for } n \in \{0, 1, \ldots, N-1\},$$

$$\mathcal{J}^{(i)}(K_N) - \mathcal{J}^{(i)}(K_i^\star) \leq \epsilon,$$

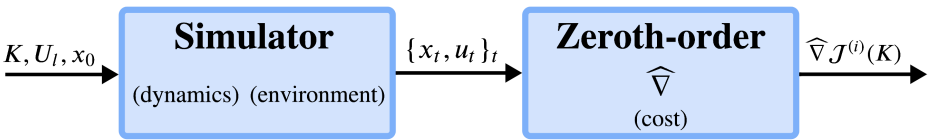
after $N = \mathcal{O}(\log(1/\epsilon))$ iterations

* Non-convex for $n_x \geq 3$.

Zeroth-order Estimation: $K_{n+1} = K_n - \eta \widehat{\nabla} \mathcal{J}^{(i)}(K_n)$

$$\texttt{ZO}(m, r, K) \to \widehat{\nabla} \mathcal{J}^{(i)}(K) := \sum_{l=1}^{m} \frac{n_x n_u (\mathcal{J}^{(i)}(K + U_l) - \mathcal{J}^{(i)}(K - U_l)) U_l}{2mr^2},$$

m (number of trajectories), r (smoothing radius) and $\|U_l\|_F = r$.



Zeroth-order Estimation: $K_{n+1} = K_n - \eta \widehat{\nabla} \mathcal{J}^{(i)}(K_n)$

$$\texttt{ZO}(m, r, K) \to \widehat{\nabla} \mathcal{J}^{(i)}(K) := \sum_{l=1}^{m} \frac{n_x n_u (\mathcal{J}^{(i)}(K + U_l) - \mathcal{J}^{(i)}(K - U_l)) U_l}{2mr^2},$$

m (number of trajectories), r (smoothing radius) and $\|U_l\|_F = r$.

Estimation Error[*]: Suppose $m = \mathcal{O}(1/\epsilon^2)$ and $r = \mathcal{O}(\epsilon)$, it holds that

$$\|\nabla \mathcal{J}^{(i)}(K) - \widehat{\nabla} \mathcal{J}^{(i)}(K)\| \leq \epsilon,$$

with high probability.

[*] Bernstein matrix inequality (Tropp (2012)).

Consider M distinct systems (sys-dyn) with different LQR objectives, i.e.,

Objective: $\bar{K}^\star := \text{argmin}_{K \in \mathcal{S}} \left\{ \bar{\mathcal{J}}(K) := \frac{1}{M} \sum_{i=1}^{M} \mathcal{J}^{(i)}(K) \right\}$,

$\bar{K}^\star$ should stabilize each (sys-dyn) and on *average* minimize their LQR objectives

Consider M distinct systems (sys-dyn) with different LQR objectives, i.e.,

Objective: $\bar{K}^\star := \text{argmin}_{K \in \mathcal{S}} \left\{ \bar{\mathcal{J}}(K) := \frac{1}{M} \sum_{i=1}^{M} \mathcal{J}^{(i)}(K) \right\}$,

$\bar{K}^\star$ should stabilize each (sys-dyn) and on *average* minimize their LQR objectives

Heterogeneity: $\max_{i \neq j} \|A^{(i)} - A^{(j)}\| \leq \epsilon_A \rightarrow$ same for ϵ_B, ϵ_Q and ϵ_R

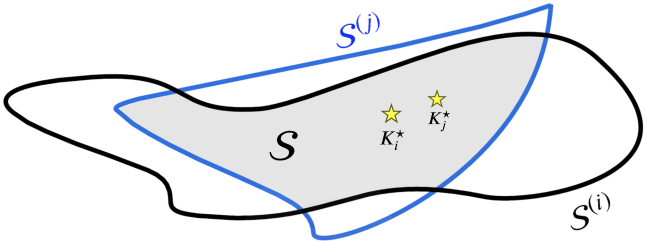
Consider M distinct systems (sys-dyn) with different LQR objectives, i.e.,

Objective: $\bar{K}^\star := \text{argmin}_{K \in \mathcal{S}} \left\{ \bar{\mathcal{J}}(K) := \frac{1}{M} \sum_{i=1}^{M} \mathcal{J}^{(i)}(K) \right\}$,

$\bar{K}^\star$ should stabilize each (sys-dyn) and on *average* minimize their LQR objectives

Heterogeneity: $\max_{i \neq j} \|A^{(i)} - A^{(j)}\| \leq \epsilon_A \rightarrow$ same for ϵ_B, ϵ_Q and ϵ_R

Stabilizing sub-level set: $\mathcal{S}^{(i)} := \left\{ K \mid \mathcal{J}^{(i)}(K) - \mathcal{J}^{(i)}(K_i^\star) \leq \gamma(\mathcal{J}^{(i)}(K_0) - \mathcal{J}^{(i)}(K_i^\star)) \right\}$,

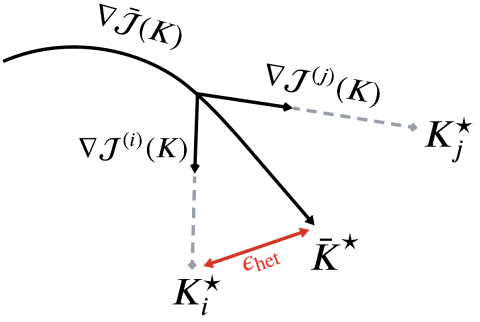


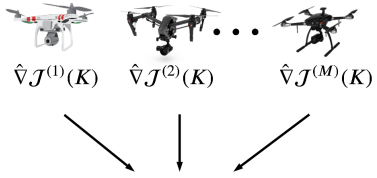
Gradient Heterogeneity: [TZAW, L4DC 2024] For any two distinct systems with different LQR objectives, and given a stabilizing controller $K \in \mathcal{S}$. It holds that,

$$\|\nabla \mathcal{J}^{(i)}(K) - \nabla \mathcal{J}^{(j)}(K)\|^2 \leq f(\epsilon_A, \epsilon_B, \epsilon_Q, \epsilon_R) = \epsilon_{\text{het}} \quad \text{(grad-het)}$$

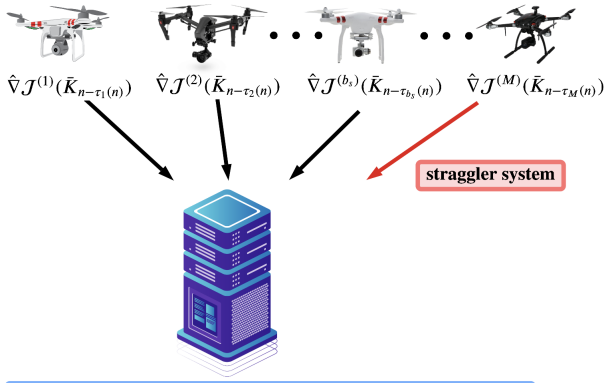
Gradient Heterogeneity: [TZAW, L4DC 2024] For any two distinct systems with different LQR objectives, and given a stabilizing controller $K \in \mathcal{S}$. It holds that,

$$\|\nabla \mathcal{J}^{(i)}(K) - \nabla \mathcal{J}^{(j)}(K)\|^2 \leq f(\epsilon_A, \epsilon_B, \epsilon_Q, \epsilon_R) = \epsilon_{\text{het}} \quad \text{(grad-het)}$$





FedLQR **[WTMA, 2023]**



Bounded staleness: $\tau_s(n) \leq \tau_{\max} \in \mathbb{N}$ denotes the staleness in the controller that system $s \in [b_s]$ possesses when locally estimating its policy gradient at step n

Asynchronous Policy Gradient LQR

Algorithm

Input: $\bar{K}_0 \in \mathcal{S}$, N, η, m, r, b_s

Initialize: $K_i = \bar{K}_0 \ \forall i \in [M]$ and $\overline{\nabla} \leftarrow 0$ **(1)**

Iteration and batch counters: $s = n = 0$

In parallel: compute and send $\widehat{\nabla}_i = \texttt{ZO}(K_i, r, m)$ to the server $\forall i \in [M]$ **(2)**

While $n < N$

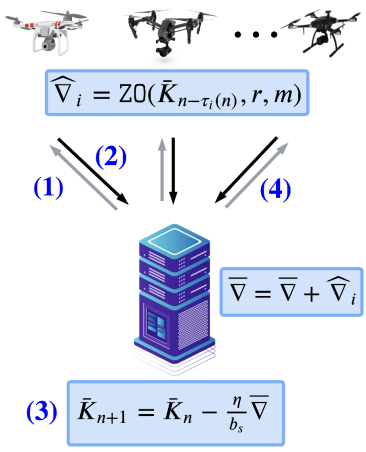
Server accumulates $\overline{\nabla} = \overline{\nabla} + \nabla_i$, s+=1

If $s = b_s$ **then**

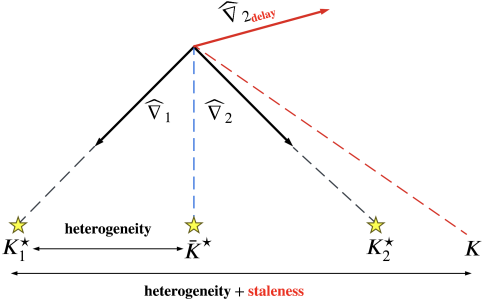
$\bar{K}_{n+1} = \bar{K}_n - \frac{\eta}{b_s}\overline{\nabla}$ **(3)**

Send $\bar{K}_{n+1}$ to the **idle** systems **(4)**

Output: $\bar{K}_N$

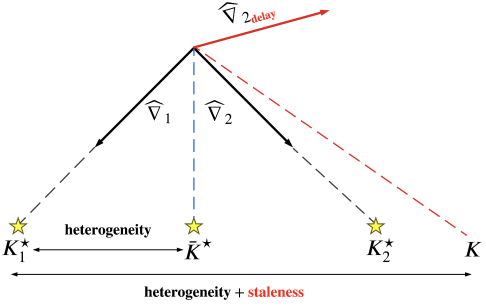


Controlling the staleness effect: $\mathbb{E}\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\|^2$



Controlling the staleness effect: $\mathbb{E}\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\|^2$

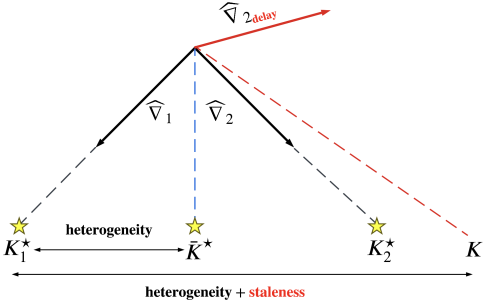
$$\mathbb{E}\left\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\right\|^2 \leq \tau_{\max} \sum_{l=n-\tau_i(n)}^{n-1} \mathbb{E}\left\|\bar{K}_{l+1} - \bar{K}_l\right\|^2$$



Controlling the staleness effect: $\mathbb{E}\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\|^2$

$$\mathbb{E}\left\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\right\|^2 \leq \tau_{\max} \sum_{l=n-\tau_i(n)}^{n-1} \mathbb{E}\left\|\bar{K}_{l+1} - \bar{K}_l\right\|^2$$

$$\mathbb{E}\left\|\bar{K}_{l+1} - \bar{K}_l\right\|^2 \leq \eta^2 \tau_{\max} \mathcal{O}\left(\epsilon_{\text{het}} + \mathbb{E}\|\nabla \mathcal{J}^{(i)}(\bar{K}_n)\|^2\right)^*$$

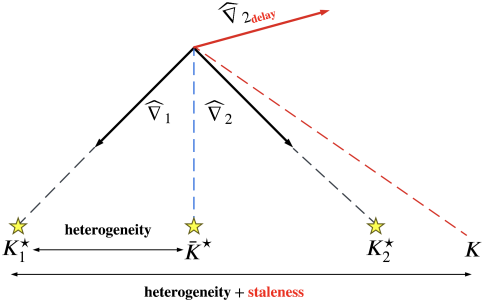


* See proof of Lemma 4.

Controlling the staleness effect: $\mathbb{E}\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\|^2$

$$\mathbb{E}\left\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\right\|^2 \leq \tau_{\max} \sum_{l=n-\tau_i(n)}^{n-1} \mathbb{E}\left\|\bar{K}_{l+1} - \bar{K}_l\right\|^2$$

$$\mathbb{E}\left\|\bar{K}_{l+1} - \bar{K}_l\right\|^2 \leq \eta^2 \tau_{\max} \mathcal{O}\left(\epsilon_{\text{het}} + \mathbb{E}\|\nabla \mathcal{J}^{(i)}(\bar{K}_n)\|^2\right)^*$$



$$\frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}\left\|\bar{K}_n - \bar{K}_{n-\tau_i(n)}\right\|^2 \lesssim \frac{\eta^2 \tau_{\max}^3 \epsilon_{\text{het}}}{b_s} + \eta^2 \tau_{\max}^3 \mathbb{E}\|\nabla \bar{\mathcal{J}}(\bar{K}_n)\|^2$$

* See proof of Lemma 4.

Goal: Control $\frac{1}{N}\sum_{n=0}^{N-1}\mathbb{E}\|\nabla\bar{\mathcal{J}}(\bar{K}_n)\|^2 \to$ **find a stationary solution**

- **Initial stabilizing controller:** $\bar{K}_0 \in \mathcal{S}$
- $\eta = \mathcal{O}\left(\sqrt{\frac{b_s}{N}}\right)$ and r sufficiently small

Goal: Control $\frac{1}{N}\sum_{n=0}^{N-1}\mathbb{E}\|\nabla\bar{\mathcal{J}}(\bar{K}_n)\|^2 \rightarrow$ **find a stationary solution**

- **Initial stabilizing controller:** $\bar{K}_0 \in \mathcal{S}$
- $\eta = \mathcal{O}\left(\sqrt{\frac{b_s}{N}}\right)$ and r sufficiently small

$$\frac{1}{N}\sum_{n=0}^{N-1}\mathbb{E}\|\nabla\bar{\mathcal{J}}(\bar{K}_n)\|_F^2 \leq \mathcal{O}\left(\frac{\bar{\Delta}_0}{\sqrt{Nb_s}} + \frac{\epsilon_{\text{het}}}{\sqrt{Nb_s}} + \frac{\tau_{\max}^2\epsilon_{\text{het}}}{N}\right)$$

- $\bar{\Delta}_0 = \mathbb{E}[\bar{\mathcal{J}}(\bar{K}_0) - \bar{\mathcal{J}}(\bar{K}^\star)] \rightarrow$ **initial optimality gap**

$\mathcal{O}\left(\frac{\tau_{\max}^2\epsilon_{\text{het}}}{N}\right) \rightarrow$ staleness effect becomes negligible when $N \gg b_s$

Goal: Control $\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \rightarrow$ **system-specific optimality gap**

- **Initial stabilizing controller:** $\bar{K}_0 \in \mathcal{S}$
- $\eta = \mathcal{O}\left(1/\tau_{\max}^{3/2}\right)$ and r sufficiently small

Goal: Control $\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \rightarrow$ **system-specific optimality gap**

- **Initial stabilizing controller:** $\bar{K}_0 \in \mathcal{S}$
- $\eta = \mathcal{O}\left(1/\tau_{\max}^{3/2}\right)$ and r sufficiently small

$$\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \leq \mathcal{O}\left(c^N \Delta_0^{(i)} + \epsilon_{\text{het}}\right)$$

- $\Delta_0^{(i)} = \mathbb{E}[\mathcal{J}^{(i)}(\bar{K}_0) - \mathcal{J}^{(i)}(\bar{K}^\star)]$, $c = 1 - \frac{\eta\lambda}{4} \in (0,1)$

Goal: Control $\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \rightarrow$ **system-specific optimality gap**

- **Initial stabilizing controller:** $\bar{K}_0 \in \mathcal{S}$
- $\eta = \mathcal{O}\left(1/\tau_{\max}^{3/2}\right)$ and r sufficiently small

$$\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \leq \mathcal{O}\left(c^N \Delta_0^{(i)} + \epsilon_{\text{het}}\right)$$

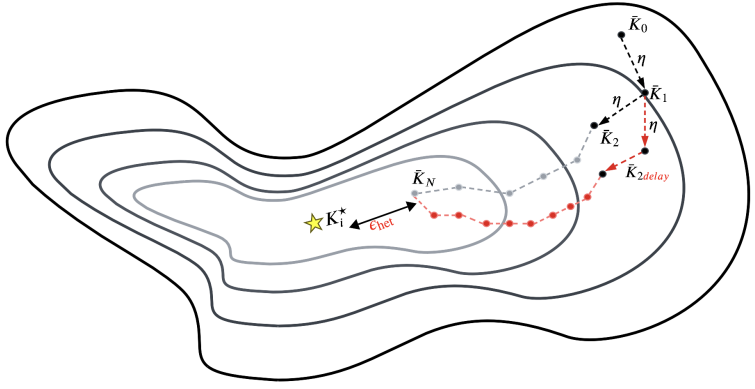
- $\Delta_0^{(i)} = \mathbb{E}[\mathcal{J}^{(i)}(\bar{K}_0) - \mathcal{J}^{(i)}(\bar{K}^\star)]$, $c = 1 - \frac{\eta\lambda}{4} \in (0,1)$

Within the number of iterations in the order of $N = \mathcal{O}\left(\tau_{\max}^{3/2} \log(1/\epsilon)\right)$ we have

$$\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \leq \mathcal{O}\left(\epsilon + \epsilon_{\text{het}}\right)$$

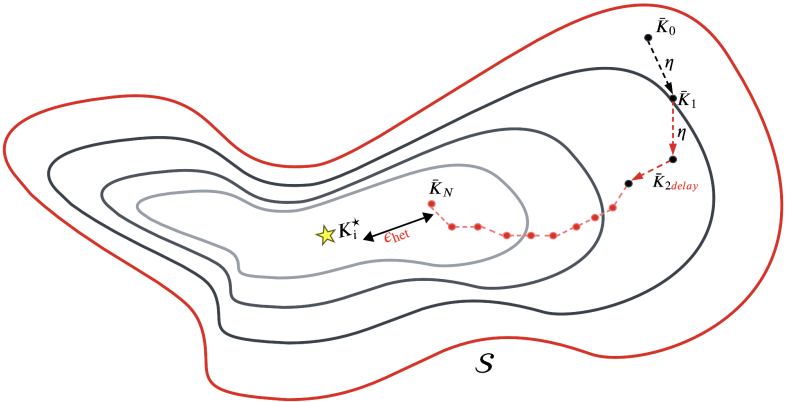
Within the number of iterations in the order of $N = \mathcal{O}\left(\tau_{\max}^{3/2} \log(1/\epsilon)\right)$ we have

$$\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \leq \mathcal{O}\left(\epsilon + \epsilon_{\text{het}}\right)$$



Goal: Show that $\bar{K}_n \in \mathcal{S}$ for any iteration $n \in \{0, 1, \ldots, N-1\}$

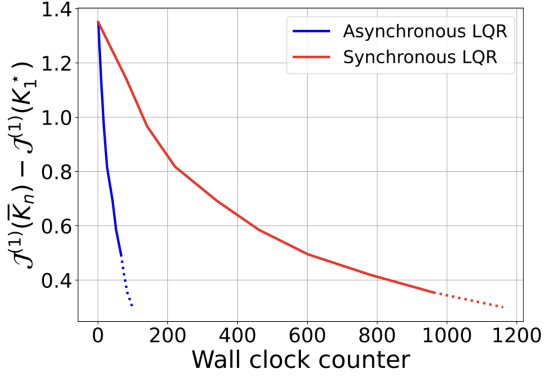
- **Initial stabilizing controller:** $\bar{K}_0 \in \mathcal{S}$
- ϵ_{het} and r sufficiently small and $\eta = \mathcal{O}(1/\tau_{\max}^{3/2})$

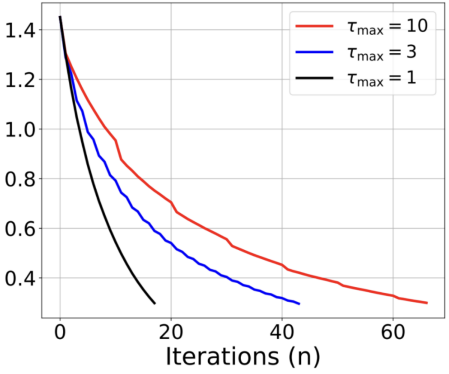


Numerical Validation

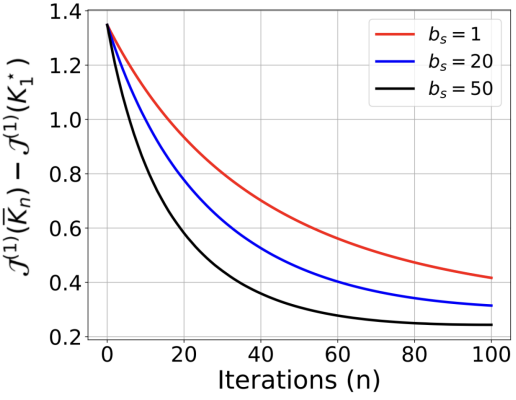
Asynchronous vs synchronous updates

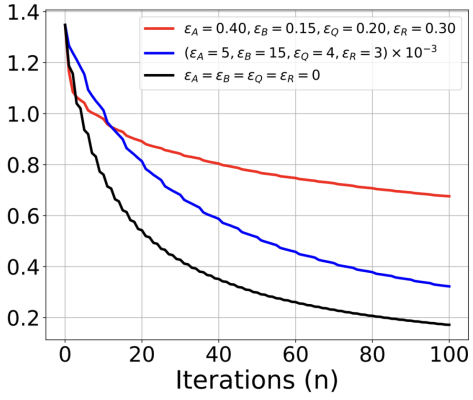
$M = 100$ heterogeneous systems with $n_x = 4$ states and $n_u = 2$ inputs





(left) $\tau_{\max} = 20,\ b_s = 20$





(left) $\tau_{\max} = 1$, **(right)** $\tau_{\max} = 5$, $b_s = 20$

Conclusions

- We studied the problem of learning linear quadratic regulators from **heterogeneous** systems with **asynchronous** policy gradient updates
- The proposed asynchronous aggregation **fully exploits** the parallelism in the distributed computation while alleviating the impact of **straggler systems**
- We provided **local** and **global** convergence guarantees, i.e.,

$$\frac{1}{N}\sum_{n=0}^{N-1}\mathbb{E}\|\nabla\bar{\mathcal{J}}(\bar{K}_n)\|_F^2 \leq \mathcal{O}\left(\frac{\bar{\Delta}_0}{\sqrt{N b_s}} + \frac{\epsilon_{\text{het}}}{\sqrt{N b_s}} + \frac{\tau_{\max}^2 \epsilon_{\text{het}}}{N}\right)$$

$$\mathbb{E}\left[\mathcal{J}^{(i)}(\bar{K}_N) - \mathcal{J}^{(i)}(K_i^\star)\right] \leq \mathcal{O}\left(\epsilon + \epsilon_{\text{het}}\right), \text{ with } N = \mathcal{O}\left(\tau_{\max}^{3/2}\log(1/\epsilon)\right)$$

- We also showed that $\bar{K}_n \in \mathcal{S}$ for any iteration $n \in \{0, 1, \ldots, N-1\}$

References

1. JA. Tropp. **"User-Friendly Tail Bounds for Sums of Random Matrices"**, FoCM 2012.
2. M. Fazel, R. Ge, S. Kakade, M. Mesbahi. **"Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator"**, ICML 2018.
3. H. Wang, LF. Toso, A. Mitra, J. Anderson. **"Model-free Learning with Heterogeneous Dynamical Systems: A Federated LQR Approach"**, 2023.
4. LF. Toso, D. Zhan, J. Anderson, H. Wang. **"Meta-Learning Linear Quadratic Regulators: A Policy Gradient MAML Approach for Model-free LQR"**, L4DC 2024.
5. X. Sha, F. Zhao, K. You. **"Asynchronous Parallel Policy Gradient Methods for the Linear Quadratic Regulator"**, 2024. $\rightarrow$ **homogeneous setting**

Collaborators

Han Wang **James Anderson**

Funding

NSF

CapitalOne

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

To find out more:

My website

Full paper

Happy to take questions!