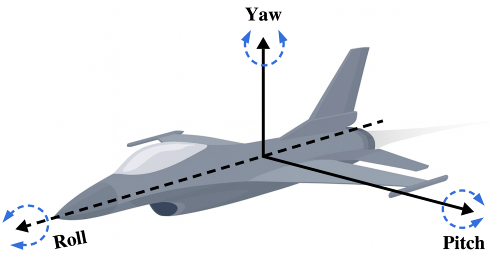
Learning Stabilizing Policies via an Unstable Subspace Representation

Leonardo F. Toso, Lintao Ye, James Anderson

Department of Electrical Engineering, Columbia University

64th IEEE Conference on Decision and Control, December 2025

Motivation: Designing a Stabilizing Control Policy

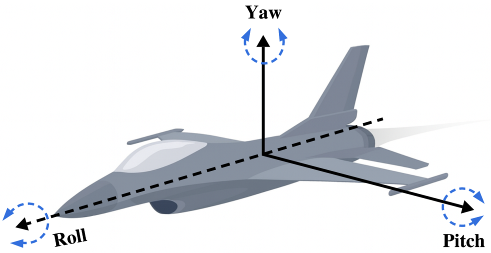


$$\texttt{state} = \begin{bmatrix} \text{pitch angle} \\ \text{roll angle} \\ \text{yaw rate} \\ \text{airspeed} \\ \text{altitude} \\ \vdots \end{bmatrix} - \underbrace{\begin{bmatrix} 0^\circ \\ 0^\circ \\ 0 \text{ rad/s} \\ V_{\text{cruise}} \\ h_{\text{desired}} \\ \vdots \end{bmatrix}}_{\text{Target}}$$

- Given $f : \mathbb{R}^{d_X} \times \mathbb{R}^{d_U} \to \mathbb{R}^{d_X}$

$$\underset{\texttt{state}}{\underset{\downarrow}{x_{t+1}}} = f(x_t, \underset{\texttt{input}}{\underset{\downarrow}{u_t}}) + \underset{\texttt{noise}}{\underset{\downarrow}{w_t}}$$

Motivation: Designing a Stabilizing Control Policy



Typically the number of states $d_X \approx 20$

With **only** two **unstable modes:**

- Longitudinal phugoid
- Lateral spiral

- Given $f : \mathbb{R}^{d_X} \times \mathbb{R}^{d_U} \to \mathbb{R}^{d_X}$

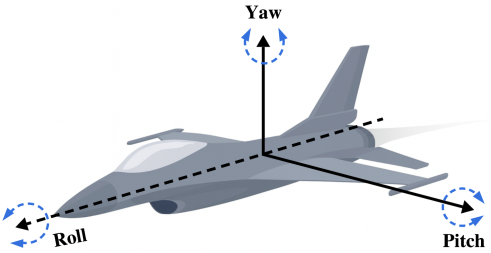
$$\underset{\text{state}}{x_{t+1}} = f(x_t, \underset{\text{input}}{u_t}) + \underset{\text{noise}}{w_t}$$

Stabilization: Design a stabilizing policy

$$\pi(x_{0:t}, u_{0:t-1}) \to u_t \text{ such that}$$

$$x_t \to 0 \text{ as } t \to \infty \text{ under } \pi(x_{0:t}, u_{0:t-1})$$

Motivation: Designing a Stabilizing Control Policy



Typically the number of states $d_X \approx 20$

With **only** two **unstable modes:**

- Longitudinal phugoid
- Lateral spiral

- Given $f : \mathbb{R}^{d_X} \times \mathbb{R}^{d_U} \to \mathbb{R}^{d_X}$

$$\underset{\text{state}}{\underset{\downarrow}{x_{t+1}}} = f(x_t, \underset{\text{input}}{\underset{\downarrow}{u_t}}) + \underset{\text{noise}}{\underset{\downarrow}{w_t}}$$

Stabilization: Design a stabilizing policy

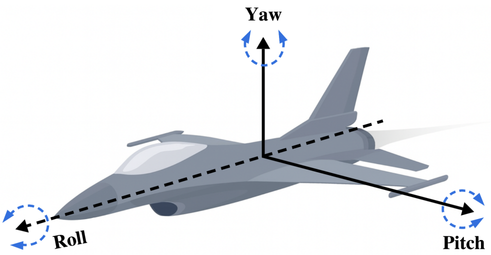
$$\pi(x_{0:t}, u_{0:t-1}) \to u_t \text{ such that}$$

$$x_t \to 0 \text{ as } t \to \infty \text{ under } \pi(x_{0:t}, u_{0:t-1})$$

- LQR (f is linear) (Kalman, 1960)
- Pole placement (Ackermann, 1972)
- Robust Control (Doyle, 1989)

Acting on all (stable and unstable) modes

Motivation: Designing a Stabilizing Control Policy



Typically the number of states $d_X \approx 20$

With **only** two **unstable modes:**

- Longitudinal phugoid
- Lateral spiral

Question: Why to act on all modes?

- Given $f : \mathbb{R}^{d_X} \times \mathbb{R}^{d_U} \to \mathbb{R}^{d_X}$

$$\underset{\text{state}}{\underset{\downarrow}{x_{t+1}}} = f(x_t, \underset{\text{input}}{\underset{\downarrow}{u_t}}) + \underset{\text{noise}}{\underset{\downarrow}{w_t}}$$

Stabilization: Design a stabilizing policy

$\pi(x_{0:t}, u_{0:t-1}) \to u_t$ such that

$x_t \to 0$ as $t \to \infty$ under $\pi(x_{0:t}, u_{0:t-1})$

- LQR (f is linear) (Kalman, 1960)
- Pole placement (Ackermann, 1972)
- Robust Control (Doyle, 1989)

Acting on all (stable and unstable) modes

Intro: Model-Free and Learning to Stabilize (LTS)

Consider $f(x_t, u_t) = Ax_t + Bu_t$, with A and B **unknown**, and **linear** feedback $u_t = Kx_t$

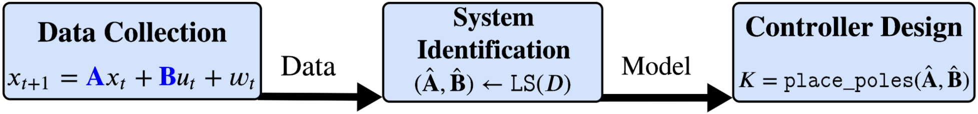
Question: How to design K such that $\rho(A + BK) < 1$ (**exponential stability**)?

Intro: Model-Free and Learning to Stabilize (LTS)

Consider $f(x_t, u_t) = Ax_t + Bu_t$, with A and B **unknown**, and **linear** feedback $u_t = Kx_t$

Question: How to design K such that $\rho(A + BK) < 1$ (**exponential stability**)?

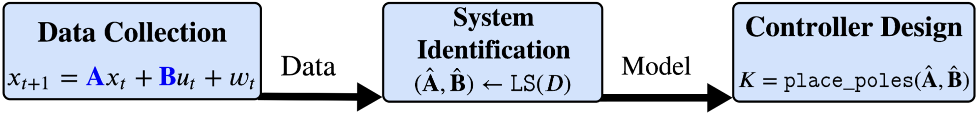
- **System Identification (sensitive to model inaccuracies)**



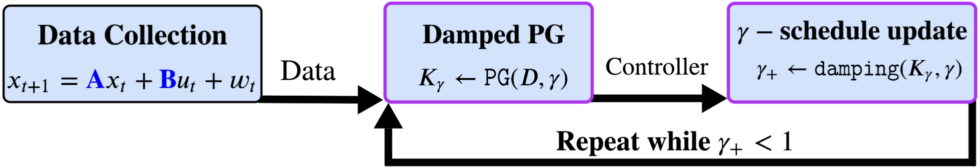
Consider $f(x_t, u_t) = Ax_t + Bu_t$, with A and B **unknown**, and **linear** feedback $u_t = Kx_t$

Question: How to design K such that $\rho(A + BK) < 1$ (**exponential stability**)?

- System Identification (sensitive to model inaccuracies)



- Policy Optimization (model-free)



Intro: Discounted Policy Gradient LQR

Consider the simple case where $w_t = 0 \; \forall t \geq 0$ (**without noise**)

$$\textbf{Discounted LQR: } \text{minimize}_K \left\{ J^\gamma(K) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t x_t^\top \left(Q + K^\top R K\right) x_t\right]\right\}$$

subject to the system dynamics $x_{t+1} = (A + BK)x_t$, where $\gamma \in (0, 1]$.

Lamperski. Computing stabilizing linear controllers via policy iteration. CDC, 2020.

Intro: Discounted Policy Gradient LQR

Consider the simple case where $w_t = 0 \; \forall t \geq 0$ (**without noise**)

$$\textbf{Discounted LQR: } \text{minimize}_K \left\{ J^\gamma(K) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t x_t^\top \left(Q + K^\top R K\right) x_t \right] \right\}$$

subject to the system dynamics $x_{t+1} = (A + BK)x_t$, where $\gamma \in (0, 1]$.

- **Equivalent Problem:** Rescaling x_t by $\gamma^{t/2}$ we have

$$\textbf{Discounted LQR: } \text{minimize}_K \left\{ J^\gamma(K) := \mathbb{E}\left[\sum_{t=0}^{\infty} x_t^\top \left(Q + K^\top R K\right) x_t \right] \right\}$$

subject to the damped dynamics $x_{t+1} = \sqrt{\gamma}(A + BK)x_t$.

- **Important:** $Q \succ 0$, $R \succ 0$ are just **artifacts**

Lamperski. Computing stabilizing linear controllers via policy iteration. CDC, 2020.

Discounted LQR: $\text{minimize}_K \left\{ J^\gamma(K) := \mathbb{E}\left[\sum_{t=0}^{\infty} x_t^\top \left(Q + K^\top R K\right) x_t\right]\right\}$

subject to the damped dynamics $x_{t+1} = \sqrt{\gamma}(A + BK)x_t$.

- **Discount method:**

1. Given $\gamma_0 < 1/\rho(A)^2$, then $K = 0$ **stabilizes** $\sqrt{\gamma_0}(A, B)$
2. **Policy Gradient** : $K \leftarrow K - \eta \widehat{\nabla} J^\gamma(K)$ such that $J^\gamma(K) \leq \bar{J}$ **(uniform bound)**
3. **Update:** $\gamma_+ \leftarrow \texttt{damping}(K, \gamma)$ such that K **stabilizes** $\sqrt{\gamma_+}(A, B)$

Intro: Discounted Policy Gradient LQR

Discounted LQR: $\text{minimize}_K \left\{ J^\gamma(K) := \mathbb{E}\left[\sum_{t=0}^{\infty} x_t^\top \left(Q + K^\top R K\right) x_t\right]\right\}$

subject to the damped dynamics $x_{t+1} = \sqrt{\gamma}(A + BK)x_t$.

- **Discount method:**

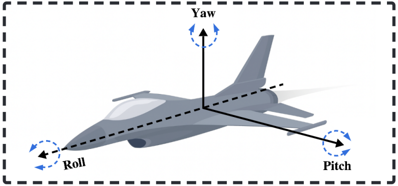
1. Given $\gamma_0 < 1/\rho(A)^2$, then $K = 0$ **stabilizes** $\sqrt{\gamma_0}(A, B)$
2. **Policy Gradient** : $K \leftarrow K - \eta \widehat{\nabla} J^\gamma(K)$ such that $J^\gamma(K) \leq \bar{J}$ **(uniform bound)**
3. **Update:** $\gamma_+ \leftarrow \texttt{damping}(K, \gamma)$ such that K **stabilizes** $\sqrt{\gamma_+}(A, B)$

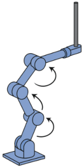
Sample Complexity: $\log(\rho(A))\mathcal{O}(d_{\mathsf{X}}^2 d_{\mathsf{U}}) \rightarrow$ prohibitive for large state dimension d_{X}

Zhao et al. Convergence and sample complexity of policy gradient methods for stabilizing linear systems. TAC, 2024.

Intro: Discounted Policy Gradient LQR

Sample Complexity: $\log(\rho(A))\mathcal{O}(d_X^2 d_U) \rightarrow$ prohibitive for large state dimension d_X





High-dimensional Robotic Arm

(balancing task) d_X is large

Few unstable modes:
- **Shoulder pitch and roll**
- **Elbow**
- **Wrist joints**

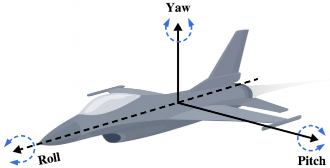
Data collection becomes:

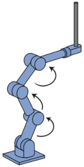
- Time consuming
- Expensive
- High risk for human interaction

Question: Can we reduce sample complexity by only acting on the **unstable** modes?

Intro: Discounted Policy Gradient LQR

Sample Complexity: $\log(\rho(A))\mathcal{O}(d_X^2 d_U) \rightarrow$ prohibitive for large state dimension d_X





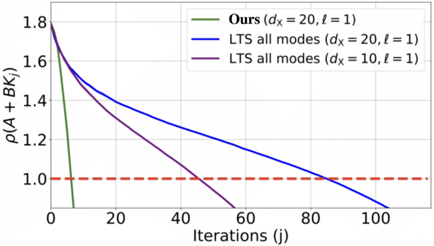
High-dimensional Robotic Arm

(balancing task) d_X is large

Few unstable modes:
- **Shoulder pitch and roll**
- **Elbow**
- **Wrist joints**

Data collection becomes:

- Time consuming
- Expensive
- High risk for human interaction



Question: Can we reduce sample complexity by only acting on the **unstable** modes?

Setup: Linear Systems

Consider the discrete-time LTI system:

$$x_{t+1} = \color{blue}{A}x_t + \color{blue}{B}u_t, \text{ for } t = 0, 1, \ldots, \text{ where } \rho(\color{blue}{A}) > 1 \textbf{ (open-loop unstable)}$$

- **Spectrum:** $\underbrace{|\lambda_1| \geq \ldots |\lambda_\ell|}_{\textbf{unstable modes}} > 1 > \underbrace{|\lambda_{\ell+1}| \geq \ldots \geq |\lambda_{d_X}|}_{\textbf{stable modes}}$, with $\ell \ll d_X$

Setup: Linear Systems

Consider the discrete-time LTI system:

$$x_{t+1} = Ax_t + Bu_t, \text{ for } t = 0, 1, \ldots, \text{ where } \rho(A) > 1 \textbf{ (open-loop unstable)}$$

- **Spectrum:** $\underbrace{|\lambda_1| \geq \ldots |\lambda_\ell|}_{\textbf{unstable modes}} > 1 > \underbrace{|\lambda_{\ell+1}| \geq \ldots \geq |\lambda_{d_X}|}_{\textbf{stable modes}}$, with $\ell \ll d_X$

- **Important:** A does not need to be **diagonalizable** (A admits a Jordan decomposition)

$$A = \Lambda J \Lambda^{-1}, \text{ with } J = \begin{bmatrix} J_u & 0 \\ 0 & J_s \end{bmatrix}, \underbrace{J_u \in \mathbb{R}^{\ell \times \ell}}_{\textbf{unstable modes}}, \underbrace{J_s \in \mathbb{R}^{(d_x - \ell) \times (d_x - \ell)}}_{\textbf{stable modes}}.$$

Goal: Design a linear policy $\pi(x_t) = u_t \triangleq Kx_t$ such that $\rho(A + BK) < 1$ (**stabilizing**)

Hu et al. On the sample complexity of stabilizing LTI systems on a single trajectory. NeurIPS, 2022.

Setup: Linear Systems

$$A = \Lambda J \Lambda^{-1}, \text{ with } J = \begin{bmatrix} J_u & 0 \\ 0 & J_s \end{bmatrix}, \underbrace{J_u \in \mathbb{R}^{\ell \times \ell}}_{\text{unstable modes}}, \underbrace{J_s \in \mathbb{R}^{(d_x - \ell)\times(d_x - \ell)}}_{\text{stable modes}}.$$

- $A\widetilde{\Phi} = J_u\widetilde{\Phi}$, where $\widetilde{\Phi}$ has ℓ **orthonormal** columns spanning the **right** unstable eigenspace
- $A^\top \Phi = J_u \Phi$, with $\Phi \in \mathbb{R}^{d_x \times \ell}$ being our **unstable subspace representation** (left)

Setup: Linear Systems

$$A = \Lambda J \Lambda^{-1}, \text{ with } J = \begin{bmatrix} J_u & 0 \\ 0 & J_s \end{bmatrix}, \underbrace{J_u \in \mathbb{R}^{\ell\times\ell}}_{\textbf{unstable modes}}, \underbrace{J_s \in \mathbb{R}^{(d_x-\ell)\times(d_x-\ell)}}_{\textbf{stable modes}}.$$

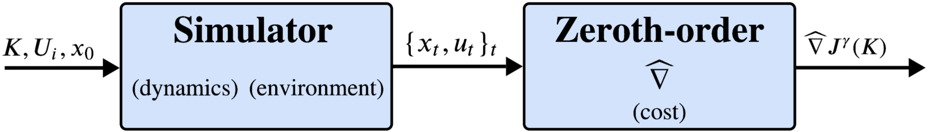
- $A\widetilde{\Phi} = J_u\widetilde{\Phi}$, where $\widetilde{\Phi}$ has ℓ **orthonormal** columns spanning the **right** unstable eigenspace
- $A^\top \Phi = J_u \Phi$, with $\Phi \in \mathbb{R}^{d_x\times\ell}$ being our **unstable subspace representation** (**left**)
- **Damped system matrices:** $A^\gamma = \sqrt{\gamma}A$, $B^\gamma = \sqrt{\gamma}B$

Policy Gradient : $K \leftarrow K - \eta\widehat{\nabla} J^\gamma(K)$, where $\widehat{\nabla} J^\gamma(K)$ is the **gradient estimation**

- **Search over** the stabilizing set $\{K \mid \rho(A^\gamma + B^\gamma K) < 1\}$ (Fazel et al., ICML 2018)

Perdomo et al. Stabilizing dynamical systems via policy gradient methods. NeurIPS, 2021.

Setup: Zeroth-order Gradient Estimation

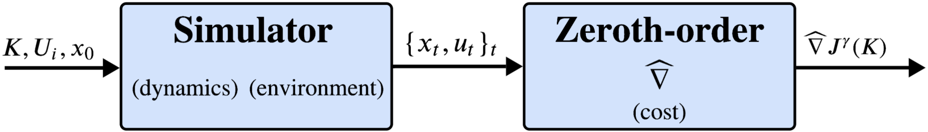


$$\mathrm{ZO}(n_s, r, \tau, K) \to \widehat{\nabla} J^\gamma(K) \triangleq \frac{1}{2rn_s} \sum_{i=1}^{n_s} \left(V^{\gamma,\tau}(K_{1,i}, x_0^i) - V^{\gamma,\tau}(K_{2,i}, x_0^i)\right) U_i,$$

- n_s : number of trajectories
- r : smoothing radius
- τ : horizon length
- $K_{\{1,2\},i} = K \pm rU_i$

Malik et al. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. JMLR, 2020.

Setup: Zeroth-order Gradient Estimation



$$\mathtt{ZO}(n_s, r, \tau, K) \to \widehat{\nabla} J^\gamma(K) \triangleq \frac{1}{2rn_s} \sum_{i=1}^{n_s} \left(V^{\gamma,\tau}(K_{1,i}, x_0^i) - V^{\gamma,\tau}(K_{2,i}, x_0^i) \right) U_i,$$

- n_s : number of trajectories
- r : smoothing radius
- τ : horizon length
- $K_{\{1,2\},i} = K \pm rU_i$

$$V^{\gamma,\tau}(K, x_0) = \sum_{t=0}^{\tau-1} \gamma^t x_t^\top \left(Q + K^\top RK \right) x_t$$

The burden in the sample complexity $\mathcal{O}(d_X^2 d_U)$ comes from this **gradient estimation**

Malik et al. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. JMLR, 2020.

Stabilizing Only the Unstable Modes

Consider first the setting where $\Omega \triangleq [\Phi \;\; \Phi_\perp]$ is a **given** orthonormal basis of $\mathbb{R}^{d_X}$

Stabilizing Only the Unstable Modes

Consider first the setting where $\Omega \triangleq [\Phi \;\; \Phi_\perp]$ is a **given** orthonormal basis of $\mathbb{R}^{d_X}$

- **Left unstable subspace decomposition:**

$$\Omega^\top A \Omega = \begin{bmatrix} A_u & \\ \Delta & A_s \end{bmatrix}, \text{ with } \underbrace{A_u \triangleq \Phi^\top A \Phi}_{\text{unstable modes}}, \quad \underbrace{\Delta \triangleq \Phi_\perp^\top A \Phi}_{\text{coupling}}, \text{ and } \underbrace{A_s \triangleq \Phi_\perp^\top A \Phi_\perp}_{\text{stable modes}}$$

Stabilizing Only the Unstable Modes

Consider first the setting where $\Omega \triangleq [\Phi \;\; \Phi_\perp]$ is a **given** orthonormal basis of $\mathbb{R}^{d_x}$

- **Left unstable subspace decomposition:**

$$\Omega^\top A \Omega = \begin{bmatrix} A_u & \\ \Delta & A_s \end{bmatrix}, \text{ with } \underbrace{A_u \triangleq \Phi^\top A \Phi}_{\text{unstable modes}}, \quad \underbrace{\Delta \triangleq \Phi_\perp^\top A \Phi}_{\text{coupling}}, \text{ and } \underbrace{A_s \triangleq \Phi_\perp^\top A \Phi_\perp}_{\text{stable modes}}$$

Policy representation: $K = \theta \Phi^\top, \mathrm{col}(\Phi) \to$ **left unstable subspace**

$$A+BK = \Omega \begin{bmatrix} A_u + B_u\theta \;\textbf{(unstable)} & \\ \Delta + B_s\theta \;\textbf{(coupling)} & A_s \;\textbf{(stable)} \end{bmatrix} \Omega^\top, \text{ with } B_u \triangleq \Phi^\top B \text{ and } B_s \triangleq \Phi_\perp^\top B$$

Stabilizing Only the Unstable Modes

Consider first the setting where $\Omega \triangleq [\Phi \;\; \Phi_\perp]$ is a **given** orthonormal basis of $\mathbb{R}^{d_x}$

- **Left unstable subspace decomposition:**

$$\Omega^\top A \Omega = \begin{bmatrix} A_u & \\ \Delta & A_s \end{bmatrix}, \text{ with } \underbrace{A_u \triangleq \Phi^\top A \Phi}_{\text{unstable modes}}, \quad \underbrace{\Delta \triangleq \Phi_\perp^\top A \Phi}_{\text{coupling}}, \text{ and } \underbrace{A_s \triangleq \Phi_\perp^\top A \Phi_\perp}_{\text{stable modes}}$$

Policy representation: $K = \theta \Phi^\top, \mathrm{col}(\Phi) \to$ **left unstable subspace**

$$A+BK = \Omega \begin{bmatrix} A_u + B_u\theta \ \textbf{(unstable)} & \\ \Delta + B_s\theta \ \textbf{(coupling)} & A_s \ \textbf{(stable)} \end{bmatrix} \Omega^\top, \text{ with } B_u \triangleq \Phi^\top B \text{ and } B_s \triangleq \Phi_\perp^\top B$$

Spectral radius: $\rho(A+BK) = \rho(A_u + B_u\theta)$, for $\theta \in \mathbb{R}^{d_u \times \ell} \to$ **"smaller" problem**

Decomposing onto the Right Unstable Subspace

Question: Why not to consider the **right** unstable subspace of A?

$$A + BK = \widetilde{\Omega} \begin{bmatrix} \widetilde{A}_u + \widetilde{B}_u \theta \ \textbf{(unstable)} & \widetilde{\Delta} \ \textbf{(coupling)} \\ \widetilde{B}_s \theta & \widetilde{A}_s \ \textbf{(stable)} \end{bmatrix} \widetilde{\Omega}^\top$$

with the **coupling** $\widetilde{\Delta} \triangleq \widetilde{\Phi}^\top A \widetilde{\Phi}_\perp$ in the "wrong" position

Decomposing onto the Right Unstable Subspace

Question: Why not to consider the **right** unstable subspace of A?

$$A + BK = \widetilde{\Omega} \begin{bmatrix} \widetilde{A}_u + \widetilde{B}_u\theta \ \textbf{(unstable)} & \widetilde{\Delta} \ \textbf{(coupling)} \\ \widetilde{B}_s\theta & \widetilde{A}_s \ \textbf{(stable)} \end{bmatrix} \widetilde{\Omega}^\top$$

with the **coupling** $\widetilde{\Delta} \triangleq \widetilde{\Phi}^\top A \widetilde{\Phi}_\perp$ in the "wrong" position

Spectral radius: $\rho(A+BK) \neq \rho(\widetilde{A}_u + \widetilde{B}_u\theta)$, unless $\widetilde{\Delta} \equiv 0$ (**symmetric** A)

Decomposing onto the Right Unstable Subspace

Question: Why not to consider the **right** unstable subspace of A?

$$A + BK = \widetilde{\Omega} \begin{bmatrix} \widetilde{A}_u + \widetilde{B}_u\theta \ \textbf{(unstable)} & \widetilde{\Delta} \ \textbf{(coupling)} \\ \widetilde{B}_s\theta & \widetilde{A}_s \ \textbf{(stable)} \end{bmatrix} \widetilde{\Omega}^\top$$

with the **coupling** $\widetilde{\Delta} \triangleq \widetilde{\Phi}^\top A \widetilde{\Phi}_\perp$ in the "wrong" position

Spectral radius: $\rho(A+BK) \neq \rho(\widetilde{A}_u + \widetilde{B}_u\theta)$, unless $\widetilde{\Delta} \equiv 0$ **(symmetric A)**

Goal: Design a low-dimensional policy $\pi(z_t) = \theta z_t$, with $z_t = \Phi^\top x_t$, such that $\rho(A_u + B_u\theta) < 1$

Decomposing onto the Right Unstable Subspace

Question: Why not to consider the **right** unstable subspace of A?

$$A + BK = \widetilde{\Omega} \begin{bmatrix} \widetilde{A}_u + \widetilde{B}_u\theta \ \textbf{(unstable)} & \widetilde{\Delta} \ \textbf{(coupling)} \\ \widetilde{B}_s\theta & \widetilde{A}_s \ \textbf{(stable)} \end{bmatrix} \widetilde{\Omega}^\top$$

with the **coupling** $\widetilde{\Delta} \triangleq \widetilde{\Phi}^\top A \widetilde{\Phi}_\perp$ in the "wrong" position

Spectral radius: $\rho(A + BK) \neq \rho(\widetilde{A}_u + \widetilde{B}_u\theta)$, unless $\widetilde{\Delta} \equiv 0$ **(symmetric A)**

Goal: Design a low-dimensional policy $\pi(z_t) = \theta z_t$, with $z_t = \Phi^\top x_t$, such that $\rho(A_u + B_u\theta) < 1$

- **Challenge:** We don't have access to Φ as A is **unknown**
- **Idea:** Learn Φ from trajectory data when $u_t \equiv 0$

Learning the Left Unstable Representation

We compute an **estimation** of $\mathit{\Phi}$ denoted by $\widehat{\mathit{\Phi}}$

- **Subspace distance:** $d(\widehat{\mathit{\Phi}}, \mathit{\Phi}) \triangleq \|\widehat{\mathit{\Phi}}^\top \mathit{\Phi}_\perp\|$ (Stewart and Sun, 1990)

Learning the Left Unstable Representation

We compute an **estimation** of $\mathit{\Phi}$ denoted by $\widehat{\mathit{\Phi}}$

- **Subspace distance:** $d(\widehat{\mathit{\Phi}}, \mathit{\Phi}) \triangleq \|\widehat{\mathit{\Phi}}^\top \mathit{\Phi}_\perp\|$ (Stewart and Sun, 1990)

Idea: We compute $\widehat{\mathit{\Phi}}$ by sampling from the autonomous **adjoint system**

Learning the Left Unstable Representation

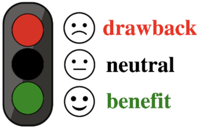
We compute an **estimation** of Φ denoted by $\widehat{\Phi}$

- **Subspace distance:** $d(\widehat{\Phi}, \Phi) \triangleq \|\widehat{\Phi}^\top \Phi_\perp\|$ (Stewart and Sun, 1990)

Idea: We compute $\widehat{\Phi}$ by sampling from the autonomous **adjoint system**

1. **Simulate the adjoint:** $x_{t+1} = A^\top x_t = \left[x_t^\top e_1^+ \dots x_t^\top e_{d_X}^+\right]^\top, e_i^+ = Ae_i$
2. **Adjoint data:** $D = [x_1, x_2, \dots, x_T] \in \mathbb{R}^{d_X \times T}$ with horizon length T
3. **Estimation:** $D = U\Sigma V^\top \rightarrow \widehat{\Phi} = [u_1, \dots, u_\ell]$

where e_i is the i-th canonical basis vector of $\mathbb{R}^{d_X}$



Learning the Left Unstable Representation

Theorem (informal). Suppose the amount of trajectory data to learn $\widehat{\Phi}$ scales as

$$T = \mathcal{O}\left(\log\left(\frac{\ell^7(d_X - \ell)}{(1 - |\lambda_{\ell+1}|)\varepsilon}\right) / \log(|\lambda_\ell|)\right),$$

then $d(\widehat{\Phi}, \Phi) \leq \varepsilon$ with high probability.

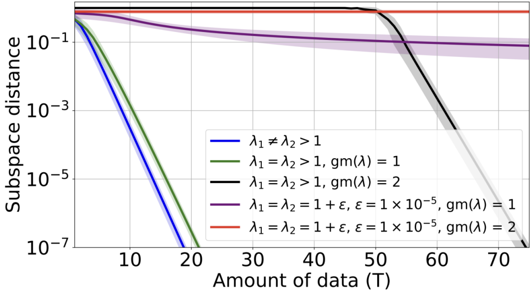
Learning the Left Unstable Representation

Theorem (informal). Suppose the amount of trajectory data to learn $\widehat{\Phi}$ scales as

$$T = \mathcal{O}\left(\log\left(\frac{\ell^7(d_X - \ell)}{(1 - |\lambda_{\ell+1}|)\varepsilon}\right) / \log(|\lambda_\ell|)\right),$$

then $d(\widehat{\Phi}, \Phi) \leq \varepsilon$ with high probability.

- **Example:** $d_X = 3$ and $\ell = 2$
- $\mathrm{gm}(\lambda)$: geometric multiplicity



Learning the Left Unstable Representation

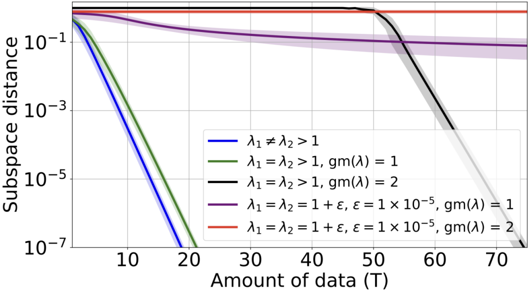
Theorem (informal). Suppose the amount of trajectory data to learn $\widehat{\Phi}$ scales as

$$T = \mathcal{O}\left(\log\left(\frac{\ell^7 (d_X - \ell)}{(1 - |\lambda_{\ell+1}|)\varepsilon}\right) / \log(|\lambda_\ell|)\right),$$

then $d(\widehat{\Phi}, \Phi) \leq \varepsilon$ with high probability.

- **Example:** $d_X = 3$ and $\ell = 2$
- $\mathrm{gm}(\lambda)$: geometric multiplicity

Estimating the unstable subspace is inconsistent when $\mathrm{gm}(\lambda) > 1$, for any unstable mode $\lambda > 1$.



Low-Dimensional Discounted LQR

Discounted LQR: $\text{minimize}_\theta \left\{ J^\gamma(\theta, \widehat{\Phi}) \triangleq \mathbb{E}\left[\sum_{t=0}^{\infty} z_t^\top \left(\widehat{\Phi}^\top Q \widehat{\Phi} + \theta^\top R \theta\right) z_t\right]\right\},$

subject to the damped low-dimensional dynamics $\hat{A}_u^\gamma \triangleq \sqrt{\gamma}\widehat{\Phi}^\top A \widehat{\Phi}$, $\hat{B}_u^\gamma \triangleq \sqrt{\gamma}\widehat{\Phi}^\top B$

Policy Gradient: $\theta \leftarrow \theta - \eta \widehat{\nabla} J^\gamma(\theta, \widehat{\Phi})$ **(low-dimensional)**

Low-Dimensional Discounted LQR

$$\textbf{Discounted LQR: } \text{minimize}_\theta \left\{ J^\gamma(\theta, \widehat{\Phi}) \triangleq \mathbb{E}\left[\sum_{t=0}^{\infty} z_t^\top \left(\widehat{\Phi}^\top Q \widehat{\Phi} + \theta^\top R \theta \right) z_t \right] \right\},$$

subject to the damped low-dimensional dynamics $\hat{A}_u^\gamma \triangleq \sqrt{\gamma}\widehat{\Phi}^\top A \widehat{\Phi}$, $\hat{B}_u^\gamma \triangleq \sqrt{\gamma}\widehat{\Phi}^\top B$

Policy Gradient: $\theta \leftarrow \theta - \eta \widehat{\nabla} J^\gamma(\theta, \widehat{\Phi})$ **(low-dimensional)**

Suppose θ is stabilizing for the low-dimensional system with A_u^γ and B_u^γ

$$\left\| \nabla J^\gamma(\theta, \Phi) - \nabla J^\gamma(\theta, \widehat{\Phi}) \right\|_F \leq C_\Phi d(\widehat{\Phi}, \Phi), \text{ with } C_\Phi = \mathcal{O}(\ell)$$

If $d(\widehat{\Phi}, \Phi)$ is **sufficiently small**, PG with $\nabla J^\gamma(\theta, \widehat{\Phi})$ **looks like** PG with $\nabla J^\gamma(\theta, \Phi)$

Discount Method on the Unstable Subspace

Initialize: γ_0 **sufficiently small** $\rightarrow \theta \equiv 0$ stabilize the damped low-dimensional system

While $\gamma_j < 1$ **do**

Initialize $\bar{\theta}_0 = \theta_j$ and **for** $n = 0, 1, \ldots, N$ **do** $\bar{\theta}_{n+1} = \bar{\theta}_n - \eta \widehat{\nabla} J^{\gamma_j}(\bar{\theta}_n, \widehat{\Phi})$

Let $\theta_{j+1} = \bar{\theta}_N$ and **compute** $\alpha_j(\theta_j, \widehat{\Phi}) = \frac{3\sigma_{\min}\left(\widehat{\Phi}^\top Q \widehat{\Phi} + \theta_j^\top R \theta_j\right)}{\frac{4}{3}\widehat{J}^{\gamma_j}(\theta_j, \widehat{\Phi}) - 3\sigma_{\min}\left(\widehat{\Phi}^\top Q \widehat{\Phi} + \theta_j^\top R \theta_j\right)}$

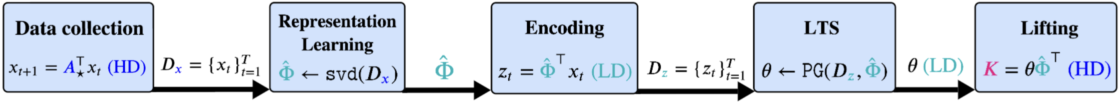
Update $\gamma_{j+1} = \mathtt{damping}(\theta_{j+1}, \gamma_j) \triangleq (1 + \xi \alpha_j(\theta_j, \widehat{\Phi}))\gamma_j$ with $\xi \in (0, 1)$

$j \leftarrow j + 1$

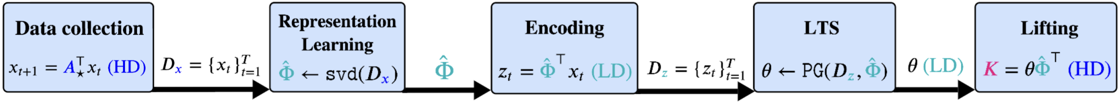
Explicit damping update: $\gamma_+ \leftarrow (1 + \xi \alpha(\theta, \widehat{\Phi}))\gamma$ **(Lyapunov Stability Analysis)**

Zhao et al. Convergence and sample complexity of policy gradient methods for stabilizing linear systems. TAC, 2024.

Sample Complexity Analysis



Sample Complexity Analysis

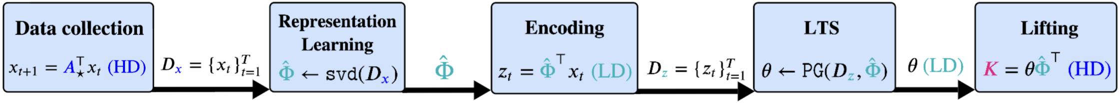


Theorem (informal). Let n_s, r, and τ be set accordingly, and the number of **adjoint** samples T be sufficiently large such that

$$d(\widehat{\Phi}, \Phi) \leq \varepsilon \triangleq \mathcal{O}\left((1 - \max\{\rho(A_u + B_u\theta_{j+1}), |\lambda_{\ell+1}|\})^{\ell}\right),$$

then $\pi(x_t) = Kx_t = \theta_{j+1}\widehat{\Phi}^\top x_t$ is **stabilizing** for the original system (A, B).

Sample Complexity Analysis



Theorem (informal). Let n_s, r, and τ be set accordingly, and the number of **adjoint** samples T be sufficiently large such that

$$d(\widehat{\Phi}, \Phi) \leq \varepsilon \triangleq \mathcal{O}\left((1 - \max\{\rho(A_u + B_u\theta_{j+1}), |\lambda_{\ell+1}|\})^{\ell} \right),$$

then $\pi(x_t) = Kx_t = \theta_{j+1}\widehat{\Phi}^\top x_t$ is **stabilizing** for the original system (A, B).

$$\begin{bmatrix} A_u + B_u\theta_{j+1}\widehat{\Phi}^\top\Phi & B_u\theta_{j+1}\widehat{\Phi}^\top\Phi_\perp \\ \Delta + B_s\theta_{j+1}\widehat{\Phi}^\top\Phi & A_s + B_s\theta_{j+1}\widehat{\Phi}^\top\Phi_\perp \end{bmatrix}$$

Learning a low-dimensional control gain $\theta \in \mathbb{R}^{d_U \times \ell}$ guarantees stabilization of a high-dimensional system (A,B) through $\widehat{\Phi}$.

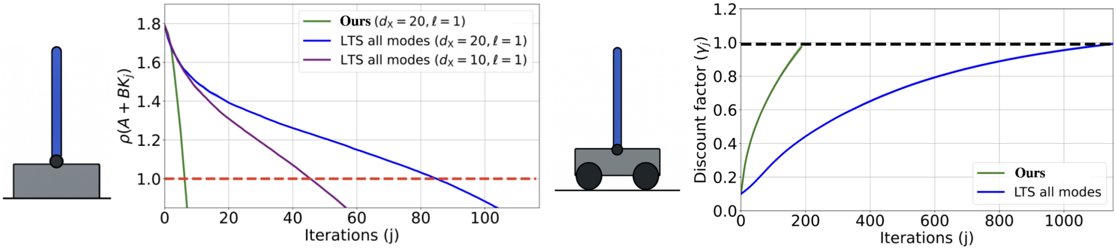
Mathias. Quadratic residual bounds for the Hermitian eigenvalue problem. SIMAX, 1998.

Sample Complexity Analysis

Corollary (informal). Let n_s, r, τ, and T be set accordingly, then a stabilizing policy $\pi(x_t) = Kx_t$ is **learned** with only $\log(\rho(A))\mathcal{O}(\ell^2 d_{\mathrm{U}})$ samples.

Sample Complexity Analysis

Corollary (informal). Let n_s, r, τ, and T be set accordingly, then a stabilizing policy $\pi(x_t) = Kx_t$ is **learned** with only $\log(\rho(A))\mathcal{O}(\ell^2 d_{\mathrm{U}})$ samples.



Augmented inverted pendulum and cartpole systems with **random** stable modes

Operate on the unstable subspace for LTS with a sufficient accurate unstable subspace representation $\widehat{\Phi}$ **reduces** sample complexity from $\mathcal{O}(d_{\mathrm{X}}^2 d_{\mathrm{U}})$ to $\mathcal{O}(\ell^2 d_{\mathrm{U}})$

Key Takeaways and Future Work

Learning to stabilize **all** modes is very **expensive** $\mathcal{O}(d_X^2)$

- We considered learning to stabilize only the ℓ **unstable modes**
- We parameterized K with a low-dimensional controller + a representation ($K \triangleq \theta \widehat{\Phi}^\top$)
- We learned $\widehat{\Phi}$ with $T = \mathcal{O}\left(\texttt{polylog}(\ell/(1-|\lambda_{\ell+1}|))/\log(|\lambda_\ell|)\right)$ **adjoint** samples
- We proved that by controlling $d(\widehat{\Phi}, \Phi)$, LTS takes only $\mathcal{O}(\ell^2)$ samples

Key Takeaways and Future Work

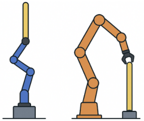
Learning to stabilize **all** modes is very **expensive** $\mathcal{O}(d_X^2)$

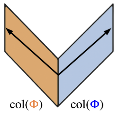
- We considered learning to stabilize only the ℓ **unstable modes**
- We parameterized K with a low-dimensional controller + a representation ($K \triangleq \theta \widehat{\Phi}^\top$)
- We learned $\widehat{\Phi}$ with $T = \mathcal{O}\left(\texttt{polylog}(\ell/(1-|\lambda_{\ell+1}|))/\log(|\lambda_\ell|)\right)$ **adjoint** samples
- We proved that by controlling $d(\widehat{\Phi}, \Phi)$, LTS takes only $\mathcal{O}(\ell^2)$ samples

What's next:

- Learn the **representation** and the **stabilizing policy** from **stochastic** data ($w_t \neq 0$)
- Learn and **refine** the representation **online** as more data becomes available
- **Multitask** setting:

Multiple high dim. systems
with aligned unstable subspace
simultaneous stab. + **adaptation**



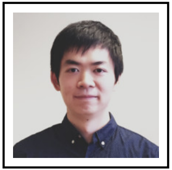


References

- Hu et al. On the sample complexity of stabilizing LTI systems on a single trajectory. NeurIPS 2022.
- Malik et al. Derivative-free methods for policy optimization: Guarantees for linear quadratic systems. JMLR 2020.
- Stewart and Sun. Matrix Perturbation Theory. Academic Press 1990.
- Mathias. Quadratic residual bounds for the Hermitian eigenvalue problem. SIMAX 1998.
- Perdomo et al. Stabilizing dynamical systems via policy gradient methods. NeurIPS 2021.
- Fazel et al. Global convergence of policy gradient methods for the linear quadratic regulator. ICML 2018.
- Zhao et al. Convergence and sample complexity of policy gradient methods for stabilizing linear systems. TAC 2024.
- Lamperski. Computing stabilizing linear controllers via policy iteration. CDC, 2020.

Acknowledgments

James Anderson



Lintao Ye

SCAN ME

Website

SCAN ME

Full Paper

Capital One

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

NSF

国家自然科学基金委员会
1986
NATIONAL NATURAL SCIENCE FOUNDATION OF CHINA

COLUMBIA UNIVERSITY
DATA SCIENCE INSTITUTE

Happy to take questions!