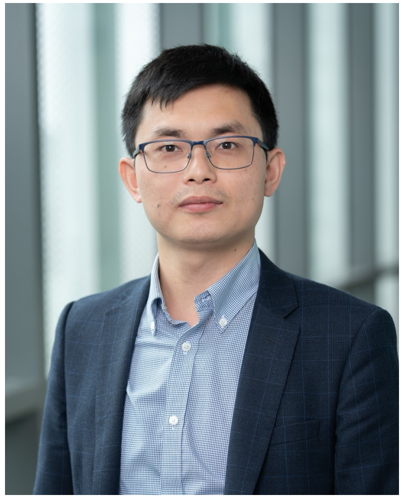
Benign Nonconvex Landscapes in Optimal and Robust Control

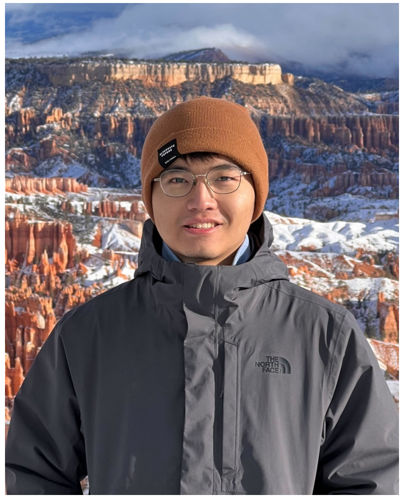
唐聿劼 Yujie Tang

北京大学 工业工程与管理系

Department of Industrial Engineering & Management, Peking University

Acknowledgement





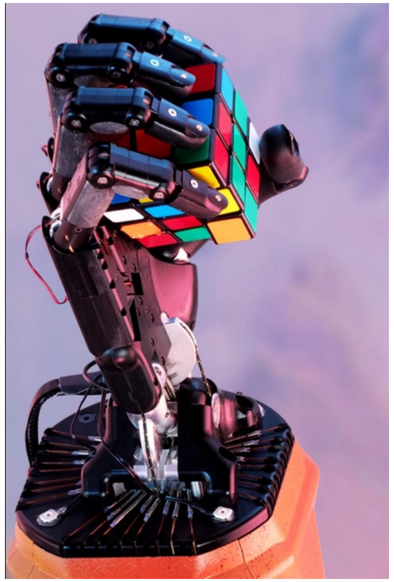
Yang Zheng
University of California San Diego

Chih-Fan (Rich) Pai
University of California San Diego

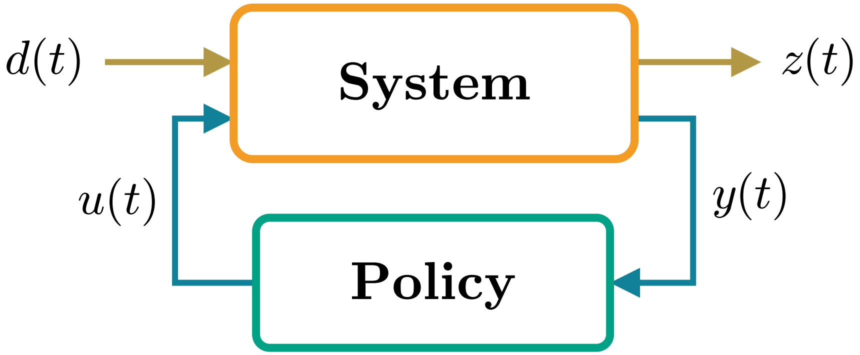
- Yujie Tang, Yang Zheng. "**On the Global Optimality of Direct Policy Search for Nonsmooth $\mathcal{H}_\infty$ Output-Feedback Control**." In Proceedings of the 62nd IEEE Conference on Decision and Control (CDC), pp. 6148-6153, 2023.
- Yang Zheng, Chih-Fan Pai, Yujie Tang. "**Benign Nonconvex Landscapes in Optimal and Robust Control, Part I: Global Optimality**." Preprint arXiv:2312.15332 (2023)
- Yang Zheng, Chih-Fan Pai, Yujie Tang. "**Benign Nonconvex Landscapes in Optimal and Robust Control, Part II: Extended Convex Lifting**." Preprint arXiv:2406.04001 (2024)

Success of Data-driven Decision Making

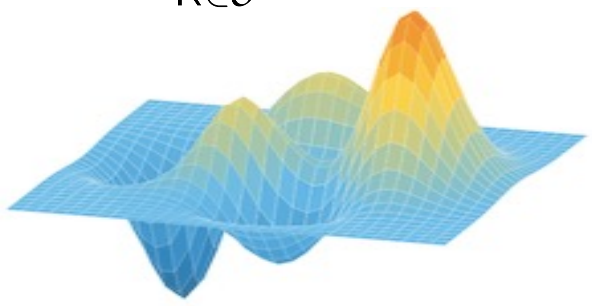
- **Data-driven decision-making** has achieved great success for complex tasks in dynamical systems, e.g., robotic manipulation/locomotion, networked systems, game playing, etc.
- **Reinforcement learning (RL)** has served as one backbone of the recent successe of data-driven decision-making.
- **Policy optimization** as one of the major workhorses of modern RL.



Policy Optimization for Control



$$\min_{\mathsf{K} \in \mathcal{C}} \; J(\mathsf{K})$$



Opportunities

- **Easy-to-implement**
- **Scalable** to high-dimensional problems
- Enable **model-free search** with rich observations

Challenges

- **Nonconvex optimization**
- Lack of principled algorithms for **optimality** (e.g., avoiding saddles/local minimizers)
- Hard to obtain **theoretical guarantees** (e.g., robustness/stability, sample efficiency)

Some Historical Background

Major approaches for optimal & robust controller synthesis:

- **Solving Riccati equations**
- **LMI-based convex reformulation**
- **Policy optimization**

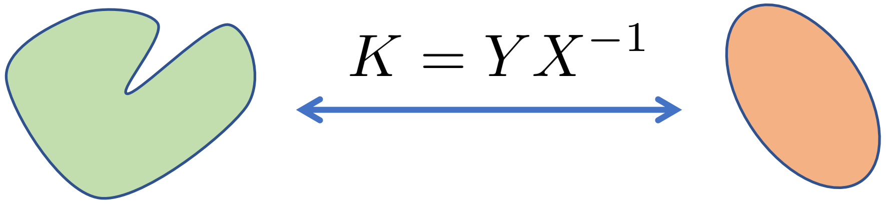
Some Historical Background

Major approaches for optimal & robust controller synthesis:

- Solving Riccati equations
- **LMI-based convex reformulation**
- **Policy optimization**

LMI-based convex reformulation

- Has became popular since 1980s due to **global guarantees** and **efficient interior point solvers**
- Relies on **re-parameterizations** (does not optimize over controller/policy directly)



- Examples: State-feedback or full-order output-feedback $\mathcal{H}_2/\mathcal{H}_\infty$ control, etc.

Policy optimization

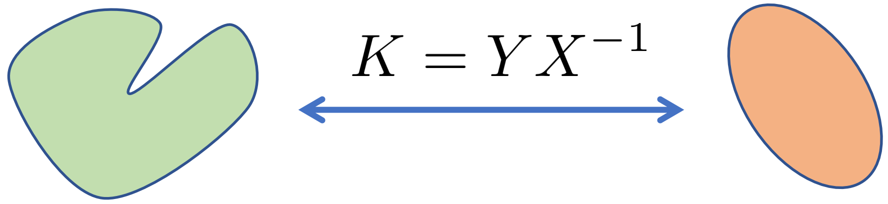
- Has a long history in control theory
  - [Apkarian & Noll, 2006] [Saeki, 2006] [Apkarian et al., 2008] [Gumussoy et al., 2009] [Arzelier et al., 2011], etc.
  - HIFOO, `hinfstruct`
- **Good empirical performance**
  - Scalability, flexibility, …
- **Weak guarantees**, unpopular among theorists

Some Historical Background

Major approaches for optimal & robust controller synthesis:

- Solving Riccati equations
- **LMI-based convex reformulation**

  - Has became popular since 1980s due to **global guarantees** and **efficient interior point solvers**
  - Relies on **re-parameterizations** (does not optimize over controller/policy directly)



  - Examples: State-feedback or full-order output-feedback $\mathcal{H}_2/\mathcal{H}_\infty$ control, etc.

- **Policy optimization**

  - **Favorable properties** have been revealed recently for a range of benchmark problems:
    - ✓ LQR
    - ✓ LQG
    - ✓ $\mathcal{H}_\infty$ state-feedback
  - A recent survey paper:

ANNUAL REVIEW OF CONTROL, ROBOTICS, AND AUTONOMOUS SYSTEMS Volume 6, 2023

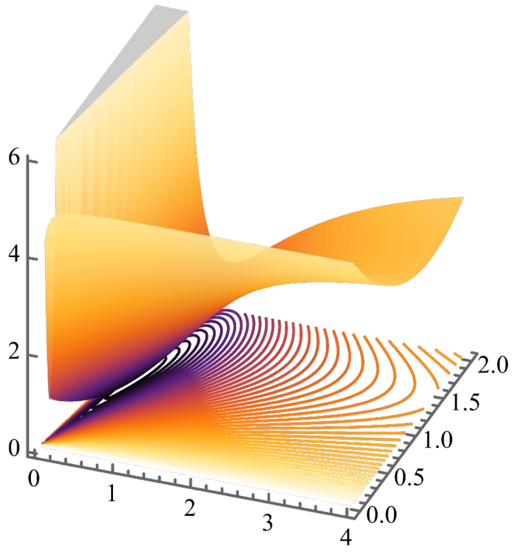
Review Article | Open Access

Toward a Theoretical Foundation of Policy Optimization for Learning Control Policies

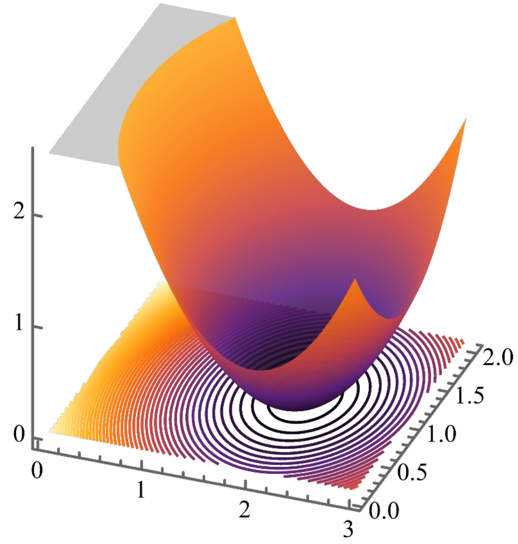
Bin Hu[1], Kaiqing Zhang[2,3], Na Li[4], Mehran Mesbahi[5], Maryam Fazel[6], and Tamer Başar[1]

Our Focus

This talk: Benign Nonconvexity in Control via Extended Convex Lifting (ECL)



Nonconvex policy optimization ⟷ **LMI-based convex reformulation**

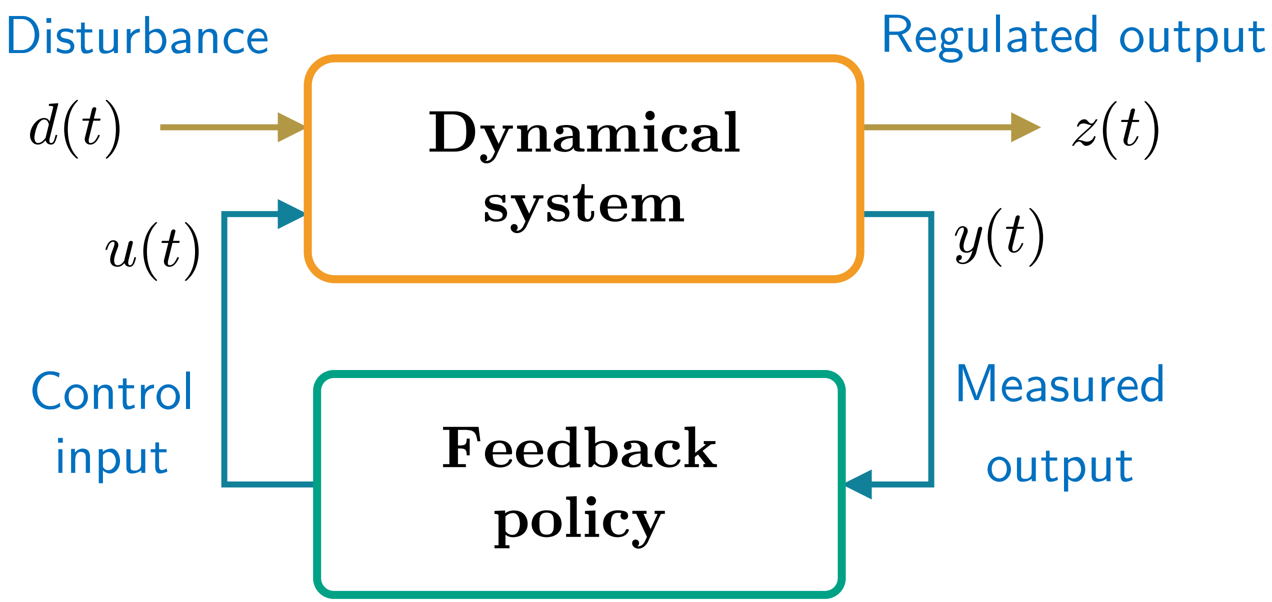


- Reconciles the gap between **nonconvex policy optimization** and **LMI-based convex reformulations**.
- For **non-degenerate** policies, **all Clarke stationary points are globally optimal**.

Outline

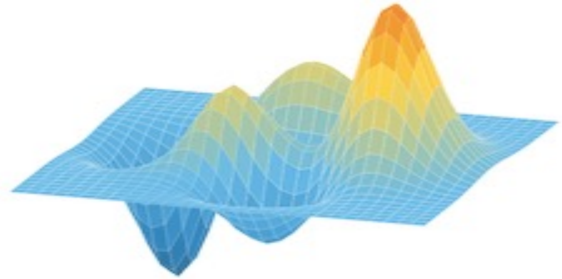
- **Problem Setup and Motivating Examples**
- Extended Convex Lifting (ECL)
- Applications for Optimal and Robust Control
- Conclusions

Problem Setup



Policy parameterization

$$\min_{\mathsf{K}} \quad J(\mathsf{K})$$
$$\text{s.t.} \quad \mathsf{K} \in \mathcal{C}$$



System dynamics

$$\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t)$$
$$y(t) = Cx(t) + D_v v(t)$$

Performance signal

$$z(t) = \begin{bmatrix} Q^{1/2}x(t) \\ R^{1/2}u(t) \end{bmatrix}$$

State feedback

$$u(t) = Kx(t)$$

Output feedback

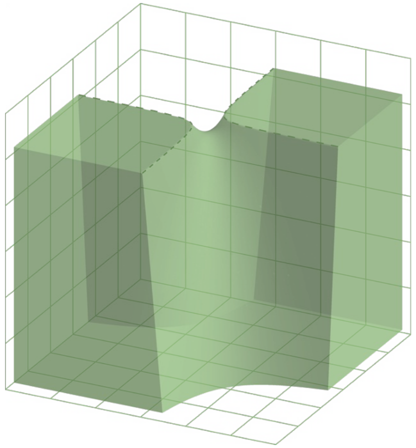
$$\frac{d\xi(t)}{dt} = A_{\mathsf{K}}\xi(t) + B_{\mathsf{K}}y(t)$$
$$u(t) = C_{\mathsf{K}}\xi(t) + D_{\mathsf{K}}y(t)$$

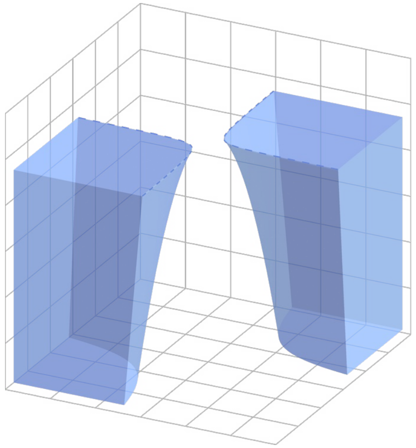
$$\mathcal{C} = \{\mathsf{K} : \text{Closed-loop system is stable}\}$$

Challenges in Policy Optimization

$$\min_{\mathsf{K}} \quad J(\mathsf{K})$$
$$\text{s.t.} \quad \mathsf{K} \in \mathcal{C}$$

Policy optimization is generally **nonconvex**!

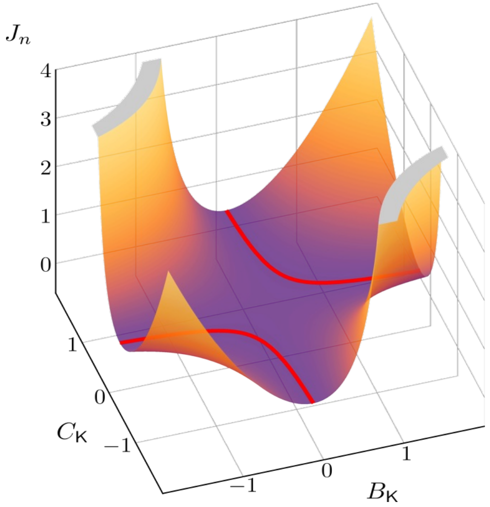


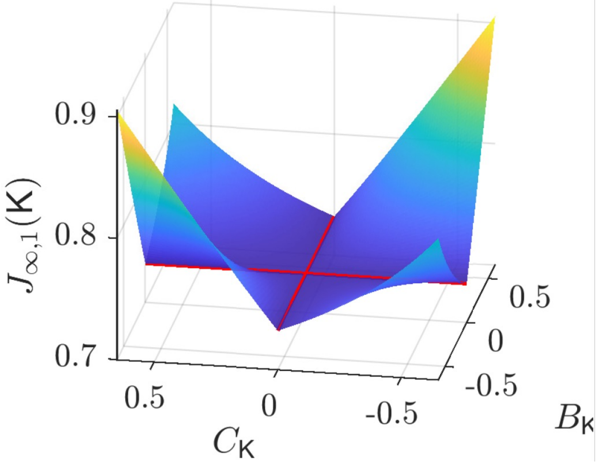


- The set of dynamic stabilizing policies is **nonconvex** and may even be **disconnected.** [Tang, Zheng, Li, 2023]

- LQR/LQG costs are **smooth** but **nonconvex**
- $\mathcal{H}_\infty$ cost are **non-smooth** and **nonconvex**

A long way to go if we want to establish theoretical guarantees!

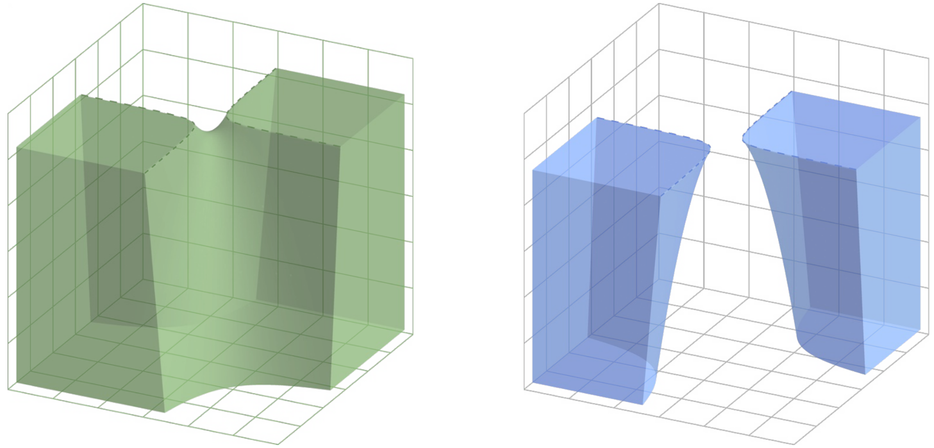




Challenges in Policy Optimization

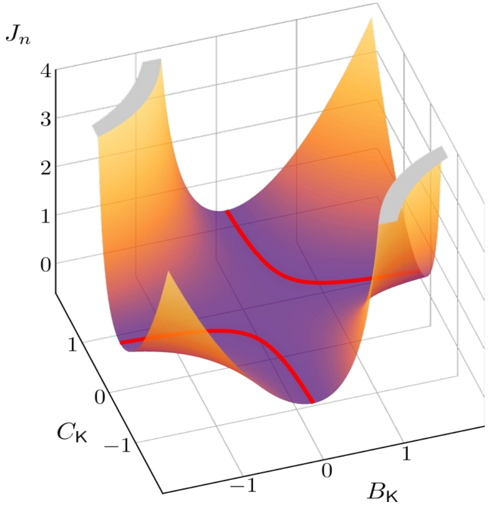
$$\min_{\mathsf{K}} \quad J(\mathsf{K})$$
$$\text{s.t.} \quad \mathsf{K} \in \mathcal{C}$$

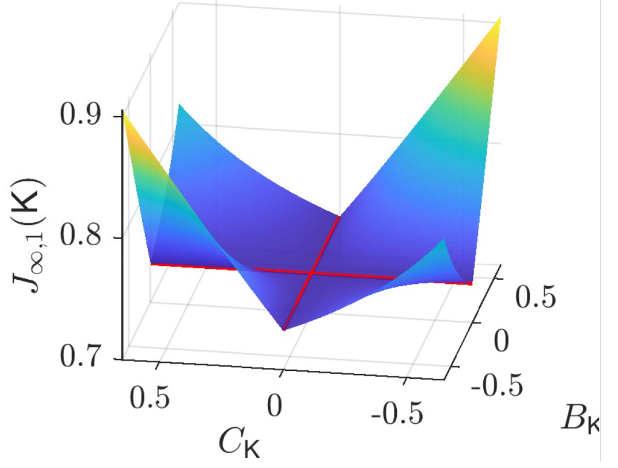
Policy optimization is generally **nonconvex**!



- The set of dynamic stabilizing policies is **nonconvex** and may even be **disconnected.** [Tang, Zheng, Li, 2023]

- LQR/LQG costs are **smooth** but **nonconvex**
- $\mathcal{H}_\infty$ cost are **non-smooth** and **nonconvex**





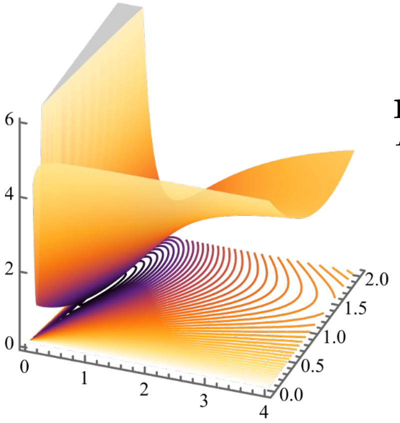
Start from the very basic:

When is a stationary point globally optimal?

Inspiration from Convex Reformulations

Our idea: Exploit **LMI-based convex reformulations** of control problems

- They reveal the **hidden convexity** of policy optimization landscapes
- Quite successful for LQR/$\mathcal{H}_\infty$ state-feedback control

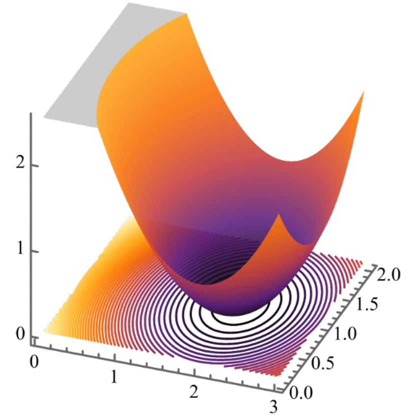


$$\begin{aligned} \min_{K,X} \quad & \operatorname{tr}\left[(Q + K^\mathsf{T} R K) X\right] \\ \text{s.t.} \quad & X = \operatorname{Lyap}(A + BK, W) \\ & X \succ 0 \end{aligned}$$

$Y = KX$

Change of variable

$$\begin{aligned} \min_{X,Y} \quad & \operatorname{tr}\left(Q + X^{-1} Y^\mathsf{T} R Y\right) \\ \text{s.t.} \quad & 0 = AX + BY \\ & \quad + XA^\mathsf{T} + Y^\mathsf{T} B^\mathsf{T} + W \\ & X \succ 0 \end{aligned}$$



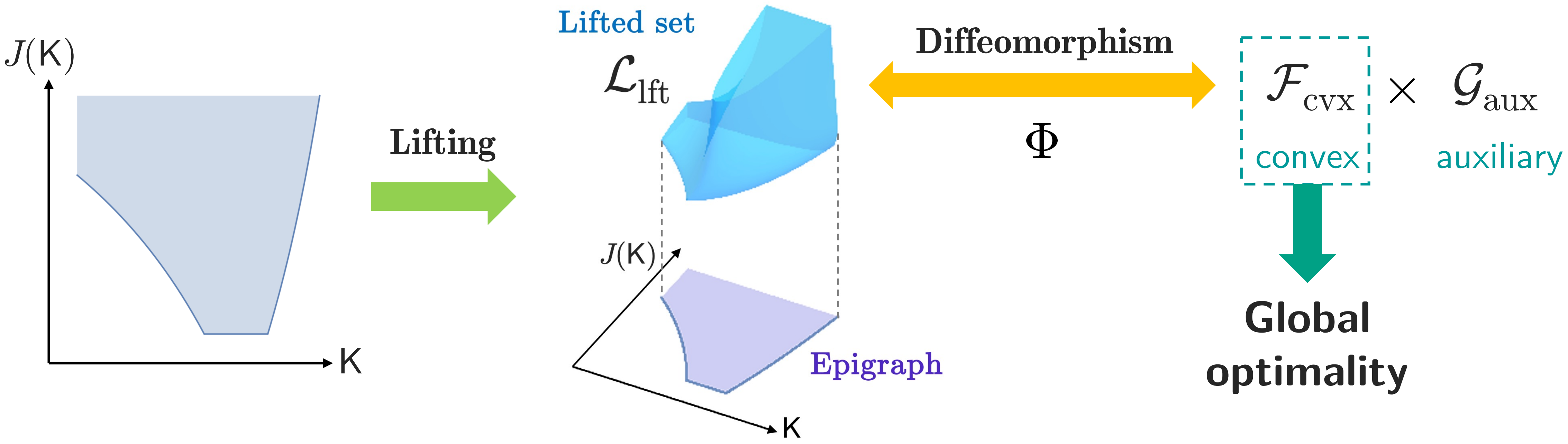
- Can we build a **general framework** for those control problems with convex reformulations?

Outline

- Problem Setup and Motivating Examples
- **Extended Convex Lifting (ECL)**
- Applications for Optimal and Robust Control
- Conclusions

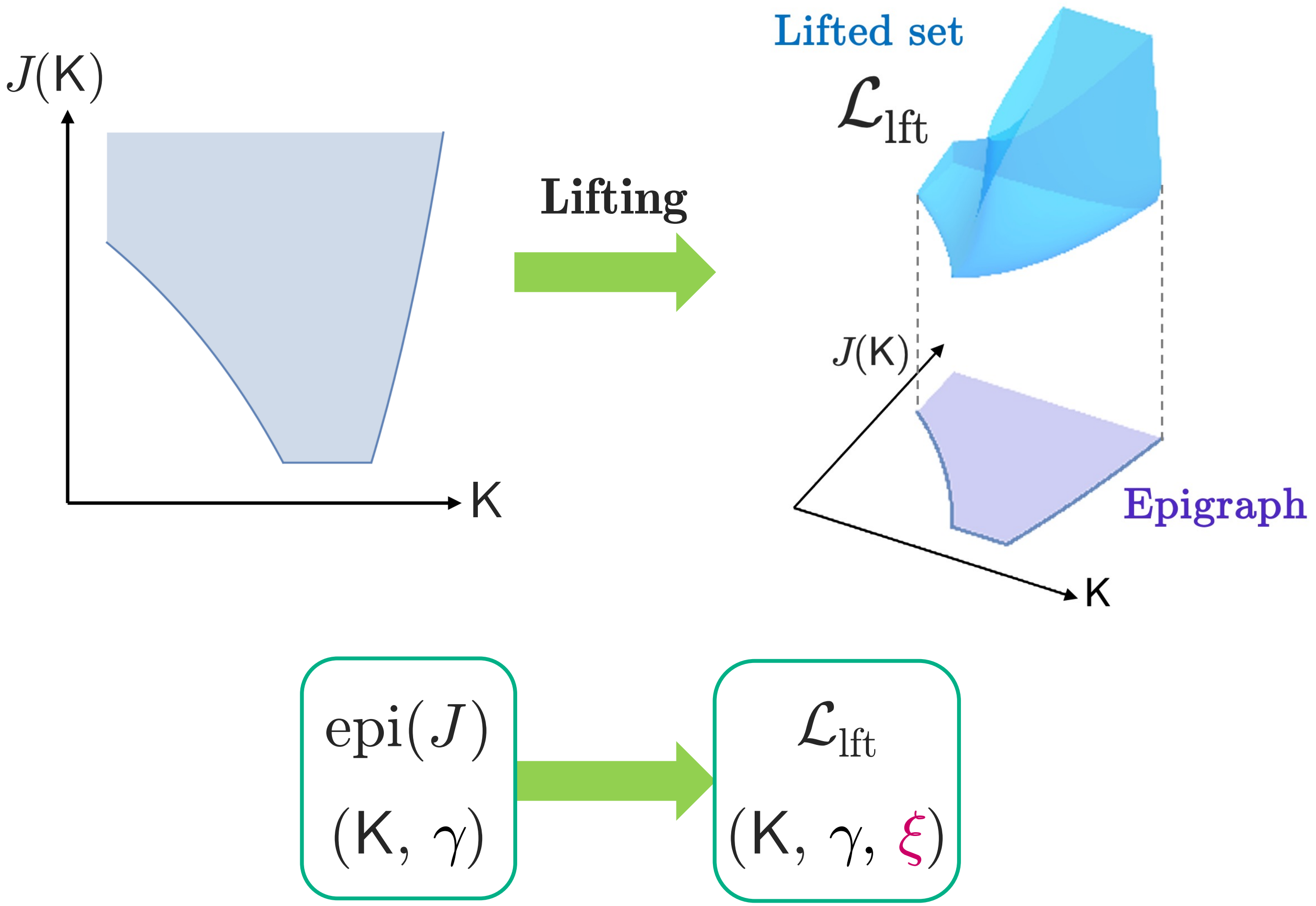
Extended Convex Lifting (ECL)

A schematic illustration of ECL:



Extended Convex Lifting (ECL)

A schematic illustration of ECL:



Why lifting?

- For many control problems, a **direct convexification is not possible**
- **A lifting procedure** corresponding to **Lyapunov variables** is necessary.

Extended Convex Lifting (ECL)

A schematic illustration of ECL:

Lifted set $\mathcal{L}_{\mathrm{lft}}$

Diffeomorphism Φ $\mathcal{F}_{\mathrm{cvx}} \times \mathcal{G}_{\mathrm{aux}}$

convex auxiliary

Why auxiliary set?

- Loosely speaking, it is related to **similarity transformations** of dynamic policies
- Needed for **output-feedback** problems

$\mathcal{L}_{\mathrm{lft}}$ $(\mathsf{K},\ \gamma,\ \xi)$ ⟷ $\mathcal{F}_{\mathrm{cvx}} \times \mathcal{G}_{\mathrm{aux}}$ $(\gamma,\ \zeta_1,\ \zeta_2)$

$$\inf_{\mathsf{K}} \ J(\mathsf{K}) \quad \text{s.t.} \ \mathsf{K} \in \mathcal{C}$$

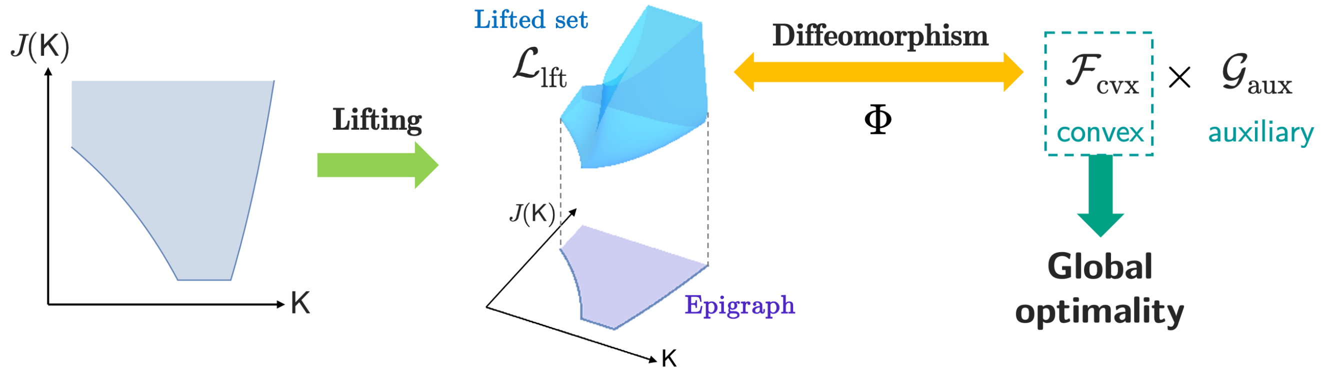
nonconvex

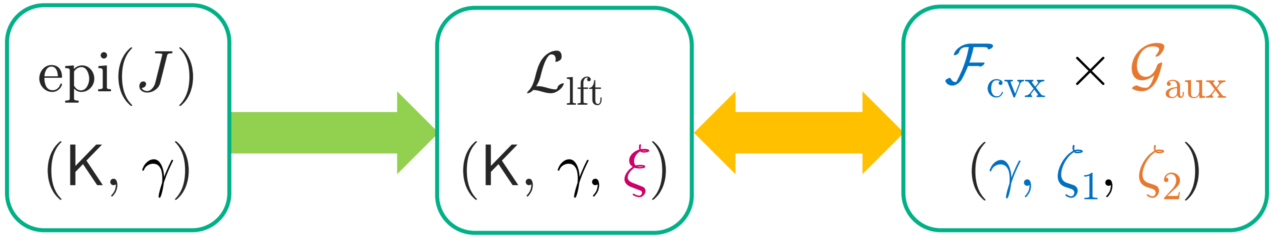
equivalent

$$\inf_{\gamma, \zeta_1} \ \gamma \quad \text{s.t.} \ (\gamma, \zeta_1) \in \mathcal{F}_{\mathrm{cvx}}$$

convex

Extended Convex Lifting (ECL)





ECL (**prototype**):

- A lifted set $\mathcal{L}_{\text{lft}}$ of the epigraph:

$$\text{epi}(J) = \pi_{\mathsf{K},\gamma}(\mathcal{L}_{\text{lft}})$$

- A diffeomorphism $\Phi : \mathcal{L}_{\text{lft}} \to \mathcal{F}_{\text{cvx}} \times \mathcal{G}_{\text{aux}}$ such that

$$\Phi(\mathsf{K}, \gamma, \xi) = (\gamma, \zeta_1, \zeta_2)$$

Existence of this ECL prototype guarantees

all Clarke stationary points are globally optimal

❖ Clarke stationary points: Generalization of stationary points to **nonsmooth functions**, based on the notion of **Clarke subdifferential**

Extended Convex Lifting (ECL)

- What could this prototype go wrong?
  - Existing convexifications of LQG and $\mathcal{H}_\infty$ output-feedback control are based on **strict LMIs**:
    - LQG : $\begin{bmatrix} A^\mathsf{T}P + PA & PB \\ B^\mathsf{T}P & -\gamma I \end{bmatrix} \prec 0, \quad \begin{bmatrix} P & C^\mathsf{T} \\ C & \Gamma \end{bmatrix} \succ 0, \quad \operatorname{trace}(\Gamma) < \gamma$
    - $\mathcal{H}_\infty$: $\begin{bmatrix} A^\mathsf{T}P + PA & PB & C^\mathsf{T} \\ B^\mathsf{T}P & -\gamma I & D^\mathsf{T} \\ C & D & -\gamma I \end{bmatrix} \prec 0$ (bounded real lemma)

    Used to construct the **lifted set $\mathcal{L}_{\mathrm{lft}}$**
  - **Strict LMIs** only characterize the **strict epigraph**

$$\mathrm{epi}_>(J) := \{(\mathsf{K}, \gamma) \mid \gamma > J(\mathsf{K})\}$$

  - Difficult to analyze Clarke stationary points only via **strict epigraphs**

Extended Convex Lifting (ECL)

- What if we turn to **non-strict versions of LMIs** to construct $\mathcal{L}_{\mathrm{lft}}$?
  - Many points in the **non-strict epigraph** (i.e., in the graph of J) will be covered

$$\mathrm{epi}_{\geq}(J) := \{(\mathsf{K}, \gamma) \mid \gamma \geq J(\mathsf{K})\}$$

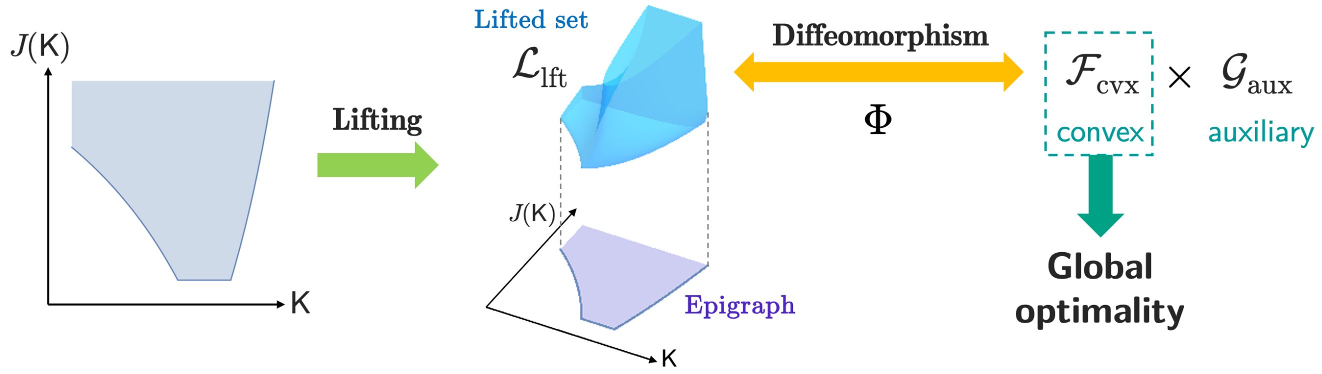
  - But some points in $\mathrm{epi}_{\geq}(J)$ will still **not be covered**

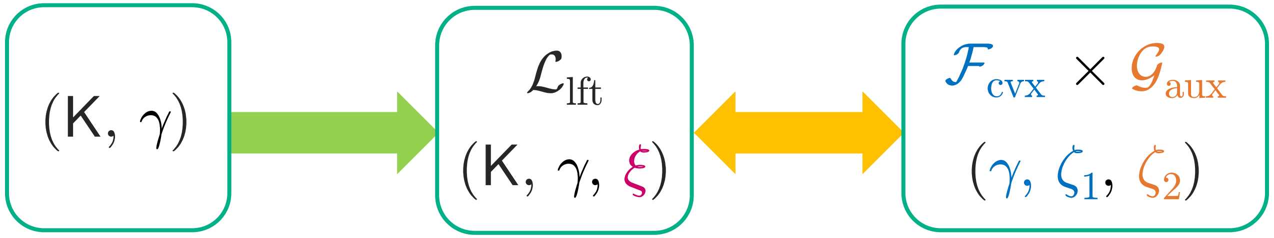
    ➡ Those points will be called **degenerate**

  - Some points **outside** $\mathrm{epi}_{\geq}(J)$ will **be covered**

    ➡ We modify the lifting procedure in the prototype

Extended Convex Lifting (ECL)





Extended Convex Lifting:

- A lifted set $\mathcal{L}_{\text{lft}}$ satisfying

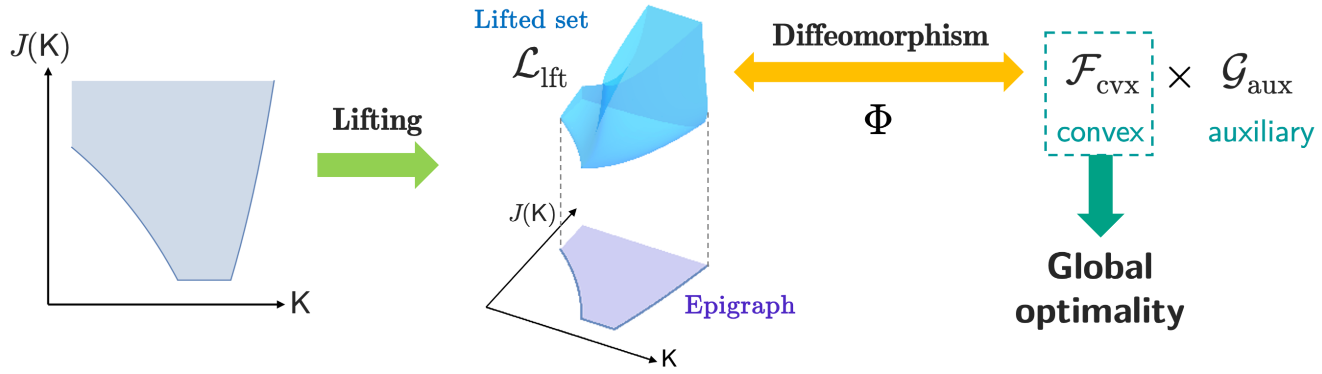
$$\text{epi}_{>}(J) \subseteq \pi_{\mathsf{K},\gamma}(\mathcal{L}_{\text{lft}}) \subseteq \text{cl}\,\text{epi}_{\geq}(J)$$

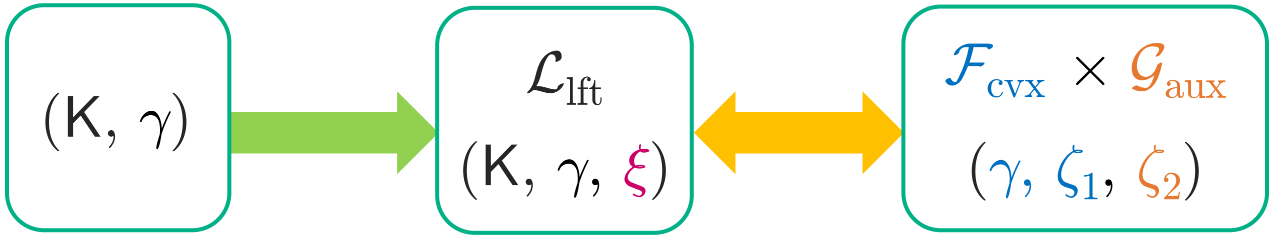
- A diffeomorphism $\Phi : \mathcal{L}_{\text{lft}} \to \mathcal{F}_{\text{cvx}} \times \mathcal{G}_{\text{aux}}$ such that

$$\Phi(\mathsf{K}, \gamma, \xi) = (\gamma, \zeta_1, \zeta_2)$$

Definition. K is called **non-degenerate** if $(\mathsf{K}, J(\mathsf{K})) \in \pi_{\mathsf{K},\gamma}(\mathcal{L}_{\text{lft}})$

Extended Convex Lifting (ECL)





Extended Convex Lifting:

- A lifted set $\mathcal{L}_{\rm lft}$ satisfying

$$\mathrm{epi}_{>}(J) \subseteq \pi_{\mathsf{K},\gamma}(\mathcal{L}_{\rm lft}) \subseteq \mathrm{cl}\,\mathrm{epi}_{\geq}(J)$$

- A diffeomorphism $\Phi : \mathcal{L}_{\rm lft} \to \mathcal{F}_{\rm cvx} \times \mathcal{G}_{\rm aux}$ such that

$$\Phi(\mathsf{K}, \gamma, \xi) = (\gamma, \zeta_1, \zeta_2)$$

Main Result

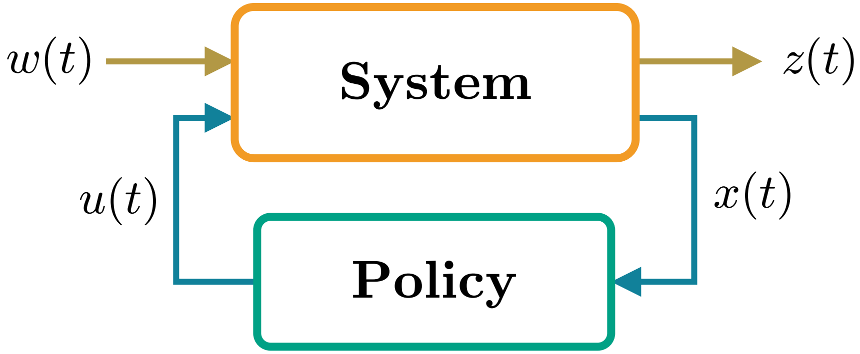
Given an ECL, under mild conditions, **all non-degenerate Clarke stationary points are globally optimal**.

Outline

- Problem Setup and Motivating Examples
- Extended Convex Lifting (ECL)
- **Applications for Optimal and Robust Control**
- Conclusions

Linear Quadratic Regulator

❑ Problem setup



Dynamics: $\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t)$

Performance: $J = \|\mathbf{T}_{zw}\|_{\mathcal{H}_2}$

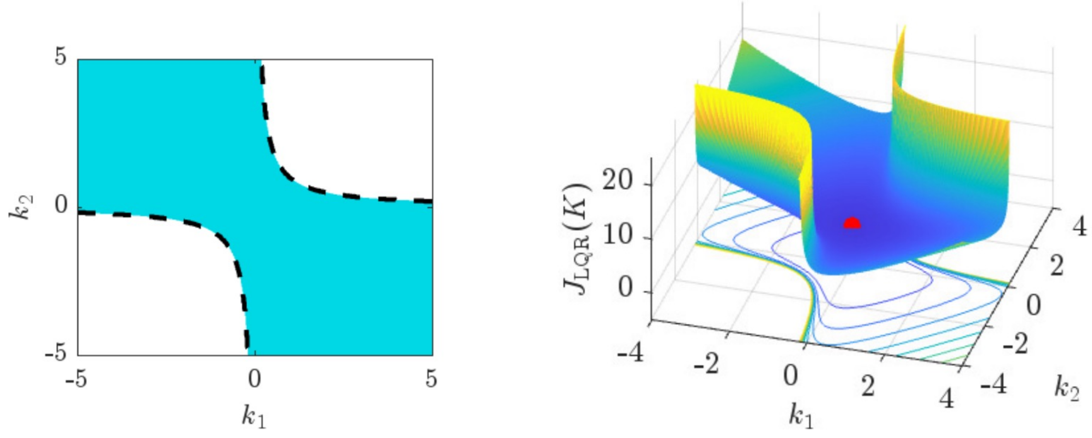
$$z(t) = \begin{bmatrix} Q^{1/2}x(t) \\ R^{1/2}u(t) \end{bmatrix}$$

Policy: $u(t) = Kx(t)$

Objective function:

$$J(K) = \mathrm{tr}\big[(Q + K^{\mathsf{T}}RK)X\big]$$

where $X = \mathrm{Lyap}(A + BK, W)$



nonconvex & smooth

Linear Quadratic Regulator

❑ Construction of ECL

Step 1: Lifting

$$\mathcal{L}_{\texttt{LQR}} = \left\{(K,\gamma,X) \mid X \succ 0, X = \mathrm{Lyap}(A+BK,W), \gamma \geq \mathrm{tr}\left[(Q+K^{\mathsf{T}}RK)X\right]\right\}$$

Step 2: Convex set

$$\mathcal{F}_{\texttt{LQR}} = \left\{(\gamma,Y,X) : X \succ 0, AX+BY+XA^{\mathsf{T}}+Y^{\mathsf{T}}B^{\mathsf{T}}+W=0, \gamma \geq \mathrm{tr}\left(QX+X^{-1}Y^{\mathsf{T}}RY\right)\right\}$$

Step 3: Diffeomorphism $\Phi(K,\gamma,X) = (\gamma,KX,X), \quad \forall (K,\gamma,X) \in \mathcal{L}_{\texttt{LQR}}$

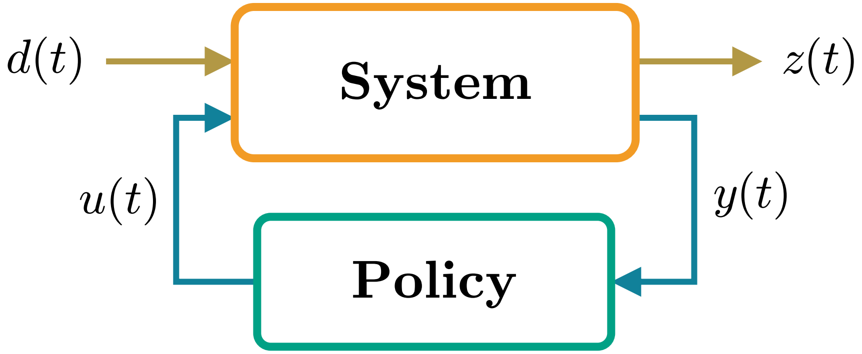
- No auxiliary set
- Lifted set satisfies $\mathrm{epi}_{\geq}(J) = \pi_{K,\gamma}(\mathcal{L}_{\texttt{LQR}})$

➡ **All policies are non-degenerate**

Theorem. Any stationary point of the LQR cost function is globally optimal.

Linear Quadratic Gaussian

❑ Problem setup



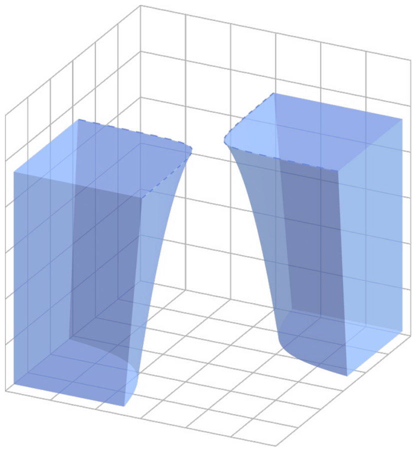
Policy:
$$\frac{d\xi(t)}{dt} = A_{\mathsf{K}}\xi(t) + B_{\mathsf{K}}y(t)$$
$$u(t) = C_{\mathsf{K}}\xi(t)$$

$$\mathsf{K} = (A_{\mathsf{K}},\ B_{\mathsf{K}},\ C_{\mathsf{K}})$$

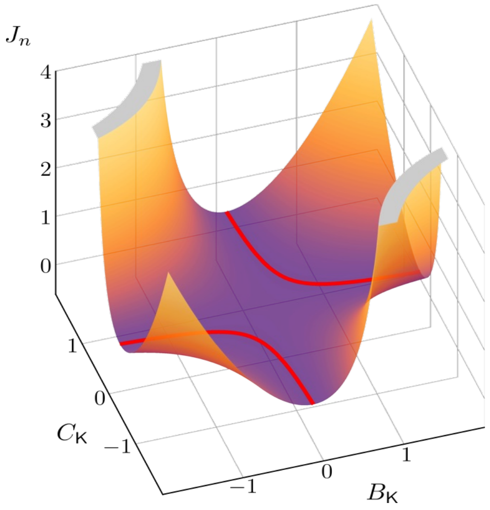
Dynamics:
$$\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t)$$
$$y(t) = Cx(t) + D_v v(t)$$

Performance: $J = \|\mathbf{T}_{zd}\|_{\mathcal{H}_2}$

$$z(t) = \begin{bmatrix} Q^{1/2}x(t) \\ R^{1/2}u(t) \end{bmatrix} \quad d(t) = \begin{bmatrix} w(t) \\ v(t) \end{bmatrix}$$



disconnected domain



multiple globally optimal points

Linear Quadratic Gaussian

- Construction of the ECL: Based on the convexification proposed in [Scherer et al., 1997]

Theorem.

1. An ECL for LQG exists, of which $\mathcal{G}_{\mathrm{aux}}$ is the set of invertible matrices.
2. A policy K is non-degenerate if and only if it is **informative** in the sense that

$$\lim_{t\to\infty} \mathbb{E}\left[x(t)\xi(t)^{\mathsf{T}}\right]$$

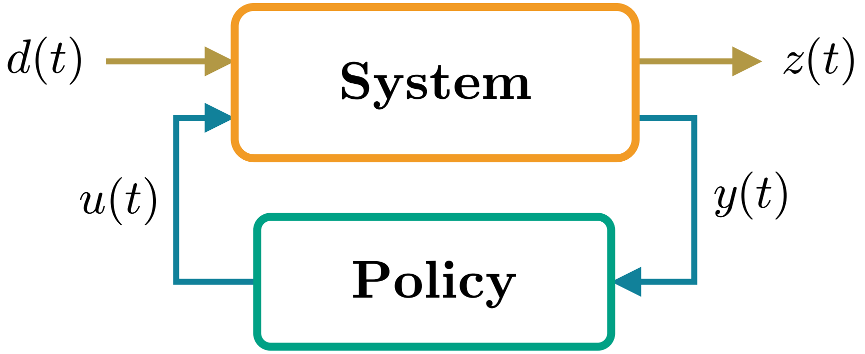
has full rank. So **any informative stationary point is globally optimal**.

3. Non-degenerate policies are **generic** in the sense that degenerate policies form a **set of measure zero**.

- Part 2 extends [Umenberger et al., 2022, Theorem 1(ii)] from Kalman filtering to LQG.
- We also show that minimal stationary policies are non-degenerate, generalizing our exisiting results in [Tang, Zheng, Li, 2023].

$\mathcal{H}_\infty$ Output-Feedback Control

❑ Problem setup



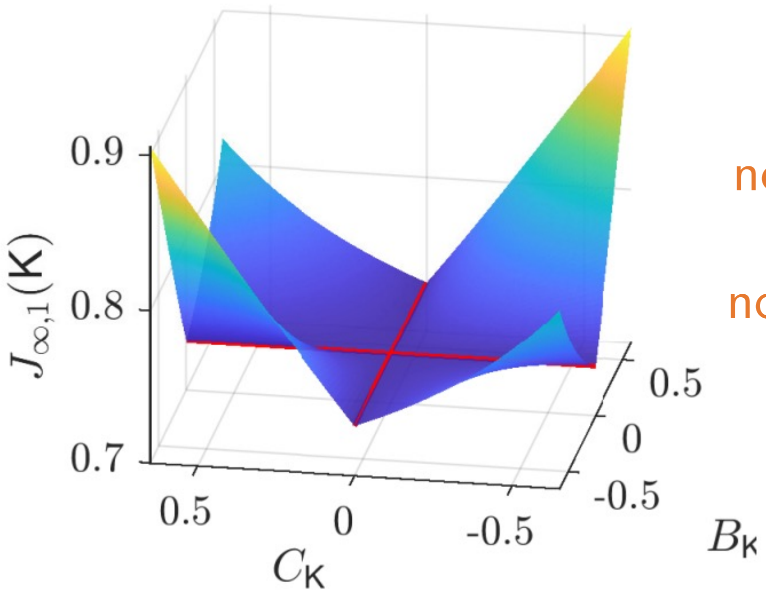
Policy: $$\frac{d\xi(t)}{dt} = A_\mathsf{K}\xi(t) + B_\mathsf{K}y(t)$$
$$u(t) = C_\mathsf{K}\xi(t) + D_\mathsf{K}y(t)$$

$$\mathsf{K} = (A_\mathsf{K}, B_\mathsf{K}, C_\mathsf{K}, D_\mathsf{K})$$

Dynamics: $$\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t)$$
$$y(t) = Cx(t) + D_v v(t)$$

Performance: $$J = \|\mathbf{T}_{zd}\|_{\mathcal{H}_\infty}$$

$$z(t) = \begin{bmatrix} Q^{1/2}x(t) \\ R^{1/2}u(t) \end{bmatrix} \quad d(t) = \begin{bmatrix} w(t) \\ v(t) \end{bmatrix}$$



nonconvex
+
nonsmooth

$\mathcal{H}_\infty$ Output-Feedback Control

❑ Construction of the ECL: Based on the convexification proposed in [Scherer et al., 1997]

Theorem. 1. An ECL for $\mathcal{H}_\infty$ output-feedback control exists.

2. A policy K is non-degenerate if and only if

a) There exists a non-strict certificate $P \succ 0$ of the $\mathcal{H}_\infty$ cost.

$$\begin{bmatrix} A_{\mathrm{cl}}^{\mathsf{T}}(\mathsf{K})P + PA_{\mathrm{cl}}(\mathsf{K}) & PB_{\mathrm{cl}}(\mathsf{K}) & C_{\mathrm{cl}}^{\mathsf{T}}(\mathsf{K}) \\ B_{\mathrm{cl}}^{\mathsf{T}}(\mathsf{K})P & -J(\mathsf{K})I & D_{\mathrm{cl}}^{\mathsf{T}}(\mathsf{K}) \\ C_{\mathrm{cl}}(\mathsf{K}) & D_{\mathrm{cl}}(\mathsf{K}) & -J(\mathsf{K})I \end{bmatrix} \preceq 0$$

b) The block P_{12} is invertible.

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^{\mathsf{T}} & P_{22} \end{bmatrix}$$

So **a Clarke stationary point is globally optimal if these conditions hold**.

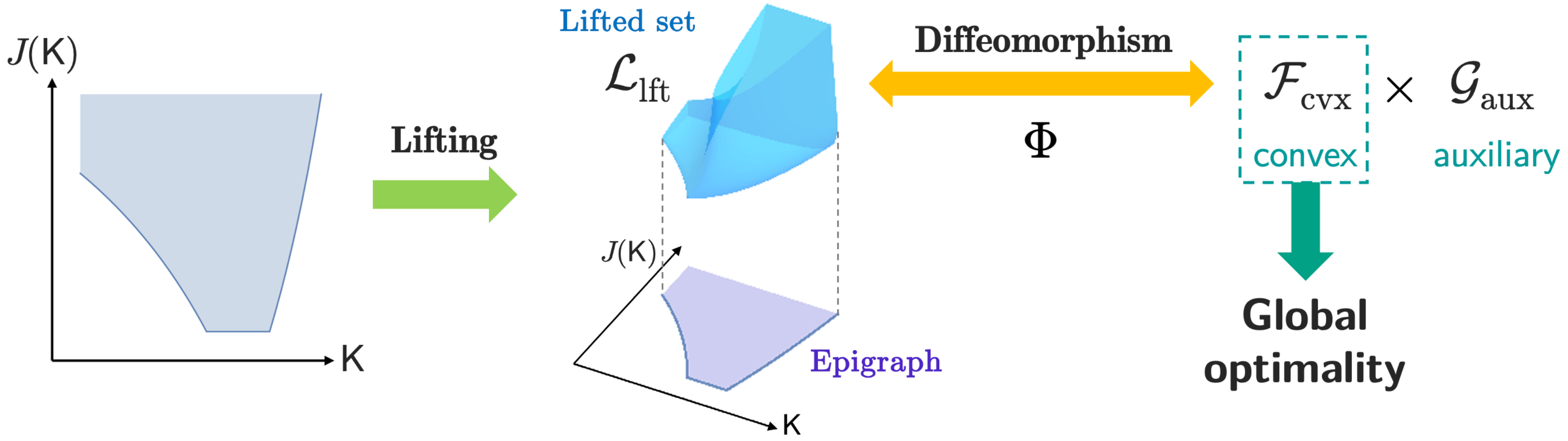
- Physical interpretation of non-degeneracy is not as clear as LQG.
- We conjecture that non-degenerate policies for $\mathcal{H}_\infty$ output-feedback control are also **generic**, with some numerical evidence, but a proof is not known yet.

Outline

- Problem Setup and Motivating Examples
- Extended Convex Lifting (ECL)
- Applications for Optimal and Robust Control
- **Conclusions**

Nonconvex Policy Optimization for Control

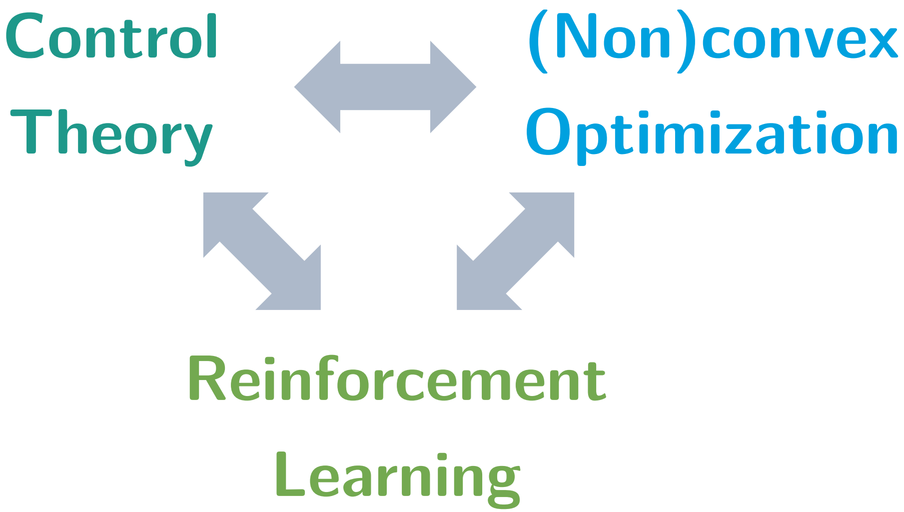
- Policy optimization in control can be **nonconvex** and **non-smooth**.
- **Extended Convex Lifting (ECL)** reveals benign nonconvexity.

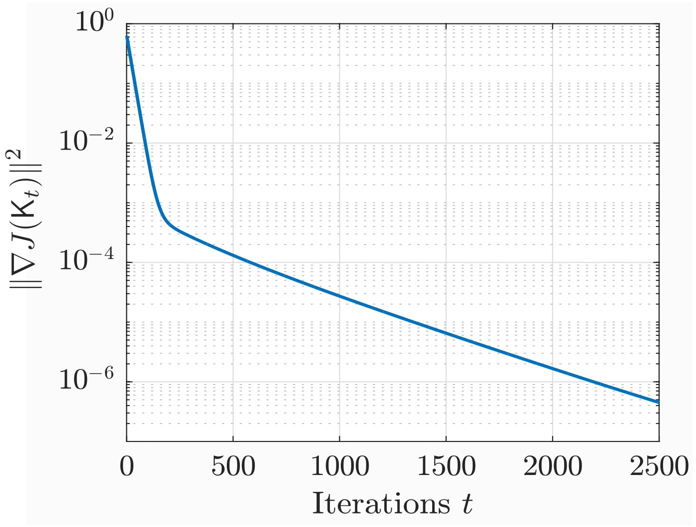


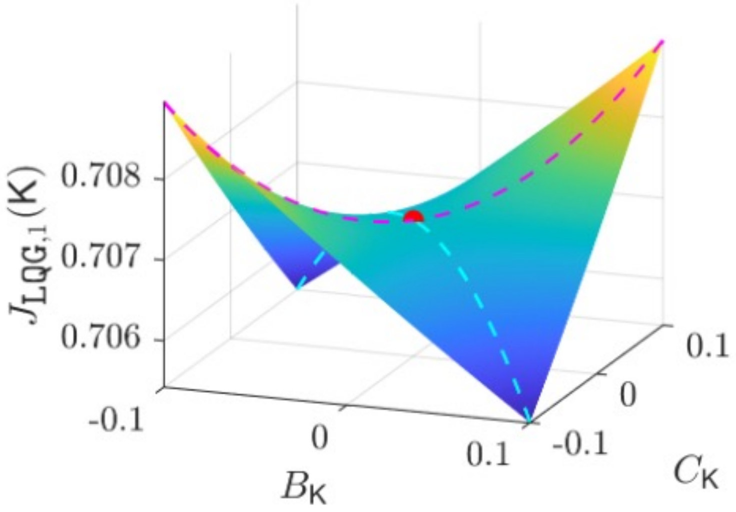
- The notion of **non-degeneracy** provides a **global optimality certificate** for Clarke stationary points.

Ongoing & Future Work

- How to incorporate finer analytical properties (e.g., weak PL inequality) in ECL?
- How to justify non-degeneracy only using data?
- How to deal with degenerate points in local policy search? Avoiding saddle points?

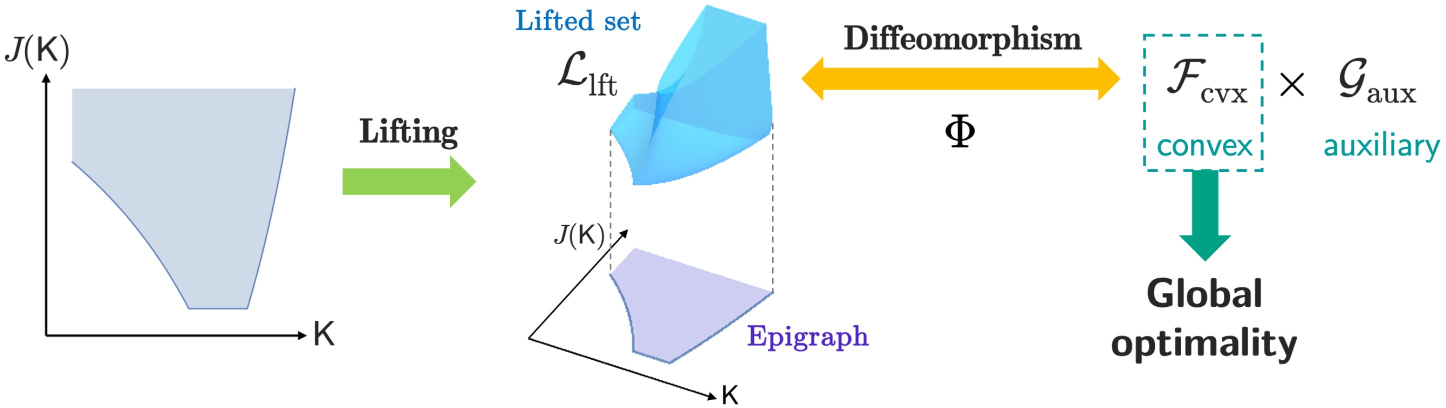






Thank you for your attention!

- Policy optimization in control can be **nonconvex** and **non-smooth.**
- **Extended Convex Lifting (ECL)** reveals benign nonconvexity.
- The notion of **non-degeneracy** provides a **global optimality certificate** for stationary points.



- Y. Tang, Y. Zheng. "On the Global Optimality of Direct Policy Search for Nonsmooth $\mathcal{H}_\infty$ Output-Feedback Control." In Proceedings of the 62nd IEEE Conference on Decision and Control (CDC), pp. 6148-6153, 2023.
- Y. Zheng, C.-F. Pai, Y. Tang. "Benign Nonconvex Landscapes in Optimal and Robust Control, Part I: Global Optimality." **arXiv:2312.15332**
- Y. Zheng, C.-F. Pai, Y. Tang. "Benign Nonconvex Landscapes in Optimal and Robust Control, Part II: Extended Convex Lifting." **arXiv:2406.04001**