Decision-making in Uncertain Dynamic Environments:

from Policy Optimization to Online Learning

Rich Pai

Advisor: Prof. Yang Zheng

Qualifying Exam Talk

Nov. 25, 2025

Optimal control problem

Control as optimization over time subject to dynamics

state → x_t; action/input → u_t; disturbance → w_t

$$\min_{\pi} \quad \sum_{t=1}^{T} \mathrm{cost}_t(x_t, u_t)$$

$$\text{s.t.} \quad x_{t+1} = \mathrm{dynamics}_t(x_t, u_t, w_t)$$

A basic formulation with *linear dynamics and quadratic costs*: **LQR**

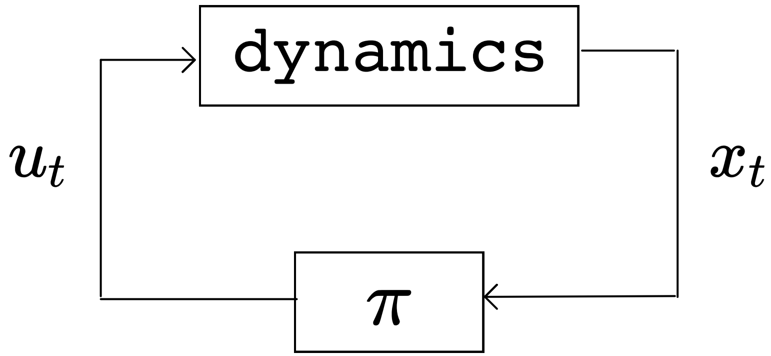
- Linear dynamics: $x_{t+1} = Ax_t + Bu_t + w_t$
- Quadratic cost: $x_t^\top Q x_t + u_t^\top R u_t$

Goal: find a policy to drive the state to the origin with small control effort

Control as online decision-making under uncertainty

$$\min_{\pi} \quad \sum_{t=1}^{T} \text{cost}_t(x_t, u_t)$$

$$\text{s.t.} \quad x_{t+1} = \text{dynamics}_t(x_t, u_t, w_t)$$



Assume **observation of x_t** and hence possibly $w_{1:t-1}$

$$u_t = \pi_t(x_{0:t-1}, u_{1:t-1}, w_{1:t-1})$$

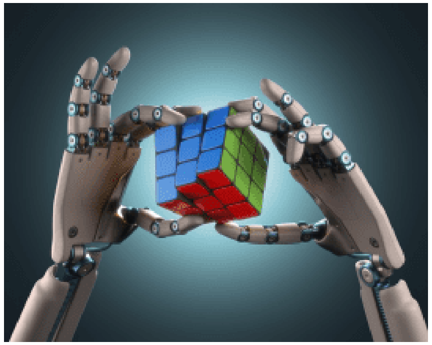
At each time step, the agent

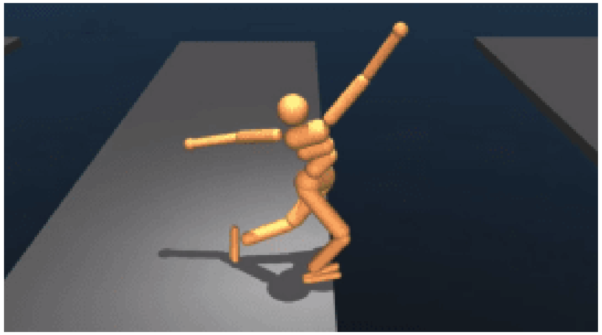
1. Picks u_t based on all available (past & current) information
2. Suffers stage cost, and x_t evolves according to dynamics

Sources of **Uncertainty**

- dynamics A, B, disturbance w_t
- or even cost: Q, R

Lots of successful applications





Refs: Silver et al., Nature 2017; Akkaya et al., 2019; Schulman et al., 2017; Bojarski et al., 2016; etc.

Talk outline

(1) Offline Planning $\to$ (2) Policy Optimization $\to$ (3) Online Learning

Part I.

Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: benign nonconvexity

(1) -> (2)

Part II.

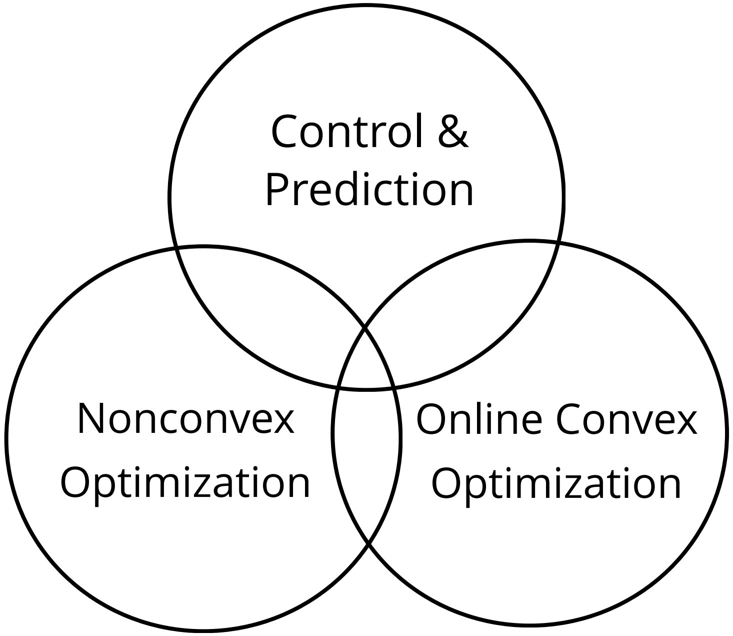
Online tracking with predictions: dynamic regret analysis of MPC

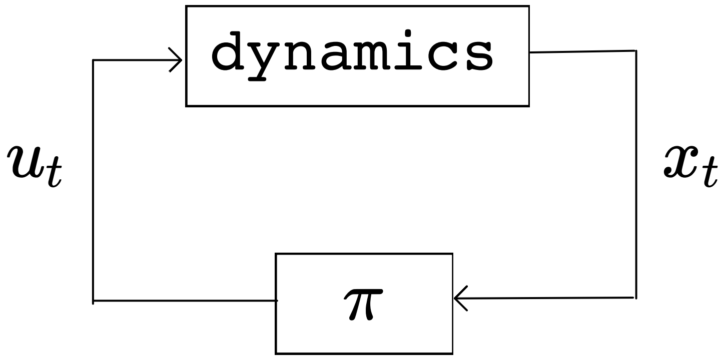
(1) -> (3)

Part III.

Online adaptive control & prediction under nonstationarity

Generalize (1) & (2)





Classical view: offline synthesis (planning)

Consider linear dynamics $x_{t+1} = Ax_t + Bu_t + w_t$

Two main frameworks under different assumptions on w_t

LQR/H2 optimal control	Hinfty robust control
• iid Gaussian noise	• worst-case bounded noise
$\min_{\pi} \mathbb{E}\left[\sum_{t=1}^{T} x_t^T Q x_t + u_t^T R u_t\right]$	$\min_{\pi} \max_{\|w_t\| \leq c} \sum_{t=1}^{T} x_t^T Q x_t + u_t^T R u_t$
• Overly optimistic	• Overly pessimistic

- Globally optimal policy [B1217, ZDG96, BB08]: linear state feedback policy

$$u_t = K_t^\star x_t$$

- Optimal gains $K_t^\star$ defined recursively by A, B, Q, R using dynamic programming

[B1217] D. Bertsekas. *Dynamic Programming and Optimal Control.* 4th edition, volumes 1 (2017) and 2 (2012). Athena Scientific

[ZDG96] K. Zhou, J. Doyle, and K. Glover. *Robust and Optimal Control.* Prentice Hall, 1996

[BB08] T. Başar, and P. Bernhard. *H-infinity optimal control and related minimax design problems: a dynamic game approach.* Springer Science & Business Media, 2008

Modern view: nonconvex policy optimization

Same disturbance models as in classical planning, so we know the form of the optimal policy

- **View control cost directly as a function of the policy parameter**
- For example, with $u_t = Kx_t$, the LQR csot $J(K) := \mathbb{E}\left[\sum_{t=1}^{T} x_t^T Q x_t + u_t^T R u_t\right]$

Benefits:
1. Model-free implementation
2. Scalable to large-scale systems

Research:
1. Structural aspect: landscape analysis
2. Algorithmic aspect: local policy search

Policy optimization of mixed H2/Hinfty control:
Benign nonconvexity and global optimality

Extended Convex Lifting (ECL):
Bridge policy optimization and classical approaches (LMI, Riccati)

M. Fazel, et al. *Global convergence of policy gradient methods for the linear quadratic regulator*. ICML, 2018.
B. Hu, et al. *Toward a theoretical foundation of policy optimization for learning control policies*. Annual Review of Control, Robotics, and Autonomous Systems, 2023

Modern view: online nonstochastic control

Instead of **planning** or **optimizing** under some specific disturbance model, we want an online method with

- **Adaptivity & instance-optimality wrt the realized disturbance**:
  - adapts efficiently to the actual nonstochastic disturbance
- **Efficient methods for general adversarial convex costs:**
  - extends beyond a given (known) quadratic cost in classical setting

Dynamic regret analysis of model predictive control (**MPC**) in online tracking

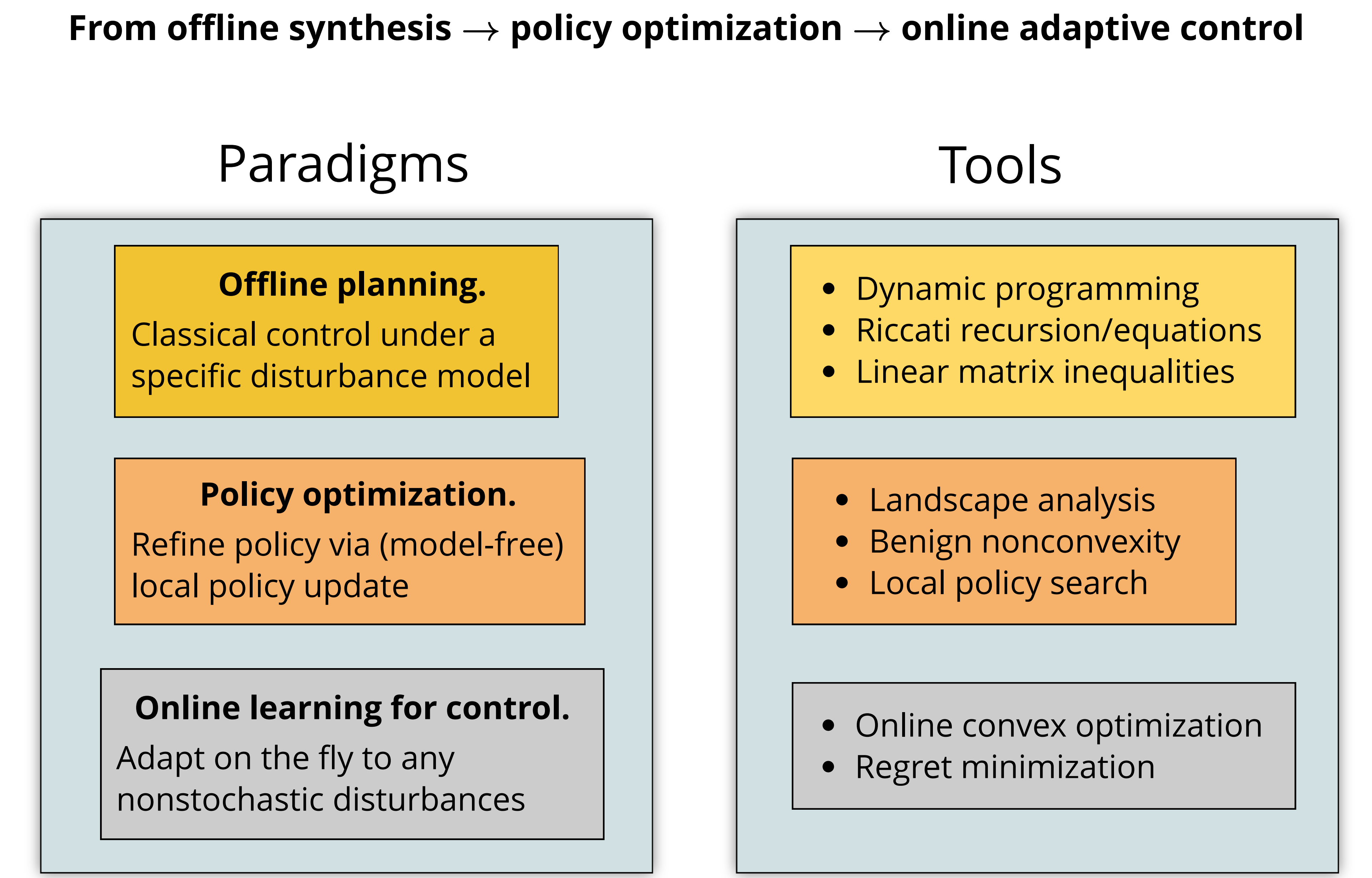
for Koopman-linearizable nonlinear systems

Online learning for control and prediction of LDS:

Adaptive regret minimization in nonstationary environments

N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh. *Online control with adversarial disturbances*. ICML, 2019

E. Hazan, and K. Singh. *Introduction to online nonstochastic control*. arXiv preprint, 2022


From offline synthesis → policy optimization → online adaptive control
Paradigms
Tools
Offline planning.
Classical control under a specific disturbance model
Policy optimization.
Refine policy via (model-free) local policy update
Online learning for control.
Adapt on the fly to any nonstochastic disturbances
Dynamic programming
Riccati recursion/equations
Linear matrix inequalities
Landscape analysis
Benign nonconvexity
Local policy search
Online convex optimization
Regret minimization

Preview & main contributions

Part I: Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control [PWTZ,ZPT25]

1. A new structural characterization of the nonconvex landscape
2. Reveal hidden convexity: every stationary point is globally optimal

Part II: Online tracking with predictions for Koopman-linearizable systems [PSQZ]

1. First dynamic regret analysis of MPC for nonlinear dynamics with a lifted linear model
2. Achieve constant regret with a logarithmically sufficiently large prediction horizon

Part III: Online adaptive control & prediction in nonstationary environments

1. Goal: problem-dependent regret guarantees for online nonstochastic control
2. Goal: online adaptive prediction for time-varying linear dynamical systems

[PWTZ] **C. Pai**, Y. Watanabe, Y. Tang, and Y. Zheng. *Policy Optimization of Mixed H2/Hinfty Control: Benign Nonconvexity and Global Optimality*. Submitted to Automatica
[ZPT25] Y. Zheng, **C. Pai**, and Y. Tang. *Extended Convex Lifting for Policy Optimization of Optimal and Robust Control.* Learning for Dynamics and Control (L4DC) 2025
[PSQZ] **C. Pai**, X. Shang, J. Qian, and Y. Zheng. *Online Tracking with Predictions for Nonlinear Systems with Koopman Linear Embedding.* Submitted to L4DC

Talk outline

Part I.

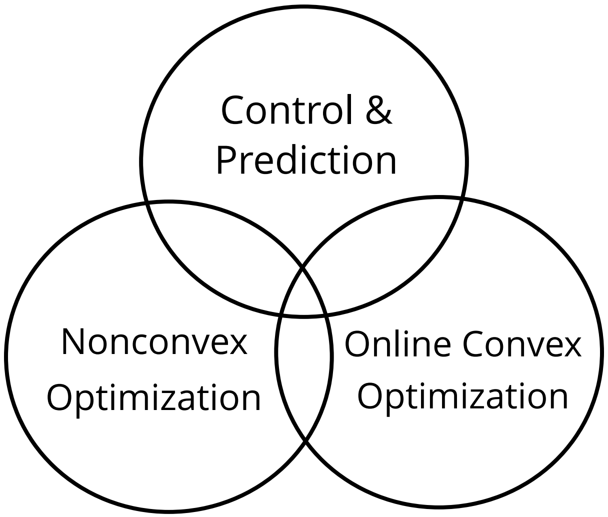
Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: benign nonconvexity

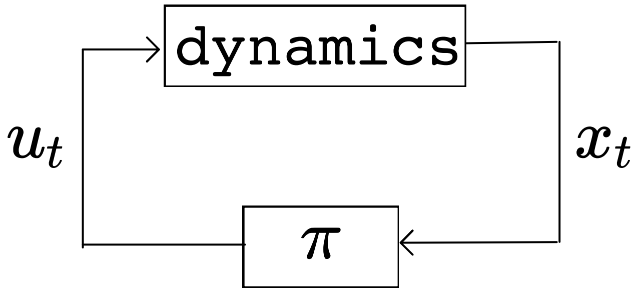
Part II.

Online tracking with predictions: dynamic regret analysis of MPC

Part III.

Online adaptive control & prediction under nonstationarity





Policy optimization of mixed H2/Hinfty control

- A fundamental formulation to balance performance and robustness
- A new theoretical characterization from a **nonconvex** optimization perspective
- Why popular? e.g., success of RL, scalability, model-free policy search

Problem Setup

1. Continuous-time dynamics: $\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t)$
2. Policy/controller parameterization: $u(t) = Kx(t)$
3. Performance signals: $z_2(t) = \begin{bmatrix} Q_2^{1/2} x(t) \\ R_2^{1/2} u(t) \end{bmatrix}$, $z_\infty(t) = \begin{bmatrix} Q_\infty^{1/2} x(t) \\ R_\infty^{1/2} u(t) \end{bmatrix}$

$$\min_K \quad \overbrace{\|T_{z_2 w}(K)\|_{\mathcal{H}_2}^2}^{\text{performance}} \leq J_{\text{mix}}(K) := \text{trace}(Q_2 + K^T R_2 K) X_K$$

$$\text{s.t.} \quad K \in \mathcal{K}_\beta := \{K : A + BK \text{ stable}, \overbrace{\|T_{z_\infty w}(K)\|_{\mathcal{H}_\infty} < \beta}^{\text{robustness}}\}$$

X_K certifys the $\mathcal{H}_\infty$ constraint, is the stabilizing solution to

$$(A + BK)X_K + X_K(A + BK)^T + \beta^{-2} X_K (Q_\infty + K^T R_\infty K) X_K + W = 0$$

Our Contributions

Mixed H2/Hinfty policy optimization

$$\min_K \quad J_{\mathrm{mix}}(K) := \operatorname{trace}(Q_2 + K^T R_2 K) X_K$$
$$\text{s.t.} \quad K \in \mathcal{K}_\beta := \{K : A + BK \text{ stable}, \|T_{z_\infty w}(K)\|_{\mathcal{H}_\infty} < \beta\}$$

- **Global optimality v.s.** Riccati eqs [BH89] or LMI **suboptimality** [KR91]
- **General two-channel v.s. single-channel** formulation [ZHB21]
- Analysis using **convex lifting** [ZPT25] **v.s. dynamic game** [ZHB21]

↓

Bridges **policy optimization** and convex **LMI** via non-strict **Riccati inequalities**

More specifically,

1. Analyze the **feasible set** $\mathcal{K}_\beta$ and precisely characterize its **boundary**
2. Identify key structural properties of the **cost function** $J_{\mathrm{mix}}(\cdot)$
3. Establish **benign nonconvexity**: every stationary point is globally optimal

[BH89] D. Bernstein and W. Haddad. *LQG control with an $\mathcal{H}_\infty$ performance bound: a Riccati equation approach.* IEEE Transactions on Automatic Control, 1989

[KR91] P. Khargonekar and M. Rotea. *Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: a convex optimization approach.* IEEE Transactions on Automatic Control, 1991

[ZHB21] K. Zhang, B. Hu, and T. Başar. *Policy optimization for $\mathcal{H}_2$ linear control with $\mathcal{H}_\infty$ robustness guarantee: Implicit regularization and global convergence*. SIAM, 2021

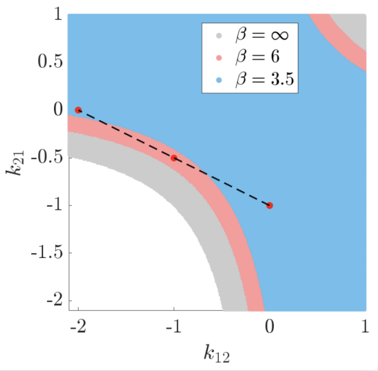
[ZPT25] Y. Zheng, C. Pai, and Y. Tang. *Extended Convex Lifting for Policy Optimization of Optimal and Robust Control.* Learning for Dynamics and Control, 2025

Feasible set and its boundary

The Hinfty-constrained domain

$$\mathcal{K}_\beta = \{K : A + BK \text{ stable}, \|T_{z_\infty w}(K)\|_{\mathcal{H}_\infty} < \beta\}$$

Open, path-connected, may be **nonconvex**, unbounded



Closure $\mathrm{cl}(\mathcal{K}_\beta) = \{K : A + BK \text{ stable}, \|T_{z_\infty w}(K)\|_{\mathcal{H}_\infty} \leq \beta\}$

The proof relies on ***fundamental properties of the state feedback Hinfty cost***

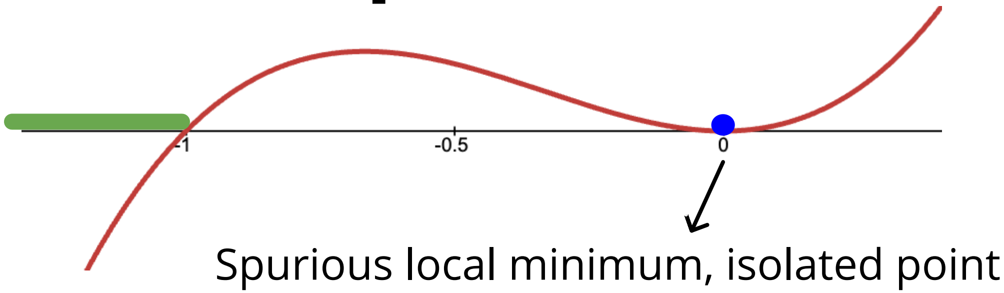
- Benign nonconvexity: no spurious stationary points (local minimum)
- Partial coercivity: cost diverges as the policy becomes marginally stabilizing

Why useful?

1. Define the extended cost over the entire closure
2. Provide more insight into the cost properties
3. Facilitate the proof of benign nonconvexity
4. Establish solvability for convex reformulations

Counter-examples

$$f(x) = x^2(x+1)$$



$$\mathcal{C}_1 = \{x : f(x) < 0\} = \{x : x < -1\}$$

$$\mathcal{C}_2 = \{x : f(x) \leq 0\} = \{x : x \leq -1, x = 0\}$$

$$\mathrm{cl}(\mathcal{C}_1) = \{x : x \leq -1\} \neq \mathcal{C}_2$$

Another example

$$\mathcal{K}_\beta = \{K : A + BK \text{ stable}, \|T_{z_\infty w}(K)\|_{\mathcal{H}_\infty} < \beta\}$$

Fortunately, this is not the case for $J_\infty(K) := \|T_{z_\infty w}(K)\|_{\mathcal{H}_\infty}$ for all stabilizing K

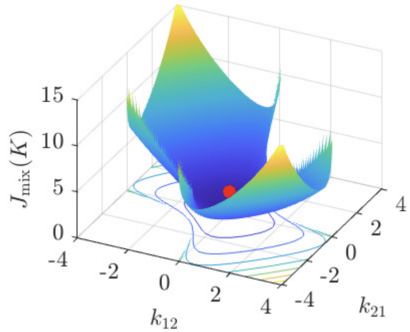
[ZPT25] Y. Zheng, **C. Pai**, and Y. Tang. *Extended Convex Lifting for Policy Optimization of Optimal and Robust Control.* Learning for Dynamics and Control (L4DC) 2025

Nonconvex landscape analysis

$$J_{\text{mix}}(K) := \text{trace}(Q_2 + K^T R_2 K) X_K$$

Minimal solution to

$$(A + BK)X_K + X_K(A + BK)^T + \beta^{-2} X_K (Q_\infty + K^T R_\infty K) X_K + W = 0$$



Properties of $J_{\text{mix}} : \text{cl}(\mathcal{K}_\beta) \to \mathbb{R}$

- Continuous on the closure
- Noncoercive, real analytic in the interior
- Explicit gradient formulas in the interior

Hidden Convexity: every stationary point is globally optimal [PWTZ]

$$\nabla J_{\text{mix}}(K) = 0 \iff K \in \arg\min_{K \in \mathcal{K}_\beta} J_{\text{mix}}(K)$$

- Recover optimality conditions (e.g., coupled Riccati equations)
- Facilitate the design of **policy iteration** algorithms
- Analysis based on **ECL** + **non-strict LMIs** and Riccati inequalities
- **Existence** and uniqueness of stationary points

[PWTZ] **C. Pai**, Y. Watanabe, Y. Tang, and Y. Zheng. "Policy Optimization of Mixed H2/Hinfty Control: Benign Nonconvexity and Global Optimality," submitted to Automatica

Policy iteration (fixed-point iteration)

A special case when $z_2 = z_\infty$: $\nabla J_{\text{mix}}(K) = 0 \Rightarrow K = -R^{-1}B^\top P_K$

1. Choose an initial policy $K_0 \in \mathcal{K}_\beta$ and let $i = 0$.
2. Policy evaluation: solve a Riccati equation to obtain P_i
3. Policy improvement: $K_{i+1} = -R^{-1}B^\top P_i$
4. Set $i \leftarrow i + 1$ and go back to Step 2.

$$\nabla J_{\text{mix}}(K) = 0 \Rightarrow K = -R_2^{-1}B^\top \Gamma_K (I + \beta^{-2}\alpha^2 X_K \Gamma_K)^{-1}$$

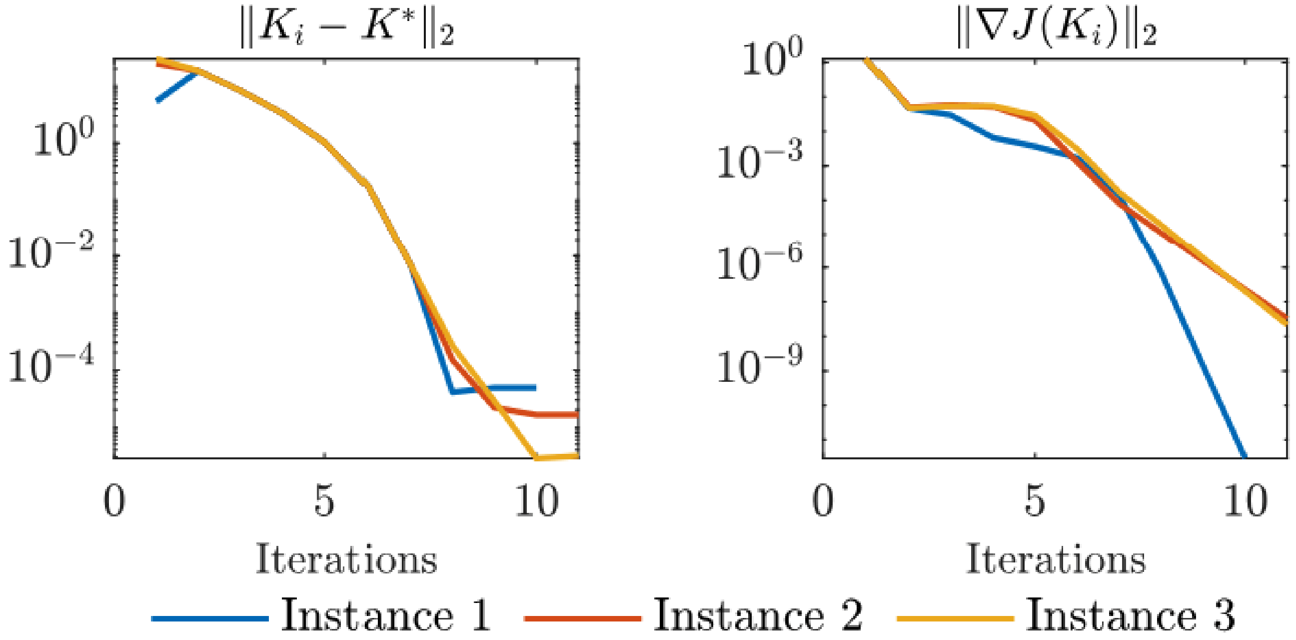
1. Choose an initial policy $K_0 \in \mathcal{K}_\beta$ and let $i = 0$.
2. Policy evaluation: solve a Riccati & Lyapunov equation to obtain X_i & Γ_i respectively
3. Policy improvement: $K_{i+1} = -R_2^{-1}B^\top \Gamma_i (I + \beta^{-2}\alpha^2 X_i \Gamma_i)^{-1}$
4. Set $i \leftarrow i + 1$ and go back to Step 2.

Can also be viewed as Gauss-Newton local update

Experiment: empirical convergence of policy iteration

The (two-channel) policy iteration works well for sufficiently large β

1. The iterate converges to a (globally optimal) stationary point.
2. The iterates always stay in the feasible set $\mathcal{K}_\beta$.



- We have shown that a stationary point exists when β is large enough
- The full convergence analysis is left for future work

Experiment: policy iteration is more scalable

- PI is much more efficient to solve higher-dimensional instances
- An order of magnitude improvement in runtime

		$K: 60 \times 60$				$K: 90 \times 90$	
		$I_1,\ \beta=10$		$I_2,\ \beta=15$		$I_3,\ \beta=20$	
		2-ch	1-ch	2-ch	1-ch	2-CH	1-CH
ARE	time	-	**0.02**	-	**0.03**	-	**0.05**
	$J_{\text{mix}}^{1/2}$	-	0.47	-	0.02	-	0.04
	$\mathcal{H}_2$	-	0.47	-	1.12	-	1.22
	$\mathcal{H}_\infty$	-	0.09	-	0.14	-	0.14
PI	time	**0.10**	0.06	**0.28**	0.09	**0.61**	0.46
	$J_{\text{mix}}^{1/2}$	0.99	0.47	1.99	0.02	0.04	0.04
	$\mathcal{H}_2$	0.99	0.47	1.99	1.12	2.44	1.22
	$\mathcal{H}_\infty$	1.98	0.09	7.72	0.14	9.27	0.14
LMI	time	0.27	0.37	11.3	20.1	89.7	143
	$J_{\text{mix}}^{1/2}$	1.00	0.47	1.20	1.12	0.04	1.22
	$\mathcal{H}_2$	0.99	0.47	1.99	1.12	2.44	1.22
	$\mathcal{H}_\infty$	1.98	0.09	7.77	0.14	9.27	0.13
hifoo	time	1.43	8.57	35.6	27.5	262	221
	$\mathcal{H}_2$	0.99	0.47	1.99	1.12	2.44	1.22
	$\mathcal{H}_\infty$	2.01	0.09	8.57	0.14	10.1	0.14

Talk outline

Part I.

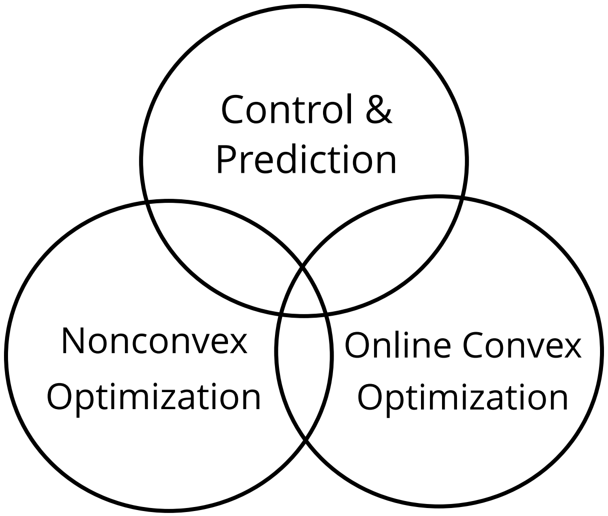
Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: benign nonconvexity

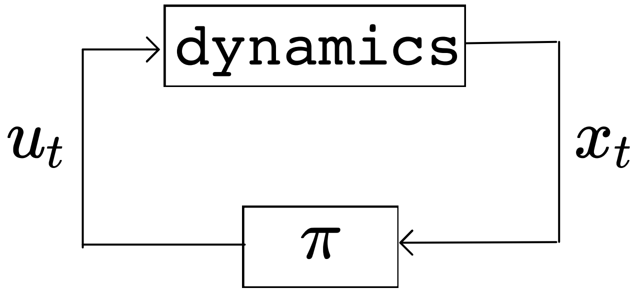
Part II.

Online tracking with predictions: dynamic regret analysis of MPC

Part III.

Online adaptive control & prediction under nonstationarity





Tracking of nonlinear systems

Nonlinear dynamics

$$z_{t+1} = f(z_t, u_t)$$

Koopman-linearizable:
there exist a lifting function ψ and A, B, C s.t. the lifted state $x_t = \psi(z_t)$ evolves linearly $x_{t+1} = Ax_t + Bu_t$ and $z_t = Cx_t$

Stage tracking cost: target trajectory

$$\ell(z_t, u_t; r_t) := \|z_t - r_t\|_{Q_z}^2 + \|u_t\|_R^2$$

$$\ell_{\text{lft}}(x_t, u_t; r_t) := \|x_t - \psi(r_t)\|_Q^2 + \|u_t\|_R^2$$

$$Q = C^\top Q_z C$$

$$\min_{u_{1:T}} \quad \sum_{i=1}^{T} \|z_t - r_t\|_{Q_z}^2 + \|u_t\|_R^2$$

$$\text{s.t.} \quad z_t = f(z_t, u_t),\ z_1 \text{ given}$$

Online tracking with predictions

Uncertainty modeling (target trajectory and its prediction)

At each time step,

1. Learner observes z_t and receives W-step predictions $r_{t:t+W-1}$
2. Learner picks u_t, and suffers the tracking cost $\ell(z_t, u_t; r_t)$
3. Adversary/environment selects target state r_{t+W}
4. State z_t evolves to z_{t+1} according to f

Goal: minimize the (restricted) dynamic regret of online policy π

$$R_T^\star(\pi) = \underbrace{\sum_{t=1}^{T} \ell(z_t, u_t; r_t)}_{\text{Tracking cost}} - \min_{u_{1:T}^\star} \sum_{t=1}^{T} \ell(z_t, u_t^\star; r_t)$$

Globally optimal **"offline non-causal policy"** with full knowledge of $r_{1:T}$ and dynamics f

Optimal (offline noncausal) policy in hindsight

$$\min_{u_{1:T}} \sum_{i=1}^{T} \ell(z_t, u_t; r_t) \quad \text{s.t.} \quad z_t = f(z_t, u_t),\ z_1 \text{ given}$$

Koopman linearizable

$$\min_{u_{1:T}} \sum_{i=1}^{T} \ell_{\text{lft}}(x_t, u_t; r_t) \quad \text{s.t.} \quad x_{t+1} = Ax_t + Bu_t,\ x_1 = \psi(z_1)$$

Riccati recursion: $P_T = Q$ and $P_t = Q + A^\top P_{t+1} A - A^\top P_{t+1} B (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A$

- Characterize the value or cost-to-go function: $V_t(x_t) = x_t^\top P_t x_t$
- Induce optimal control gains: $K_t^\star = (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A$
- State transition matrix $A_{\text{cl},t_1 \to t_2} := A_{\text{cl},t_2} A_{\text{cl},t_2-1} \cdots A_{\text{cl},t_1+1}$ with $A_{\text{cl},t} := A - BK_t^\star$

Globally optimal time-varying policy [FS20, ZLL21, GH22]:

$$\pi_t^\star(x_t; \mathbf{r}) = u_t^\star = \underbrace{-K_t^\star(x_t - \psi(r_t))}_{\text{feedback}} - \underbrace{\textstyle\sum_{i=t}^{T-1} K_{t\to i}^\star (A\psi(r_i) - \psi(r_{i+1}))}_{\text{feedforward}}$$

[FS20] D. Foster and M. Simchowitz. *Logarithmic regret for adversarial online control*. ICML, 2020

[ZLL21] R. Zhang, Y. Li, and Na Li. *On the regret analysis of online LQR control with predictions*. IEEE ACC, 2021

[GH22] G. Goel and B. Hassibi. *The power of linear controllers in LQR control*. IEEE CDC, 2022

Model Predictive Control (MPC)

The most widely used approach for online control with predictions

At each step t, solve a **shorter-horizon** optimization problem and apply the **first** action from the optimized action sequence

Knowledge of $r_{t:t+W-1}$ from prediction

$$\min_{u_{1:W|t}} \sum_{i=1}^{W} \ell(z_{i|t}, u_{i|t}; r_{t+i-1}) \quad \text{s.t.} \quad z_{i+1|t} = f(z_{i|t}, u_{i|t}), \quad z_{1|t} = z_t$$

Koopman linearizable

$$\min_{u_{1:W|t}} \sum_{i=1}^{W} \ell_{\text{lft}}(x_{i|t}, u_{i|t}; r_{t+i-1}) \quad \text{s.t.} \quad x_{i+1|t} = Ax_{i|t} + Bu_{i|t}, \quad x_{1|t} = x_t = \psi(z_t)$$

Some rationales. model, disturbance, cost; full-horizon DP

- Model-based approach to tackle uncertainty and computational difficulty
- Can be viewed as (multistep lookahead) policy iteration for approximate DP

MPC as an online feedback policy

Generally, MPC (implicitly) defines a **time-varying** state **feedback** policy

At each t, solve

$$u^\star_{1:W|t}(x_t) = \arg\min_{u_{1:W|t}} \left\{ \sum_{i=1}^{W} \ell_{\mathrm{lft}}(x_{i|t}, u_{i|t}; r_{t+i-1}) : x_{i+1|t} = Ax_{i|t} + Bu_{i|t},\ x_{1|t} = x_t \right\}$$

In our case, MPC defines a time-varying policy (due to $r_{1:T}$)

$$\pi_t^{\mathrm{MPC}}(x_t) = u^\star_{1|t} = \underbrace{-K_1^{\mathrm{MPC}}(x_t - \psi(r_t))}_{\text{feedback}} - \underbrace{\sum_{i=t}^{t+W-2} K^{\mathrm{MPC}}_{1\to i-t+1}(A\psi(r_i) - \psi(r_{i+1}))}_{\text{feedforward}}$$

Stationary gains: $K_1^{\mathrm{MPC}} = K^\star_{T-W}$ and $\{K^{\mathrm{MPC}}_{1\to k}\}_{k=0}^{W} = \{K^\star_{T-W\to T-W+k}\}_{k=0}^{W}$

Dynamic regret guarantee

Main result (informal) [PSQZ]

As long as W is large enough, the dynamic regret satisfies

$$R_T^\star(\mathrm{MPC}) = O(W^2 \lambda^{2W} T) \text{ where } \lambda \in (0,1)$$

- Grows linearly with T
- Decays exponentially with W

1. No terminal cost design, but a sufficiently long W is required.
2. The power of predictions + the exponential convergence of stable linear dynamics.

$\lambda = \max\{\gamma_\infty, \rho_\infty\}$

- Factor $\gamma_\infty := \frac{1}{2}(1 + \rho(A_{\mathrm{cl},\infty}))$ captures stability of $A_{\mathrm{cl},t_1 \to t_2}$
- Factor ρ_∞: $\|P_t - P_\infty\| = O(\rho_\infty^{T-t})$, $\|K_t - K_\infty\| = O(\rho_\infty^{T-t})$
- $W \geq \Delta_{\mathrm{stab}} = O(\log(1 - \rho(A_{\mathrm{cl},\infty}))^{-1})$

[PSQZ] **C. Pai**, X. Shang, J. Qian, and Y. Zheng. *Online Tracking with Predictions for Nonlinear Systems with Koopman Linear Embedding.* Submitted to L4DC

Dynamic regret analysis

Performance difference lemma: $R_T^\star(\mathrm{MPC}) = \sum_{t=1}^{T} \|u_t^{\mathrm{MPC}} - u_t^\star\|_{\Sigma_t}^2$

$$u_t^{\mathrm{MPC}} - u_t^\star = \underbrace{(K_t^\star - K_1^{\mathrm{MPC}})(x_t - \psi(r_t))}_{\text{Feedback}} + \underbrace{\sum_{i=t}^{t+W-2} (K_{t\to i}^\star - K_{1\to i-t+1}^{\mathrm{MPC}}) w_i}_{\text{Feedforward}} + \underbrace{\sum_{i=t+W-1}^{T-1} K_{t\to i}^\star w_i}_{\text{Truncation}}$$

$$w_t := A\psi(r_i) - \psi(r_{i+1})$$

$$R_T^\star(\mathrm{MPC}) \le (1) + (2) + (3)$$

(1) Truncation deviation: $\sum_{t=1}^{T-W} \left\|\sum_{i=t+W-1}^{T-1} K_{t\to i}^\star w_i\right\|_{\Sigma_t}^2 = O(\gamma_\infty^{2W} T)$

(2) Feedback deviation: $\sum_{t=1}^{T-W} \left\|(K_t^\star - K_1^{\mathrm{MPC}})(x_t - \psi(r_t))\right\|_{\Sigma_t}^2 = O(\rho_\infty^{2W} T)$

(3) Feedforward deviation: $\sum_{t=1}^{T-W} \left\|\sum_{i=t}^{t+W-2} (K_{t\to i}^\star - K_{1\to i-t+1}^{\mathrm{MPC}}) w_i\right\|_{\Sigma_t}^2 = O(\lambda_\infty^{2W} T)$

Experiment: tracking a reference sinusoid

Koopman-linearizable **nonlinear** dynamics:

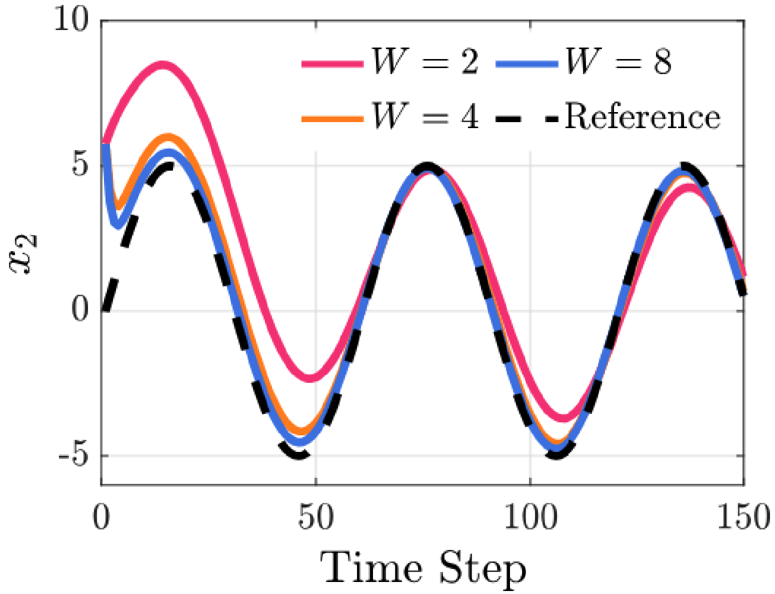
$$\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix} = \begin{bmatrix} 0.99 z_{1,t} \\ 0.9 z_{2,t} + z_{1,t}^2 + z_{1,t}^3 + z_{1,t}^4 + u_t \end{bmatrix}$$

Linear dynamics in the lifted space:

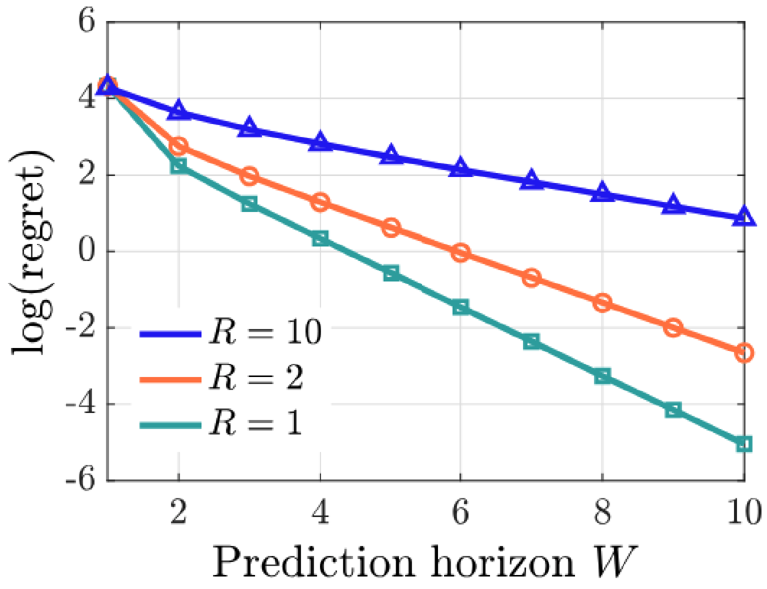
$$x_{t+1} = \begin{bmatrix} 0.99 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 1 & 1 & 1 \\ 0 & 0 & 0.99^2 & 0 & 0 \\ 0 & 0 & 0 & 0.99^3 & 0 \\ 0 & 0 & 0 & 0 & 0.99^4 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t$$

Lifted state: $x := [z_1, z_2, z_1^2, z_1^3, z_1^4]^\top$, with state recovery $z_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} x_t$

$r_{2,t} = 5\sin(\pi t/30), T = 200$



$R_T^\star$(MPC) exp decays



Experiment: two-wheeled robots

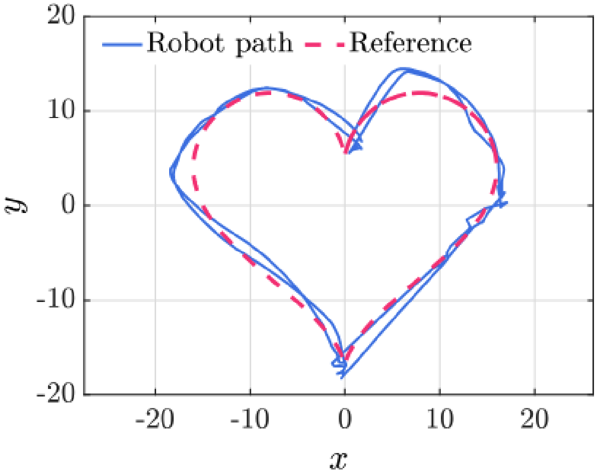
Non-Koopman-linearizable nonlinear dynamics

$$z_{x,t+1} = z_{x,t} + \Delta t \cdot \cos(z_{\delta,t}) \cdot v_t\,,$$
$$z_{y,t+1} = z_{y,t} + \Delta t \cdot \sin(z_{\delta,t}) \cdot v_t\,,$$
$$z_{\delta,t+1} = z_{\delta,t} + \Delta t \cdot w_t\,,$$

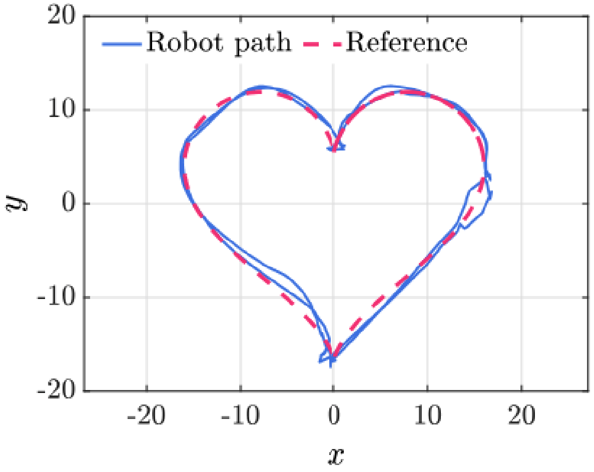
A heart-shaped reference trajectory:

$$r_{x,t} = 16\sin^3(t-6),$$
$$r_{y,t} = 13\cos(t) - 5\cos(2t-12) - 2\cos(3t-18) - \cos(4t-24),$$
$$r_{\delta,t} = \arctan\left(\frac{r_{y,t+1} - r_{y,t}}{r_{x,t+1} - r_{x,t}}\right).$$

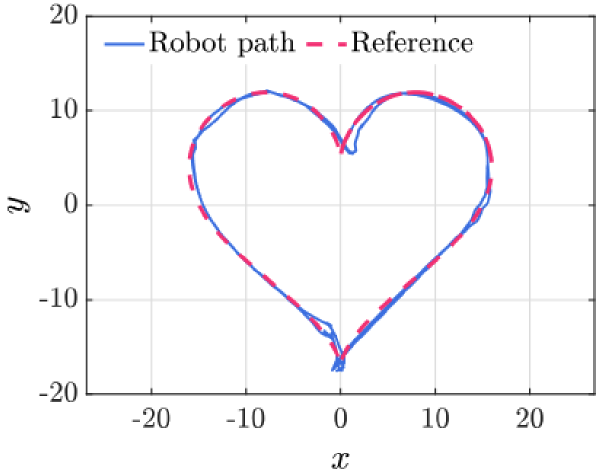
$W = 6$



$W = 9$



$W = 12$



Talk outline

Part I.

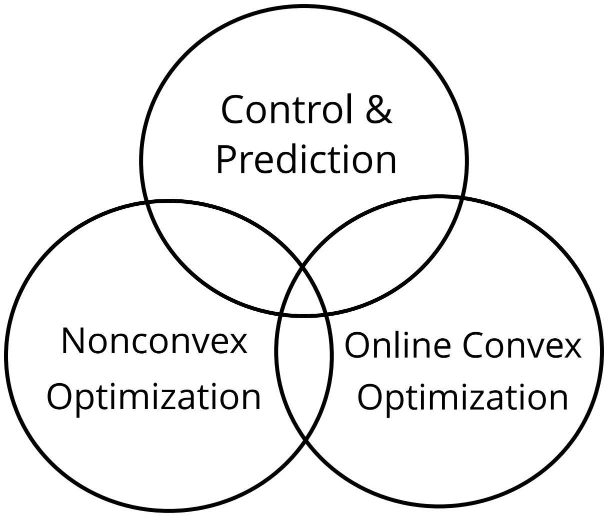
Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: benign nonconvexity

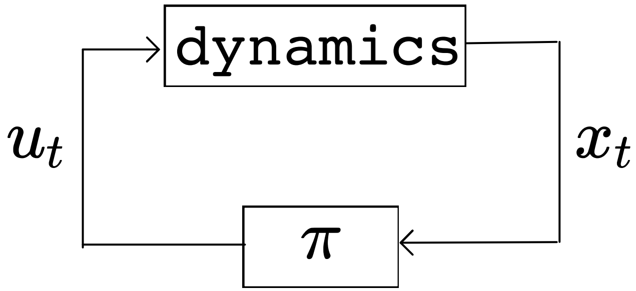
Part II.

Online tracking with predictions: dynamic regret analysis of MPC

Part III.

Online adaptive control & prediction under nonstationarity





Online learning / convex optimization

A repeated game between learner & environment (adversary)

for $t = 1, 2, \ldots, T$, learner

- selects $x_t \in \mathcal{X}$
- receives convex $f_t : \mathcal{X} \to \mathbb{R}$
- suffers loss $f_t(x_t)$

The goal of the learner is to minimize $\sum_{t=1}^{T} f_t(x_t)$

The adversarial nature of f_t hinders a prior computation of optimal decisions

Goal: minimize (static) regret

the best fixed comparator w in hindsight

$$R_T(w) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(w)$$

How to connect online learning with control?

Some excellent treatment of online learning:

[E16] E. Hazan. *Introduction to online convex optimization*. Foundations and Trends in Optimization, 2016.

[F19] O. Francesco. *A modern introduction to online learning*. arXiv preprint, 2019.

Online control under nonstochastic disturbance

for $t = 1, 2, \ldots, T$, learner

- observes x_t, selects $u_t \in \mathcal{U}$
- receives convex $c_t : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ and disturbance w_t
- suffers loss $c_t(x_t, u_t)$ and state evolves

Goal: minimize (static) policy regret [ABHKS19]

$$R_T(\pi) = \max_{\|w_t\| \leq 1} \left(\sum_{t=1}^{T} c_t(x_t, u_t) - \min_{\pi \in \Pi} \sum_{t=1}^{T} c_t(\hat{x}_t, \pi(\hat{x}_t)) \right)$$

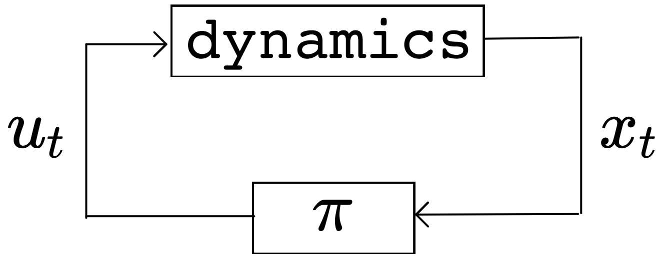
Main challenges: nonconvexity, trajectory mismatch

Techniques: OCO with memory [ABHKS19] or OCO with delayed feedback [FS20]

[ABHKS19] N. Agarwal, B. Bullins, E. Hazan, S. Kakade, and K. Singh. *Online control with adversarial disturbances*. ICML, 2019
[FS20] D. Foster and M. Simchowitz. *Logarithmic regret for adversarial online control.* ICML, 2020.

$$
\min_{\pi} \quad \sum_{t=1}^{T} \mathrm{cost}_t(x_t, u_t)
$$

$$
\text{s.t.} \quad x_{t+1} = \mathrm{dynamics}_t(x_t, u_t, w_t)
$$



Offline Synthesis

1. Specific disturbance models (H2/H-infty) and quadratic cost
2. Simple, explicit, closed-form globally optimal policy
3. Absolute optimality wrt the disturbance model

Online Learning

1. Arbitrary disturbance sequences and convex cost functions
2. Generally intractable to find a globally optimal policy
3. Relative optimality: compete with a certain policy class
4. Instance-optimality wrt the actual realized disturbance and cost

Generalizations: three layers of adaptivity

Static regret: adaptive to adversary

$$R_T(x) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(x) = \mathcal{O}(\sqrt{T})$$

Universal dynamic regret: adaptive to any nonstationarity

$$R_T(w_{1:T}) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(w_t) = \mathcal{O}(\sqrt{T(1+P_T)}), \quad P_T = \sum_{t=2}^{T} \|w_t - w_{t-1}\|_2$$

Problem-dependent regret: adaptive to any problem instances

$$R_T(w_{1:T}) = \sum_{t=1}^{T} f_t(x_t) - \sum_{t=1}^{T} f_t(w_t) = \mathcal{O}(\sqrt{(1 + P_T + \min\{V_T, F_T\})(1+P_T)})$$

$$V_T = \sum_{t=2}^{T} \sup_{x \in \mathcal{X}} \|\nabla f_t(x) - \nabla f_{t-1}(x)\|_2^2, \quad F_T = \sum_{t=1}^{T} f_t(w_t)$$

Ongoing and future work

Online adaptive control and prediction under nonstationarity

1. Problem-dependent regret minimization for online nonstochastic control

2. Online time-series prediction for time-varying linear dynamical systems

Main challenges for control and prediction:

Nonconvexity, trajectory mismatch (memory)

Tools/techniques from online learning:

convex relaxation, meta-base structure [ZLZ18], switching cost regularization [ZYWZ23], tailored optimism [ZZZZ24]

[ZLZ18] L. Zhang, S. Lu, and Z. Zhou. *Adaptive online learning in dynamic environments*. NeurIPS, 2018.

[ZYWZ23] P. Zhao, Y. Yan, Y. Wang, and Z. Zhou. *Non-stationary online learning with memory and non-stochastic control*. JMLR, 2023

[ZZZZ24] P. Zhao, Y. Zhang, L. Zhang, and Z. Zhou. *Adaptivity and non-stationarity: Problem-dependent dynamic regret for online convex optimization*. JMLR, 2024.

Conclusion

Offline Planning $\rightarrow$ Policy Optimization $\rightarrow$ Online Learning

Part I.

Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ policy optimization

Offline planning -> Policy optimization

[PWTZ, ZPT25]

Part II.

Dynamic regret analysis of MPC

Offline planning -> Online learning

[PSQZ]

Part III.

Online adaptive control & prediction

Generalize offline planning & policy optimization

[PWTZ] **C. Pai**, Y. Watanabe, Y. Tang, and Y. Zheng. *Policy Optimization of Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ Control: Benign Nonconvexity and Global Optimality.* Submitted to Automatica

[ZPT25] Y. Zheng, **C. Pai**, and Y. Tang. *Extended Convex Lifting for Policy Optimization of Optimal and Robust Control.* Learning for Dynamics and Control (L4DC) 2025

[PSQZ] **C. Pai**, X. Shang, J. Qian, and Y. Zheng. *Online Tracking with Predictions for Nonlinear Systems with Koopman Linear Embedding.* Submitted to L4DC

Thanks for your attention!

Q&A

Some other relevant projects I was involved in:

[ZPT1] Y. Zheng, **C. Pai**, and Y. Tang. *Benign Nonconvex Landscapes in Optimal and Robust Control, Part I: Global Optimality.* Submitted to IEEE TAC

[ZPT2] Y. Zheng, **C. Pai**, and Y. Tang. *Benign Nonconvex Landscapes in Optimal and Robust Control, Part II: Extended Convex Lifting.* Submitted to IEEE TAC

[WPZ25] Y. Watanabe, **C. Pai**, and Y. Zheng. *Semidefinite Programming Duality in Infinite-Horizon Linear Quadratic Differential Games.* IEEE CDC, 2025