Decision-Making in Uncertain Dynamic Environments
From policy optimization to online learning

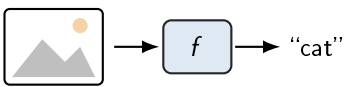
Chih-Fan (Rich) Pai

Defense Talk, ECE, UC San Diego
November 4, 2026

Advisor: Prof. Yang Zheng
Committee: Profs. Massimo Franceschetti, Sicun Gao, Jorge Poveda

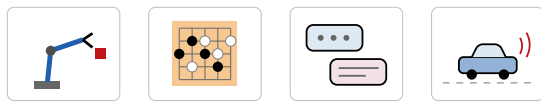
Machine learning in dynamical systems

Standard machine learning



data samples are independent (i.i.d.)

Dynamical systems



data are coupled over time through the dynamics

Several fundamental challenges:

- optimizing feedback/observer gains often yields nonconvex problems
- how to model and address uncertainty and disturbances
- the system dynamics may be unknown, partially observed, or marginally stable

My PhD research: principled methods for **learning and optimization in dynamical systems**

The world is full of dynamical systems



robotics



power grids



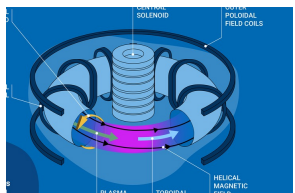
chemical processes



autonomous driving



warehouse logistics



nuclear fusion

Images from public sources, for illustration only.

From policy optimization to online learning

- Two foundational tasks in dynamical systems: **control** and **prediction**
- Both problems involve nonconvex optimization over gain-parameterized policies:

$$\text{Control (feedback gain } K\text{):} \quad u_t = Kx_t$$

$$\text{Prediction (observer gain } L\text{):} \quad \hat{x}_{t+1} = A\hat{x}_t + L(y_t - C\hat{x}_t), \quad \hat{y}_t = C\hat{x}_t$$

- Beyond nonconvexity, the appropriate solution/guarantee depends heavily on what information is available and how we model uncertainty
- **Part I, policy optimization for control:** known system and disturbance models
  - can we find globally optimal feedback policies despite nonconvexity?
- **Part II, online learning for prediction:** arbitrary disturbances
  - learn from the actual data sequence, to compete with the best observer in hindsight

Outline

Part I: Policy optimization for control

Part II: Online learning for prediction

Part III: Learning and duality in control

Conclusion and future directions

Outline

Part I: Policy optimization for control

- Problem formulation
- Our contributions and main result
- Analysis: Extended Convex Lifting (ECL)
- Policy iteration and numerical experiments

Part II: Online learning for prediction

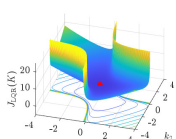
Part III: Learning and duality in control

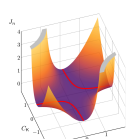
Conclusion and future directions

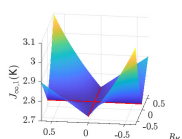
Policy optimization (PO) for classical control

Search the policy K directly: a (constrained) nonconvex optimization problem

- Classical control problems revisited from the PO perspective:
  - LQR [Fazel et al., 2018; Mohammadi et al., 2022], LQG [Tang et al., 2021], $\mathcal{H}_\infty$ control [Guo, Hu, 2022], LQ games [Zhang et al., 2019], Kalman filter [Umenberger et al., 2022]
- Challenge: does local search converge to the global optimal policy?
  - favorable landscape properties + structure of stationary points
- Part I: mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control + general conditions that certify global optimality







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

- System $\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t)$ with linear feedback $u(t) = Kx(t)$
- **Two channels:** $w \to z_2$ (average, $\mathcal{H}_2$ norm) and $w \to z_\infty$ (worst case, $\mathcal{H}_\infty$ norm)

$$\inf_K \; J(K) := \overbrace{\mathrm{Tr}\left((Q_2 + K^\mathsf{T} R_2 K) X_K\right)}^{\text{enhance system performance}} \;\geq\; \|\mathsf{T}_{z_2 w}(K)\|_{\mathcal{H}_2}^2$$

$$\text{s.t.} \;\; K \in \mathcal{K}_\beta := \{K : \underbrace{A + BK \text{ Hurwitz}}_{\text{enforce stability}}, \; \underbrace{\|\mathsf{T}_{z_\infty w}(K)\|_{\mathcal{H}_\infty} < \beta}_{\text{enhance system robustness}}\}$$

- Adjusting $\beta \to \infty$ or $\beta \to \inf_K \|\mathsf{T}_{z_\infty w}(K)\|_{\mathcal{H}_\infty}$ gives optimal $\mathcal{H}_2$ or $\mathcal{H}_\infty$ control
- Our goal is to understand the optimization landscape of this nonconvex problem.

Outline

Part I: Policy optimization for control
- Problem formulation
- Our contributions and main result
- Analysis: Extended Convex Lifting (ECL)
- Policy iteration and numerical experiments

Our contributions

- **Global optimality vs sub-optimality**:
  - [Khargonekar, Rotea, 1991] only focused on suboptimal solutions, with no certificate of global optimality.
- **General two-channel vs single-channel formulation**:
  - [Zhang et al., 2021] certified global optimality for the single-channel case only.
- **Analysis techniques**:
  - **Extended convex lifting (ECL)** [Zheng, **Pai**, Tang, 2026] **vs dynamic game approach** [Başar, Bernhard, 2008] used in [Zhang et al., 2021]

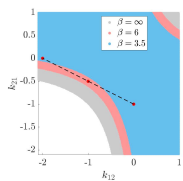
	ECL	Dynamic game
Two-channel	✓	?
Single-channel	✓	✓

Main result: benign nonconvexity

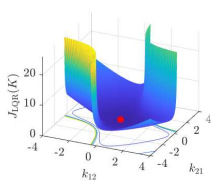
> **Theorem (No spurious stationary points).** *Every stationary point $K \in \mathcal{K}_\beta$ of J is a global minimizer.*

We also characterize

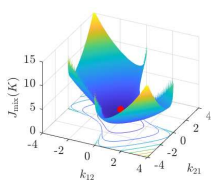
- the feasible set $\mathcal{K}_\beta$ and the cost J, including their boundary behavior
- existence and uniqueness of stationary points



$\mathcal{K}_\beta$ is nonconvex



J_{LQR}: nonconvex, coercive



J: nonconvex, not coercive

Paper: **C.-F. Pai**, Y. Watanabe, Y. Tang, Y. Zheng. Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: benign nonconvexity and global optimality. *Automatica*, provisionally accepted. arXiv:2603.04843.

Outline

Part I: Policy optimization for control

- Problem formulation
- Our contributions and main result
- Analysis: Extended Convex Lifting (ECL)
- Policy iteration and numerical experiments

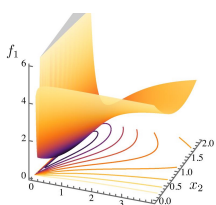
Hidden convexity: change of variables

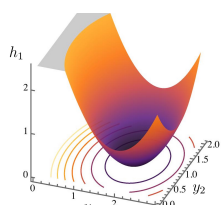
- Consider nonconvex f_1 with $x^\star = (\frac{1}{2}, 1)$:

$$f_1(x_1, x_2) = \left(\frac{x_2}{x_1} - 2\right)^2 + \left(x_2 - 1\right)^2, \qquad \forall x_1, x_2 > 0$$

- Through diffeomorphism $g(x_1, x_2) = (\frac{x_2}{x_1}, x_2)$, $y^\star = (2, 1)$:

$$h_1(y_1, y_2) := f_1\left(g^{-1}(y_1, y_2)\right) = \left(y_1 - 2\right)^2 + \left(y_2 - 1\right)^2, \qquad \forall y_1, y_2 > 0$$



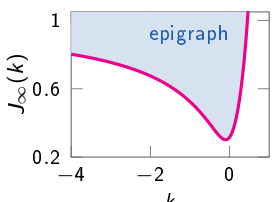


Hidden convexity: a lifting procedure

- A simple nonconvex $\mathcal{H}_\infty$ instance:

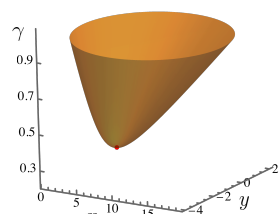
$$J_\infty(k) = \|\mathsf{T}_{zw}(k)\|_{\mathcal{H}_\infty} = \frac{\sqrt{0.1 + k^2}}{1 - k}, \qquad k < 1$$

- Lift the nonconvex epigraph of J_∞ to a higher-dimensional set, which then admits a diffeomorphism to a convex set.



nonconvex $J_\infty(k)$





convex $\mathcal{F}_\infty$

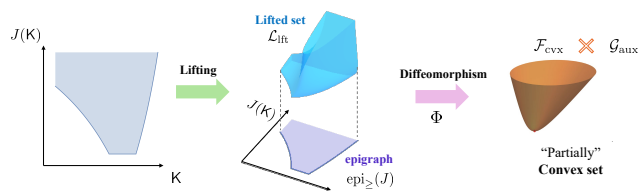
Extended Convex Lifting and its guarantees [Zheng, **Pai**, Tang, 2026]

Definition (ECL). *For a continuous $f : \mathcal{D} \to \mathbb{R}$, $(\mathcal{L}_{\mathrm{lft}}, \mathcal{F}_{\mathrm{cvx}}, \Phi)$ is an ECL of f if*

1. $\mathrm{epi}_{>}(f) \subseteq \pi_{x,\gamma}(\mathcal{L}_{\mathrm{lft}}) \subseteq \mathrm{cl}\,\mathrm{epi}_{\geq}(f)$
2. $\mathcal{F}_{\mathrm{cvx}}$ *convex;* $\Phi : \mathcal{L}_{\mathrm{lft}} \to \mathcal{F}_{\mathrm{cvx}}$ C^2 *diffeomorphism*
3. $\Phi(x, \gamma, \xi) = (\gamma, \zeta_1)$, ξ*: extra variable*

Theorem (ECL guarantees). *If f is differentiable with an ECL:*

1. $\inf_{\mathcal{D}} f = \inf_{\mathcal{F}_{\mathrm{cvx}}} \gamma$ *(convex reformulation);*
2. *non-degenerate stationary points are globally optimal.*



Paper: Y. Zheng, **C.-F. Pai**, Y. Tang. Benign nonconvex landscapes in optimal and robust control, Part II: Extended convex lifting. *IEEE TAC*, 71(8):5350–5365, 2026.

ECL for mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control?

Step 1: lifted set

$$\mathcal{L}_{\text{lft}} = \left\{ (K,\gamma,X,Z) \;\middle|\; \begin{array}{l} X \succ 0,\ Z \succeq KXK^{\mathsf{T}},\ \gamma \geq \mathrm{Tr}(Q_2 X) + \mathrm{Tr}(R_2 Z), \\ (A+BK)X + X(A+BK)^{\mathsf{T}} + W + \beta^{-2} X(Q_\infty + K^{\mathsf{T}} R_\infty K)X \preceq 0 \end{array} \right\}$$

Step 2: diffeomorphism $\Phi(K,\gamma,X,Z) = (\gamma, X, KX, Z)$

Step 3: convex set

$$\mathcal{F}_{\text{cvx}} = \left\{ (\gamma, X, Y, Z) \;\middle|\; \begin{array}{l} X \succ 0,\ \begin{bmatrix} X & Y^{\mathsf{T}} \\ Y & Z \end{bmatrix} \succeq 0,\ \gamma \geq \mathrm{Tr}(Q_2 X) + \mathrm{Tr}(R_2 Z), \\ \begin{bmatrix} AX + XA^{\mathsf{T}} + BY + Y^{\mathsf{T}} B^{\mathsf{T}} + W & (C_\infty X + D_\infty Y)^{\mathsf{T}} \\ C_\infty X + D_\infty Y & -\beta^2 I \end{bmatrix} \preceq 0 \end{array} \right\}$$

$\mathcal{F}_{\text{cvx}}$ is (almost) the classical LMI formulation; ECL connects this convex formulation back to the original nonconvex PO landscape.

ECL construction and guarantees for mixed control

Theorem (ECL construction for the mixed cost J).

1. $\mathrm{epi}_{\geq}(J) \subseteq \pi_{K,\gamma}(\mathcal{L}_{\mathrm{lft}}) \subseteq \mathrm{cl}\,\mathrm{epi}_{\geq}(J)$;
2. Φ *is a* C^{∞} *diffeomorphism from* $\mathcal{L}_{\mathrm{lft}}$ *to* $\mathcal{F}_{\mathrm{cvx}}$.

By direct application of the ECL guarantees:

Convex reformulation:

$$\inf_{K \in \mathcal{K}_\beta} J(K) = \inf_{(\gamma,X,Y,Z) \in \mathcal{F}_{\mathrm{cvx}}} \gamma$$

Every stationary point of the mixed policy optimization is **a global minimizer**.

Outline

Part I: Policy optimization for control
- Problem formulation
- Our contributions and main result
- Analysis: Extended Convex Lifting (ECL)
- Policy iteration and numerical experiments

Two-channel policy iteration

A fixed-point iteration on the stationarity condition:

$$\nabla J(K) = 0 \;\Rightarrow\; K = -R_2^{-1} B^{\mathsf{T}} \Gamma_K (I + \beta^{-2}\alpha^2 X_K \Gamma_K)^{-1}$$

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

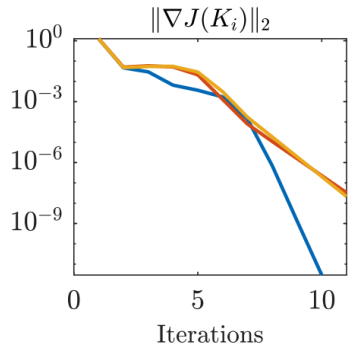
A direct policy-based alternative to solving the classical LMI problem

Policy iteration scales better

Runtime to the global optimum

Policy size	LMI (MOSEK)	Policy iteration
60×60	$> 11\,\mathrm{s}$	$< 1\,\mathrm{s}$
90×90	$89\,\mathrm{s}$	$< 1\,\mathrm{s}$

- Riccati, PI, and LMI return the same policy.
- LMI: X has $O(n^2)$ entries; PI updates only K.



$\|\nabla J(K_i)\|_2$ along policy iteration

Paper: **C.-F. Pai**, Y. Watanabe, Y. Tang, Y. Zheng. Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: benign nonconvexity and global optimality. *Automatica*, provisionally accepted. arXiv:2603.04843.

Outline

Part I: Policy optimization for control

Part II: Online learning for prediction

- Formalizing online nonstochastic prediction
- Reduction to online regression
- Online regression with optimism
- Designing hints with bounded errors
- Main result and numerical experiments

Part III: Learning and duality in control

Conclusion and future directions

From classical filtering to online learning

- Classical filters [Kalman, 1960; Hassibi et al., 1999]: designed **in advance** for a **noise model**



- Question: what if the **actual disturbances** are arbitrary or adversarial?
- Online learning: learn from the actual seq, compete with the best predictor in hindsight
  - can we do almost as well online, without knowing the future? $\Rightarrow$ regret
- Goal: can we get low regret under the challenging combination of
  - **marginal stability**, **partial observation**, and **nonstochastic disturbances**?
- Part II: formalize online nonstochastic prediction + achieve polylogarithmic regret

Paper: **C.-F. Pai**, Y. Zheng. Online nonstochastic prediction: logarithmic regret via predictive online least squares. arXiv:2605.04364, 2026.

Sequential output prediction in an LDS

- Consider a partially observed linear dynamical system (LDS):

$$x_{t+1} = Ax_t + w_t, \qquad y_t = Cx_t + v_t.$$

We assume that A is marginally stable ($\rho(A) \le 1$), and disturbances w_t, v_t are nonstochastic (bounded but otherwise arbitrary).

Interactive protocol. Consider T-round prediction. At each round t,

1. learner predicts $\hat{y}_t$ from past outputs $y_1, \dots, y_{t-1}$;
2. **nature** picks w_{t-1}, v_t (possibly adversarially), and the **LDS** reveals y_t;
3. learner suffers $\|\hat{y}_t - y_t\|^2$.

- Unlike standard online prediction:
  - **Opportunity:** y_t keeps the **LDS structure**, which we want to exploit
  - **Difficulty:** marginal stability $\Rightarrow$ $\|y_t\|$ may grow like t^r (unbounded)

Compared to what? The best Luenberger observer in hindsight

Comparator class $\mathcal{L}_{\kappa,\gamma}$: stable Luenberger observers [Luenberger, 1964]

$\hat{x}_{t+1}^{\mathrm{LB}} = A\hat{x}_t^{\mathrm{LB}} + L(y_t - C\hat{x}_t^{\mathrm{LB}}), \qquad \hat{y}_t^{\mathrm{LB}}(L) = C\hat{x}_t^{\mathrm{LB}}, \qquad A - LC$ strongly stable

Luenberger regret (against the best observer gain $L^\star$ in hindsight):

$$\mathrm{Reg}_T^{\mathrm{LB}} = \sum_{t=1}^{T} \|\hat{y}_t - y_t\|^2 - \min_{L \in \mathcal{L}_{\kappa,\gamma}} \sum_{t=1}^{T} \|\hat{y}_t^{\mathrm{LB}}(L) - y_t\|^2$$

- **Why this class?** Classical (steady state) filters are Luenberger observers, with L **chosen in advance from a noise model**:

$$\textit{Kalman}: \min_L \mathbb{E}_W \sum_{t=1}^{T} \|\hat{y}_t^{\mathrm{LB}}(L) - y_t\|^2 \qquad \mathcal{H}_\infty: \min_L \sup_W \frac{\sum_{t=1}^{T} \|\hat{y}_t^{\mathrm{LB}}(L) - y_t\|^2}{\|W\|^2}$$

What does low Luenberger regret mean?

- Imagine after observing the entire sequence $y_{1:T}$, we could go back and select the best fixed observer gain:

$$L^\star = \arg \min_{L \in \mathcal{L}_{\kappa,\gamma}} \sum_{t=1}^{T} \|\hat{y}_t^{\mathrm{LB}}(L) - y_t\|^2$$

  - $L^\star$ is unavailable during online prediction, because it depends on future observations.

- Low (sublinear) regret: in the long run, we predict as well as $L^\star$ **as if we had known the entire disturbances/observations in advance**
  - $L^\star$ tuned to the actual disturbances, while **Kalman**/$\mathcal{H}_\infty$ to a **model in advance**

- Learning is needed even with (A, C): $L^\star$ depends on the actual W, unknown in advance

- But **directly learning $L^\star$ is hard**, since $\hat{y}_t^{\mathrm{LB}}(L)$ has infinite memory and is nonconvex in L

$$\hat{y}_t^{\mathrm{LB}}(L) = \sum_{k=0}^{t-1} C(A - LC)^k L\, y_{t-k-1}$$

Comparison with prior work

	Nonstochastic disturbances	Partial observation	Marginal stability
Agnostic prediction [Hazan et al., 2017; Kozdoba et al., 2019]	✓	✓	✗ (bounded y_t)
Learning Kalman filters (polylog) [Tsiamis, Pappas, 2023; Qian, Zheng, 2025]	✗ (stochastic)	✓	✓
Online state prediction (sublinear) [Ghai et al., 2020]	✓	✗ ($y_t = x_t$)	✓
This work [Pai, Zheng, 2026]	✓	✓	✓

- **Our contribution:** $O(\log^2 T)$ Luenberger regret on every bounded disturbance sequence, extending polylog regret from stochastic to adversarial noise, and from $y_t = x_t$ to partial observation

Roadmap: two obstacles

1. How to compete with nonconvex, infinite-memory observers?
   - comparator $\hat{y}_t^{\mathrm{LB}}(L) = \sum_{k=0}^{t-1} C(A - LC)^k L\, y_{t-k-1}$ depends on the whole past

2. How to guarantee low regret when observations grow?
   - under marginal stability and bounded disturbances, $\|y_t\|$ may grow like t^r

Outline

Part II: Online learning for prediction

- Formalizing online nonstochastic prediction
- Reduction to online regression
- Online regression with optimism
- Designing hints with bounded errors
- Main result and numerical experiments

Obstacle 1: how to compete with nonconvex, infinite-memory observers?

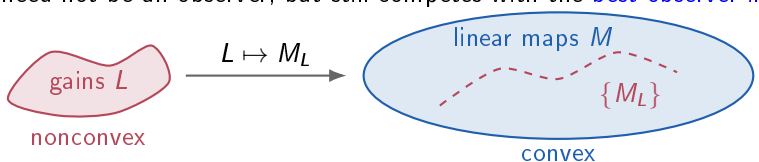
Idea 1: truncate, then relax (improper learning)

- **Truncate (AR model** [Anava et al., 2013]**):** $C(A-LC)^k L$ decays geometrically:

$$\hat{y}_t^{\mathrm{LB}}(L) = \sum_{k=0}^{t-1} C(A-LC)^k L\, y_{t-k-1} \;\approx\; M_L z_t, \qquad z_t = [y_{t-1}; \ldots; y_{t-H}]$$

$$M_L = [CL,\ \ldots,\ C(A-LC)^{H-1}L]$$

- **Relaxation (improper learning):** instead of learning nonconvex L directly, we learn any linear regression matrix M (convex)
  - M need not be an observer, but still competes with the best observer in hindsight



Lemma (Truncation). *With* $H = O(\log T)$: $\mathrm{Reg}_T^{\mathrm{LB}}(L) \le \mathrm{Reg}_T(M_L) + O(1),\ \forall L.$

Outline

Part II: Online learning for prediction

- Formalizing online nonstochastic prediction
- Reduction to online regression
- Online regression with optimism
- Designing hints with bounded errors
- Main result and numerical experiments

Obstacle 2: how to guarantee low regret when observations grow?

Why can't we use standard OLS?

After Idea 1, online regression on the past H outputs

- Can we simply use online least squares (OLS)?
  - at each round t, minimize the cumulative loss so far

$$M_t^{\mathrm{OLS}} = \arg\min_M \ \lambda\|M\|_F^2 + \sum_{s<t} \|Mz_s - y_s\|^2, \qquad \hat{y}_t = M_t^{\mathrm{OLS}} z_t$$

- Precisely, OLS regret scales with its own prediction error. Here, $\|y_t\| = O(t^r)$.

$$\mathrm{Reg}_T(M) = O\big(H \log T \cdot \max_t \|M_t^{\mathrm{OLS}} z_t - y_t\|^2\big)$$

- We cannot guarantee that this prediction error term is bounded. To target polylog regret, can we obtain a regret bound that depends on something different (Idea 2: optimism)?

Idea 2: predictive online least squares (POLS)

Optimism in online learning [Rakhlin, Sridharan, 2013a,b]: at each t (before y_t is revealed), can we incorporate a guess $\tilde{y}_t$ into our objective, so that the regret bound depends on the quality of this guess, i.e., $y_t - \tilde{y}_t$?

- POLS: online least squares with predictive hints
  - Optimistic OLS: one extra predictive term in OLS

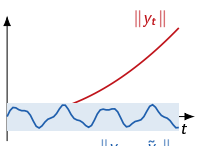
$$M_t^{\mathrm{POLS}} = \arg\min_M \underbrace{\lambda\|M\|_F^2 + \sum_{s<t}\|Mz_s - y_s\|^2}_{\text{standard OLS}} + \underbrace{\|Mz_t - \tilde{y}_t\|^2}_{\text{agree with hint}}, \qquad \hat{y}_t = M_t^{\mathrm{POLS}} z_t$$

  - VAW [Vovk, 2001; Azoury, Warmuth, 2001]: a special case with zero hint $\tilde{y}_t = 0$

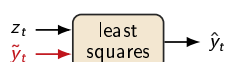
POLS regret and a two-step interpretation

Proposition (Regret of POLS). *For any hints $\tilde{y}_t$, $\mathrm{Reg}_T(M) = O(H \log T \cdot \Delta_{\max}^2)$, where $\Delta_{\max} = \max_{t \le T} \|y_t - \tilde{y}_t\|$ is the largest hint error.*

- A **two-step procedure** for prediction:
  1. **Guess** $\tilde{y}_t$ from the LDS structure (baseline prediction)
  2. **Adapt via POLS**, to compete with the best-in-hindsight observer gain



1. guess: $\|y_t\|$ grows, hint error should not



2. adapt: regress with the hint

Outline

Part II: Online learning for prediction

- Formalizing online nonstochastic prediction
- Reduction to online regression
- Online regression with optimism
- Designing hints with bounded errors
- Main result and numerical experiments

Next: can the hint error stay bounded as y_t grows?

Idea 3: designing hints with bounded errors

Goal: design a hint $\tilde{y}_t$ with $\Delta_{\max} = \max_{t \le T} \|y_t - \tilde{y}_t\| = O(1)$ even when y_t grows

Combined with POLS (Idea 2): $\Delta_{\max} = O(1) \;\Rightarrow\; \mathrm{Reg}_T^{\mathrm{LB}} = O(\log^2 T)$

- **With a model** (A, C): run any stabilizing observer, $\tilde{y}_t = C\hat{x}_t^{\mathrm{LB}}(\tilde{L})$
  - works for **any** marginal A, since the error dynamics are stable:

$$e_{t+1} = (A - \tilde{L}C)\, e_t + w_t - \tilde{L} v_t, \qquad y_t - \tilde{y}_t = C e_t + v_t$$

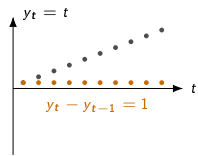
  - stable error dynamics + bounded disturbances $\Rightarrow \Delta_{\max} = O(1)$
- **Without a model:** can we construct hints from the past observations alone?
  - a linear filter of past outputs, $\tilde{y}_t = -\sum_i c_i\, y_{t-i}$
  - for certain classes of systems, the answer is yes

A simple scalar example

Consider the scalar system $x_{t+1} = a x_t + w_t, \quad y_t = x_t \quad (v_t = 0)$, with unknown $a \in \{+1, -1\}$.

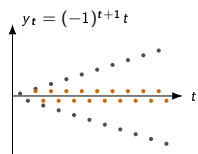
Which will work? $\tilde{y}_t = y_{t-1}$ or $\tilde{y}_t = y_{t-2}$?

- $a = +1$: $w_t = 1 \Rightarrow y_t = t$ grows linearly
  - $\tilde{y}_t = y_{t-1}$: error $= w_{t-1} = 1$, bounded
  - $\tilde{y}_t = y_{t-2}$: error $= w_{t-1} + w_{t-2} = 2$, bounded



$a = +1$, $w_t = 1$

- $a = -1$: $w_t = (-1)^t \Rightarrow y_t = (-1)^{t+1} t$ (the sign flips)
  - $|y_t| = t$ grows linearly
  - $\tilde{y}_t = y_{t-1}$: one flip, error $= \pm(2t - 1)$, grows
  - $\tilde{y}_t = y_{t-2}$: two flips cancel, error $= \pm 2$, bounded



$a = -1$, $w_t = (-1)^t$

$$y_t - y_{t-2} = (a^2 - 1)\, x_{t-2} + w_{t-1} + a\, w_{t-2}, \quad \text{where} \quad a^2 - 1 = 0 \text{ for } a = \pm 1$$

Model-free hint design: 2-lag hint

- **Unknown** (A, C): the hint error decomposition

$$y_t - y_{t-2} = \underbrace{v_t - v_{t-2}}_{\text{noise}} + \underbrace{C(w_{t-1} + A\,w_{t-2})}_{\text{last 2 steps}} + \underbrace{C\sum_{s\geq 0} A^s(A^2 - I)\,w_{t-3-s}}_{= C(A^2-I)\,x_{t-2}}$$

> Lemma (Two-step cancellation). *If $A = S\Lambda S^{-1}$ has real spectrum in $[-1, 1]$,*
> $$\sum_{s\geq 0} \left\|A^s(A^2 - I)\right\| \leq 2n\kappa_A\,, \qquad \kappa_A = \|S\|\|S^{-1}\|.$$

- The factor $(A^2 - I)$ removes the marginal modes and makes $\sum_s \|(A^2 - I)A^s\|$ summable (though $\sum_s \|A^s\|$ diverges when $\rho(A) = 1$)

> Theorem (2-lag hint error). *Without knowing (A, C),*
> $$\max_{t\leq T} \|y_t - y_{t-2}\| \leq 2C_v + \left(1 + \kappa_A + 2n\kappa_A\right)\|C\|C_w.$$

Outline

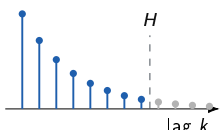
Part II: Online learning for prediction

- Formalizing online nonstochastic prediction
- Reduction to online regression
- Online regression with optimism
- Designing hints with bounded errors
- Main result and numerical experiments

Putting it together: key ideas of FM-POLS

1. Truncation

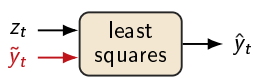
gains decay: regress on $H = O(\log T)$ past outputs



handles infinite memory, nonconvexity

2. Optimism (POLS)

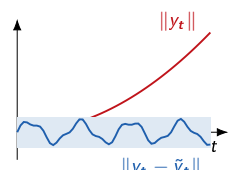
regret depends on the hint error, not on $\|y_t\|$



reduces growth to the hint error

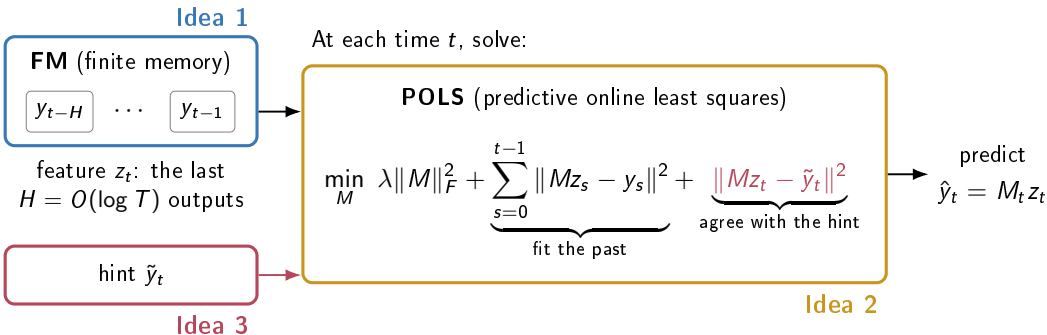
3. Hint design

hint error stays bounded as y_t grows



bounds the hint error, with or without a model

Our algorithm: FM-POLS



- **Streaming data and algorithm**: one rank-one update per round, as in recursive least squares

Main result

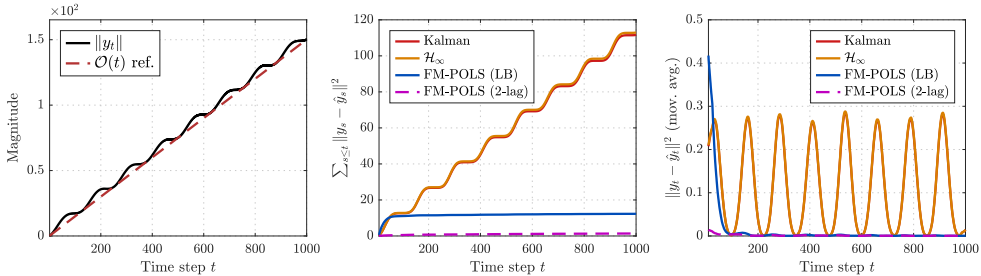
Theorem (Polylogarithmic regret of FM-POLS). *Run POLS on $z_t = [y_{t-1}; \dots; y_{t-H}]$ with $H = O(\log T)$ and hints $\tilde{y}_t$. For every $L \in \mathcal{L}_{\kappa,\gamma}$,*

$$\mathrm{Reg}_T^{\mathrm{LB}}(L) \le \underbrace{\mathrm{Reg}_T(M_L)}_{O(H \log T \cdot \Delta_{\max}^2)\ \text{(Idea 2)}} + \underbrace{O(1)}_{\text{Idea 1}}.$$

With the hints of Idea 3, $\Delta_{\max} = O(1)$: $\mathrm{Reg}_T^{\mathrm{LB}} = O(\log^2 T)$ on every bounded disturbance sequence, even as y_t grows.

- One $\log T$ from the memory length H (Idea 1), one from online least squares (Idea 2), and Idea 3 keeps the hint error bounded

Learning versus fixed-gain filtering



(a) outputs $\|y_t\|$ (b) cumulative loss (c) per-step loss

Settings: symmetric system (± 1 eigenvalues), bias-plus-sinusoid disturbances.

- Fixed-gain filters keep a persistent periodic error; FM-POLS's per-step loss quickly decays to near zero.
- The model-based hint is not necessarily better than the model-free 2-lag hint. (system dependent)

Outline

Part I: Policy optimization for control

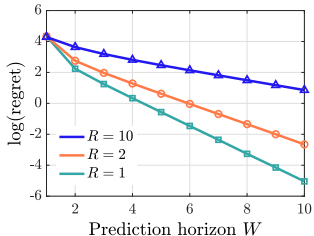
Part II: Online learning for prediction

Part III: Learning and duality in control

Conclusion and future directions

Dynamic regret analysis of data-driven predictive control

- **Task:** track a bounded reference with a W-step preview; unknown nonlinear dynamics with a Koopman linear embedding
- **Controller:** data-driven predictive control from offline data (Willems' lemma); no model or lifting identified
- **Benchmark:** the clairvoyant controller, $\mathrm{Reg}_T = J_T(\pi) - J_T^\star$



log regret vs. preview W

Theorem (Dynamic regret). *If* $W \geq \Delta_{\mathrm{stab}}$: $\mathrm{Reg}_T = O\big(W^2 \lambda_\infty^{2W}\, T\big)$, *with* $\lambda_\infty < 1$; *so* $W = \Theta(\log T) \Rightarrow \mathrm{Reg}_T = O(1)$.

Paper: **C.-F. Pai**, X. Shang, J. Qian, Y. Zheng. Online tracking with predictions for nonlinear systems with Koopman linear embedding. *L4DC*, PMLR 331:1575–1600, 2026.

Benign nonconvexity and duality in control

- **Benign nonconvex landscapes, Part I: global optimality.** Certify global optimality of policy optimization for output-feedback LQG and $\mathcal{H}_\infty$ control.
- **Part II: extended convex lifting (ECL).** The lifting framework behind Part I of this talk: hidden convexity via a lifted set and a diffeomorphism.
- **ECL for policy optimization** (L4DC 2025): a short version of the ECL framework and its applications.
- **LQ differential games:** semidefinite programming duality for infinite-horizon LQ games.

Papers:

Y. Zheng, **C.-F. Pai**, Y. Tang. Benign nonconvex landscapes in optimal and robust control, Part I: Global optimality, 71(8):5334–5349; Part II: Extended convex lifting, 71(8):5350–5365. *IEEE TAC*, 2026.
Y. Zheng, **C.-F. Pai**, Y. Tang. Extended convex lifting for policy optimization of optimal and robust control. *L4DC*, PMLR 283:392–404, 2025.
Y. Watanabe, **C.-F. Pai**, Y. Zheng. Semidefinite programming duality in infinite-horizon LQ differential games. *CDC*, 3964–3970, 2025.

Outline

Part I: Policy optimization for control

Part II: Online learning for prediction

Part III: Learning and duality in control

Conclusion and future directions

Conclusion

	Part I: control	**Part II: prediction**
Setting	mixed $\mathcal{H}_2/\mathcal{H}_\infty$; system and disturbance model known	arbitrary disturbances, system model known or unknown
Challenge	nonconvex in the gain K	nonconvex in L, infinite memory, growing observations
Approach	hidden convexity (ECL)	finite-memory regression + optimism + predictive hint design
Result	no spurious stationary points; fast policy iteration	$\mathrm{Reg}_T^{\mathrm{LB}} = O(\log^2 T)$ vs. the best observer in hindsight

Both parts: nonconvex gain-parameterized policies.
Part I: reveal hidden convexity; Part II: convex relaxation + optimism

These results show how exploiting the structure of dynamical systems can lead to strong guarantees.

Open directions

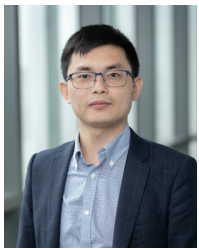
- **Mixed $\mathcal{H}_2/\mathcal{H}_\infty$.** Convergence of policy iteration; dynamic output feedback; connection to LQ games and risk-averse control; model-free extension.
- **Online prediction.** Model-free hints for complex spectrum; lower bounds; time-varying/structured dynamics; stronger comparators; more broadly, next-token prediction.
- **Predictive control.** Closed-loop analysis of MPC and model-based RL; regret under *approximate* Koopman embeddings.

Publications

- **C.-F. Pai**, Y. Zheng. Online nonstochastic prediction: logarithmic regret via predictive online least squares. arXiv:2605.04364, 2026.
- **C.-F. Pai**, Y. Watanabe, Y. Tang, Y. Zheng. Policy optimization of mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: benign nonconvexity and global optimality. *Automatica*, provisionally accepted. arXiv:2603.04843.
- **C.-F. Pai**, X. Shang, J. Qian, Y. Zheng. Online tracking with predictions for nonlinear systems with Koopman linear embedding. *L4DC*, PMLR 331:1575–1600, 2026.
- Y. Zheng, **C.-F. Pai**, Y. Tang. Benign nonconvex landscapes in optimal and robust control, Part I: Global optimality. *IEEE TAC*, 71(8):5334–5349, 2026.
- Y. Zheng, **C.-F. Pai**, Y. Tang. Benign nonconvex landscapes in optimal and robust control, Part II: Extended convex lifting. *IEEE TAC*, 71(8):5350–5365, 2026.
- Y. Watanabe, **C.-F. Pai**, Y. Zheng. Semidefinite programming duality in infinite-horizon linear quadratic differential games. *CDC*, 3964–3970, 2025.
- Y. Zheng, **C.-F. Pai**, Y. Tang. Extended convex lifting for policy optimization of optimal and robust control. *L4DC*, PMLR 283:392–404, 2025.

Acknowledgments

Advisor



Yang Zheng

Committee



Massimo Franceschetti



Sicun Gao



Jorge Poveda

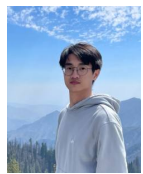
Thank you for your guidance, feedback, and time.

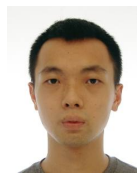
Supported by the J. Yang Scholarship, the UC San Diego Graduate Scholarly Impact Award, and NSF ECCS-2154650 and NSF CAREER 2340713.

Acknowledgments

Collaborators and labmates







Yujie Tang Yuto Watanabe Xu Shang Jiachen Qian Feng-Yi Liao

... and my friends and family, for their love and support.

Thank you for your attention!

Questions?

References: Part I

[1] T. Başar and P. Bernhard. *$\mathcal{H}_\infty$-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Springer, 2008.

[2] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *ICML*, 2018.

[3] X. Guo and B. Hu. Global convergence of direct policy search for state-feedback $\mathcal{H}_\infty$ robust control: A revisit of nonsmooth synthesis with Goldstein subdifferential. In *NeurIPS*, 2022.

[4] P. P. Khargonekar and M. A. Rotea. Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control: a convex optimization approach. *IEEE Trans. Autom. Control*, 36(7):824–837, 1991.

[5] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović. Convergence and sample complexity of gradient methods for the model-free linear–quadratic regulator problem. *IEEE Trans. Autom. Control*, 67(5):2435–2450, 2022.

[6] Y. Tang, Y. Zheng, and N. Li. Analysis of the optimization landscape of linear quadratic Gaussian (LQG) control. In *L4DC*, 2021.

[7] J. Umenberger, M. Simchowitz, J. Perdomo, K. Zhang, and R. Tedrake. Globally convergent policy search for output estimation. In *NeurIPS*, 2022.

[8] K. Zhang, Z. Yang, and T. Başar. Policy optimization provably converges to Nash equilibria in zero-sum linear quadratic games. In *NeurIPS*, 2019.

[9] 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 J. Control Optim.*, 59(6):4081–4109, 2021.

[10] Y. Zheng, C.-F. Pai, and Y. Tang. Benign nonconvex landscapes in optimal and robust control, Part II: Extended convex lifting. *IEEE Trans. Autom. Control*, 71(8):5350–5365, 2026.

References: Part II

[11] O. Anava, E. Hazan, S. Mannor, and O. Shamir. Online learning for time series prediction. In *COLT*, 2013.

[12] K. S. Azoury and M. K. Warmuth. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Mach. Learn.*, 43(3):211–246, 2001.

[13] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge Univ. Press, 2006.

[14] U. Ghai, H. Lee, K. Singh, C. Zhang, and Y. Zhang. No-regret prediction in marginally stable systems. In *COLT*, 2020.

[15] B. Hassibi, A. H. Sayed, and T. Kailath. *Indefinite-Quadratic Estimation and Control: A Unified Approach to $\mathcal{H}^2$ and $\mathcal{H}^\infty$ Theories*. SIAM, 1999.

[16] E. Hazan, K. Singh, and C. Zhang. Learning linear dynamical systems via spectral filtering. In *NeurIPS*, 2017.

[17] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82(1):35–45, 1960.

[18] M. Kozdoba, J. Marecek, T. Tchrakian, and S. Mannor. On-line learning of linear dynamical systems: Exponential forgetting in Kalman filters. In *AAAI*, 2019.

[19] D. G. Luenberger. Observing the state of a linear system. *IEEE Trans. Mil. Electron.*, 8(2):74–80, 1964.

[20] J. Qian and Y. Zheng. Logarithmic regret and polynomial scaling in online multi-step-ahead prediction. arXiv:2511.12467, 2025.

[21] A. Rakhlin and K. Sridharan. Online learning with predictable sequences. In *COLT*, 2013.

[22] A. Rakhlin and K. Sridharan. Optimization, learning, and games with predictable sequences. In *NeurIPS*, 2013.

[23] A. Tsiamis and G. J. Pappas. Online learning of the Kalman filter with logarithmic regret. *IEEE Trans. Autom. Control*, 2023.

[24] V. Vovk. Competitive on-line statistics. *Int. Stat. Rev.*, 69(2):213–248, 2001.