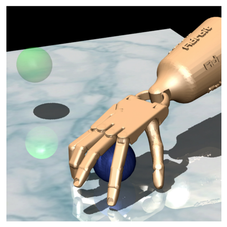
Policy Gradient Descent for Control: Global Optimality via Convex Parameterization

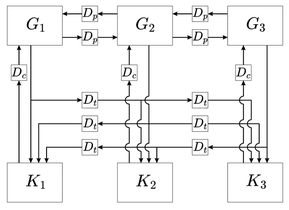
Yue Sun, Maryam Fazel

University of Washington, Seattle

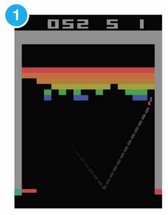
Control & ML



Robot



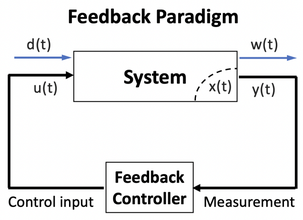
Decentralized control

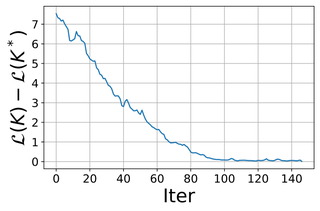


Game

Joint interest in **Control and ML(Reinforcement Learning)**
ML: Policy gradient descent on nonconvex objectives
Control theory: Reparameterize the problem to obtain a convex objective
Question: If the optimal control problem can be made convex, does policy gradient descent converge to global optimum?

Feedback control & Policy gradient





Left: The controller observes the measurement y and send feedback control u. The goal is related to the transfer function (e.g., minimizing the linear quadratic cost, $\mathcal{H}_2, \mathcal{H}_\infty$ norms) from d to w.

Right: Zeroth order policy gradient descent for minimizing $\mathcal{L}_2$ gain. The cost function is nonconvex in policy and the iterates converge to a global minimum.

Summary

- **Generality:** For a family of optimal control problems that can be convexified, policy gradient descent converges to the **global** optimum despite the nonconvexity.
- **Connection between domains:** We propose a concise proof that bridges the nonconvex landscape with the convex parameterizations.

Y. Sun and M. Fazel, "Learning Optimal Controllers by Policy Gradient: Global Optimality via Convex Parameterization," 2021 60th IEEE Conference on Decision and Control (CDC).

Continuous time linear quadratic regulator

- Cost:

$$\mathcal{L}(u(t)) := \mathbf{E}_{x_0 \sim \mathcal{N}(0,\Sigma)} \int_0^\infty (x(t)^\top Q x(t) + u(t)^\top R u(t)) dt$$

Learn u that minimizes the loss.

- Optimal controller: state feedback by Riccati equations.

Solve P_R: $AP_R + P_R A^\top + Q - P_R B R^{-1} B P_R = 0$,
Solve K^*: $u = K^* x = -R^{-1} B^\top P_R x$.

Learn K^*.

- Convex formulation

$$\min_{Z,L,P} \; f(L,P,Z) := \mathbf{Tr}(QP) + \mathbf{Tr}(ZR)$$
$$\text{s.t., } \mathcal{A}(P) + \mathcal{B}(L) + \Sigma = 0, \begin{bmatrix} Z & L^\top \\ L & P \end{bmatrix} \succeq 0$$

$\mathcal{A}(P) = AP + (AP)^\top$, $\mathcal{B}(L) = BL + (BL)^\top$. And $K^* = L^* P^{*-1}$.

Main theorem

$$\min_K \; \mathcal{L}(K), \quad \text{s.t., } K \text{ stabilizes} \quad \Rightarrow \quad \min_{Z,L,P} \; f(L,P,Z), \quad \text{s.t., } (L,P,Z) \in \mathcal{S}$$

Assumptions:

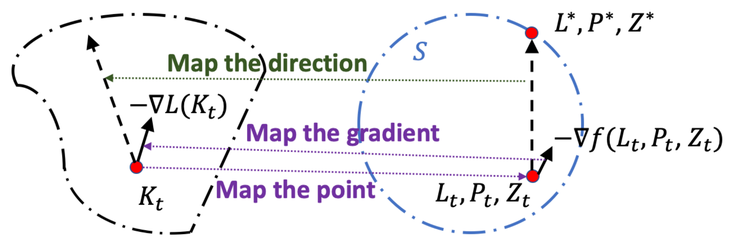
- $\mathcal{S}$ is convex. $f(L,P,Z)$ is convex on $\mathcal{S}$.
- We can express $\mathcal{L}(K)$ as:

$$\mathcal{L}(K) = \min_{L,P,Z} \; f(L,P,Z), \quad \text{s.t.,} (L,P,Z) \in \mathcal{S}, \; K = LP^{-1}.$$

Theorem: With the assumptions, we have $\nabla \mathcal{L}(K) = 0 \;\Leftrightarrow\; K = K^*$.

Policy gradient descent converges to a global optimum.

Interaction between spaces



Gradient dominance of nonconvex cost $\Leftarrow$
Gradient dominance of convex cost
+ Diffeomorphism between two spaces

Discrete time Markov jump system

- Dynamics: $x(t+1) = A_{w(t)} x(t) + B_{w(t)} u(t), \; w(t) \in \{1, ..., N\}$.
$N = 1$: Discrete time linear system
- Probabilistic model for transition

$$\mathbf{Pr}(w(t+1) = j | w(t) = i) = \rho_{ij} \in [0,1], \; \forall t \geq 0.$$

- Cost: let $K = [K_1, ..., K_N]$,

$$\min_K \; \mathcal{L}(K) := \mathbf{E}_{w,x_0} \sum_{t=0}^\infty x(t)^\top Q x(t) + u(t)^\top R u(t), \; u(t) = K_{w(t)} x(t).$$

- Convex formulation: Initial distribution $\mathbf{Pr}(w(0) = i) = p_i$,

$$\min \; \mathbf{Tr}(QX_0) + \mathbf{Tr}(Z_0 R),$$
$$\text{s.t. } X_0 = \sum_{i=1}^N X_i, \; Z_0 = \sum_{i=1}^N Z_i, \; \begin{bmatrix} Z_i & L_i \\ L_i^\top & X_i \end{bmatrix} \succeq 0,$$
$$X_i - p_i \Sigma = \sum_{j=1}^N U_{ji}, \; \begin{bmatrix} \rho_{ji}^{-1} U_{ji} & A_j X_j + B_j L_j \\ (A_j X_j + B_j L_j)^\top & X_j \end{bmatrix} \succeq 0.$$

Minimizing $\mathcal{L}_2$ gain

- Dynamics

$$\dot{x}(t) = Ax(t) + Bu(t) + B_w w(t), \; y = Cx(t) + Du(t)$$

x is state, u is input, w is a perturbation. Find the optimal state feedback controller $u(t) = K^* x(t)$.

- Cost: $\mathcal{L}(K) := \sup_{\|w\|_2 = 1} \|y\|_2$.
- Convex formulation

$$\min_{L,P,\gamma} \; f(L,P,\gamma) := \gamma, \; \text{s.t., } \gamma \geq 0,$$
$$\begin{bmatrix} AP + PA^\top + BL + L^\top B^\top + B_w B_w^\top & (CP + DL)^\top \\ CP + DL & -\gamma^2 I \end{bmatrix} \preceq 0.$$

Finite horizon time varying discrete time LQR

- Dynamics: $x(t+1) = A(t)x(t) + B(t)u(t) + w(t)$.
Controller: $u(t) = \sum_{i=0}^t K(t, t-i) x(i)$, $\mathcal{K}$ stacks K.
- Cost: $\min_{\mathcal{K}} \; \mathcal{L}(\mathcal{K}) := \sum_{t=0}^T x(t)^\top Q(t) x(t) + u(t)^\top R(t) u(t)$.
- Convex formulation: Let Z be the constant shifting matrix,

$$\min_{\Phi_X, \Phi_U} \; f(\Phi_X, \Phi_U) = \left\| \mathrm{diag}(\mathcal{Q}^{1/2}, \mathcal{R}^{1/2}) \begin{bmatrix} \Phi_X \\ \Phi_U \end{bmatrix} \Sigma^{1/2} \right\|_F^2,$$
$$\text{s.t., } \begin{bmatrix} I - Z\mathcal{A} & -Z\mathcal{B} \end{bmatrix} \begin{bmatrix} \Phi_X \\ \Phi_U \end{bmatrix} = I.$$