Convex Parameterization for Linear Dynamical Systems

Sham Kakade and Wen Sun

CS 6789: Foundations of Reinforcement Learning

Recap

LQR:

minimize $E\left[x_H^\top Q x_H + \sum_{t=0}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t)\right]$

such that $x_{t+1} = Ax_t + Bu_t + w_t\,, \quad x_0 \sim D,\ w_t \sim N(0,\sigma^2 I)\,,$

Recap

LQR:

minimize $E\left[x_H^\top Q x_H + \sum_{t=0}^{H-1}(x_t^\top Q x_t + u_t^\top R u_t)\right]$

such that $x_{t+1} = Ax_t + Bu_t + w_t, \quad x_0 \sim D,\ w_t \sim N(0,\sigma^2 I),$

Optimal control and Ricatii equation:

Set $P_H = Q,\ P_t = A^\top P_{t+1} A + Q - A^\top P_{t+1} B (B^\top P_{t+1} B + R)^{-1} B^\top P_{t+1} A$

Recap

LQR:

minimize $E\left[x_H^\top Q x_H + \sum_{t=0}^{H-1}(x_t^\top Q x_t + u_t^\top R u_t)\right]$

such that $x_{t+1} = Ax_t + Bu_t + w_t, \quad x_0 \sim D,\ w_t \sim N(0, \sigma^2 I),$

Optimal control and Ricatii equation:

Set $P_H = Q,\ P_t = A^\top P_{t+1} A + Q - A^\top P_{t+1} B (B^\top P_{t+1} B + R)^{-1} B^\top P_{t+1} A$

$\pi^\star(x_t) = -K_t^\star x_t$ where $K_t^\star = (B^\top P_{t+1} B + R)^{-1} B^\top P_{t+1} A$

Recap

LQR:

minimize $E\left[x_H^\top Q x_H + \sum_{t=0}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t)\right]$

such that $x_{t+1} = Ax_t + Bu_t + w_t\,, \quad x_0 \sim D,\ w_t \sim N(0,\sigma^2 I)\,,$

Optimal control and Ricatii equation:

Set $P_H = Q,\ P_t = A^\top P_{t+1} A + Q - A^\top P_{t+1} B (B^\top P_{t+1} B + R)^{-1} B^\top P_{t+1} A$

$\pi^\star(x_t) = -K_t^\star x_t$ where $K_t^\star = (B^\top P_{t+1} B + R)^{-1} B^\top P_{t+1} A$

$$V_t^\star(x) = x^\top P_t x + \sigma^2 \sum_{h=t+1}^{H} \mathrm{Trace}(P_h)$$

Today’s main Question:

What if the cost function is arbitrary convex function, and how to handle adversarial noise?

Setting: Linear Dynamical System and Convex Cost functions

$$x_{t+1} = Ax_t + Bu_t + w_t, x_0 \sim \mathcal{D}, w_t \sim \mathcal{N}(0, \sigma^2 I)$$

Setting: Linear Dynamical System and Convex Cost functions

$$x_{t+1} = Ax_t + Bu_t + w_t, x_0 \sim \mathscr{D}, w_t \sim \mathcal{N}(0,\sigma^2 I)$$

$$\text{min} \quad \cancel{\max} \, \mathbb{E}\left[\sum_{h=0}^{H-1} c(x_h, a_h) \middle| \pi\right]$$

Convex Function!

Setting: Linear Dynamical System and Convex Cost functions

$$x_{t+1} = Ax_t + Bu_t + w_t, x_0 \sim \mathcal{D}, w_t \sim \mathcal{N}(0,\sigma^2 I)$$

$$\max \mathbb{E}\left[\sum_{h=0}^{H-1} c(x_h, a_h) \middle| \pi\right]$$

Convex Function!

We cannot guarantee the optimal policy is linear anymore…

Setting: Linear Dynamical System and Convex Cost functions

$$x_{t+1} = Ax_t + Bu_t + w_t, x_0 \sim \mathcal{D}, w_t \sim \mathcal{N}(0,\sigma^2 I)$$

$$\max \mathbb{E}\left[\sum_{h=0}^{H-1} c(x_h, a_h) \middle| \pi\right]$$

Convex Function!

We cannot guarantee the optimal policy is linear anymore…

But we consider a restricted policy class

$$\Pi = \{\pi = \{-K_t\}_{t=0}^{H-1} : K_t \in \mathcal{K}, \forall t\}$$

Setting: Linear Dynamical System and Convex Cost functions

$$x_{t+1} = Ax_t + Bu_t + w_t, x_0 \sim \mathcal{D}, w_t \sim \mathcal{N}(0,\sigma^2 I)$$

$$\max \mathbb{E}\left[\sum_{h=0}^{H-1} c(x_h, a_h) \,\middle|\, \pi\right]$$

Convex Function!

We cannot guarantee the optimal policy is linear anymore…

But we consider a restricted policy class

$$\Pi = \{\pi = \{-K_t\}_{t=0}^{H-1} : K_t \in \mathcal{K}, \forall t\}$$

Goal: find the best linear controllers: $\{-K_t^\star\}_{t=0}^{H-1} := \arg\min_{\pi\in\Pi} \mathbb{E}\left[\sum_{t=0}^{H-1} c(x_t, u_t) \,|\, \pi\right]$

A New Convex Parameterization of Controllers

Unfortunately, even with quadratic cost, $\max_{\pi} \mathbb{E}\left[\sum_{h=0}^{H-1} c(x_h, a_h) \,\middle|\, \pi\right]$ is not convex with respect to $\pi := \{-K_t\}_{t=0}^{H-1}$

A New Convex Parameterization of Controllers

Unfortunately, even with quadratic cost, $\max_{\pi} \mathbb{E}\left[\sum_{h=0}^{H-1} c(x_h, a_h) \,\middle|\, \pi\right]$ is not convex with respect to $\pi := \{-K_t\}_{t=0}^{H-1}$

Let's search for a different parameterization that is equivalent to this $u = -K_t x$ parameterization

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

Assume we roll out π and the system is at x_t, we can compute all previous noises:

$$w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}, w_{t-2} = x_{t-1} - Ax_{t-2} - Bu_{t-2}, \ldots$$

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

Assume we roll out π and the system is at x_t, we can compute all previous noises:

$$w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}, w_{t-2} = x_{t-1} - Ax_{t-2} - Bu_{t-2}, \ldots$$

$u_t = -K_t x_t$ $\Leftarrow$ $u_t = M_0 x_0 + M_1 w_0 + M_2 w_1 + \cdots M_t w_{t-1}$

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

Assume we roll out π and the system is at x_t, we can compute all previous noises:

$$w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}, w_{t-2} = x_{t-1} - Ax_{t-2} - Bu_{t-2}, \ldots$$

$$u_t = -K_t x_t$$

$$= -K_t w_{t-1} - K_t(A - BK_{t-1})x_{t-1}$$

$$x_t = A x_{t-1} + B \underline{u_{t-1}} + w_{t-1}$$

$$= \underline{A x_{t-1} - B K_{t-1} x_{t-1}}$$

$$x_{t-1} = A x_{t-2} - B K_{t-2} x_{t-2} + w_{t-2}$$

$$= (A - B K_{t-2}) x_{t-2}$$

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

Assume we roll out π and the system is at x_t, we can compute all previous noises:

$$w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}, w_{t-2} = x_{t-1} - Ax_{t-2} - Bu_{t-2}, \ldots$$

$$u_t = -K_t x_t$$

$$= -K_t w_{t-1} - K_t(A - BK_{t-1})x_{t-1}$$

$$= -K_t w_{t-1} - K_t(A - BK_{t-1})(\underbrace{Ax_{t-2} - BK_{t-2}x_{t-2} + w_{t-2}}_{:= x_{t-1}})$$

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

Assume we roll out π and the system is at x_t, we can compute all previous noises:

$$w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}, w_{t-2} = x_{t-1} - Ax_{t-2} - Bu_{t-2}, \dots$$

$$u_t = -K_t x_t$$

$$= -K_t w_{t-1} - K_t(A - BK_{t-1})x_{t-1}$$

$$= -K_t w_{t-1} - K_t(A - BK_{t-1})(Ax_{t-2} - BK_{t-2}x_{t-2} + w_{t-2})$$

$$= \underbrace{-K_t}_{:=M_{t-1;t}} w_{t-1} \underbrace{-K_t(A - BK_{t-1})}_{:=M_{t-2;t}} w_{t-2} \underbrace{-K_t(A - BK_{t-1})(A - BK_{t-2})}_{:=M_{t-3;t}} x_{t-2}$$

$$(A - BK_{t-3})x_{t-3} + w_{t-3}$$

$$\Rightarrow M_{t-3;t} \cdot w_{t-3} + \dots$$

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

Assume we roll out π and the system is at x_t, we can compute all previous noises:

$$w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}, w_{t-2} = x_{t-1} - Ax_{t-2} - Bu_{t-2}, \ldots$$

$$u_t = - K_t x_t$$

$$= - K_t w_{t-1} - K_t(A - BK_{t-1})x_{t-1}$$

$$= - K_t w_{t-1} - K_t(A - BK_{t-1})(Ax_{t-2} - BK_{t-2}x_{t-2} + w_{t-2})$$

$$= \underbrace{-K_t}_{:=M_{t-1;t}} w_{t-1} \underbrace{-K_t(A - BK_{t-1})}_{:=M_{t-2;t}} w_{t-2} \underbrace{-K_t(A - BK_{t-1})(A - BK_{t-2})}_{:=M_{t-3;t}} x_{t-2}$$

$$= \underbrace{\left[-K_t\left(\prod_{\tau=1}^{t}(A - BK_{t-\tau})\right)\right]}_{M_t} \underset{1}{x_0} + \sum_{\tau=0}^{t-1} \underbrace{\left[-K_t \prod_{h=1}^{t-1-\tau}(A - BK_{t-h})\right]}_{M_{\tau;t}} w_\tau$$

A New Convex Parameterization of Controllers

Let us consider an arbitrary linear controller $\pi := \{-K_t\}_{t=0}^{H-1}$

Assume we roll out π and the system is at x_t, we can compute all previous noises:

$$w_{t-1} = x_t - Ax_{t-1} - Bu_{t-1}, w_{t-2} = x_{t-1} - Ax_{t-2} - Bu_{t-2}, \ldots$$

$$u_t = -K_t x_t \quad \Longleftrightarrow \quad u_t = M_t x_0 + M_{0;t} w_0 + M_{1;t} w_1 + \ldots + M_{t-1;t} w_{t-1}$$

$$= -K_t w_{t-1} - K_t(A - BK_{t-1})x_{t-1}$$

$$= -K_t w_{t-1} - K_t(A - BK_{t-1})(Ax_{t-2} - BK_{t-2}x_{t-2} + w_{t-2})$$

$$= \underbrace{-K_t}_{:=M_{t-1;t}} w_{t-1} \underbrace{-K_t(A - BK_{t-1})}_{:=M_{t-2;t}} w_{t-2} \underbrace{-K_t(A - BK_{t-1})(A - BK_{t-2})}_{:=M_{t-3;t}} x_{t-2}$$

$$= \underbrace{\left[-K_t\left(\prod_{\tau=1}^{t}(A - BK_{t-\tau})\right)\right]}_{M_t} x_0 + \sum_{\tau=0}^{t-1} \underbrace{\left[-K_t \prod_{h=1}^{t-1-\tau}(A - BK_{t-h})\right]}_{M_{\tau;t}} w_\tau$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

Computing u_t

$$u_t = M_t \cdot x_0 + M_{t-1;t} \cdot w_{t-1} + \cdots M_{0;t} \cdot w_0$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

x_0

x_0

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

$$x_0 \longrightarrow u_0 = -K_0 x_0$$

$$x_0 \longrightarrow u_0 = \underset{-}{M_0}\underset{/\!/}{x_0}$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

$$x_0 \longrightarrow u_0 = -K_0 x_0 \xrightarrow{w_0} x_1$$

$$x_0 \longrightarrow u_0 = M_0 x_0 \xrightarrow{w_0} x_1, w_0$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\Big\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\Big\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

$$x_0 \longrightarrow u_0 = -K_0 x_0 \longrightarrow x_1 \longrightarrow u_1 = -K_1 x_1$$

$$x_0 \longrightarrow u_0 = M_0 x_0 \longrightarrow x_1, w_0$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

$$x_0 \longrightarrow u_0 = -K_0 x_0 \longrightarrow x_1 \longrightarrow u_1 = -K_1 x_1$$

$$\begin{tikzcd}
x_0 \arrow[r] \arrow[rrr, bend left] & u_0 = M_0 x_0 \arrow[r] & x_1, w_0 \arrow[r] & u_1 = M_1 x_0 + M_{0;1} w_0
\end{tikzcd}$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

$$x_0 \longrightarrow u_0 = -K_0x_0 \longrightarrow x_1 \longrightarrow u_1 = -K_1x_1 \longrightarrow x_2$$

$$\begin{tikzcd}
x_0 \arrow[r] \arrow[rrr, bend left] & u_0 = M_0x_0 \arrow[r] & x_1, w_0 \arrow[r] & u_1 = M_1x_0 + M_{0;1}w_0 \arrow[r, "w_1"] & x_2, w_1
\end{tikzcd}$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

$$\begin{tikzcd} x_0 \arrow[r] & u_0 = -K_0x_0 \arrow[r] & x_1 \arrow[r] & u_1 = -K_1x_1 \arrow[r] & x_2 \arrow[r] & u_2 = -K_2x_2 \end{tikzcd}$$

$$\begin{tikzcd} x_0 \arrow[r] \arrow[rrr, bend left] & u_0 = M_0x_0 \arrow[r] & x_1, w_0 \arrow[r] & u_1 = M_1x_0 + M_{0;1}w_0 \arrow[r] & x_2, w_1 \end{tikzcd}$$

A New Convex Parameterization of Controllers

[Claim] For any linear controllers $\pi := \{-K_t\}_{t=0}^{H-1}$, there exists a parameterization $\left\{\{M_t, M_{t-1;t}, \ldots M_{0;t}\}\right\}_{t=0}^{H-1}$, that generates the same sequence trajectory, given any fixed x_0, and fixed noise $w_0, \ldots, w_{H-1}$.

$$\begin{tikzcd}
x_0 \arrow[r] & u_0 = -K_0 x_0 \arrow[r] & x_1 \arrow[r] & u_1 = -K_1 x_1 \arrow[r] & x_2 \arrow[r] & u_2 = -K_2 x_2
\end{tikzcd}$$

$$\begin{tikzcd}
x_0 \arrow[r] \arrow[rrr, bend left] \arrow[rrrrr, bend right] & u_0 = M_0 x_0 \arrow[r] & \cancel{x_1}, w_0 \arrow[r] \arrow[rrr, bend right] & u_1 = M_1 x_0 + M_{0;1} w_0 \arrow[r] & \cancel{x_2}, w_1 \arrow[r] & u_2 = M_2 x_0 + M_{0;2} w_0 + M_{1;2} w_1
\end{tikzcd}$$

A New Convex Parameterization of Controllers

[Claim] Given controller $\widetilde{\pi} := \left\{ \{ M_t, M_{0;t}, \ldots M_{t-1;t} \} \right\}_{t=0}^{H-1}$, u_t & x_t are all linear with respect to the parameters, $\forall t$

A New Convex Parameterization of Controllers

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$**,** u_t & x_t **are all linear with respect to the parameters,** $\forall t$

$$x_0, \underbrace{u_0 = M_0 x_0}_{\text{linear wrt } M}$$

A New Convex Parameterization of Controllers

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$, u_t & x_t are all linear with respect to the parameters, $\forall t$

$$x_0, u_0 = M_0 x_0$$

$$x_1 = A x_0 + \underbrace{B M_0 x_0}_{u_0} + w_0$$

A New Convex Parameterization of Controllers

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$, u_t & x_t are all linear with respect to the parameters, $\forall t$

$$x_0, u_0 = M_0 x_0$$

$$x_1 = A x_0 + B M_0 x_0$$

$$u_1 = \underline{M_1 x_0 + M_{0;1} w_0}$$

linear wrt M_s

A New Convex Parameterization of Controllers

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$, u_t & x_t are all linear with respect to the parameters, $\forall t$

$$x_0, u_0 = M_0 x_0$$

$$x_1 = Ax_0 + BM_0 x_0$$

$$u_1 = M_1 x_0 + M_{0;1} w_0$$

$$x_2 = \underbrace{Ax_1}_{A} + \underbrace{BM_1 x_0 + BM_{0;1}}_{B \cdot u_1} w_0 + w_1$$

A New Convex Parameterization of Controllers

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$, u_t & x_t are all linear with respect to the parameters, $\forall t$

$$x_0, u_0 = M_0 x_0$$

$$x_1 = A x_0 + B M_0 x_0$$

$$u_1 = M_1 x_0 + M_{0;1} w_0$$

$$x_2 = A x_1 + B M_1 x_0 + B M_{0;1} w_0$$

$\{x, u\}$ are linear wrt these M matrices

A New Convex Parameterization of Controllers

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$, u_t & x_t are all linear with respect to the parameters, $\forall t$

$$x_0, u_0 = M_0 x_0$$

$$x_1 = Ax_0 + BM_0 x_0$$

$$u_1 = M_1 x_0 + M_{0;1} w_0$$

$$x_2 = Ax_1 + BM_1 x_0 + BM_{0;1} w_0$$

$\{x, u\}$ are linear wrt these M matrices

linear wrt Ms

linear wrt Ms

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$, $\sum_{t=0}^{H-1} c(x_t, u_t)$ is convex with respect the parameters, $\forall t$

Convexity and Gradient Descent

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$**,**
$\mathbb{E}\left[\sum_{t=0}^{H-1} c(x_t, u_t) \,|\, \pi \right]$ **is convex with respect the parameters,** $\forall t$

Convexity allows to perform Gradient Descent directly on parameters

$$\left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$$

Convexity and Gradient Descent

[Claim] Given controller $\widetilde{\pi} := \left\{ \{ M_t, M_{0;t}, \ldots M_{t-1;t} \} \right\}_{t=0}^{H-1}$**,**

$\mathbb{E}\left[\sum_{t=0}^{H-1} c(x_t, u_t) \,|\, \pi \right]$ **is convex with respect the parameters,** $\forall t$

Convexity allows to perform Gradient Descent directly on parameters

$$\left\{ \{ M_t, M_{0;t}, \ldots M_{t-1;t} \} \right\}_{t=0}^{H-1}$$

Algorithm:

Convexity and Gradient Descent

[Claim] Given controller $\widetilde{\pi} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$,

$\mathbb{E}\left[\sum_{t=0}^{H-1} c(x_t, u_t) \,|\, \pi \right]$ **is convex with respect the parameters,** $\forall t$

Convexity allows to perform Gradient Descent directly on parameters

$$\left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$$

Algorithm:

1. Execute the control $\left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$ to generate a trajectory $(x_0, u_0, \ldots, x_{H-1}, u_{H-1})$

Convexity and Gradient Descent

[Claim] Given controller $\widetilde{\pi} := \left\{ \{ M_t, M_{0;t}, \ldots M_{t-1;t} \} \right\}_{t=0}^{H-1}$,

$\mathbb{E} \left[\sum_{t=0}^{H-1} c(x_t, u_t) \,|\, \pi \right]$ is convex with respect the parameters, $\forall t$

Convexity allows to perform Gradient Descent directly on parameters

$$\left\{ \{ M_t, M_{0;t}, \ldots M_{t-1;t} \} \right\}_{t=0}^{H-1}$$

Algorithm:

1. Execute the control $\left\{ \{ M_t, M_{0;t}, \ldots M_{t-1;t} \} \right\}_{t=0}^{H-1}$ to generate a trajectory $(x_0, u_0, \ldots, x_{H-1}, u_{H-1})$

2. Compute gradient of $\sum_{h=0}^{H-1} c(x_h, u_h)$ wrt all parameters M, perform gradient descent

$$\nabla_x c(x, u) \cdot \frac{dx}{dM_s}$$

Extension to adversarial online control

Online Control Setting

Online Control Setting

On the k-th day,

Online Control Setting

On the k-th day,

1. adversary decides sequence of noises $\{w_0^k, \ldots, w_{H-1}^k\}$ and (convex) cost function $c^k(x, u)$,

Online Control Setting

On the k-th day,

1. adversary decides sequence of noises $\{w_0^k, \ldots, w_{H-1}^k\}$ and (convex) cost function $c^k(x, u)$,
2. Without knowing noises and cost, learner proposes a sequence of controllers $\{\{M_t^k, M_{0;t}^k, \ldots M_{t-1;t}^k\}\}_{t=0}^{H-1}$

Online Control Setting

On the k-th day,

1. adversary decides sequence of noises $\{w_0^k, \ldots, w_{H-1}^k\}$ and (convex) cost function $c^k(x, u)$,
2. Without knowing noises and cost, learner proposes a sequence of controllers $\{\{M_t^k, M_{0;t}^k, \ldots M_{t-1;t}^k\}\}_{t=0}^{H-1}$
3. Learner executes controllers, and suffers total cost $\sum_{h=0}^{H} c^k(x_h^k, \overset{u_h^k}{\cancel{a_h^k}})$

$$x_{h+1}^k = A x_h^k + B u_h^k + \underline{\underline{w_h^k}}$$

Online Control Setting

On the k-th day,

1. adversary decides sequence of noises $\{w_0^k, \ldots, w_{H-1}^k\}$ and (convex) cost function $c^k(x, u)$,
2. Without knowing noises and cost, learner proposes a sequence of controllers $\{\{M_t^k, M_{0;t}^k, \ldots M_{t-1;t}^k\}\}_{t=0}^{H-1}$
3. Learner executes controllers, and suffers total cost $\sum_{h=0}^{H} c^k(x_h^k, a_h^k)$

Goal: No-Regret $\sum_{k=0}^{K-1} \sum_{h=0}^{H} c^k(x_h^k, a_h^k) - \min_{\{-K_h^\star\}_{h=0}^{H-1}} \sum_{k=0}^{K-1} J^k(\{-K_h^\star\}_{h=0}) = o(K),$

Online Control Setting

$$\min_{\pi} \max_{w_0 \dots w_{H-1}} \sum_{h=0}^{H-1} c(x_h, u_h; w, \pi)$$

On the k-th day,

1. adversary decides sequence of noises $\{w_0^k, \ldots, w_{H-1}^k\}$ and (convex) cost function $c^k(x, u)$,
2. Without knowing noises and cost, learner proposes a sequence of controllers $\{\{M_t^k, M_{0;t}^k, \ldots M_{t-1;t}^k\}\}_{t=0}^{H-1}$
3. Learner executes controllers, and suffers total cost $\sum_{h=0}^{H} c^k(x_h^k, a_h^k)$

Goal: No-Regret $$\sum_{k=0}^{K-1} \sum_{h=0}^{H} c^k(x_h^k, a_h^k) - \boxed{\min_{\{-K_h^\star\}_{h=0}^{H-1}} \sum_{k=0}^{K-1} \underset{\Delta}{J^k(\{-K_h^\star\}_{h=0})}} = o(K),$$

$$\left(J^k(\pi) = \sum_{h=0}^{H-1} c^k(x_h, u_h), \text{where } x_{h+1} = Ax_h + Bu_t + w_t, u_t = \pi(x_t) \right)$$

The online gradient descent algorithm

Online Learning setting:

At iteration k:

The online gradient descent algorithm

Online Learning setting:

At iteration k:

1. **Adversary** picks a loss function $\ell_k : \overset{\subseteq \mathbb{R}^d}{\mathcal{Z}} \mapsto \mathbb{R}$

The online gradient descent algorithm

Online Learning setting:

At iteration k:

1. **Adversary** picks a loss function $\ell_k : \mathcal{Z} \mapsto \mathbb{R}$
2. **Without knowing** ℓ_k, learner has to propose $z_k \in \mathcal{Z}$

The online gradient descent algorithm

Online Learning setting:

At iteration k:

1. **Adversary** picks a loss function $\ell_k : \mathcal{Z} \mapsto \mathbb{R}$

2. **Without knowing** ℓ_k, learner has to propose $z_k \in \mathcal{Z}$

3. The loss ℓ_t is revealed, the learner suffers loss $\ell_t(z_t)$

The online gradient descent algorithm

Online Learning setting:

At iteration k:

1. **Adversary** picks a loss function $\ell_k : \mathscr{Z} \mapsto \mathbb{R}$

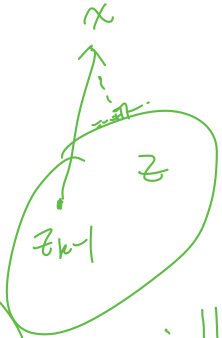
2. **Without knowing** ℓ_k, learner has to propose $z_k \in \mathscr{Z}$

3. The loss ℓ_t is revealed, the learner suffers loss $\ell_t(z_t)$

The OGD Algorithm:

At iteration t, learner proposes $\boxed{z_k := P_{\mathscr{Z}}\left(z_{k-1} - \eta \nabla \ell_{k-1}(z_{k-1})\right)}$

Δ ~~argmin~~



$\operatorname{argmin}_{z \in Z} \|z - x\|_2^2$

The online gradient descent algorithm

The OGD Guarantee:

ℓ_k

Although the OGD learner makes choice $\cancel{z_i}$ without seeing loss $\cancel{\ell_i}$, it is no-regret:

z_k

The online gradient descent algorithm

The OGD Guarantee:

Although the OGD learner makes choice z_t without seeing loss ℓ_t, it is no-regret:

Assume $\mathcal{X}$ is convex, and bounded $\max_{z,y \in \mathcal{Z}} \|z - y\|_2 \leq F$, and loss is G-Lipschitz, then OGD has the following regret:

$$\sum_{k=1}^{K} \ell_k(z_k) - \min_{x \in \mathcal{Z}} \sum_{k=1}^{K} \ell_k(z) \leq O\left(\left(F^2 + G^2\right)\sqrt{K}\right)$$

loss we suffer

Reduce Online Control to Online Learning

We define parameter $\mathbf{z} := \left\{ \{ M_t, M_{0;t}, \ldots M_{t-1;t} \} \right\}_{t=0}^{H-1}$

Reduce Online Control to Online Learning

We define parameter $\mathbf{z} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$

At iteration t, given $w_0^k, \ldots w_{H-1}^k$ and c^k which are determined by adversary, we define loss function $\ell_t(\mathbf{z})$ as:

Reduce Online Control to Online Learning

We define parameter $\mathbf{z} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$

At iteration t, given $w_0^k, \ldots w_{H-1}^k$ and c^k which are determined by adversary, we define loss function $\ell_t(\mathbf{z})$ as:

$$\ell_k(\mathbf{z}) := \sum_{h=0}^{H-1} c^k(x_h, u_h)$$

Convex wrt z

Where:

$$x_0, u_0 = M_0 x_0 \qquad x_1 = Ax_0 + BM_0x_0 \qquad u_1 = M_1x_0 + M_{0;1}w_0^k \qquad x_2 = Ax_1 + BM_1x_0 + BM_{0;1}w_0^k$$

$\Rightarrow$ x, u are linear in z

Reduce Online Control to Online Learning

We define parameter $\mathbf{z} := \left\{ \{M_t, M_{0;t}, \ldots M_{t-1;t}\} \right\}_{t=0}^{H-1}$

At iteration t, given $w_0^k, \ldots w_{H-1}^k$ and c^k which are determined by adversary, we define loss function $\ell_t(\mathbf{z})$ as:

$$\ell_k(\mathbf{z}) := \sum_{h=0}^{H-1} c^k(x_h, u_h)$$

Where:

$$x_0, u_0 = M_0 x_0 \quad x_1 = Ax_0 + BM_0 x_0 \quad u_1 = M_1 x_0 + M_{0;1} w_0^k \quad x_2 = Ax_1 + BM_1 x_0 + BM_{0;1} w_0^k$$

Thus, we can run any no-regret online learning algorithm (e.g., OGD) on the loss sequence $\{\ell_k\}$

Summary

x_0 $\quad$ w_{t-1}

$K_t \Longleftrightarrow M_t, M_{t-1;t} \cdots M_{0;t}$

w_0

1. LQR formulation, DP for LQR (Riccati Equation), and SDP formulation

3. Another form of over-parameterized controllers which leads to a convex parameterization (hence we can do gradient descent).

3. Online Control (contrast it w/ min-max robust control)