learnability of LQR

Artificial Intelligence and Language Processing

David Waltz Editor

A Theory of the Learnable 1984

L. G. VALIANT

PAC Adaptive Control of Linear Systems 1997

Claude-Nicolas Fiechter
Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260
fiechter@cs.pitt.edu

On the Sample Complexity of the Linear Quadratic Regulator 2019

Sarah Dean[♯], Horia Mania[♯], Nikolai Matni[†], Benjamin Recht[♯], and Stephen Tu[♯]
[♯] University of California, Berkeley
[†] California Institute of Technology

October 3, 2017, Revised: December 17, 2018

- consider discounted LQR problem with disturbance

$$\min_{u} E\left[\sum_{t=0}^{\infty} \gamma^t \cdot (x_t^T Q x_t + u_t^T R u_t)\right], \quad x \in \mathbb{R}^n, \quad u \in \mathbb{R}^m, \quad \gamma \in (0,1)$$

$$\text{s.t. } x_t^+ = A x_t + B u_t + w_t \quad \text{where } w_t \sim \mathcal{N}(0, \sigma^2 I) \text{ iid}$$

<u>recall</u>: solution is $u = \pi^*(x) = -K^* x$ if (A,B) controllable & $(A, \sqrt{Q})$ observable

where $K^* = \gamma(\gamma B^T P^* B + R)^{-1} B^T P^* A$,

$$P^* = A^T(\gamma P^* - \gamma^2 P^* B(\gamma B^T P^* B + R)^{-1} B^T P^*) A + Q$$

- we will consider 2 classes of learning algorithms:
  1°. Fiechter 1997 : PAC learnability
  2°. Dean et al 2019 : high-dimensional statistics

1°. <u>idea</u>: use linear regression to estimate A & B then use to compute π

1'. let $\theta = [B \ A]^T \in \mathbb{R}^{(n+m) \times n}$, $\varphi = \begin{bmatrix} u \\ x \end{bmatrix} \in \mathbb{R}^{n+m}$, $y^T = x$

1¹. let $\Theta = [B \ A]^T \in \mathbb{R}^{(n+m)\times n}$, $\varphi = \begin{bmatrix} u \\ x \end{bmatrix} \in \mathbb{R}^{n+m}$, $y^T = x$

so that $x^+ = Ax + Bu + w \iff y = \varphi^T \Theta + w$

1². given trajectory data $x : [0,T] \to \mathbb{R}^n$, $u : [0,T] \to \mathbb{R}^m$

create data matrices $Y_T = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}$, $\Phi_T = \begin{bmatrix} \varphi_1^T \\ \vdots \\ \varphi_T^T \end{bmatrix}$, $W_T = \begin{bmatrix} w_1^T \\ \vdots \\ w_T^T \end{bmatrix}$

so that $Y_T = \Phi_T \cdot \Theta + W_T$ measured, $\hat{Y}_T = \Phi_T \cdot \hat{\Theta}$ predicted

1³. solve $\min_{\hat{\Theta}} \| Y_T - \hat{Y}_T \|$ to obtain $\hat{\Theta} = (\Phi_T^T \Phi_T)^{-1} \Phi_T^T Y_T$,

the minimum-variance estimate of Θ : $\mathrm{Cov}[\hat{\Theta}_i, \hat{\Theta}_j] = \sigma^2 P_T$ where

the matrix $P_T = (\Phi_T^T \Phi_T)^{-1} = \left[\sum_{t=1}^{T} \varphi_t \varphi_t^T \right]^{-1}$ is <u>dispersion</u>, $\hat{\Theta} = [\hat{\Theta}_1, \dots, \hat{\Theta}_n]$

1⁴. show that this is a <u>probably approximately</u> correct (PAC) algorithm:

$$\forall \delta, \varepsilon > 0 : P\left(\forall x : \frac{|v^{\hat{\pi}}(x) - v^{\pi^*}(x)|}{1 + \|x\|^2} \leq \varepsilon \right) \geq 1 - \delta$$

- error in $\hat{A}, \hat{B}$ inversely proportional to $\lambda = \min \mathrm{spec}\, P_T^{-1}$
- for large T, λ increases linearly with T
- error $v^{\hat{\pi}} - v^{\pi^*}$ bounded by T

2°. begins similarly to (1°), but with "batches" of data

$$\{ (x_t^\ell, u_t^\ell) : \ell \in \{1, \dots, N\}, t \in [0,T] \}$$

2¹. the least-squares data is

$$Y = \begin{bmatrix} y_1^1 \\ \vdots \\ y_T^N \end{bmatrix} \quad \Phi = \begin{bmatrix} (\varphi_T^1)^T \\ \vdots \\ (\varphi_T^N)^T \end{bmatrix} \quad W = \begin{bmatrix} (w_T^1)^T \\ \vdots \\ (w_T^N)^T \end{bmatrix}$$

only use <u>final</u> sample!

but the equation remains $Y = \Phi \cdot \Theta + W$

2². show that least-squares estimate together with

$y = \Phi \theta + w$

2^2. show that least-squares estimate together with SLS LQR controller achieves error bound

$$P\left(\frac{V^{\hat{\pi}} - V^{\pi^*}}{V^{\pi^*}} \le O\left(C_{LQR} \sqrt{\frac{(n+m)\log(1/\delta)}{N}} \right) \right) \geq 1-\delta$$

where C_{LQR} depends on $T, A, B, Q, R, \sigma_u^2, \sigma_w^2$:

$$C_{LQR} \propto \sigma_w \left(\frac{1}{\sqrt{\underbrace{\lambda_G}_{\text{eigval of controllability grammians}}}} + \frac{\|K^*\|_2}{\sigma_u} \right) \| \underbrace{zI - (A - BK^*)}_{\text{xfer func from disturbance to state}} \|_\infty$$

2^3. importantly, controller is robust; also consider FIR implementation, other performance criteria