Implementation of Adaptive LQR and Thompson Sampling Algorithms

Saurav Kumar 21D070063
Guide: Prof. Debraj Chakraborty

November 27, 2024

Overview

- Introduction
- Adaptive LQ algorithm
- Least Square Estimation
- Confidence Bound
- Control policy via Riccati equation
- Simulation
- Thompson Sampling

Introduction

- Exploration-exploitation trade-off in estimating and controlling the linear system with quadratic cost over state and control.
- Both methods rely on estimating system parameters while optimizing the control policy.
- However, Regret of the order of $\sqrt{T}$ for Adaptive LQ algorithm and $T^{2/3}$ for Thompson Sampling.

Problem Formulation

Linear Quadratic Control Problem:

$$x_{t+1} = A^* x_t + B^* u_t + w_{t+1}$$
$$c_t = x_t^\top Q x_t + u_t^\top R u_t$$

- **State** $x_t \in \mathbb{R}^n$, **Control** $u_t \in \mathbb{R}^d$.
- Matrices $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{d \times d}$ are positive definite.
- Unknown parameters $A^* \in \mathbb{R}^{n \times n}$, $B^* \in \mathbb{R}^{d \times d}$.

Objective: Minimize cumulative cost:

$$J(u_0, u_1, \ldots) = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}[c_t].$$

Least Squares Estimation

- The system dynamics can be represented as:

$$X_{t+1} = Z_t \Theta_t + W_{t+1}$$

- Here:

$$Z_t = \begin{bmatrix} x_0^\top & u_0^\top \\ \vdots & \vdots \\ x_{t-1}^\top & u_{t-1}^\top \end{bmatrix}, \quad X_{t+1} = \begin{bmatrix} x_1^\top \\ \vdots \\ x_t^\top \end{bmatrix}$$

- Size of matrices increases iteratively

Regularized Least Squares (RLS)

$$\theta_t = \begin{bmatrix} A^\top \\ B^\top \end{bmatrix}, \quad \hat{\theta}_t = (Z_t^\top Z_t + \lambda I)^{-1} Z_t^\top X_{t+1} \tag{1}$$

- $\theta_t \in \mathbb{R}^{(n+d)\times n}$: combined system matrix A and B
- $Z_t \in \mathbb{R}^{t\times(n+d)}$: State and input matrix
- $X_{t+1} \in \mathbb{R}^{t\times n}$: Observed state transitions
- λ: Regularization parameter
- $V_t = (Z_t^\top Z_t + V_0) \in \mathbb{R}^{(n+d)\times(n+d)}$: Design Matrix
- $V_0 = \lambda I$

Admissible Set S

The unknown parameter Θ^* is a member of the set S such that:

$$S \subseteq S_0 \cap \left\{ \Theta \in \mathbb{R}^{(n+d)\times n} \mid \mathrm{trace}(\Theta^\top \Theta) \leq S^2 \right\},$$

where:

$$S_0 = \left\{ \Theta^T = (A, B) \in \mathbb{R}^{n\times(n+d)} \mid (A, B) \text{ is controllable,} \right.$$

(A, M) is observable, where Q $=$ M$^\top M$.

Confidence set $\mathcal{C}_t(\delta)$

For any $0 < \delta < 1$, with probability at least $1 - \delta$:

$$\text{trace}\left((\hat{\Theta}_t - \Theta^*)^\top V_t(\hat{\Theta}_t - \Theta^*)\right) \leq \beta_t(\delta).$$

In particular:

$$P\left(\Theta^* \in C_t(\delta),\ t = 1, 2, \dots\right) \geq 1 - \delta,$$

where:

$$C_t(\delta) = \left\{\Theta^T \in \mathbb{R}^{n\times(n+d)} : \text{trace}\left((\Theta - \hat{\Theta}_t)^\top V_t(\Theta - \hat{\Theta}_t)\right) \leq \beta_t(\delta)\right\}.$$

$$\beta_t(\delta) = \left(nL_s\sqrt{2\log\left(\frac{\sqrt{\det V_t}}{\delta\sqrt{\det \lambda I}}\right)} + \lambda^{1/2}S\right)^2 \tag{2}$$

Optimism in the Face of Uncertainty (OFU) Principle

- At each time step t, the algorithm selects the most optimistic parameter $\tilde{\Theta}_t$ from the confidence set $\mathcal{C}_t(\delta) \cap S$

$$J(\tilde{\Theta}_t) \leq \inf_{\Theta \in \mathcal{C}_t(\delta) \cap S} J(\Theta) + \frac{1}{\sqrt{t}} \tag{3}$$

Control Policy via Riccati Equation

$$P = Q + A^\top PA - A^\top PB(R + B^\top PB)^{-1}B^\top PA \tag{4}$$

$$K_t = -(R + B^\top PB)^{-1}B^\top PA \tag{5}$$

$$J(\theta) = trace(P) \tag{6}$$

- Solve the discrete Riccati equation iteratively.
- Control input: $\mathbf{u}_t = K_t\mathbf{x}_t$.

Simulation Parameters

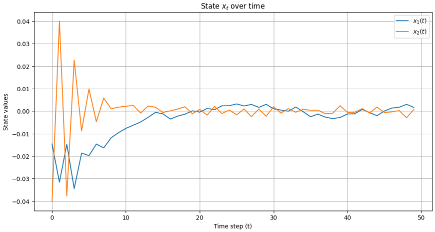
- State dimension: $n = 2$
- Control dimension: $d = 1$
- Cost matrices: $Q = I_n$, $R = I_d$
- Regularization Parameter: $\lambda = 10^{-4}$
- Bound on Process Noise: $L_s = 0.1$
- Bound on System Paramete: $S = 1.0$
- Confidence level: $\delta = 0.1$
- Time steps: $T = 50$

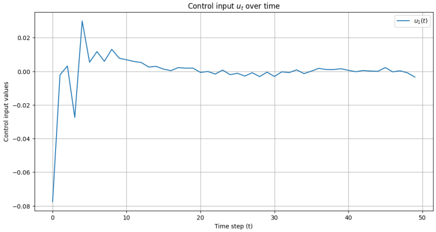
$$\mathbf{A} = \begin{bmatrix} 1.0 & 0.40 \\ 0.005 & -0.99 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$$

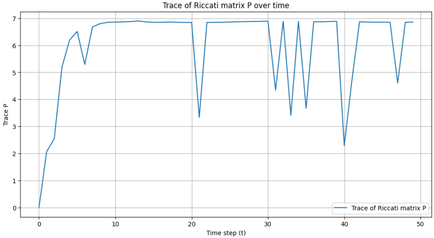
Estimates

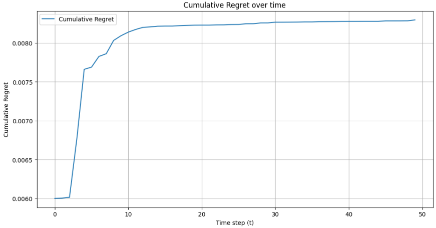
$$\tilde{\mathbf{A}} = \begin{bmatrix} 0.9986 & 0.3997 \\ 0.0054 & -0.9896 \end{bmatrix}, \quad \tilde{\mathbf{B}} = \begin{bmatrix} 0.2012 \\ 0.4997 \end{bmatrix}$$

Simulation Graphs









Thompson Sampling

$$\tilde{\theta}_t = \hat{\theta}_t + (\sqrt{\beta_t(\delta')}) W_t \eta_t, \quad \eta_t \sim \mathcal{N}(0, I) \tag{7}$$

- $\eta_t \in \mathbb{R}^{(n+d)\times n}$
- $\beta_t(\delta))$: Confidence bound
- $W_t = V_t^{-1/2}$ Cholesky decomposition of the design matrix inverse
- Perturbed parameter $\tilde{\theta}_t$ is used for policy computation.

Parameters

- State Dimension: $n = 2$
- Control Dimension: $d = 1$
- True System Matrices:

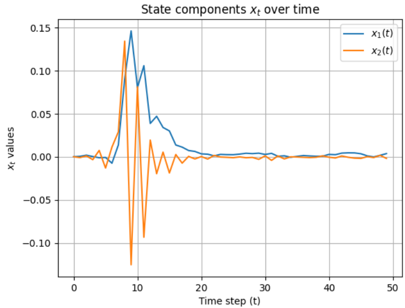
$$\mathbf{A} = \begin{bmatrix} 1.0 & 0.40 \\ 0.005 & -0.99 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$$

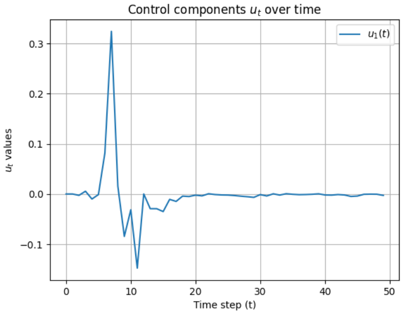
- Time Steps: $T = 50$
- Cost Matrices: $Q = I_n$, $R = I_d$ (identity matrices)
- Regularization Parameter: $\lambda = 10^{-4}$
- Bound on Process Noise: $L_s = 0.001$
- Bound on System Parameter: $S = 1.0$
- Confidence Level Parameter: $\delta = 0.1$
- Episode length: $\tau = 1.0$

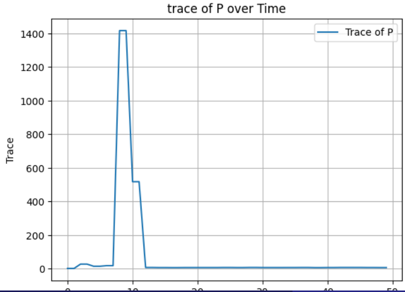
Estimates

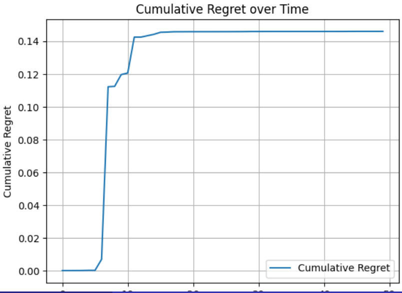
$$\tilde{\mathbf{A}} = \begin{bmatrix} 1.0057 & 0.3844 \\ 0.0054 & -0.9749 \end{bmatrix}, \quad \tilde{\mathbf{B}} = \begin{bmatrix} 0.2058 \\ 0.4901 \end{bmatrix}$$

Simulation Graph









Conclusion

- Adaptive LQR performs better in terms of cumulative regret minimization w.r.t. time steps than the Thompson Sampling algorithm.
- Future work corresponds to implementation Robust Adaptive LQ algorithm which also has cumulative regret of the order of $\sqrt{T}$.

References

- Y. Abbasi-Yadkori and C. Szepesv ´ari, “Regret bounds for the adaptive control of linear quadratic systems,” in Proceedings of the 24th Annual Conference on Learning Theory, pp. 1–26, JMLR Workshop and Conference Proceedings, 2011.
- M. Abeille and A. Lazaric, “Thompson sampling for linear-quadratic control problems,” in Artificial intelligence and statistics, pp. 1246–1254, PMLR, 2017.