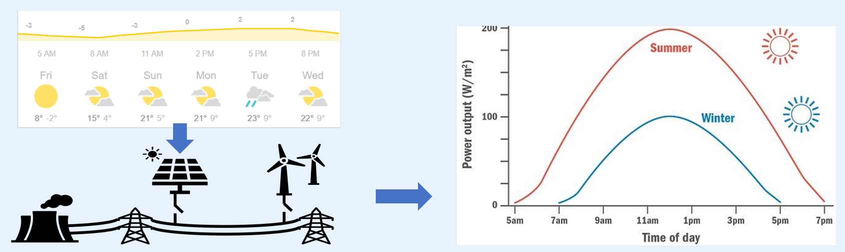
Identification and Adaptive Control of Markov Jump Systems: Sample Complexity and Regret Bounds

Yahya Sattar[1], Zhe Du[2], Davoud Ataee Tarzanagh[2], Necmiye Ozay[2], Laura Balzano[2], Samet Oymak[1]

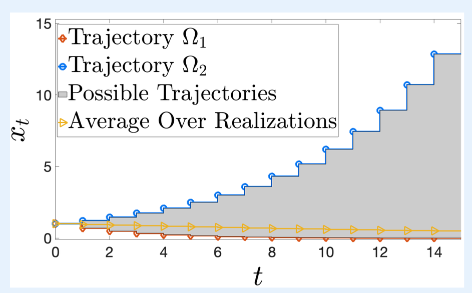
[1]Department of Electrical and Computer Engineering, University of California, Riverside, [2]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor

Introduction

- Markov jump systems (MJS) are powerful for modeling a richer set of problems where the underlying dynamics can abruptly change over time.

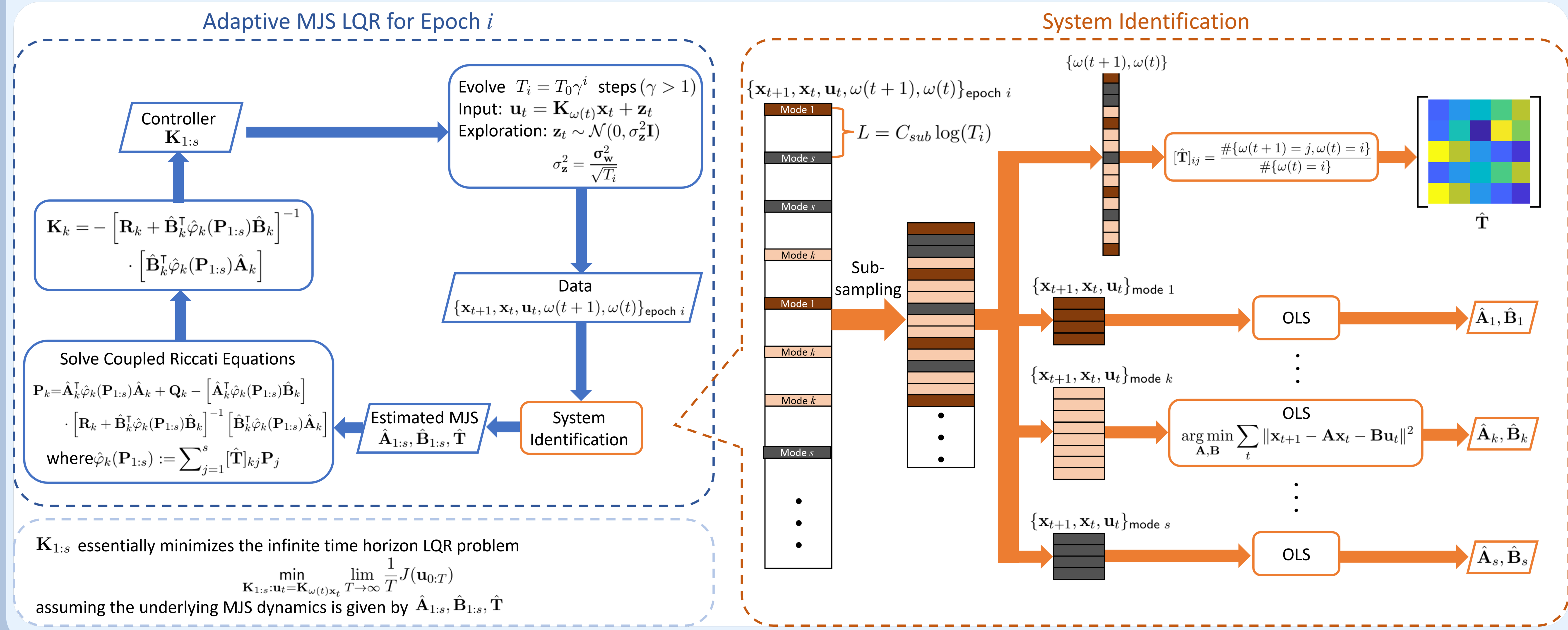


- The statistical analysis for MJS is challenging:
  - The common stability of MJS is in the **mean-square** sense.
  - The convergence of an MJS trajectory depends heavily on the mode switching sequence. For example. state trajectories of a two-mode MJS $\begin{cases} x_{t+1} = 0.7x_t \\ x_{t+1} = 1.2x_t \end{cases}$ with $\mathbf{T}=\begin{bmatrix}0.7 & 0.3\\ 0.4 & 0.6\end{bmatrix}$, $\mathbf{x}_0$=1, $\Omega_1 = \{1,1,\ldots\}$, and $\Omega_2 = \{2,2,\ldots\}$.



  - Mean-square stability $\not\Leftrightarrow$ each individual mode being stable.
  - Potential unstable realizations make non-asymptotic statistical analysis in learning and data-driven control difficult, e.g. no available light-tail inequality.
- **Goal**: address these challenges to provide guarantees for non-asymptotic identification and adaptive control of MJS's.

Our Approach



Problem Formulation

- **Markov Jump System (MJS($\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}$))**

$$\mathbf{x}_{t+1} = \mathbf{A}_{\omega(t)}\mathbf{x}_t + \mathbf{B}_{\omega(t)}\mathbf{u}_t + \mathbf{w}_t$$

  - Active mode index: $\omega(t) \sim$ Markov Chain($\mathbf{T}$), ergodic $\mathbf{T} \in \mathbb{R}^{s \times s}$
  - Mode k: state matrix $\mathbf{A}_k \in \mathbb{R}^{n \times n}$, input matrix $\mathbf{B}_k \in \mathbb{R}^{n \times p}, \forall k \in [s]$
  - Process noise: $\mathbf{w}_t \sim \mathcal{N}(0, \sigma_\mathbf{w}^2\mathbf{I})$
- **Assumptions**
  - MJS dynamics $\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}$ are *unknown*.
  - **Mean Square Stabilizability (MSS)** — exists $\mathbf{K}_{1:s}$ s.t. input $\mathbf{u}_t = \mathbf{K}_{\omega(t)}\mathbf{x}_t$ gives $\|\mathbb{E}[\mathbf{x}_t] - \mathbf{x}_\infty\| \to 0, \|\mathbb{E}[\mathbf{x}_t\mathbf{x}_t^\intercal] - \mathbf{\Sigma}_\infty\| \to 0$ for some $\mathbf{x}_\infty$ and $\mathbf{\Sigma}_\infty$.
- **Linear Quadratic Regulator (LQR($\mathbf{Q}_{1:s}, \mathbf{R}_{1:s}$))**

$$\underset{\mathbf{u}_{0:T}}{\text{minimize}} \quad J(\mathbf{u}_{0:T}) := \sum_{t=0}^{T} \mathbb{E}\left[\mathbf{x}_t^\intercal \mathbf{Q}_{\omega(t)}\mathbf{x}_t + \mathbf{u}_t^\intercal \mathbf{R}_{\omega(t)}\mathbf{u}_t\right]$$

  - Cost matrices: $\mathbf{Q}_k \succ 0, \mathbf{R}_k \succ 0, \forall k \in [s]$. At time t, $\mathbf{x}_t$ and $\omega(t)$ are observed.
- **Goal**
  - Identify MJS dynamics $\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}$ and solve LQR in *real time*.
  - Performance guarantee
    - Estimation error: $\|\hat{\mathbf{A}}_k - \mathbf{A}_k\|, \|\hat{\mathbf{B}}_k - \mathbf{B}_k\|, \|\hat{\mathbf{T}} - \mathbf{T}\|_\infty$
    - Regret: $J - J^\star$

Theory

- **System Identification** — under certain conditions, in epoch i, with prob. $1-\delta$

$$\max \left\{ \begin{matrix} \|\hat{\mathbf{A}}_k - \mathbf{A}_k\|, \\ \|\hat{\mathbf{B}}_k - \mathbf{B}_k\| \end{matrix} \right\} \leq \tilde{\mathcal{O}}\left(\frac{\sigma_\mathbf{z} + \sigma_\mathbf{w}}{\sigma_\mathbf{z}\pi_{\min}} \frac{(n+p)\log(T_i)}{\sqrt{T_i}}\right),$$

$$\|\hat{\mathbf{T}} - \mathbf{T}\|_\infty \leq \tilde{\mathcal{O}}\left(\frac{1}{\pi_{\min}}\sqrt{\frac{\log(T_i)}{T_i}}\right).$$

  - Can be applied to generic system identification problem outside of LQR setting.
- **LQR** — under certain conditions, with prob. $1-\delta$

$$\text{Regret}(T) \leq \tilde{\mathcal{O}}(\log^2(T)\sqrt{T}).$$

- When $\mathbf{B}_{1:s}$ are known, no exploration is needed, i.e. $\sigma_\mathbf{z} = 0$, and guarantees improve:

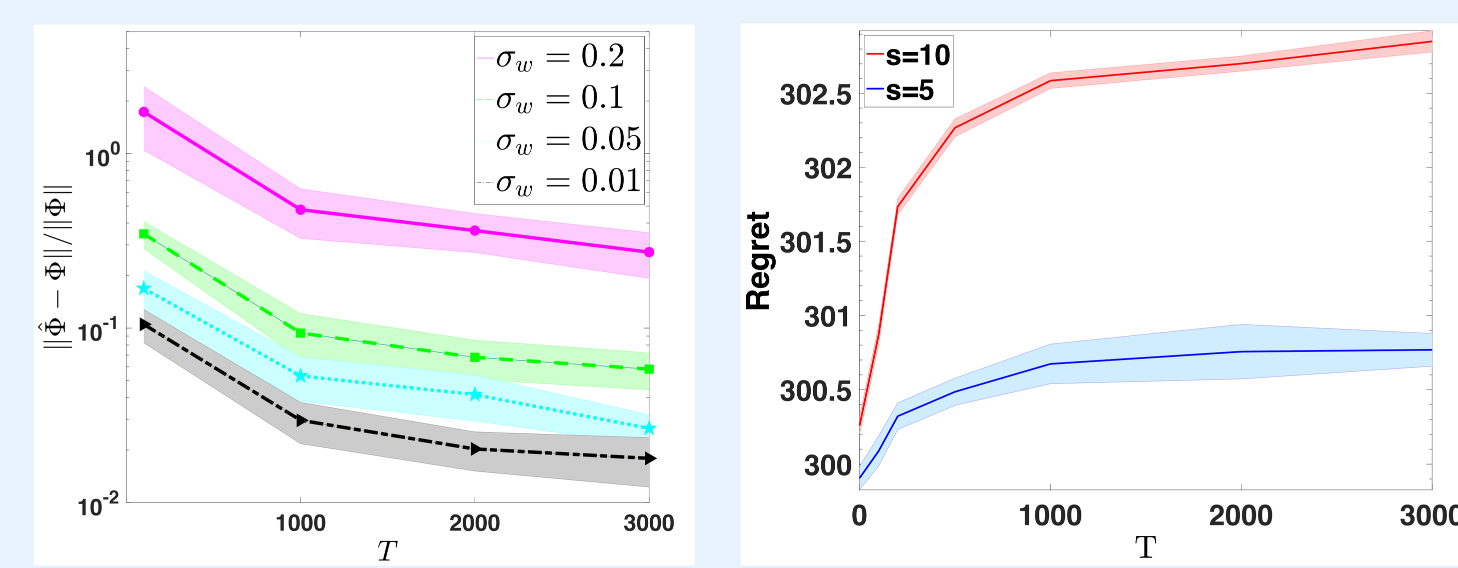
$$\|\hat{\mathbf{A}}_k - \mathbf{A}_k\| \leq \tilde{\mathcal{O}}\left(\frac{(n+p)\log(T)}{\pi_{\min}\sqrt{T}}\right),$$

$$\text{Regret}(T) \leq \tilde{\mathcal{O}}(\log^3(T)).$$

Experiments

The depicted results are averaged over 50 independent runs.

- **System Identification** (left)
  - Consider MJS with $n = 10$, $p = 8$, and $s = 10$. ($\mathbf{A}_{1:s}, \mathbf{B}_{1:s}, \mathbf{T}$) are generated randomly.
  - $\hat\Psi_k := [\hat{\mathbf{A}}_k, \hat{\mathbf{B}}_k]$, $\Psi_k := [\mathbf{A}_k, \mathbf{B}_k]$, $\|\hat\Psi - \Psi\|/\|\Psi\| := \max_{k\in[s]} \|\hat\Psi_k - \Psi_k\|/\|\Psi_k\|$.
- **Adaptive MJS-LQR** (right)
  - Consider MJS with $n = 10$, $p = 8$, and $s \in \{5, 10\}$.
  - Set the number of epochs to five and epoch increment ratio to $\gamma = 2$.



(a) System identification with varying $\sigma_\mathbf{w}$. (b) Regret evaluation for adaptive MJS-LQR control.

Postdoc position available in RL: Please drop an email to Samet at oymak@ece.ucr.edu

ysatt001@ucr.edu, zhedu@umich.edu