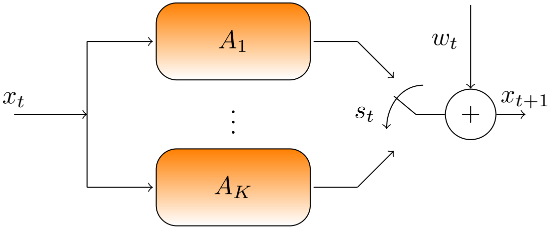
Consistency and Rate of Convergence of Switched Least Squares System Identification for Autonomous Markov Jump Linear Systems

Borna Sayedana[1]
Joint work with: Mohammad Afshari[2], Peter E. Caines[1] , Aditya Mahajan[1]

McGill University[1], CIM[1], GERAD[1],Georgia Institute of Technology[2]

IEEE Conference on Decision and Control
December 2022

Motivation



- Markov jump linear systems are suitable in modeling dynamical systems with abrupt changes.

Application of Markov jump linear systems

- Network Control Systems (NCS)
- Power systems
- Cyber-physical systems
- Time series analysis (e.g. in stock markets)

- **Question**: How we can perform system identification?

System Identification for Linear Systems

- We use a least squares based algorithm

Convergence of Least squares method (a.s. results)

1. Convergence of general regression: [Lai and Wei, 1982]
2. Autonomous linear systems: [Lai and Wei, 1985]
3. Least squares for ARMAX models : [Chen and Guo, 1986, Chen and Guo, 1987, Lai and Wei, 1986, Caines, 2018]

Convergence of Least squares method (high probability bounds)

1. [Abbasi-Yadkori and Szepesvári, 2011, Faradonbeh et al., 2018, Simchowitz et al., 2018, Oymak and Ozay, 2019, Lale et al., 2020]

Literature Review

- Adaptive control for switched linear systems:

Asymptotically stable adaptive controllers

1. [Caines and Chen, 1985, Caines and Zhang, 1995, Xue and Guo, 2001]

- The system identification of Markov/switched least squares problem is less explored.

System identification for switched linear systems

1. Switched linear systems: [Ozay et al., 2015, Hespanhol and Aswani, 2020, Sarkar et al., 2019, Shi et al., 2023]
2. Markov jump linear systems: [Sattar et al., 2021]

Contributions

Our results

- *Switched least squares* for system Identification of MJS.
- Converges *almost surely* and establish *strong consistency*.
- Data-dependent and data-independent rate of convergence.

Compared to the literature

1. Convergence guarantees:
   - Almost sure convergence/high probability convergence
2. Assumption on stability:
   - Stability on the average sense/Mean square stability

An overview of the results

1. Theorem 1: (Strong consistency)

$$\lim_{T\to\infty}\left\|\hat{A}_{i,T}-A_i\right\|_\infty=0,\quad \text{a.s.}\quad \forall i$$

2. Theorem 2: (Rate of convergence)
   - Data-dependent.
   - Data-independent.

$$\left\|\hat{A}_{i,T}-A_i\right\|_\infty\leq\mathcal{O}\left(\sqrt{\log(T)/T}\right),\quad \text{a.s.}\quad \forall i$$

Autonomous Markov jump linear systems

- states : (s_t, x_t), $s_t \in \{1, \ldots, k\}$, $x_t \in \mathbb{R}^n$
- set of possible modes : $\mathcal{A} = \{A_1, \ldots, A_k\}$.

Continuous state dynamics

$$x_{t+1} = A_{s_t} x_t + w_t, \quad t \geq 0,$$

where $\{w_t\}_{t\geq 0}$, $w_t \in \mathbb{R}^n$, is a noise process.

Discrete state dynamics

- $\{s_t\}_{t\geq 0}$ is a time-homogeneous irreducible and aperiodic Markov chain.
- $\pi_\infty = (\pi_\infty(1), \ldots, \pi_\infty(k))$: stationary distribution $(\pi(i) \neq 0)$.

Assumption on the noise process

- Let $\mathcal{F}_{t-1} = \sigma(x_{0:t}, s_{0:t})$.

Assumption on the noise

1. $\{w_t\}_{t \geq 0}$ is a martingale difference sequence w.r.t. $\{\mathcal{F}_t\}_{\geq 0}$ i.e.:

$$\mathbb{E}[|w_t|] < \infty \quad \text{and} \quad \mathbb{E}[w_t \mid \mathcal{F}_{t-1}] = 0$$

2. Furthermore, $\exists\ \alpha > 2$ such that : $\sup_{t \geq 0} \mathbb{E}[\|w_t\|^{\alpha} \mid \mathcal{F}_{t-1}] < \infty \quad$ a.s.
3. There exists $0 \prec C \in \mathbb{R}^{n \times n}$ such that:

$$\liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} w_t w_t^{\intercal} = C$$

- Noise process can be non-stationary and heavy tailed process.

Assumption on the stability

Assumption on the stability of system

The MJS system is stable in the *average sense* if:

$$\sum_{t=1}^{T} \|x_t\|^2 = \mathcal{O}(T) \quad \text{a.s.}$$

Assumption: The MJS system is stable in the average sense.

Switched Least Squares Method

- Global Least squares optimization:

$$\hat{\theta}_T^{\intercal} = \underset{\theta^{\intercal}=[A_1,\ldots,A_k]}{\arg\min} \sum_{t=0}^{T-1} \|x_{t+1} - A_{s_t} x_t\|^2.$$

- Suppose: $\mathcal{T}_{i,T} = \{t \leq T | s_t = i\}$.

Switched Least Squares Problem

- Decoupling:

$$\hat{A}_{i,T} = \underset{A_i \in \mathbb{R}^{n\times n}}{\arg\min} \sum_{t\in\mathcal{T}_{i,T}} \|x_{t+1} - A_i x_t\|^2, \quad \forall i \in \{1, \cdots, k\}.$$

A few definitions

- $X_{i,T} = \sum_{t \in \mathcal{T}_{i,T}} x_t x_t^\intercal$: i-th mode empirical covariance
- The least squares update:

$$\hat{A}_{i,T+1} = \hat{A}_{i,T} + \left[\frac{X_{i,T}^{-1} x_T (x_{T+1} - \hat{A}_{i,T} x_T)^\intercal}{1 + x_T^\intercal X_{i,T}^{-1} x_T} \right] \mathbb{1}\{s_{T+1} = i\}$$

where: $X_{i,T+1} = X_{i,T} + \left[x_{T+1} x_{T+1}^\intercal \right] \mathbb{1}\{s_{T+1} = i\}$.

Strong consistency

- An estimator $\hat{\theta}_T$ is called strongly consistent if:

$$\lim_{T \to \infty} \hat{\theta}_T = \theta \quad \text{a.s.}$$

Main Results (Data dependent)

Theorem

- *The switched least squares estimates $\{\hat{A}_{i,T}\}_{i=1}^{k}$ are strongly consistent, i.e.,:*

$$\lim_{T\to\infty} \left\| \hat{A}_{i,T} - A_i \right\|_\infty = 0, \quad a.s. \quad \forall i$$

- Rate of convergence:

$$\left\| \hat{A}_{i,T} - A_i \right\|_\infty \leq \mathcal{O}\left(\sqrt{\frac{\log\left[\lambda_{\max}(X_{i,T})\right]}{\lambda_{\min}(X_{i,T})}} \right), \quad \text{a.s.}$$

Main Results (Data independent)

Theorem

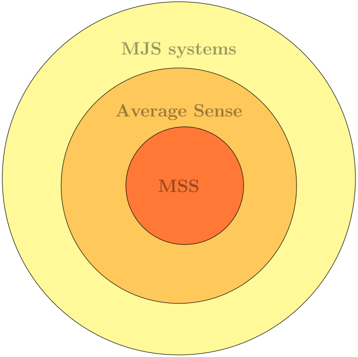
1. *The rate of convergence of the switched least squares is upper bounded by:*

$$\left\|\hat{A}_{i,T} - A_i\right\|_\infty \leq \mathcal{O}\left(\sqrt{\log(T)/\pi_\infty(i)T}\right), \quad a.s.$$

Mean square stability

- System is Mean square stable (MSS) if $\exists\, x_\infty, Q_\infty$:

$$\lim_{\tau\to\infty} \left\| \mathbb{E}[x_\tau] - x_\infty \right\| \to 0 \quad \text{and} \quad \lim_{\tau\to\infty} \left\| \mathbb{E}[x_\tau x_\tau^\intercal] - Q_\infty \right\| \to 0.$$



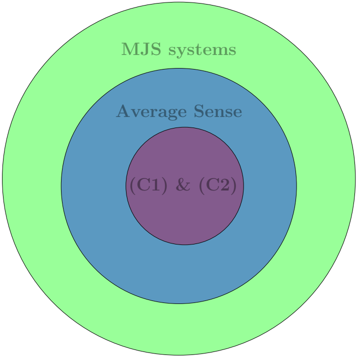
- $\sum_{t=1}^{T} \|x_t\|^2 = \mathcal{O}(T)$ a.s.

Results

(i) Average sense stability $\prec$ MSS
(ii) MSS $\implies$ Consistency

Almost sure stability

- Suppose π_∞ satisfies **(C1)** $\pi_\infty(i) \neq 0 \quad \forall i$, **(C2)** $\prod_{i=1}^{k} \sigma_i^{\pi_\infty(i)} < 1$.

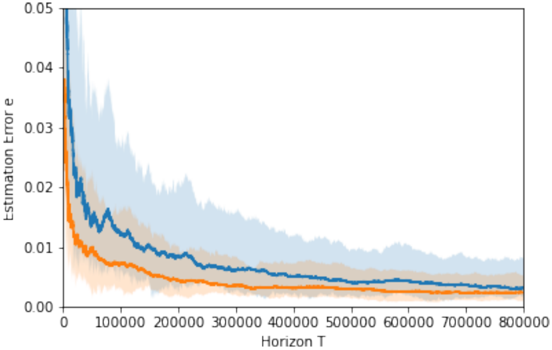


- $\sum_{t=1}^{T} \|x_t\|^2 = \mathcal{O}(T) \quad$ a.s.

Results

(i) Average sense stability $\prec$ (C1) & (C2)
(ii) (C1) & (C2) $\implies$ Consistency

Numerical simulation



- $A_1 = \begin{bmatrix} 1.5 & 0 \\ 0 & 0.2 \end{bmatrix}$, and $A_2 = \begin{bmatrix} 0.01 & 0.1 \\ 0.1 & 0.1 \end{bmatrix}$,
- transition matrix $P = \begin{bmatrix} 1/12 & 11/12 \\ 3/4 & 1/4 \end{bmatrix}$ with $\pi_\infty = [0.45, 0.55]$
- *Not* MSS and (C1) and (C2)

Conclusion

- *Strong* consistency of switched least squares for MJS systems.
- Data-independent rate of convergence of $\sqrt{\log(T)/T}$ almost surely.
- Stability in the average sense $\implies$ almost sure convergence of the method.
- Mean square stability $\implies$ almost sure convergence of the method .

Thank you!

Bibliography I

Abbasi-Yadkori, Y. and Szepesvári, C. (2011).
Regret bounds for the adaptive control of linear quadratic systems.
In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26. JMLR Workshop and Conference Proceedings.

Caines, P. E. (2018).
Linear stochastic systems, volume 77.
SIAM.

Caines, P. E. and Chen, H.-F. (1985).
Optimal adaptive lqg control for systems with finite state process parameters.
IEEE Transactions on Automatic Control, 30(2):185–189.

Bibliography II

Caines, P. E. and Zhang, J.-F. (1995).
On the adaptive control of jump parameter systems via nonlinear filtering.
SIAM journal on control and optimization, 33(6):1758–1777.

Chen, H.-F. and Guo, L. (1986).
Convergence rate of least-squares identification and adaptive control for stochastic systems.
International Journal of Control, 44(5):1459–1476.

Chen, H.-F. and Guo, L. (1987).
Optimal adaptive control and consistent parameter estimates for armax model with quadratic cost.
SIAM Journal on Control and Optimization, 25(4):845–867.

Bibliography III

Faradonbeh, M. K. S., Tewari, A., and Michailidis, G. (2018).
Finite time identification in unstable linear systems.
Automatica, 96:342–353.

Hespanhol, P. and Aswani, A. (2020).
Statistical consistency of set-membership estimator for linear systems.
IEEE Control Systems Letters, 4(3):668–673.

Lai, T. L. and Wei, C. Z. (1982).
Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems.
The Annals of Statistics, 10(1):154–166.

Lai, T. L. and Wei, C. Z. (1985).
Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems.
Multivariate Analysis VI, pages 375–393.

Bibliography IV

Lai, T. L. and Wei, C. Z. (1986).
Extended least squares and their applications to adaptive control and prediction in linear systems.
IEEE Transactions on Automatic Control, 31(10):898–906.

Lale, S., Azizzadenesheli, K., Hassibi, B., and Anandkumar, A. (2020).
Logarithmic regret bound in partially observable linear dynamical systems.
arXiv preprint arXiv:2003.11227.

Oymak, S. and Ozay, N. (2019).
Non-asymptotic identification of LTI systems from a single trajectory.
In *2019 American control conference (ACC)*, pages 5655–5661. IEEE.

Bibliography V

Ozay, N., Lagoa, C., and Sznaier, M. (2015).
Set membership identification of switched linear systems with known number of subsystems.
Automatica, 51:180–191.

Sarkar, T., Rakhlin, A., and Dahleh, M. (2019).
Nonparametric system identification of stochastic switched linear systems.
In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3623–3628. IEEE.

Sattar, Y., Du, Z., Tarzanagh, D. A., Balzano, L., Ozay, N., and Oymak, S. (2021).
Identification and adaptive control of Markov jump systems: Sample complexity and regret bounds.
arXiv preprint arXiv:2111.07018.

Bibliography VI

Shi, S., Mazhar, O., and De Schutter, B. (2023).
Finite-sample analysis of identification of switched linear systems with arbitrary or restricted switching.
IEEE Control Systems Letters, 7:121–126.

Simchowitz, M., Mania, H., Tu, S., Jordan, M. I., and Recht, B. (2018).
Learning without mixing: Towards a sharp analysis of linear system identification.
In *Conference On Learning Theory*, pages 439–473. PMLR.

Xue, F. and Guo, L. (2001).
Necessary and sufficient conditions for adaptive stablizability of jump linear systems.
Communications in Information and Systems, 1(2):205–224.