System Identification Under Bounded Noise: Optimal Rates Beyond Least Squares

Xiong Zeng

Ph.D. Candidate
Electrical and Computer Engineering

Joint work with Prof. Jing Yu and Prof. Necmiye Ozay

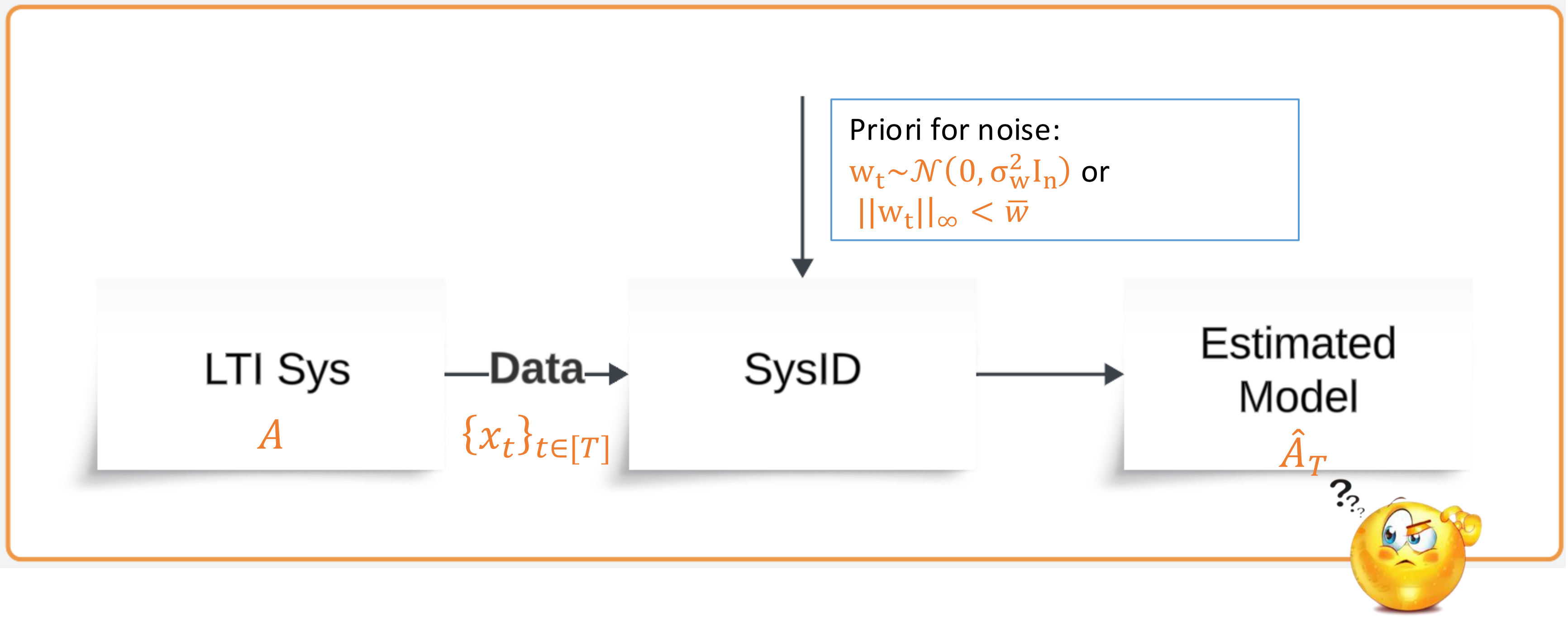
System identification is important for physical intelligence

ANYmal from ETH
LeCAR lab from CMU

Consider a linear time-invariant (LTI) system:

$$x_{t+1} = Ax_t + w_t,$$

where A is unknown, $x_t, w_t \in \mathbb{R}^n$.



Consider a LTI system:

$$x_{t+1} = Ax_t + w_t,$$

where A is unknown, $x_t, w_t \in \mathbb{R}^n$.

Theorem 1 (Simchowitz et al., COLT 2018). When the noise w_t is Gaussian, the ordinary least squares (OLS) estimator is minimax optimal.

In safety-critical scenarios, bounded noise is more common.

Two questions:

1. Is OLS still minimax optimal for bounded noise?

2. Which estimator is minimax optimal for bounded noise?

Consider an LTI system:

$$\mathrm{x}_{\mathrm{t}+1} = \mathrm{A}\mathrm{x}_{\mathrm{t}} + \mathrm{w}_{\mathrm{t}},$$

where A is unknown, $\mathrm{x}_{\mathrm{t}}, w_t \in \mathbb{R}^{\mathrm{n}}$. Assume:

- **(bounded noise)** $||\mathrm{w}_{\mathrm{t}}||_{\infty} < \bar{w}\ and\ \mathrm{w}_{\mathrm{t}}$ are i.i.d. ∀t,
- (**stable**) $\rho(A) < 1$, where ρ(·) := the spectral radius
- (**boundary probability**) $\forall \epsilon \in [0, \bar{w}]$, $\exists C > 0$, such that $\forall j \in [n]$,

$$\max\left(P\left(w_t^{(j)} < -\bar{w} + \epsilon\right), P\left(w_t^{(j)} > \bar{w} - \epsilon\right)\right) < C\epsilon.$$

Theorem 2 (**our work**). Given a single trajectory $\{x_t\}_{t\in[T]}$. $\mathcal{F}_T$:= the σ-algebra generated by $\{x_t\}_{t\in[T]}\ and\ \hat{A}_T$:= a $\mathcal{F}_T$ -measurable estimator of A. Then, $\forall \delta \in (0,1)$ and ∀ small $\epsilon > 0$,

$$\sup_{\hat{A}_T} \inf_{A} P(||\hat{A}_T - A||_2 < \epsilon) \geq 1 - \delta \quad \text{only if} \quad T > \frac{1}{4\bar{w}C\epsilon}\left(1 - \frac{2\delta}{n}\right).$$

(The best estimator can achieve $\Omega(\frac{1}{\epsilon})$)

The ordinary least squares (OLS) for scalar case:

$$\hat{a}_T^{OLS} = \mathrm{argmin}_{\mathrm{a}} \sum_{t=1}^{T-1} \left|\left|x_{t+1} - a x_t\right|\right|^2$$

Theorem 3 (our work). Assume $|a| < 1$. Then, $\forall \delta \in (0,1)$ and small $\epsilon > 0$,

$$P\left(\left|\left|\hat{a}_T^{OLS} - a\right|\right|_2 < \epsilon\right) \geq 1 - \delta, \quad \text{only if } \; T > \Omega\left(\frac{1}{\epsilon^2}\right).$$

($\Omega\left(\frac{1}{\epsilon^2}\right) > \Omega\left(\frac{1}{\epsilon}\right)$, OLS is not optimal for bounded noise)

The set membership estimator (SME) based on $\{x_t\}_{t\in[T]}$

$$\mathcal{S}_{\mathrm{T}} = \left\{A \in \mathbb{R}^{n\times n}: \left|\left|x_{t+1} - A x_t\right|\right|_\infty \leq \bar{w}, \forall t \in [T-1]\right\}.$$

Theorem 4 (Li & Yu et al., ICML 2024). $\forall \delta \in (0,1)$ and small $\epsilon > 0$,

$$\forall \hat{A}_T \in \mathcal{S}_{\mathrm{T}}, \; P\left(\left|\left|\hat{A}_T - A\right|\right|_2 < \epsilon\right) \geq 1 - \delta, \quad \text{if} \quad T > \Omega\left(\frac{1}{\epsilon}\right).$$

(SME achieves the optimal $\Omega\left(\frac{1}{\epsilon}\right)$)

Basic Intuition

For two length-T trajectories of two different systems that differ by ϵ,

- **Gaussian** → KL divergence = $O(T\epsilon^2)$;
- **Bounded** → Total variation (TV) distance = $O(T\epsilon)$.

(ChatGPT can't handle it ☺)

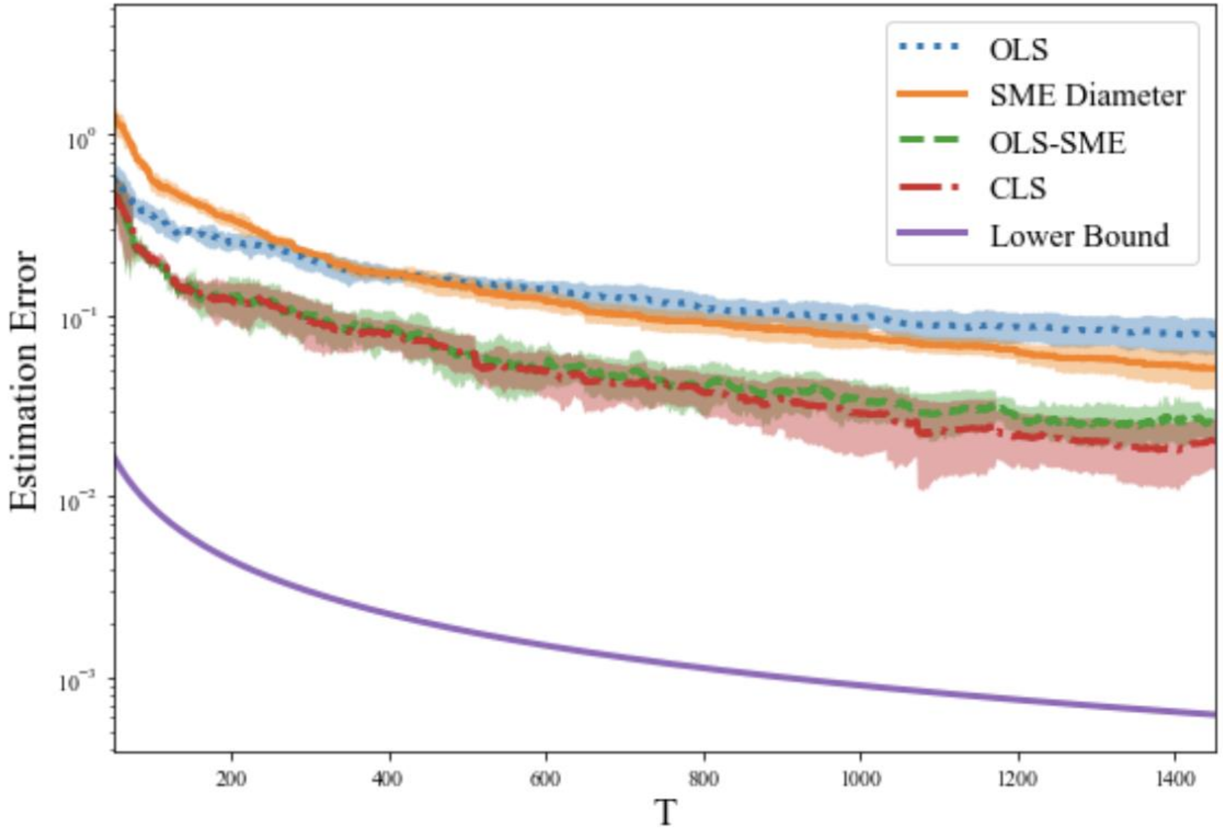
Simulation

OLS: $\min_{A} \sum_{t=1}^{T-1} ||x_{t+1} - Ax_t||^2$

OLS-SME: $\min_{A \in \mathcal{S}_T} ||A - \hat{A}_T^{OLS}||$

Constrained Least Square (CLS):

$$\min_{A \in \mathcal{S}_T} \sum_{t=1}^{T-1} ||x_{t+1} - Ax_t||^2$$



OLS achieves $O(\frac{1}{\sqrt{T}})$ ☹

SME, OLS-SME, and CLS achieve $O(\frac{1}{T})$ ☺

CLS > OLS-SME > SME > OLS

Summary and Future Work:

		Minimax Lower Bound	Lower Bound for OLS
Regression	Gaussian	$\Omega(1/\sqrt{T})$ (Wainwright, 2019)	$\Omega(1/\sqrt{T})$ (Mourtada, 2022)
	Bounded	$\Omega(1/T)$ (Yi & Neykov, 2024)	$\Omega(1/\sqrt{T})$ (Rudelson & Vershynin, 2008)
LTI Sys Id	Gaussian	$\Omega(1/\sqrt{T})$ (Jedra & Proutiere, 2019)	$\Omega(1/\sqrt{T})$ (Tu et al., 2024)
	Bounded	$\Omega(1/T)$ (**Thm. 2**)	$\Omega(1/\sqrt{T})$ (**Thm. 3**)

1. The state dimension n and the failure probability δ dependencies of our minimax lower bound are loose.

2. In experiments, CLS > SME. The tight sample complexity upper bound of CLS is unclear.

(ChatGPT might handle them ☺)

Thank you!

Zeng, X., Yu, J., & Ozay, N. (2025). System Identification Under Bounded Noise: Optimal Rates Beyond Least Squares. IEEE Control Systems Letters 2025 (with presentation at CDC 2025).