



# System Identification Under Bounded Noise: Optimal Rates Beyond Least Squares



Xiong Zeng, Jing Yu, and Necmiye Ozay

University of Michigan, Ann Arbor

**Introduction:** For linear system identification, the ordinary least squares (**OLS**) estimator is **minimax optimal** under **Gaussian** noise. In safety-critical scenarios, **bounded** noise is very common.

?

**Questions:**



1. Is **OLS** estimator still **minimax optimal** for bounded noise?
2. **Which** estimator is **minimax optimal** for bounded noise?

**Problem Setup:** Given a discrete-time linear time-invariant (LTI) system:

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

where  $x_t, w_t \in \mathbb{R}^n$ . Assume

- 1). (**bounded noise**)  $\|w_t\|_\infty < \bar{w}$  and  $w_t$  are i.i.d. for all  $t$ ;
  - 2). (**stable**)  $\rho(A) < 1$ ;
  - 3). (**boundary probability**)  $\forall \epsilon \in [0, \bar{w}], \exists C > 0$ , s. t.  $\forall j \in [n]$ ,
- $$\max \left( P(w_t^{(j)} < -w + \epsilon), P(w_t^{(j)} > w - \epsilon) \right) < C\epsilon.$$

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

$$\sup_{\hat{A}_T} \inf_{A \in \mathbb{R}^{n \times n}} P(\|\hat{A}_T - A\|_2 < \epsilon) \geq 1 - \delta$$

only if

$$T > \frac{1}{4wC\epsilon} (1 - \frac{2\delta}{n}).$$

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

		Minimax Lower Bound	Lower Bound for OLS
Regression	Gaussian	$\Omega(1/\sqrt{T})$ (Wainwright '19)	$\Omega(1/\sqrt{T})$ (Mourtada '22)
	Bounded	$\Omega(1/T)$ (Yi & Neykov '24)	$\Omega(1/\sqrt{T})$ (Rudelson & Vershynin '08)
LTI Sys Id	Gaussian	$\Omega(1/\sqrt{T})$ (Jedra & Proutiere '19)	$\Omega(1/\sqrt{T})$ (Tu et al. '24)
	Bounded	$\Omega(1/T)$ (Thm. 1)	$\Omega(1/\sqrt{T})$ (Thm. 2)

The ordinary least squares (**OLS**) for scalar case:

$$\hat{a}_T^{OLS} = \operatorname{argmin}_a \sum_{t=1}^{T-1} \|x_{t+1} - ax_t\|^2$$

**Theorem 2 (our work).** Assume  $|a| < 1$ . Then,  $\forall \delta \in (0,1)$  and small  $\epsilon > 0$ ,

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

(OLS only achieves  $\Omega(\frac{1}{\epsilon^2})$ )

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

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

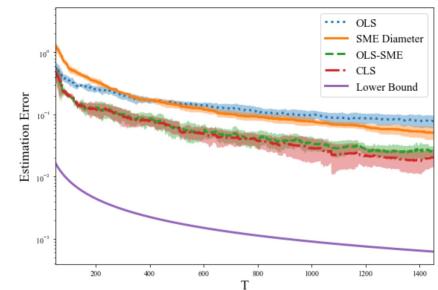
**Theorem 3 ([2]).**  $\forall \delta \in (0,1)$  and small  $\epsilon > 0$ , with or without knowing  $\bar{w}$ ,

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

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

**Algorithm Comparison:**

$$\begin{aligned} \text{OLS: } & \min_A \sum_{t=1}^{T-1} \|x_{t+1} - Ax_t\|^2 \\ \text{OLS-SME: } & \min_{A \in \mathcal{S}_T} \|A - \hat{A}_T^{OLS}\| \\ \text{CLS: } & \min_{A \in \mathcal{S}_T} \sum_{t=1}^{T-1} \|x_{t+1} - Ax_t\|^2 \end{aligned}$$



OLS achieves  $\Omega(\frac{1}{\sqrt{T}})$  ☺

SME, OLS-SME, and CLS achieves  $\Omega(\frac{1}{\sqrt{T}})$  ☺

**References:**

- [1] Zeng, X., Yu, J., & Ozay, N. (2025). System Identification Under Bounded Noise: Optimal Rates Beyond Least Squares, L-CSS and CDC 2025.
- [2]. Y. Li, J. Yu, L. Conger, T. Kargin, and A. Wierman, "Learning the uncertainty sets of linear control systems via set membership: A non-asymptotic analysis," ICML 2024.