Proceedings of Machine Learning Research vol :1–2, 2026 39th Annual Conference on Learning Theory

Separating Oblivious and Adaptive Models of Variable Selection

Ziyun Chen ZIYUNCC@CS.WASHINGTON.EDU
Jerry Li JERRYZLI@CS.WASHINGTON.EDU
University of Washington

Kevin Tian KJTIAN@CS.UTEXAS.EDU
Yusong Zhu ZHUYS@UTEXAS.EDU
University of Texas at Austin

Editors: Steve Hanneke and Tor Lattimore

In this work, we study sparse recovery with ℓ_∞ error guarantees, a formulation motivated by *variable selection*, where the goal is to recover the support of a k-sparse signal in $\mathbb{R}^d$. Our main contribution is a sharp separation between the *oblivious* and *adaptive* models of ℓ_∞ sparse recovery.

Problem 1 (ℓ_∞ sparse recovery) *Let $(n, d) \in \mathbb{N}^2$ and $k \in [d]$. Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ have i.i.d. $\mathcal{N}(0, 1/n)$ entries, and let $(\boldsymbol{\theta}^\star, \boldsymbol{\xi}) \in \mathbb{R}^d \times \mathbb{R}^n$ be unknown with $\mathrm{nnz}(\boldsymbol{\theta}^\star) \le k$. Given observations $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}^\star + \boldsymbol{\xi}$, the goal is to output $\boldsymbol{\theta} \in \mathbb{R}^d$ such that $\mathrm{nnz}(\boldsymbol{\theta}) \le k$ and, for a universal constant $C > 0$,*

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_\infty \le C \left\|\mathbf{X}^\top \boldsymbol{\xi}\right\|_\infty . \tag{1}$$

Model 1 (Oblivious model) *The pair $(\boldsymbol{\theta}^\star, \boldsymbol{\xi})$ is chosen independently of $\mathbf{X}$.*

Model 2 (Adaptive model) *No independence assumption is imposed between $(\boldsymbol{\theta}^\star, \boldsymbol{\xi})$ and $\mathbf{X}$.*

Our results show that these two models have fundamentally different sample complexities. In the oblivious model, the optimal ℓ_∞ guarantee is attainable with $n \approx k \log d$ samples and nearly-linear running time. In particular, when $\boldsymbol{\xi} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$, this gives $\widetilde{O}(\sigma)$ error in (1). To our knowledge, this is the first such estimator at an $\widetilde{O}(k)$ sample complexity. We also prove that current best analysis for LASSO in Wainwright (2019) is strictly worse than ours when $n = o(k^2)$.

Theorem 1 *Let $n = \Omega(k \log d)$. There exists a nearly-linear-time estimator which solves Problem 1 under Model 1, with high probability.*

We further show that the adaptive model exhibits a different behavior: the required sample complexity is quadratic in k. We give matching upper and lower bounds up to logarithmic factors.

Theorem 2 *Let $n = \Omega\left(k^2 \log \frac{d}{k}\right)$. There exists a nearly-linear-time estimator which solves Problem 1 under Model 2, with high probability. Conversely, if $n = o(k^2)$, then no algorithm can solve Problem 1 under Model 2 with probability greater than $1/2$.*

Thus, up to logarithmic factors, Theorem 2 characterizes both the statistical and computational complexity of adaptive ℓ_∞ sparse recovery. This establishes, to our knowledge, the first separation between oblivious and adaptive sparse recovery in the ℓ_∞ sense. In the standard ℓ_2 setting, no such quadratic separation occurs, and $n \approx k \log d$ samples suffice even in adaptive formulations.[1]

Keywords: sparse recovery, variable selection, high-dimensional statistics

1. Extended abstract. Full version appears as Chen et al. (2026): `https://arxiv.org/abs/2602.16568v1`.

References

Ziyun Chen, Jerry Li, Kevin Tian, and Yusong Zhu. Separating oblivious and adaptive models of variable selection, 2026. URL https://arxiv.org/abs/2602.16568.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.