Lower Bounds on the Performance of Polynomial-time Algorithms for Sparse Linear Regression

Yuchen Zhang  Martin J. Wainwright  Michael I. Jordan
University of California, Berkeley

Statistical optimality vs. Computation tractability
Active research differentiating “optimal rate” and “computable optimal rate” in statistical estimation and machine learning:

Impossible to achieve Combinatorial algorithm Polynomial-time algorithm
Error Rate

• Sparse PCA detection (Berthet and Rigollet, 2013).
• Submatrix detection (Ma and Wu, 2013).
• Learning halfspaces (Daniely, Linial & Shalev-Shwartz, 2013)
• Sparse linear regression (our work)

Sparse Linear Regression: A Classical Problem
Observe a design matrix $X\in\mathbb{R}^{n\times d}$ and a response vector $y = X\theta^* + w$ such that

$\bullet w \sim N(0,\sigma^2 I_{d\times d}).$

$\bullet \theta^* \in \mathbb{R}^d$ is $k$-sparse, $k \ll d$.

Goal: find a $k$-sparse estimator $\hat{\theta}$ of $\theta^*$ such that the prediction loss $\mathbb{E}[\|X(\hat{\theta} - \theta^*)\|^2]$ is small.

Application: signal processing, financial data analysis, bioinformatics, imaging technology, etc.

Algorithms and Upper Bounds

Combinatorial Algorithm (NP-hard): $\ell_0$-based estimator $\hat{\theta}_0 := \arg\min_{\theta \in \mathbb{B}_0(k)} \|X\theta - y\|_2$:

$$\sup_{\theta^* \in \mathbb{B}_0(k)} \mathbb{E}\left[ \frac{1}{n} \|X(\hat{\theta}_0 - \theta^*)\|^2 \right] \lesssim \frac{\sigma^2 k \log(d)}{n}$$

Poly-time Algorithm: run Lasso, then truncate it to be $k$-sparse.

$$\sup_{\theta^* \in \mathbb{B}_0(k)} \mathbb{E}\left[ \frac{1}{n} \|X(\hat{\theta}_0 - \theta^*)\|^2 \right] \lesssim \frac{\sigma^2 k \log(d)}{\gamma^2 n}$$

where $\gamma < 1$ is the Restricted Eigenvalue of matrix $X$.

Key Observation: A $1/\gamma^2$ performance gap between the combinatorial estimator and the (known) poly-time estimator. Is there a better poly-time estimator?

Main Theoretical Result

Take-home Message

$\bullet$ There is a fundamental performance gap between poly-time algorithms and exponential-time algorithms for sparse linear regression.

$\bullet$ Gap is characterized by the restricted eigenvalue.

Theorem Assume NP $\not\subseteq$ P/poly. For any $\gamma > 0$ and any $(d, n, k)$ relation, there is an $X \in \mathbb{R}^{n\times d}$ with restricted eigenvalue $\gamma$, such that any $k$-sparse poly-time estimator $\hat{\theta}_{poly}$ satisfies:

$$\sup_{\theta^* \in \mathbb{B}_0(k)} \mathbb{E}\left[ \frac{1}{n} \|X(\hat{\theta}_{poly} - \theta^*)\|^2 \right] \gtrsim \frac{\sigma^2 k \log d}{\gamma^2 n}.$$