High Dimensional Robust Sparse Regression

Liu Liu Yanyao Shen Tianyang Li Constantine Caramanis

The University of Texas at Austin

Robust and High-Dimensional Stat. Workshop at Simons Institute.

November 1, 2018

Problem setup: robust estimation for sparse regression

Sparse regression model:

- dimensions: $n \ll d$.
- iid Gaussian X: $\Sigma = I_d$.
- $y_i = \boldsymbol{x}_i{}^T \boldsymbol{\beta}^* + \xi_i$.
- noise: $\xi_i \sim \mathcal{N}(0, \sigma^2)$.
- $\boldsymbol{\beta}^* \in \mathbb{R}^d$ is k-sparse.

Contamination model:

- We observe $z_i = (y_i, \boldsymbol{x}_i)$.
- $\{z_1, \cdots, z_n\} \sim (1-\epsilon)P + \epsilon Q$.
- P: sparse regression model.
- Q: *arbitrary* distribution.
- ϵ: *const fraction* of outliers.

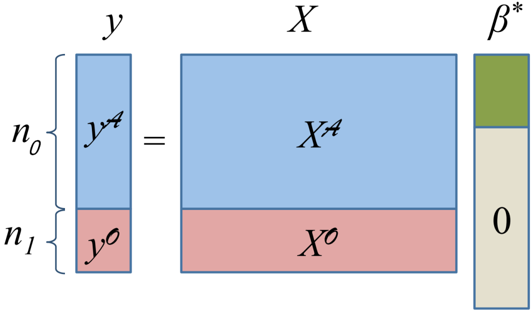
Problem setup: robust estimation for sparse regression

Sparse regression model:

- dimensions: $n \ll d$.
- iid Gaussian X: $\Sigma = I_d$.
- $y_i = \boldsymbol{x}_i{}^T \boldsymbol{\beta}^* + \xi_i$.
- noise: $\xi_i \sim \mathcal{N}(0, \sigma^2)$.
- $\boldsymbol{\beta}^* \in \mathbb{R}^d$ is k-sparse.

Contamination model:

- We observe $z_i = (y_i, \boldsymbol{x}_i)$.
- $\{z_1, \cdots, z_n\} \sim (1-\epsilon)P + \epsilon Q$.
- P: sparse regression model.
- Q: *arbitrary* distribution.
- ϵ: *const fraction* of outliers.



Approach

Algorithmic idea:

Iterative Hard Thresholding

$+$

Robust Sparse Mean Estimation on gradients

*[DKS16]: statistical query-based l.b. of $\Omega(k^2)$ on rob. sparse mean estimation.

Approach

Algorithmic idea:

Iterative Hard Thresholding

$+$

Robust Sparse Mean Estimation on gradients

Robust Sparse Mean Estimation:

Given ϵ-corrupted set of n samples from a d dimensional Gaussian $\mathcal{N}(\mu, I_d)$, how can we estimate μ, if μ is k-sparse?

---

*[DKS16]: statistical query-based l.b. of $\Omega(k^2)$ on rob. sparse mean estimation.

Approach

Algorithmic idea:

Iterative Hard Thresholding

$+$

Robust Sparse Mean Estimation on gradients

Robust Sparse Mean Estimation:

Given ϵ-corrupted set of n samples from a d dimensional Gaussian $\mathcal{N}(\mu, I_d)$, how can we estimate μ, if μ is k-sparse?

[BDLS17] shows that an efficient algorithm obtains $\|\widehat{\mu} - \mu\|_2 \leq O(\epsilon)$, with $n = \Omega(k^2 \log d/\epsilon^2)^*$. This is based on the ellipsoid algorithm in [DKK$^+$16].

---

*[DKS16]: statistical query-based l.b. of $\Omega(k^2)$ on rob. sparse mean estimation.

Our contribution

- Sparse regression algorithm that is resilient to a constant fraction of arbitrary outliers. Our algorithm requires $n = \Omega(k^2 \log d)$ samples.

† [Gao17]: this error rate is minimax optimal under the ϵ-contamination model.

Our contribution

- Sparse regression algorithm that is resilient to a constant fraction of arbitrary outliers. Our algorithm requires $n = \Omega(k^2 \log d)$ samples.
- Meta-theorem which allows the use of any robust sparse mean estimation subroutine:
  - By ellipsoid algorithm in [BDLS17], we can recover $\boldsymbol{\beta}^*$ within additive error $O(\epsilon\sigma)$.[†] **But this is computationally expensive**.

[†] [Gao17]: this error rate is minimax optimal under the ϵ-contamination model.

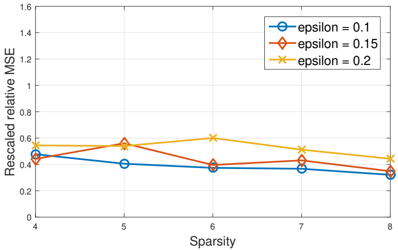
Our contribution

- Sparse regression algorithm that is resilient to a constant fraction of arbitrary outliers. Our algorithm requires $n = \Omega(k^2 \log d)$ samples.
- Meta-theorem which allows the use of any robust sparse mean estimation subroutine:
  - By ellipsoid algorithm in [BDLS17], we can recover $\boldsymbol{\beta}^*$ within additive error $O(\epsilon\sigma)$.† **But this is computationally expensive**.
- Efficient filtering algorithm for robust sparse mean estimation.
  - By this algorithm, we can recover $\boldsymbol{\beta}^*$ within additive error $O(\sqrt{\epsilon}\sigma)$.
  - The filtering algorithm is practical and **faster by at least** d^2.
- In particular: **exact recovery as** $\sigma \to 0$.

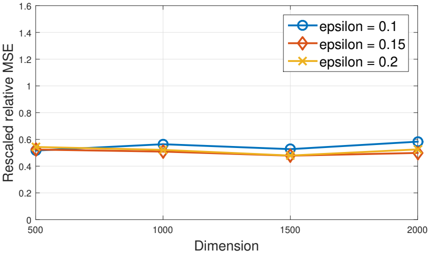
† [Gao17]: this error rate is minimax optimal under the ϵ-contamination model.

Experimental results I: robust sparse mean estimation

We generate authentic samples through $\boldsymbol{g}_i = \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{G}$, where $\boldsymbol{G}$ is k-sparse. The rescaled relative MSE: $\|\widehat{\boldsymbol{G}} - \boldsymbol{G}\|_2^2 / (\epsilon \|\boldsymbol{G}\|_2^2)$ should be independent of the parameters $\{\epsilon, k, d\}$.



(a) Rescaled relative MSE vs. sparsity.

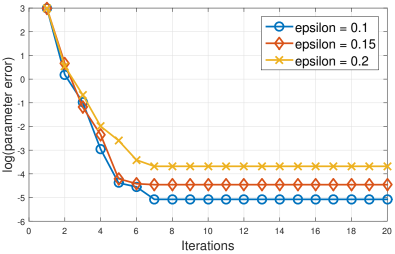


(b) Rescaled relative MSE vs. dimension.

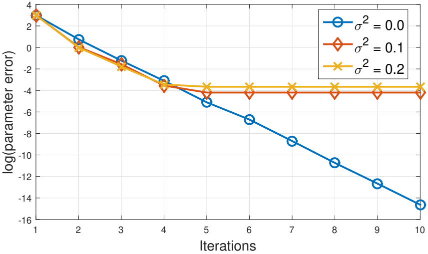
Figure 1: Sample complexity $n \propto k^2 \log(d)/\epsilon$. Different curves for $\epsilon \in \{0.1, 0.15, 0.2\}$ are the average of 15 trials.

Experimental results II: robust sparse regression

We use filtering algorithm as our RSGE, and generate authentic samples $y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^* + \xi_i$. As expected, the convergence is linear, and flatten out at the level of the final error.



(a) $\log(\|\boldsymbol{\beta}^t - \boldsymbol{\beta}^*\|_2^2)$ vs. iterates.



(b) $\log(\|\boldsymbol{\beta}^t - \boldsymbol{\beta}^*\|_2^2)$ vs. iterates.

Figure 2: In all cases, we fix $k = 5$, $d = 500$, and choose the sample complexity $n \propto 1/\epsilon$. (2a) has fixed $\sigma^2 = 0.1$. (2b) has fixed $\epsilon = 0.1$.

For more information please refer to our paper

Liu Liu, Yanyao Shen, Tianyang Li, Constantine Caramanis.
High Dimensional Robust Sparse Regression.
https://arxiv.org/abs/1805.11643

And please contact me if you have any questions

liuliu@utexas.edu

References

[BDLS17] Sivaraman Balakrishnan, Simon S. Du, Jerry Li, and Aarti Singh. Computationally efficient robust sparse estimation in high dimensions. In *Proceedings of the 2017 Conference on Learning Theory*, 2017.

[DKK+16] Ilias Diakonikolas, Gautam Kamath, Daniel M Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high dimensions without the computational intractability. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 655–664. IEEE, 2016.

[DKS16] Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. *arXiv preprint arXiv:1611.03473*, 2016.

[Gao17] Chao Gao. Robust regression via mutivariate regression depth. *arXiv preprint arXiv:1702.04656*, 2017.