NeurIPS 2026

Efficient algorithms for linear regression with heteroskedastic errors

Siddhant Chaudhary[1] Aditya Bhaskara[1]
Silvio Lattanzi[2] Sergei Vassilvitskii[2]

[1]University of Utah [2]Google Research





Model and goal

- Linear regression with independent Gaussian errors:

$$y_i = \langle x_i, \beta^* \rangle + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2). \tag{1}$$

- An unknown set S of at least m reliable sources has $\sigma_i \leq 1$. All other variances may be arbitrarily large, aka *Subset-of-Signals*.
- Covariates satisfy $\|x_i\|_2 \leq 1$ and $\epsilon_i \perp x_i$. The set S is fixed before covariates are sampled.
- Goal: efficiently recover $\beta^* \in \mathbb{R}^d$, even when $m \ll n$.

Motivation: sources have different noise levels

- Sensors and crowdsourcing produce observations of varying quality.
- We observe one sample per source; each source's variance is unknown.
- Ordinary least squares can amplify very noisy observations.
- Weighted least squares requires knowledge of the variances.
- Can a small reliable subset still allow accurate regression?

Approach 1: ℓ_1 regression

- Minimize the sum of absolute residuals:

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} |y_i - x_i^\top \beta|. \tag{2}$$

- A sample's gradient contribution depends on its residual's sign.
- Away from β^*, reliable samples favor steps toward β^* on average.
- The new guarantee requires well-spread random covariates on S; other covariates only need the norm bound.

Paper: Theorems 1.1 / 3.2.

Approach 2: iterative truncation

- Compute residuals from the current estimate: $r_i = y_i - x_i^\top \beta$.
- Clip residuals to $[-\delta, \delta]$: $\widetilde{r}_i = \mathrm{clip}(r_i, \pm\delta)$.
- Refit with ordinary least squares on corrected responses: $\widetilde{y}_i = x_i^\top \beta + \widetilde{r}_i$.
- Repeat within each stage, then halve δ. The error contracts toward a statistical floor.
- Requires a known initial error bound, comparable subset covariance, and fourth moment conditions on the reliable subset. Fixed designs are allowed.

Paper: Algorithm 1 / Theorem 4.1.

Recovery guarantees

- ℓ_1 regression, with well-spread reliable covariates:

$$\|\widehat{\beta} - \beta^*\|_2 \leq \widetilde{O}\left(\frac{d\sqrt{n}}{m}\right). \tag{3}$$

- Iterative truncation, with deterministic design conditions and initialization:

$$\|\widehat{\beta} - \beta^*\|_{\Sigma_n} \leq \widetilde{O}\left(\frac{\sqrt{dn}}{m}\right). \tag{4}$$

- If $\Sigma_n \asymp I/d$, both yield ℓ_2 error $\widetilde{O}(d\sqrt{n}/m)$.
- This error is subconstant when $m \gg d\sqrt{n}$, up to logarithmic factors.

$\Sigma_n = \frac{1}{n}\sum_i x_i x_i^\top$; $\|v\|_{\Sigma_n}^2 = \frac{1}{n}\sum_i (x_i^\top v)^2$. Bounds hold with high probability; $\widetilde{O}$ hides logarithmic factors.

Theorem conditions: $n \gtrsim d \log n$, $m \gtrsim \sqrt{nd}\log n$ for ℓ_1; $m \gtrsim \sqrt{dn}\ln n$ for truncation.

Takeaways

- Accurate regression is possible with unknown noise variances, even when some variances are unbounded.
- ℓ_1 regression: a new analysis of a standard estimator.
- Iterative truncation: residual clipping and stagewise contraction.
- A sufficiently large, informative reliable minority can give vanishing error.

Thank you.