Randomized tests for high-dimensional regression

Yue Li[1], Ilmun Kim[2], Yuting Wei[1]

Carnegie Mellon University[1]; University of Cambridge[2]

Background and Problem

- **Global testing problem.** $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)$ observations generated i.i.d. from a linear model

$$y_i = \langle \boldsymbol{x}_i, \boldsymbol{\beta} \rangle + \sigma z_i, \quad \boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$$

  for some unknown vector $\boldsymbol{\beta} \in \mathbb{R}^p$.

$$\textbf{Goal :} \quad H_0 : \boldsymbol{\beta} = \boldsymbol{0} \quad \text{versus} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{0}.$$

  Setting: n/p of constant order and $\boldsymbol{\beta}$ non-sparse.
- **Challenges.** Classical F-test does not work when $p \geq n$!
  - Impose specific structure assumptions [**ZC11, CGZ18, JJ14, JM14a, ACCP11**].
  - Can we find an *adaptive* and *general* approach?
- Solution: **random projection/sketching**
  - Widely studied in reducing computational cost and preserving privacy [**BM01, LKR05, Sar06, PW17**].
  - Statistical behaviors have been less studied.
- Close to our work: Kernel regression [**YPW17**], two-sample test [**LJW11**].

Contributions

- Propose a sketched F-test which does not restrain the size of n, p.
- Provide a systematic way of selecting the projection dimension based on the underlying intrinsic dimension.
- Characterize situations where our test enjoys better power than existing competitors.

Examples: choice of k ($k \leq r$)

With SVD $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, define $\widetilde{\boldsymbol{\beta}} = \boldsymbol{U}^\top \boldsymbol{\beta}$. Then

- **α-polynomial decay:** $\lambda_j \propto j^{-\alpha}$ with $\alpha > 1$ and homogeneous $\widetilde{\beta}_i$. We have
$$r \lesssim (\log p)^{\frac{1}{\alpha - 1}}.$$
- **γ-exponential decay:** $\lambda_j \propto \exp(-j^\gamma)$ with $\gamma > 0$ and homogeneous $\widetilde{\beta}_i$. We have
$$r \lesssim (\log \log p)^{\frac{1}{\gamma}}.$$
- structured coefficient: $0 < c_1 \leq \widetilde{\beta}_i \sqrt{i} \leq c_2$ and $\lambda_j \propto j^{-1}$. We have
$$r \lesssim (\log p)^3.$$

Algorithm

Algorithm Sketched F-test

Input: data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, response vector $\boldsymbol{y} \in \mathbb{R}^n$, a sketching dimension $k < n$
Output: global testing result for the linear model.
Step 1: generate a sketching matrix $S_k \in \mathbb{R}^{p \times k}$ with i.i.d. $\mathcal{N}(0,1)$ entries;
Step 2: compute the least square regression estimate $\widehat{\boldsymbol{\beta}}^S := (S_k^T \boldsymbol{X}^\top \boldsymbol{X} S_k)^{-1} S_k^\top \boldsymbol{X}^\top \boldsymbol{y}$;
Step 3: calculate the sketched F-test statistic

$$F(S_k) := \frac{\boldsymbol{y}^\top \boldsymbol{X} S_k \widehat{\boldsymbol{\beta}}^S / k}{\|\boldsymbol{y} - \boldsymbol{X} S_k \widehat{\boldsymbol{\beta}}^S\|_2^2 / (n-k)};$$

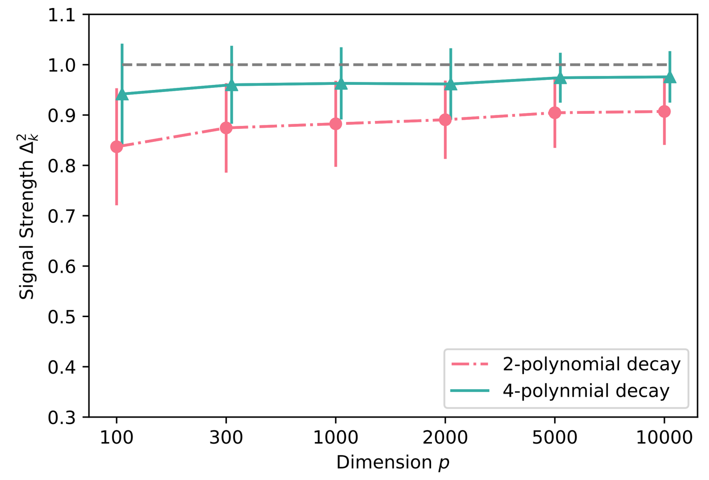
Step 4: if $F(S_k) \geq q_{\alpha,k,n-k}$, reject H_0; otherwise accept H_0.

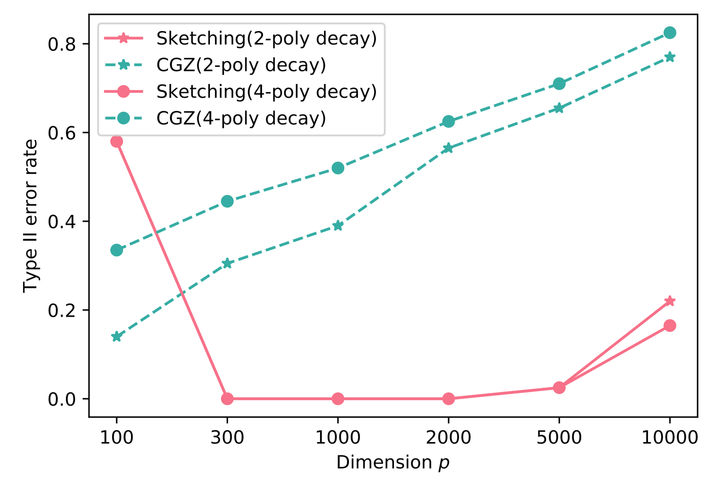
Numerical results

Numerical comparisons with CGZ18,ZC11 with decaying patterns: slow-decay (log) and fast-decay (polynomial) with $k = \lfloor n/2 \rfloor$.

$\Vert\boldsymbol{\Sigma}\Vert_F = 100$		H_0: $\Vert\boldsymbol{\beta}\Vert_2 = 0$	$\Vert\boldsymbol{\beta}\Vert_2 = 1$	$\Vert\boldsymbol{\beta}\Vert_2 = 5$
slow-decay	**Sketching**	3.2%	**1.4%**	**0.0%**
	CGZ	6.0%	5.4%	4.6%
	ZC	**2.1%**	16.8%	0.6%
fast-decay	**Sketching**	**4.0%**	**1.4%**	**2.4%**
	CGZ	6.2%	10.4%	12.4%
	ZC	4.2%	14.7%	6.3%

Asymptotic Behavior. With the structure design and optimal choice of k, we plot the signal strength and error v.s. feature dimension p.





Characterization of Power

The power of the proposed test is determined by

$$\Delta_k^2 := \boldsymbol{\beta}^\top \boldsymbol{\Sigma} S_k (S_k^\top \boldsymbol{\Sigma} S_k)^{-1} S_k^\top \boldsymbol{\Sigma} \boldsymbol{\beta}.$$

Theorem 1. *When the data and noise both follow Gaussian distributions and are independent to each other, the power of the proposed test satisfies*

$$\Psi_n^S(S_k) - \Phi\left(-z_\alpha + \sqrt{\frac{(1-\rho)n}{2\rho}} \frac{\Delta_k^2}{\sigma^2}\right) \to 0,$$

where $\rho_n = k/n \to \rho$.

Comparisons. Assuming the normalized vector $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}/\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}\|_2$ is uniformly distributed on the p-dimensional unit sphere independent of S_k, the proposed test has higher power than [**CGZ18**] *w.h.p.* if

$$\frac{4}{\sqrt{\rho(1-\rho)}} \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}} \frac{1}{\sqrt{n}} \leq 1.$$

Optimal choice of k. In this case, choose $k = \lfloor n/2 \rfloor$.

Optimality under Structure Design

- The model class with **intrinsic dimension** up to r is defined as, with some $\eta = o(1)$ and $\lambda_1 \geq \cdots \geq \lambda_p$ being eigenvalues of $\boldsymbol{\Sigma}$,

$$\sum_{i=r+1}^{p} \lambda_i \leq \eta \sum_{i=1}^{p} \lambda_i \quad \text{and} \quad r\lambda_{r+1} \leq \eta \sum_{i=1}^{p} \lambda_i.$$

- When we choose k proportional to the intrinsic dimension r of the model, we can **fully preserve the signal** *w.h.p.*!

Theorem 2. *Within the r-intrinsic dimensional model class, the proposed test is minimax rate optimal with radius*

$$\epsilon_n^2 = \frac{r^{1/2}}{n},$$

and the upper bound is reached by choosing sketching dimension $k = O(r)$.

References

[1] Y. Yang, M. Pilanci, and M. J. Wainwright, "Randomized sketches for kernels: Fast and optimal nonparametric regression," *The Annals of Statistics*, 2017.
[2] M. Lopes, L. Jacob, and M. J. Wainwright, "A more powerful two-sample test in high dimensions using random projection," in *Advances in Neural Information Processing Systems*, 2011, pp. 1206–1214.
[3] H. Cui, W. Guo, and W. Zhong, "Test for high-dimensional regression coefficients using refitted cross-validation variance estimation," *The Annals of Statistics*, 2018.