Shifting Inequality and Recovery of Sparse Signals

Lie Wang

Introduction

- The problem of recovering a high dimensional sparse signal based on a small number of measurements has attracted much attention recently.
- Model selection.
- Construction approximation.
- Compressive sensing.

Introduction

- Main model:

$$y = F\beta + z$$

- F is an n by p matrix, where n could be much less than p.
- Z is the vector of measurement error.
- β is the unknown vector of coefficients, our goal is to reconstruct β .

Introduction

- The error vector z can either be zero (noiseless case), bounded, or Gaussian (i.i.d. standard normal).
- β is assumed to be sparse, usually in terms of L_0 norm (number of nonzero coefficients).
- L_0 minimization is computationally undoable.

Methods

- In many cases the sparse solution can be found through L_1 minimization.
- This L_1 minimization problem has been studied, for example, in Fuchs (2004), Candes and Tao (2005) and Donoho (2006).

$$(P) \quad \min \|\gamma\|_1 \quad \text{subject to} \quad F\gamma = y.$$

Methods

- Noisy case, two L_1 minimization methods.
- Under L_2 constraint of residuals.

$$\min \|\gamma\|_1 \quad \text{subject to} \quad \|y - F\gamma\|_2 \leq \epsilon.$$

- Dantzig selector, by Candes and Tao

$$\min_{\gamma} \|\gamma\|_1 \quad \text{subject to} \quad \|F^T(y - F\gamma)\|_\infty \leq \lambda.$$

Conditions

- It is clear that regularity conditions are needed in order for these methods to be well behaved. Near orthogonal condition.
- Restricted Isometry Property (RIP).
- Candes and Tao considered sparse recovery problems in the RIP framework .

Conditions

- k-restricted isometry constant δ_k of F

$$\sqrt{1-\delta_k}\|c\|_2 \leq \|Fc\|_2 \leq \sqrt{1+\delta_k}\|c\|_2$$

  for any k sparse vector c.

- k k'-restricted orthogonality constant $\theta_{k,k'}$

$$|\langle Fc, Fc'\rangle| \leq \theta_{k,k'}\|c\|_2\|c'\|_2$$

  for any k and k' sparse vectors c, c' with disjoint support.

Conditions

- Different conditions on δ and θ have been used in the literature. For example, Candes and Tao (2007) imposes

$$\delta_{2k} + \theta_{k,2k} < 1$$

- Candes (2008) uses

$$\delta_{2k} < \sqrt{2} - 1.$$

- Actually, the second condition is stronger.

Noiseless Case

- Understanding the noiseless case is not only of interest on its own right, it also provides deep insight into the problem of reconstructing sparse signals in the noisy case.
- In this case, we need to recover the sparse signal exactly.

Noiseless Case

- (Candes and Tao) Let F be an n*p matrix. Suppose k>1 satisfies

$$\delta_k + \theta_{k,k} + \theta_{k,2k} < 1.$$

- Let β be a k-sparse vector and Y=F β . Then β is the unique minimizer to

$$\min \|\gamma\|_1 \quad subject\ to \quad F\gamma = y.$$

Unified Argument

- We found that all those results can be derived from the following elementary inequality (called shifting inequality):
- Suppose r⩽q ⩽3r, and

$$a_1 \geq a_2 \geq \cdots a_r \geq b_1 \geq \cdots b_q \geq c_1 \geq \cdots \geq c_r \geq 0$$

then $$\sqrt{\sum_{i=1}^{q} b_i^2 + \sum_{i=1}^{r} c_i^2} \leq \frac{\sum_{i=1}^{r} a_i + \sum_{i=1}^{q} b_i}{\sqrt{q+r}}.$$

Noiseless Case

- Our result:
- Let F be an n*p matrix. Suppose k>1 satisfies

$$\delta_{1.25k} + \theta_{1.25k,k} < 1.$$

and Y=F β . Then ,the minimizer to

$$\min \|\gamma\|_1 \quad subject\ to \quad F\gamma = y$$

satisfies

$$\|\hat{\beta} - \beta\|_2 \le C_0 k^{-\frac{1}{2}} \|\beta_{-\max(k)}\|_1$$

Noiseless Case

- Suppose the largest k element of β are the first k elements. Suppose $h = \hat{\beta} - \beta$

$$h_0 = (h(1), h(2), \cdots, h(k))$$

$$|h(k+1)| \geq |h(k+2)| \geq \cdots \geq |h(p)|$$

- We will use the following simple result:

$$||h_0||_1 \geq ||h_0^c||_1.$$

Cutting the error into pieces

$h = \hat{\beta}_q - \beta$:	h_0	h_*	h_1	h_2	h_3	$\cdots$	$\cdots$	$\cdots$
Length:	k	$\frac{k}{4}$	k	k	k	k	k	

- Cutting the error vector into pieces.
- Bound $\|h\|_2$ by $h(1)^2 + \cdots + h(k+r)^2$
- Bound $h(1)^2 + \cdots + h(k+r)^2$ by calculating

$$< Fh, F(h(1), h(2), \cdots, h(k+r), 0, \cdots, 0) >$$

Cutting the error into pieces

- First

$$< Fh, F(h(1), h(2), \cdots, h(k+r), 0, \cdots, 0) >= 0$$

- On the other hand

$$\begin{aligned}
&< Fh, F(h(1), h(2), \cdots, h(k+r), 0, \cdots, 0) > \\
&\geq \|F(h_0 + h_*)\|_2^2 - \sum_{i \geq 1} |< F(h_0 + h_*), Fh_i >| \\
&\geq \|h_0 + h_*\|_2^2 (1 - \delta_{k+r} - \theta_{k,k+r} \frac{\sum \|h_i\|_2}{\|h_0 + h_*\|_2})
\end{aligned}$$

Bounded Noise Case

- Suppose $y = F\beta + z$ and z belongs to some bounded set B.

$$\min \|\gamma\|_1 \quad \text{subject to} \quad y - F\gamma \in \mathcal{B}.$$

1. $B = \{z : \|F' z\|_\infty \le \eta\}$
2. $B = \{z : \|z\|_2 \le \eta\}$

Bounded Noise Case

- Our results improve Candes and Tao (2005, 2007) in the first case.

$$\|\hat{\beta}^{DS} - \beta\|_2 \leq \frac{\sqrt{10}}{1 - \delta_{1.25k} - \theta_{k,1.25k}} \cdot \sqrt{k}\eta.$$

- And improve Donoho, Elad and Temlyakov (2006) in the second case.

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \leq \frac{2\sqrt{2(1 + \delta_{1.25k})}}{1 - \delta_{1.25k} - \theta_{k,1.25k}} \cdot \eta.$$

Gaussian Noise Case

- We can apply the previous results to the Gaussian noise case.
- With high probability, the Gaussian noise vector belongs to $\{z : \|F^T z\|_\infty \leq \lambda\}$ with $\lambda = \sigma\sqrt{2 \log p}$
- With high probability, the Gaussian noise vector belongs to $\{z : \|z\|_2 \leq \epsilon\}$ with $\varepsilon = \sigma\sqrt{n + 2\sqrt{n \log n}}$.

Gaussian Noise Case

- We have the following results:
- With probability $P \geq 1 - \frac{1}{2\sqrt{\pi \log p}}$

$$\|\hat{\beta}^{DS} - \beta\|_2 \leq \frac{\sqrt{10}}{1 - \delta_{1.25k} - \theta_{k,1.25k}} \sqrt{k}\sigma\sqrt{2\log p}$$

- With probability $1 - \frac{1}{n}$

$$\|\hat{\beta}^{\ell_2} - \beta\|_2 \leq \frac{2\sqrt{2(1 + \delta_{1.25k})}}{1 - \delta_{1.25} - \theta_{k,1.25k}} \sigma\sqrt{n + 2\sqrt{n\log n}}.$$

Oracle Inequality

- We can also derive the oracle type of results.

$$y = X\beta + z \text{ with } \|X'z\|_\infty \leq \sqrt{2\log p},$$

- Suppose $\hat{\beta}$ is the minimizer to

$$\min \|\gamma\|_1 \quad \text{subject to} \quad \|X'(X\gamma - y)\|_\infty \leq \lambda_p$$

Oracle Inequality

- Then with high probability

$$\|\hat{\beta} - \beta\|_2^2 \le C^2 \lambda_p^2 \left(\sigma^2 + \sum_{i=1}^{p} \min(\beta^2(i), \sigma^2) \right)$$

- The idea of the proof is still the same, the application of our elementary inequality.

MIC

- Mutual Incoherent

$$\mu = \max |< F_i, F_j >|$$

- Instead of using RIC, we can put conditions on mutual incoherent. This type of condition is generally stronger, but much easier to check.

MIC

This type of conditions has been studied. For example, In Donoho, Elad, and Temlyakov (2006),

$$k < \frac{1}{4}(\frac{1}{\mu} + 1)$$

In Tseng (2009),

$$k < (\frac{1}{2} - O(\mu))\frac{1}{\mu} + 1.$$

MIC

We can improve the condition to

$$k < \frac{1}{2}\left(\frac{1}{\mu} + 1\right)$$

For the L_2 bounded noise case

$$\|\hat{\beta} - \beta\|_2 \leq C(\eta + \epsilon), \quad where\ C = \frac{\sqrt{3}}{1-(2k-1)\mu}.$$

For the L_∞ bounded noise case

$$\|\hat{\beta} - \beta\|_2 \leq C(\eta + \epsilon), \quad where\ C = \frac{\sqrt{2k+1}}{1-(2k-1)\mu}.$$

Future Work

- Further improve the condition, what is the best?
- For MIC, without any other constraint on F, our condition cannot be improved.
- For RIC, there is still room for improvement.
- Other type of conditions.