On the Fundamental Limits of Adaptive Sensing

Ery Arias-Castro, Emmanuel Candès and Mark Davenport

Lekshmi Ramesh

Indian Institute of Science
Bangalore

December 15, 2018

Introduction

- It is possible to reliably recover sparse signals from very few linear measurements
- Conventional schemes are non-adaptive, the measurement matrix is fixed beforehand
- Can adaptive schemes provide any advantage? For example, can we reduce the MSE if the rows of the measurement matrix are chosen adaptively?

Introduction

- Main result: For any adaptive sensing scheme and estimation procedure, it is impossible to significantly outperform random projection followed by ℓ_1 minimization

- Model
$$y = Ax + z$$
with $A \in \mathbb{R}^{m \times d}$ and $z \in \mathcal{N}(0, \sigma^2 I)$

- Let A be a random projection matrix with unit norm rows and let $\hat{x}$ be the output of the Dantzig selector. Then,
$$\mathbb{E}\|\hat{x} - x\|_2^2 \leq c\frac{k}{m}d\sigma^2 \log d$$
provided $m \geq k \log \frac{d}{k}$.
This scaling is optimal.

- Let $a_i^\top$ be the rows of A and recall

$$y_i = a_i^\top x + z_i, \quad i \in [m]$$

Adaptive scheme: a_i is a (possibly random) function of $(a_1, y_1, \ldots, a_{i-1}, y_{i-1})$.

- Main result (formal). Let $d \geq 2$, $k < \frac{d}{2}$ and consider any m. Then,

$$\inf_{\hat{x}} \sup_{\|x\|_0 \leq k} \mathbb{E}\|\hat{x} - x\|_2^2 \geq c\frac{k}{m} d\sigma^2.$$

Proof strategy

- First lower bound the MSE for x drawn from the following prior:

$$x_i = \begin{cases} \mu, & \text{w.p. } k/d \\ 0, & \text{w.p. } 1 - k/d. \end{cases}$$

  - Show lower bound for support recovery
  - Extend to MSE lower bound
- Extend to lower bound for arbitrary x

Support recovery error

- Can restrict to a_i that are deterministic functions of $(y_1, \ldots, y_{i-1})$
- Assumptions: $\|a_i\|_2 \leq 1$, $\sigma = 1$
- We first look at error in support recovery when adaptive schemes are allowed
- Error metric: expected Hamming distance

$$\mathbb{E}|\hat{S} \Delta S| = \sum_{i=1}^{d} \mathbb{P}(\hat{S}_i \neq S_i)$$

Support recovery error

- Result: Suppose x is sampled from the Bernoulli prior with $k < d/2$. Then any estimate $\hat{S}$ obeys

$$\mathbb{E}|\hat{S} \Delta S| \geq k\left(1 - \frac{\mu}{2}\sqrt{\frac{m}{d}}\right)$$

- If signal amplitude is low (say $\mu \leq \sqrt{d/m}$), then large number of errors ($\mathbb{E}|\hat{S} \Delta S| \geq k/2$)

Support recovery error

- Let $P_{1,j} = P(\cdot|x_j \neq 0)$ and $P_{0,j} = P(\cdot|x_j = 0)$ for any $j \in [d]$
- Let $\pi_1 = k/d$ and $\pi_0 = 1 - k/d$. Then

$$\mathbb{P}(\hat{S}_j \neq S_j) = \pi_1 P_{1,j}(\hat{S}_j = 0) + \pi_0 P_{0,j}(\hat{S}_j \neq 0)$$

- Optimizing over all tests, the Bayes risk is

$$B \geq \min(\pi_0, \pi_1)(1 - d_{TV}(P_{0,j}, P_{1,j}))$$

where d_{TV} denotes the total variation distance

Support recovery error

- Expected Hamming distance

$$
\begin{aligned}
\mathbb{E}|\hat{S}\Delta S| &= \sum_{j=1}^{d} P(\hat{S}_j \neq S_j) \\
&\geq \sum_{j=1}^{d} B_j \\
&\geq \pi_1 \sum_{j=1}^{d} (1 - d_{TV}(P_{0,j}, P_{1,j})) \\
&\geq k\left(1 - \frac{1}{\sqrt{d}}\sqrt{\sum_{j=1}^{d} d_{TV}^2(P_{0,j}, P_{1,j})}\right)
\end{aligned}
$$

Support recovery error

- Using

$$\mathbb{E}|\hat{S}\Delta S| \geq k\left(1 - \frac{1}{\sqrt{d}}\sqrt{\sum_{j=1}^{d} d_{TV}^2(P_{0,j}, P_{1,j})}\right)$$

and

$$\sum_{j=1}^{d} d_{TV}^2(P_{0,j}, P_{1,j}) \leq \frac{m\mu^2}{4}$$

we get the final result.
We now prove the second inequality.

Support recovery error

- We first upper bound $d_{TV}(P_{0,j}, P_{1,j})$. Using Pinsker's inequality

$$d_{TV}^2(P_{0,j}, P_{1,j}) \leq \frac{\pi_0}{2} D(P_{0,j} \| P_{1,j}) + \frac{\pi_1}{2} D(P_{1,j} \| P_{0,j})$$

- Let $P_{0,j} = P_0$ and $P_{1,j} = P_1$ and note

$$\begin{aligned} P_0 &= \sum_{x'} P(x') P(y_1, \ldots, y_m | x', x_j = 0) \\ &=: \sum_{x'} P(x') P_{0,x'} \end{aligned}$$

  where $x' = (x_1, \ldots, x_{j-1}, x_{j+1}, x_d)$

- Similar expression for P_1

Support recovery error

- Thus

$$D(P_0 \| P_1) = D(\sum_{x'} P(x') P_{0,x'} \| \sum_{x'} P(x') P_{1,x'})$$
$$\leq \sum_{x'} P(x') D(P_{0,x'} \| P_{1,x'})$$

using joint convexity of KL divergence

Support recovery error

- What are $P_{0,x'}$ and $P_{1,x'}$?
- Recall that

$$y_i = a_i^\top x + z_i = \sum_{l=1}^{d} a_{il} x_l + z_i$$

- Let $j \in [d]$. For $x_j = 0$,

$$y_i = c_i + z_i$$

and for $x_j = \mu$

$$y_i = c_i + \mu a_{ij} + z_i$$

where $c_i = \sum_{l \neq j} a_{il} x_l$

Support recovery error

- Thus,

$$P(y_i|x', x_j = 0) \equiv \mathcal{N}(c_i, \sigma^2)$$

  and

$$P(y_i|x', x_j = 1) \equiv \mathcal{N}(c_i + \mu a_{ij}, \sigma^2)$$

Support recovery error

- This gives

$$\begin{aligned} D(P_{0,x'} \| P_{1,x'}) &= D(P(\underline{y}|x', x_j = 0) \| P(\underline{y}|x', x_j = 1)) \\ &= \mathbb{E}_{P_{0,x'}} \sum_{i=1}^{m} \log \frac{P(y_i|x', x_j = 0)}{P(y_i|x', x_j = 1)} \\ &= \mathbb{E}_{P_{0,x'}} \sum_{i=1}^{m} \frac{1}{2} \Big((y_i - c_i - \mu a_{ij})^2 - (y_i - c_i)^2 \Big) \\ &= \frac{\mu^2}{2} \sum_{i=1}^{m} \mathbb{E}_{P_{0,x'}} a_{ij}^2 \end{aligned}$$

Support recovery error

- Finally,

$$D(P_0 \| P_1) \leq \sum_{x'} P(x') D(P_{0,x'} \| P_{1,x'})$$
$$\leq \sum_{x'} P(x') \frac{\mu^2}{2} \sum_{i=1}^{m} \mathbb{E}(a_{ij}^2 | x_j = 0)$$

- Similarly,

$$D(P_1 \| P_0) \leq \sum_{x'} P(x') \frac{\mu^2}{2} \sum_{i=1}^{m} \mathbb{E}(a_{ij}^2 | x_j \neq 0)$$

Support recovery error

- Relating to the total variation distance,

$$\sum_{j=1}^{d} d_{TV}^2(P_{1,j}, P_{0,j}) \leq \frac{\mu^2}{4} \sum_{i,j} \mathbb{E} a_{ij}^2$$
$$= \frac{m\mu^2}{4}$$

Support recovery error

- Lower bound on expected Hamming distance

$$\mathbb{E}|\hat{S}\Delta S| \geq k\left(1 - \frac{1}{\sqrt{d}}\sqrt{\sum_{j=1}^{d} d_{TV}^2(P_{1,j}, P_{0,j})}\right)$$

$$\geq k\left(1 - \frac{\mu}{2}\sqrt{\frac{m}{d}}\right)$$

Connecting to MSE

- Let $S = \mathrm{supp}(x)$ and $\hat{S} = \{j : \hat{x}_j \geq \mu/2\}$. Then,

$$\begin{aligned} \|\hat{x} - x\|_2^2 &= \sum_{j \in S} (\hat{x}_j - x_j)^2 + \sum_{j \in S^C} \hat{x} J^2 \\ &\geq \frac{\mu^4}{2} |S \backslash \hat{S}| + \frac{\mu^2}{4} |\hat{S} \backslash S| \\ &= \frac{\mu^2}{4} |\hat{S} \Delta S| \end{aligned}$$

Connecting to MSE

- Taking expectation,

$$\mathbb{E}\|\hat{x} - x\|_2^2 \geq \frac{\mu^2}{4}\mathbb{E}|\hat{S}\Delta S| \geq \frac{\mu^2}{4}k\left(1 - \frac{\mu}{2}\sqrt{\frac{m}{d}}\right)$$

- Final step: bound for arbitrary x

Reference

- *E. A.-Castro, E. J. Candès and M. A. Davenport*. On the Fundamental Limits of Adaptive Sensing. In IEEE Transactions on Information Theory, January 2013.