Sparse solutions of underdetermined systems

I-Liang Chern

September 22, 2016

Outline

- Sparsity and Compressibility: the concept for measuring sparsity and compressibility of data
- Minimum measurements to recover sparse data: Find a measurement $\mathbf{A}$ so that $\mathbf{A}\mathbf{x} = \mathbf{y}$ is solvable
  - For general sparse data
  - For a specific sparse data $\mathbf{x}$

This is a note from S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing, Springer 2013.

Sparsity

- Notations:
  - $[N] := \{1, ..., N\}$, $\bar{S} := [N] \backslash S$;
  - A vector $\mathbf{x} \in \mathbb{C}^N$, its support is
    Supp $(\mathbf{x}) := \{j \,|x_j \neq 0\}$.
  - For $S \subset [N]$, $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{x}_S$ is defined by

$$x_{S,i} = \begin{cases} x_i & \text{if } i \in S \\ 0 & \text{otherwise.} \end{cases}$$

  - s-sparse vectors $\Sigma_s = \{\mathbf{x} \in \mathbb{C}^N, \mathbf{x}$ is s-sparse.$\}$
- Definitions:
  - $\|\mathbf{x}\|_0 := \#$ Supp $(\mathbf{x})$;
  - $\|\mathbf{x}\|_p := \left(\sum_{j=1}^N |x_j|^p\right)^{1/p}$, $0 < p < \infty$; $\|\mathbf{x}\|^p \to \|\mathbf{x}\|_0$ as $p \to 0$.

Sparsity

- Non increasing rearrangement of $\mathbf{x}$ is defined by

$$x_1^* \geq x_2^* \geq \cdots \geq x_N^* \geq 0$$

  where $x_j^* = |x_{\pi(j)}|$ and π is a permutation on $[N]$.
- Given $\mathbf{x}$, we can find its s-sparse projection $\mathbf{x}_s^*$, the part of component of $\mathbf{x}$ which contains s largest absolute values of $\mathbf{x}$. Such $\mathbf{x}_s^*$ may not be unique.
- A vector $\mathbf{x}$ is s-sparse if $\mathbf{x} = \mathbf{x}_s^*$
- Best s-term approximation: An $\mathbf{x}_s^*$ of $\mathbf{x}$.

Compressibility

- Best s-term approximation error in p-norm ($p \geq 1$)

$$\sigma_s(\mathbf{x})_p := \inf\{\|\mathbf{x} - \mathbf{z}\|_p |\ \mathbf{z} \in \Sigma_s\}$$

- A vector is called compressible if $\sigma_s(\mathbf{x})_p$ decays fast in s. (e.g. $\sigma_s(\mathbf{x})_p = O(s^{-r})$ for some $p \geq 1$ and some $r > 1$.)
- Many data are compressible. The following theorem says that $\sigma_s(\mathbf{x})_q$ can be controlled by $\|\mathbf{x}\|_p$ with $0 < p < q$.

Compressibility of a vector

Theorem

The following inequality holds

$$\boxed{\sigma_s(\mathbf{x})_q \leq \frac{c_{p,q}}{s^{1/p-1/q}} \|\mathbf{x}\|_p,}$$

where $0 < p < q$ *and*

$$c_{p,q} := \left[\left(\frac{p}{q}\right)^{p/q} \left(1 - \frac{p}{q}\right)^{1-p/q} \right]^{1/p} \leq 1.$$

Remarks

1. For $q = 2$, $p \leq 1$, we have

$$\sigma_s(\mathbf{x})_2 \leq \frac{1}{s^{1/p-1/2}} \|\mathbf{x}\|_p.$$

   This suggests that the unit ball in ℓ_p quasi-norm for $p \leq 1$ are good models for compressible vectors.

2. In particular, $p = 1$,

$$\sigma_s(\mathbf{x})_2 \leq \frac{1}{2\sqrt{s}} \|\mathbf{x}\|_1.$$

   The vectors with $\|\mathbf{x}\|_1 \leq 1$ can also serve as compressible vectors.

Proof.

1. Set $\alpha_j = (x_j^*)^p/\|\mathbf{x}\|_p^p$, where $\mathbf{x}^*$ is a non-increasing rearrangement of $\mathbf{x}$, then the theorem is equivalent to:

$$(\alpha_1 \geq \cdots \geq \alpha_N, \quad \alpha_1 + \cdots + \alpha_N \leq 1)$$

implies

$$\alpha_{s+1}^{q/p} + \cdots + \alpha_N^{q/p} \leq \frac{c_{p,q}^q}{s^{q/p-1}}.$$

This is a constrained optimization problem: let $r = q/p > 1$, we want to maximize

$$f(\alpha_1, ..., \alpha_N) := \alpha_{s+1}^r + \cdots + \alpha_N^r$$

over a convex polygon

$$\mathcal{C} := \{(\alpha_1, ..., \alpha_N) | \alpha_1 \geq ... \geq \alpha_N \geq 0, \alpha_1 + \cdots + \alpha_N \leq 1\}$$

2 The maximum occurs at corners.

- If $\alpha_1 = \cdots = \alpha_N = 0$, then $f(\alpha) = 0$.
- If $\alpha_1 + \cdots + \alpha_N = 1$,
  $\alpha_1 = \cdots = \alpha_k > \alpha_{k+1} = \cdots = \alpha_N = 0$, $1 \leq k \leq s$, then $f(\alpha) = 0$.
- If $\alpha_1 + \cdots + \alpha_N = 1$,
  $\alpha_1 = \cdots = \alpha_k > \alpha_{k+1} = \cdots = \alpha_N = 0$, $s+1 \leq k \leq N$, then $\alpha_i = 1/k$ for $1 \leq i \leq k$ and $f(\alpha) = (k-s)/k^r$.

It follows that

$$\max_{\alpha \in \mathcal{C}} f(\alpha) = \max_{s+1 \leq k \leq N} \frac{k-s}{k^r}$$

3 Taking k as a continuous variable, we obtain maximum at $k^* = (r/(r-1))s$. Hence

$$\max_{\alpha \in \mathcal{C}} f(\alpha) \leq \frac{1}{r}\left(1 - \frac{1}{r}\right)^{r-1} \frac{1}{s^{r-1}} = c_{p,q}^q \frac{1}{s^{q/p-1}}$$

Minimal Measurements for Reconstructing Sparse Vectors

- Linear measurement: $\mathbf{y} = A\mathbf{x}$. A is $m \times N$.
- Question: minimal number of linear measurements needed to reconstruct s-sparse vectors from these measurements, regardless of the practicality of the reconstruction scheme.
- Two meanings:
  - Reconstruction of all s-sparse vectors
  - Given a specific s-sparse $\mathbf{x}$, construct a measurement A such that it can recover $\mathbf{x}$ exactly.
- Minimal measurements:
  - For the first, $m = 2s$.
  - For the second, $m = s + 1$.

Remarks

- The meaning for the second is in the measure sense, which means that the Lebesgue measure of the set of matrices which cannot recover a given specific s-sparse vector $\mathbf{x}$ is zero.
- For the recovery of all vectors, however, the reconstruction may not be stable, in the sense that a small perturbation of $\mathbf{y}$ causes large change of $\mathbf{x}$. If we add the stability requirement, then the minimal measurements become

$$m \geq Cs \ln(N/s),$$

where C depends on the stability criterion.

Recover all s-sparse vectors

Theorem

Given $A \in \mathbb{C}^{m \times N}$, the following properties are equivalent:

(a) *Every s-sparse vector $x \in \mathbb{C}^N$ is the unique s-sparse solution of $Az = Ax$, that is, if $Ax = Az$ and $x, z \in \Sigma_s$, then $x = z$.*

(b) *The null space $ker A$ does not contain any $2s$-sparse vector other than the zero vector,that is, $ker A \cap \Sigma_{2s} = \{0\}$.*

(c) *For every $S \subset [N]$ with $|S| \leq 2s$, the submatrix A_S is injective as a map from $\mathbb{C}^S$ to $\mathbb{C}^m$.*

(d) *Every set of $2s$ columns of A is linearly independent.*

Proof

1. (b)$\Rightarrow$ (a): If $Az = Ax$, then $z - x \in ker A$. Since both z and x are s-sparse, we get $z - x$ is $2s$-sparse. From (b), $z - x = 0$.
2. (a) $\Rightarrow$ (b): For any $\mathbf{v} \in ker A \cap \Sigma_{2s}$, we can split $\mathbf{v} = \mathbf{z} - \mathbf{x}$ such that $\operatorname{supp} \mathbf{z} \cap \operatorname{supp} \mathbf{x} = \emptyset$ and both $\mathbf{x}$ and $\mathbf{z}$ are s-sparse. By (a), $\mathbf{z} = \mathbf{x}$. Thus, $\mathbf{v} = 0$.
3. (b)$\Leftrightarrow$(c)$\Leftrightarrow$(d): For any $S \subset [N]$, noting that $A\mathbf{v} = A_S \mathbf{v}_S$. Use $|S| = dim\, Dom(A_S) = dim\, ker A_S + dim\, Ran A_S$, we get that $ker A = \{0\}$ if and only if $dim\, Ran A_S = |S|$.

Recover all s-sparse vectors

- We look for a measurement matrix with $2s \times N$ which can recover all s-sparse vectors.
- Answer 1: Given any $0 < t_1 < \cdots < t_N$, define

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_N \\ t_1^2 & t_2^2 & \cdots & t_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1^{2s-1} & t_2^{2s-1} & \cdots & t_N^{2s-1} \end{pmatrix}$$

Let $S \subset [N]$ with $\text{card}(S) = 2s$. Any column sub matrix A_S is a Vandermonde matrix, which is invertible.

Recover all s-sparse vectors

- Answer 2: Fourier sub matrix, we choose $\omega_j = e^{2\pi i (j-1)/N}$

$$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \omega_1 & \omega_2 & \cdots & \omega_N \\ \omega_1^2 & \omega_2^2 & \cdots & \omega_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ \omega_1^{2s-1} & \omega_2^{2s-1} & \cdots & \omega_N^{2s-1} \end{pmatrix}$$

Recover a specific s-sparse vector

- Given an s-sparse vector, find an $(s+1) \times N$ matrix A so that the measurement $\mathbf{y} = A\mathbf{x}$ can recover $\mathbf{x}$ exactly via (P0).
- Answer: The set of all $(s+1) \times N$ matrices which cannot do so has measure zero.