AMATH 840: Advanced Numerical Methods for Computational and Data Sciences

Giang Tran

Department of Applied Mathematics, University of Waterloo

Jan 26, 2021

Lecture 07: Sparse Recovery of Basis Pursuit (cont'd)

- Given $\mathbf{y} \in \mathbb{C}^n$ and $A \in \mathbb{C}^{m \times n}$, solve the basis pursuit (BP) problem:

$$\min_{\mathbf{z} \in \mathbb{C}^n} \|\mathbf{z}\|_1 \quad s.t. \quad \|\mathbf{y} - A\mathbf{z}\| \leq \eta, \quad \text{where } \eta \geq 0.$$

- Goal: Study conditions on A that provide reconstruction guarantees of (BP).
  - Last time: Null Space Property
  - Today:
    - Coherence Condition
    - RIP Condition

From now on, we use the Theorems' numbers from the reference.

Recall: Null Space Property, $\ell_1 \Rightarrow \ell_0$

- The ℓ_1-error of best s-term approximation of a vector $\mathbf{x} \in \mathbb{K}^n$:

$$\sigma_s(\mathbf{x})_1 = \inf\{\|\mathbf{x} - \mathbf{z}\|_1 : \ \mathbf{z} \in \mathbb{K}^n \text{ and } \mathbf{z} \text{ is } s\text{-sparse}\}$$

- A matrix $A \in \mathbb{K}^{m\times n}$ is said to satisfy the null space property of order s

$$\begin{aligned}
&\Leftrightarrow \|v_S\|_1 < \|v_{S^c}\|_1, \ \forall \ v \in \ker A \setminus \{0\}, \ \forall S \subset [n] \text{ with } |S| \le s. \\
&\Leftrightarrow 2\|v_S\|_1 < \|v\|_1, \ \forall v \in \ker A \setminus \{0\} \text{ and} \\
&\qquad\qquad\qquad S = \{\text{indices of } s \text{ largest ab.entries of } v\}. \\
&\Leftrightarrow \|v\|_1 < 2\sigma_s(v)_1, \quad \text{for all } v \in \ker A \setminus \{0\}.
\end{aligned}$$

- **Theorems 4.4** [1]. Given $A \in \mathbb{K}^{m\times n}$, every s-sparse vector $w \in \mathbb{K}^n$ is the unique solution of

$$\min_{z\in\mathbb{K}^n} \|z\|_1 \quad \text{s.t.} \quad Aw = Az$$

if and only if A satisfies the null space property of order s.

[1] A Mathematical Introduction to Compressive Sensing, by S. Foucart & H. Rauhut.

Recall: Stable and Robust Null Space Property

- **Theorem 4.19.** Let $A \in \mathbb{C}^{m \times n}$ and $\|\cdot\|$ be a norm on $\mathbb{C}^m$. Suppose there exist constants $\rho \in (0,1)$ and $\tau > 0$ s.t.

$$\|v_S\|_1 \leq \rho \|v_{S^c}\|_1 + \tau \|Av\| \quad \forall v \in \mathbb{C}^n,\ \forall S \subset [n] \text{ with } |S| \leq s. \quad (1)$$

  Let $\mathbf{w} \in \mathbb{C}^n$ and $\mathbf{y} = A\mathbf{w} + \mathbf{e}$ with $\|e\| \leq \eta$. Then a solution $\mathbf{w}^\#$ of the ℓ_1-minimization problem

$$\min_{z \in \mathbb{C}^n} \|\mathbf{z}\|_1 \quad s.t. \quad \|y - A\mathbf{z}\| \leq \eta$$

$\|v_S\|_1 \lesssim \sqrt{s}\|v_S\|_2$

  approximates the vector $\mathbf{w}$ with ℓ_1-error:

$$\|w - w^\#\|_1 \leq \frac{2(1+\rho)}{1-\rho}\sigma_s(w)_1 + \frac{4\tau}{1-\rho}\eta.$$

- **Theorem 4.12.** If $\eta = 0$, $\tau = 0$, and we only require that condition (1) holds for $v \in \ker A$, we have the stable sparse recovery result.

Recall: ℓ_2-Robust Null Space Property

- A matrix $A \in \mathbb{C}^{m\times n}$ is said to satisfy the ℓ_2-robust null space property of order s w.r.t. $\|\cdot\|$ with constants $0 < \rho < 1$ and $\tau > 0$ if

$$\|v_S\|_2 \le \frac{\rho}{s^{1/2}}\|v_{S^c}\|_1 + \tau\|Av\| \quad \forall v \in \mathbb{C}^n,\ \forall S \subset [n] \text{ with } |S| \le s.$$

- **Theorem 4.22.** Suppose the matrix $A \in \mathbb{C}^{m\times n}$ is said to satisfy the ℓ_2-robust null space property of order s w.r.t. $\|\cdot\|_2$ with constants $0 < \rho < 1$ and $\tau > 0$. Then for any $w \in \mathbb{C}^n$, a solution $w^\#$ of the BPDN:

$$\min_z \|z\|_1 \quad s.t. \quad \|Az - y\|_2 \le \eta,$$

with $y = Aw + e$ and $\|e\|_2 \le \eta$ approximates the vector w with:

$$\|w - w^\#\|_1 \le C\,\sigma_s(w)_1 + D\sqrt{s}\,\eta,$$

$$\|w - w^\#\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(w)_1 + D\eta,$$

for some constants C, D depending only on ρ and τ.

If $\eta \to 0$ and $\sigma_s(w)_1 \to 0$, then $w^\# \approx w$.

Training and Generalization Errors Estimation

From the error estimations on the solution, we can derive the corresponding generalization error. For example,

Suppose $A \in \mathbb{C}^{m\times n}$ satisfies the ℓ_2-robust NSP of order s with constants $\rho \in (0,1)$ and $\tau > 0$. Given $\mathbf{y} = A\mathbf{w} + \mathbf{e}$ with $\|e\|_2 \leq \eta$. From Theorem 6.8., any solution $\mathbf{w}^{\#}$ of the ℓ_1-minimization problem

$$\min_{\mathbf{z}\in\mathbb{C}^n} \|\mathbf{z}\|_1 \quad s.t. \quad \|y - A\mathbf{z}\|_2 \leq \eta$$

Recall $p,q \geq 1$

$$\|A\|_{p\to q} := \sup_{x\neq 0} \frac{\|Ax\|_q}{\|x\|_p}$$

approximates the vector $\mathbf{w}$ with

$$\|\mathbf{w} - w^{\#}\|_1 \leq C\sigma_s(w)_1 + D\sqrt{s}\eta,$$

for some constants $C, D > 0$ depending only on ρ and τ, Therefore,

$$\|\mathbf{y} - A\mathbf{w}^{\#}\|_2 \leq \|\mathbf{y} - A\mathbf{w}\|_2 + \|A\mathbf{w} - A\mathbf{w}^{\#}\|_2 = \|\mathbf{e}\|_2 + \|A\|_{1\to 2}\|\mathbf{w} - \mathbf{w}^{\#}\|_1$$
$$\leq \eta + \|A\|_{1\to 2}\left(C\sigma_s(w)_1 + D\sqrt{s}\eta\right).$$

Coherence Condition for Basis Pursuit

Recall: Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns $a_1, \ldots, a_n$.

- The ℓ_1-coherence function μ_1 of A is defined for $s \in [n-1]$ by

$$\mu_1(s) := \max_{k \in [n]} \max \left\{ \sum_{j \in S} |\langle a_k, a_j \rangle|, S \subset [n], |S| = s, k \notin S \right\}.$$

- The coherence $\mu = \mu(A)$ of the matrix A is defined as

$$\mu = \mu(A) := \max_{1 \leq k \neq j \leq n} |\langle a_k, a_j \rangle|.$$

Theorem 5.15. Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns. If

$$\mu_1(s) + \mu_1(s-1) < 1,$$

then every s-sparse vector $\mathbf{w} \in \mathbb{C}^n$ is exactly recovered from the measurement $\mathbf{y} = A\mathbf{w}$ via basis pursuit.

Coherence Condition for Basis Pursuit

Sketch Proof. We will prove that if $\mu_1(s) + \mu_1(s-1) < 1$, then A satisfies the NSP:

$$\|v_S\|_1 < \|v_{S^c}\|_1 \quad \forall v \in \ker A \setminus \{0\},\ \forall S \subset [n] \text{ with } |S| \le s.$$

Take $v = (v_1, \ldots, v_n)^T \in \ker A \setminus \{0\}$ and $S \subset [n]$ with $|S| \le s$. Then,

$$0 = Av = \sum_{j=1}^{n} v_j a_j$$

$$0 = \langle 0, v_k \rangle = \left\langle \sum_{j=1}^{n} v_j a_j, a_k \right\rangle = v_k + \sum_{j=1, j\neq k}^{n} v_j \langle a_j, a_k \rangle$$

$$v_k = -\sum_{j=1, j\neq k}^{n} v_j \langle a_j, a_k \rangle = -\sum_{j\in S^c, j\neq k} v_j \langle a_j, a_k \rangle - \sum_{\ell\in S, \ell\neq k} v_\ell \langle a_\ell, a_k \rangle$$

$$\sum_{k\in S} |v_k| \le \cdots$$

Some Properties of Coherence

$$\mu_1(s) := \max_{k\in[n]} \max\left\{ \sum_{j\in S} |\langle a_k, a_j\rangle|, S \subset [n], |S| = s, k \notin S \right\}$$

$$\mu = \mu(A) := \max_{1\le k\neq j\le n} |\langle a_k, a_j\rangle|.$$

Theorem 5.3. Let $A \in \mathbb{C}^{m\times n}$ be a matrix with ℓ_2-normalized columns and let $s \in [n]$. Then

1. $\mu \le \mu_1(s) \le s\mu \le s$, for all $1 \le s \le n-1$.
2. $\max\{\mu_1(s), \mu_1(t)\} \le \mu_1(s+t) \le \mu_1(s) + \mu_1(t)$, for all $1 \le s, t \le n-1$ with $s+t \le n-1$.
3. For all s-sparse vector $\mathbf{x} \in \mathbb{C}^n$, we have

$$(1 - \mu_1(s-1))\|\mathbf{x}\|_2^2 \le \|A\mathbf{x}\|_2^2 \le (1 + \mu_1(s-1))\|\mathbf{x}\|_2^2.$$

Proof. See Lecture0304 slides or Theorem 5.3. in the Reference.

Some Properties of Coherence

Theorems 5.7& 5.8 (Welch bound). Let $A \in \mathbb{C}^{m \times n}$ be a matrix with $n \geq m$ and ℓ_2-normalized columns and let $s \in [n]$. Then

1. The coherence of A satisfies

$$\mu \geq \sqrt{\frac{n-m}{m(n-1)}}.$$

2. The ℓ_1-coherence of A satisfies

$$\mu_1(s) \geq s\sqrt{\frac{n-m}{m(n-1)}} \quad \text{whenever} \quad s < \sqrt{n-1}.$$

Equality holds iff there exist constants $c \geq 0$ and $\lambda > 0$ s.t.

$$|\langle a_i, a_j \rangle| = c, \quad \forall i, j \in [n],\ i \neq j; \quad \text{and} \quad AA^* = \frac{1}{\lambda}\mathrm{Id}_m.$$

If those conditions are satisfied, we say the columns of A form an equiangular tight frame.

Some Properties of Coherence

Proof. 1. See Theorem 5.7. in the Reference.
The main idea is to evaluate $\mathrm{tr}(A^*A)$ and $\mathrm{tr}(AA^*)$ using the following properties: For any matrix $H \in \mathbb{C}^{m\times m}$ and $B \in \mathbb{C}^{n\times m}$,

$$\mathrm{tr}(H) = \langle H, \mathrm{Id}_m\rangle_F \leq \|H\|_F \|\mathrm{Id}_m\|_F = \sqrt{m}\sqrt{\mathrm{tr}(HH^*)}.$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

2. See Theorem 5.8. in the Reference.

Remark. If the equality of the Welch bound holds, m cannot be arbitrarily large (see Theorem 5.10. in Reference). Indeed, let $A \in \mathbb{K}^{m\times n}$ be a matrix with $n \geq m$ and ℓ_2-normalized columns. If the columns of A form an equiangular tight frame, then

1. $n = \dfrac{m(m+1)}{2}$ when $\mathbb{K} = \mathbb{R}$.
2. $n = m^2$ when $\mathbb{K} = \mathbb{C}$.

Number of Measurements for Basis Pursuit Using Coherence Condition

Summary: Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns.

- Coherence condition: If $\mu_1(s) + \mu_1(s-1) < 1$, then every s-sparse vector $\mathbf{w} \in \mathbb{C}^n$ is exactly recovered from the measurement $\mathbf{y} = A\mathbf{w}$ via basis pursuit.
- Welch bound:

$$\mu(A) \geq \sqrt{\frac{n-m}{m(n-1)}}.$$

$$1 > \mu_1(s) + \mu_1(s-1) \geqslant (2s-1)\mu(A) \geqslant (2s-1)\sqrt{\frac{n-m}{m(n-1)}}$$

$$m(n-1) > (2s-1)^2(n-m)$$

$$m(n-1+(2s-1)^2) > n(2s-1)^2$$

$$m\left(1+\frac{4s^2-4s}{n}\right) > (2s-1)^2 \quad \text{So } m \geqslant Cs^2 \quad (\text{if } n \to \infty)$$

Number of Measurements for Basis Pursuit Using Coherence Condition

$(2\mu-1)s \leq \frac{2\tilde{c}}{\sqrt{m}} s < 1$ for $\tilde{c} = \sqrt{\frac{C}{5}}$

$\tilde{c}$ Then

- So, if $m \geq Cs^2$ and $\mu \leq \frac{c}{\sqrt{m}}$, every s-sparse vector $w \in \mathbb{K}^n$ is is exactly recovered from the measurement $\mathbf{y} = A\mathbf{w}$ via basis pursuit.

Remark. Using coherence condition for (BP), we cannot relax the quadratic in $m \geq Cs^2$. For example, choose

$$m = (2s-1)^2/2, \quad n \geq 2m, \quad s \leq \sqrt{n-1}.$$

Then

$$\mu_1(s) + \mu_1(s-1) > 1.$$

Restricted Isometry Property

+) If $\delta_{2s} < \frac{4}{\sqrt{41}}$, then $m \geq Cs\log(\frac{n}{s})$.

+) If $m \geq Cs\log(\frac{n}{s})$ + A random matrix $\Rightarrow \delta_{2s} < \frac{4}{\sqrt{41}}$

Definition. The s^{th} restricted isometry constant $\delta_s = \delta_s(A)$ of a matrix $A \in \mathbb{C}^{m\times n}$ is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2$$

for all s-sparse vector $\mathbf{x} \in \mathbb{C}^n$. Equivalently,

$$\delta_s = \max_{S\subset[n],|S|\leq s} \|A_S^* A_S - \mathrm{Id}\|_{2\to 2}.$$

If $\delta \approx 0$, $\|Ax\|_2^2 \approx \|x\|_2^2$ $\forall$ s-sparse vector $x \in \mathbb{C}^n$.

RIP Theorems

Theorem 6.12. Suppose $A \in \mathbb{C}^{m \times n}$ satisfies the RIP condition:

BP	IHT	HTP	OMP
$\delta_{2s} < \frac{4}{\sqrt{41}}$ ≈ 0.6246	$\delta_{6s} < \frac{1}{\sqrt{3}}$ ≈ 0.5773	$\delta_{6s} < \frac{1}{\sqrt{3}}$ ≈ 0.5773	$\delta_{13s} < \frac{1}{6}$ ≈ 0.1666

Then for any $w \in \mathbb{C}^n$ and $y \in \mathbb{C}^m$ with $\|y - Aw\|_2 \leq \eta$, a solution $w^\#$ of the ℓ_1-minimization:

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad s.t. \quad \|Az - y\|_2 \leq \eta,$$

approximates the vector w with errors (C, D depend only on δ_{2s}):

$$\|w - w^\#\|_1 \leq C\sigma_s(w)_1 + D\sqrt{s}\eta.$$

$$\|w - w^\#\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(w)_1 + D\eta.$$

RIP Theorems

Theorem 6.21. Suppose $A \in \mathbb{C}^{m\times n}$ satisfies the RIP condition:

BP	IHT	HTP	OMP
$\delta_{2s} < \frac{4}{\sqrt{41}} \approx 0.6246$	$\delta_{6s} < \frac{1}{\sqrt{3}} \approx 0.5773$	$\delta_{6s} < \frac{1}{\sqrt{3}} \approx 0.5773$	$\delta_{13s} < \frac{1}{6} \approx 0.1666$

$\rho \in (0,1)$

$\rho = \rho(\delta_{6s})$

Then for any $\mathbf{w} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$ with $\mathbf{y} = A\mathbf{w} + \mathbf{e}$, the iteration $\mathbf{w}^n$ of the IHT and HTP for $\mathbf{y} = A\mathbf{w} + \mathbf{e}$, $\mathbf{w}^0 = 0$ and s is replaced by $2s$ satisfies

$$\|\mathbf{w} - \mathbf{w}^n\|_1 \leq C\sigma_s(\mathbf{w})_1 + D\sqrt{s}\|\mathbf{e}\|_2 + 2\rho^n\sqrt{s}\|\mathbf{w}\|_2.$$

IHT, HTP

$$\|\mathbf{w} - \mathbf{w}^n\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(w)_1 + D\|\mathbf{e}\|_2 + 2\rho^n\|\mathbf{w}\|_2$$

Reference

Chapters 4, 5, and 6, A Mathematical Introduction to Compressive Sensing, by S. Foucart and H. Rauhut.