AMATH 840: Advanced Numerical Methods for Computational and Data Science

Giang Tran
Department of Applied Mathematics, University of Waterloo
Winter 2024

Part 1: Sparse Optimization and Compressive Sensing

1.4: Recovery Guarantees for Sparse Optimization Problems

Winter 2024

1.4.1: Recovery Guarantees for ℓ_0- Algorithms

Problem: Given $A \in \mathbb{C}^{m\times n}$ and $\mathbf{y} \in \mathbb{C}^m$. Find

$$\mathbf{w} \in \mathbb{C}^n \quad \text{such that } \mathbf{y} = A\mathbf{w} \text{ and } \|w\|_0 \leq s.$$

We will go over recovery guarantees for OMP, IHT, HTP:

- Exact Recovery Condition for OMP
- Conditions based on Coherence of the Measurement Matrix.

We use the same numbers of Theorems, Lemmas, Propositions from "A Mathematical Introduction to Compressive Sensing", by S. Foucart and H. Rauhut.

Recall: Orthogonal Matching Pursuit

Orthogonal Matching Pursuit

Input: Measurement matrix $A \in \mathbb{C}^{m \times n}$ with ℓ_2-normalized columns, measurement vector $\mathbf{y} \in \mathbb{C}^m$, sparsity level s or tolerance ε.

Initialization: $\mathcal{S}^0 = \emptyset, \mathbf{w}^0 = 0$.

Iteration: Repeat until Stopping Criterion is met.

$$j_{k+1} := \arg\max_{j \in [n]} \{ |(A^*(\mathbf{y} - A\mathbf{w}^k))_j| \}$$

$$\mathcal{S}^{k+1} := \mathcal{S}^k \cup \{j_{k+1}\}$$

$$\text{Find } \mathbf{w}^{k+1} \ s.t \ A^*_{\mathcal{S}^{k+1}} \mathbf{y} = A^*_{\mathcal{S}^{k+1}} A_{\mathcal{S}^{k+1}} \mathbf{w}^{k+1}_{\mathcal{S}^{k+1}}$$

Output: The sparse vector $\mathbf{w}^{\#}$.

```
import numpy as np
from sklearn.linear_model import OrthogonalMatchingPursuit as omp
```

Orthogonal Matching Pursuit: Exact Recovery Condition

Proposition 3.5. (Exact Recovery Condition). Given $A \in \mathbb{C}^{m\times n}$ with ℓ_2-normalized columns. The following statements are equivalent:

1. Every nonzero vector $\mathbf{w} \in \mathbb{C}^n$ supported on a set $\mathcal{S}$ of size s is recovered from $\mathbf{y} = A\mathbf{w}$ after at most s iterations of OMP
2. $A_{\mathcal{S}}$ is injective and

$$\max_{j\in\mathcal{S}} |\langle \mathbf{r}, \mathbf{a}_j\rangle| > \max_{\ell\in\mathcal{S}^c} |\langle \mathbf{r}, \mathbf{a}_\ell\rangle|$$

   for all nonzero $\mathbf{r} \in \{A\mathbf{z}, \mathrm{supp}(\mathbf{z}) \subseteq \mathcal{S}\}$.
3. $\|A_{\mathcal{S}}^{\dagger} A_{\mathcal{S}^c}\|_{1\to 1} < 1$, where $A_{\mathcal{S}}^{\dagger} = (A_{\mathcal{S}}^* A_{\mathcal{S}})^{-1} A_{\mathcal{S}}^*$.

Recall: For any $p, q \geq 1$, define

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

Coherence

Definition 4.1. Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns. The ℓ_1-coherence function μ_1 of A is defined for $s \in [n-1]$ by

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

Definition 4.2. Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns $a_1, \ldots, a_n$. The coherence $\mu = \mu(A)$ of the matrix A is defined as

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

Lemma: For $1 \leq s \leq n-1$, $\mu \leq \mu_1(s) \leq s\mu$ and $\mu \leq 1$.

Gershgorin's Disk Theorem

Recall the Gershgorin's disk theorem, which states the locations of the eigenvalues of a square matrix.

Gershgorin's Theorem. Let λ be an eigenvalue of a square matrix $A \in \mathbb{C}^{n \times n}$. Then there exists $j \in [n]$ such that

$$|\lambda - A_{jj}| \leq \sum_{\ell \in [n], \ell \neq j} |A_{j,l}|.$$

Coherence (cont'd)

Theorem 5.3. Let $A \in \mathbb{C}^{m\times n}$ be a matrix with ℓ_2-normalized columns and let $s \in [n]$. For all s-sparse vector $x \in \mathbb{C}^n$, we have

$$(1-\mu_1(s-1))\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\mu_1(s-1))\|x\|_2^2.$$

Equivalently, for each set $S \subseteq [n]$ with $\mathrm{card} S \leq s$, the eigenvalues of $A_S^* A_S$ lie in the interval $[1-\mu_1(s-1), 1+\mu_1(s-1)]$.

In particular, if $\mu_1(s-1) < 1$, then $A_S^* A_S$ is invertible.

Proof

Proof Sketch.

- Since
$$\|Ax\|_2^2 = \langle Ax, Ax\rangle = \langle A_S x_S, A_S x_S\rangle = x_S^* A_S^* A_S x_S,$$
we have
$$\max_{\|x\|_2=1, \operatorname{supp} x\subseteq S} \|Ax\|_2^2 = \lambda_{\max}(A_S^* A_S) \quad \text{and} \quad \min_{\|x\|_2=1, \operatorname{supp} x\subseteq S} \|Ax\|_2^2 = \lambda_{\min}(A_S^* A_S).$$
- The diagonal entries of $(A_S^* A_S)$ are 1, since the columns of A are unit vectors.
- Using Gershgorin's disk theorem, the eigenvalues of $(A_S^* A_S)$ are contained in the union of disks centered at 1 with radii
$$r_j = \sum_{\ell\in S, \ell\neq j} |(A_S^* A_S)_{j,\ell}| = \sum_{\ell\in S, \ell\neq j} |\langle a_\ell, a_j\rangle \leq \mu_1(s-1), \quad j \in S.$$

□

Bounds on Coherence

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

$$\mu \geq \sqrt{\frac{n-m}{m(n-1)}}, \qquad \text{(Theorem 5.7)}$$

and

$$\mu_1(s) \geq s\sqrt{\frac{n-m}{m(n-1)}}, \quad \text{whenever} \quad s < \sqrt{n-1}. \qquad \text{(Theorem 5.8)}$$

Analysis of OMP

Using Proposition 3.5 and the definition of coherences, we obtain the following result:

Theorem 5.14. Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns. If

$$\mu_1(s) + \mu_1(s-1) < 1, \quad \left(\text{in particular, if } \mu(A) < \frac{1}{2s-1}\right),$$

then every s-sparse vector $w \in \mathbb{C}^n$ is exactly recovered from the measurement vector $y = Aw$ after at most s iterations of OMP.

Analysis of OMP

Question: Fix $s < n$. What is the computational complexity to verify a given matrix $A \in \mathbb{C}^{m \times n}$ with ℓ_2-normalized columns satisfies the coherence condition $\mu_1(s) + \mu_1(s-1) < 1$?

Analysis of IHT

Initialization: s-sparse vector w^0, typically $w^0 = 0$.

Iterations: Repeat until a stopping criterion is met:

$$w^{k+1} = H_s(w^k + A^*(y - Aw^k)).$$

Theorem. Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns. If $\mu_1(2s) < 1/2$, (in particular, if $\mu < (1/4s)$), then every s-sparse vector $w \in \mathbb{C}^n$ is recovered from $y = Aw$ via iterative hard thresholding.

[1]

[1] Iterative hard thresholding for compressed sensing, by T. Blumensath and M.E. Davies.

Analysis of HTP

Initialization: s-sparse vector w^0, typically $w^0 = 0$.

Iteration: Repeat until a stopping criterion is met.

$$S^{k+1} = L_s(w^k + A^*(y - Aw^k)),$$
$$w^{k+1} = \underset{z \in \mathbb{C}^n}{\operatorname{argmin}}\{\|y - Az\|_2, \quad \operatorname{supp}(z) \subseteq S^{k+1}\}$$

Theorem 5.17. Let $A \in \mathbb{C}^{m \times n}$ be a matrix with ℓ_2-normalized columns. If

$$2\mu_1(s) + \mu_1(s-1) < 1,$$

then every s-sparse vector $w \in \mathbb{C}^n$ is exactly recovered from $\mathbf{y} = A\mathbf{w}$ after at most s iterations of hard thresholding pursuit.

2

[2] Hard thresholding pursuit: an algorithm for compressive sensing, by S. Foucart.

Analysis of Basis Pursuit

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 $w \in \mathbb{C}^n$ is exactly recovered from the measurement vector $\mathbf{y} = A\mathbf{w}$ via basis pursuit:

$$\min_{\mathbf{x} \in \mathbb{C}^n} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = A\mathbf{x}.$$

[3]

[3] Hard thresholding pursuit: an algorithm for compressive sensing, by S. Foucart.

Summary

- Given $A \in \mathbb{C}^{m\times n}$ with unit columns and $y \in \mathbb{C}^m$, find a s-sparse vector $w \in \mathbb{C}^n$ s.t. $y = Aw$.
- $\mu_1(s) := \max_{k\in[n]} \max\left\{ \sum_{j\in S} |\langle a_k, a_j\rangle|, S \subset [n], |S| = s, k \notin S \right\}$.
- $\mu(A) := \max_{1\le k\neq j\le n} |\langle a_k, a_j\rangle|$.
- If Coherence condition, then every s-sparse vector $w \in \mathbb{C}^n$ is exactly recovered from $y = Aw$ after at most s iterations of the method.
  - For OMP: $\mu_1(s) + \mu_1(s-1) < 1$ or $\mu(A) < \dfrac{1}{2s-1}$.
  - For IHT: $\mu_1(2s) < 1$ or $\mu(A) < \dfrac{1}{4s}$.
  - For HTP: $2\mu_1(s) + \mu_1(s-1) < 1$ or $\mu(A) < \dfrac{1}{3s-1}$.

Summary (cont'd)

Theorem (Union of Bases)

Suppose that a dictionary is a union of $q+1$ orthonormal bases, i.e.,

$$A = (B_0 \ B_1 \ \cdots \ B_q),$$

where $B_i, i = 0, 1, \ldots, q$, are orthonormal bases of $\mathbb{R}^n$. If a vector

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}^0 & \mathbf{w}^1 & \cdots & \mathbf{w}^q \end{bmatrix}^T \in \mathbb{R}^{(q+1)n}$$

satisfies

$$0 < \|\mathbf{w}^0\|_0 \leq \|\mathbf{w}^1\|_0 \leq \cdots \leq \|\mathbf{w}^q\|_0,$$

and

$$\sum_{l=1}^{q} \frac{\mu\|\mathbf{w}^l\|_0}{1+\mu\|\mathbf{w}^l\|_0} < \frac{1}{2(1+\mu\|\mathbf{w}^0\|_0)} \tag{1}$$

then both OMP and BP can recover from A and $\mathbf{y} = A\mathbf{w}$.

4 5

[4] Sparse representation in pairs of bases, by Gribonval and Nielsen;
[5] Greed is good: Algorithmic results for sparse approximation, by Tropp.

Summary (cont'd)

With the same settings as in Theorem of union of bases, we have

- If $s < \left(\sqrt{2} - 1 + \frac{1}{2q}\right)\frac{1}{\mu(A)}$, the condition (1) is satisfied. [6]
- In particular, for $q = 1$, if $s < \left(\sqrt{2} - \frac{1}{2}\right)\frac{1}{\mu(A)}$, the condition (1) is satisfied and sharp. [7]
- For $q \geq 2$, if $s < \left(\frac{1}{2} + \frac{1}{4q-2}\right)\frac{1}{\mu(A)}$, the condition (1) is satisfied. [8]

---

[6] Sparse representation in pairs of bases, by Gribonval and Nielsen.
[7] A generalized uncertainty principle and sparse representation in pairs of bases, by Elad and Bruckstein
[8] On the stability of the basis pursuit in the presence of noise, by Donoho and Elad.

References

- A mathematical Introduction to Compressive Sensing, by S. Foucart and H. Rauhut. Chapter 3 and Chapter 5.
- Greed is Good: Algorithmic Results for Sparse Approximation, by J. Tropp.
- Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit, by J. Tropp and A. Gilbert. $m = \mathcal{O}(s \ln n)$.
- Iterative hard thresholding for compressed sensing, by T. Blumensath and M.E. Davies. Error estimation.
- Hard thresholding pursuit: an algorithm for compressive sensing, by S. Foucart.

1.4.2: Recovery Guarantees for ℓ_1-Optimization Problems

- Models:
  - Basis pursuit:

$$\min_{\mathbf{z}\in\mathbb{C}^n} \|\mathbf{z}\|_1 \quad s.t. \quad A\mathbf{z} = \mathbf{y}. \qquad (BP)$$

  - Basis pursuit denoising:

$$\min_{\mathbf{z}\in\mathbb{C}^n} \|\mathbf{z}\|_1 \quad s.t. \quad \|A\mathbf{z} - \mathbf{y}\|_2 \le \eta, \qquad (BP_\eta)$$

or

$$\min_{\mathbf{z}\in\mathbb{C}^n} \frac{1}{2}\|A\mathbf{z} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{z}\|_1. \qquad (QP_\lambda)$$

  - Lasso:

$$\min_{\mathbf{z}\in\mathbb{C}^n} \frac{1}{2}\|A\mathbf{z} - \mathbf{y}\|_2^2 \quad s.t. \quad \|\mathbf{z}\|_1 \le \tau. \qquad (LS_\tau)$$

- With suitable η, λ, τ, the solutions of $BP_\eta, QP_\lambda, LS_\tau$ coincide.
- BP vs QP_λ: $\lim_{\lambda\to 0^+} \mathbf{z}_{QP_\lambda} = \mathbf{z}_{bp}$, provided that the (BP) has a unique solution $\mathbf{z}_{bp}$.
- Algorithms: SPGL1, SpaRSA, Primal-Dual, FISTA, Nesterov's 2nd method, Augmented Lagrangian/Split-Bregman, coordinate descent,...

Basis Pursuit: Reconstruction Guarantees

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

Question: Study conditions on A that ensure exact reconstruction of every sparse vector $\mathbf{w} \in \mathbb{C}^n$ as a solution of (BP) with the vector $y \in \mathbb{C}^m$ obtained as $\mathbf{y} = A\mathbf{w}$.

Answer: We will study the following verification criteria:

- Coherence Condition
- Null Space Property (NSP)
- Stable Null Space Property
- Robust Null Space Property
- Restricted Isometry Properties

Null Space Property

- Provide a necessary and sufficient condition for exact recovery of sparse vectors via basis pursuit.

Definition 4.1. A matrix $A \in \mathbb{K}^{m \times n}$ is said to satisfy

- The null space property (NSP) relative to a set $\mathcal{S} \subset [n]$ if

$$\|\mathbf{v}_{\mathcal{S}}\|_1 < \|\mathbf{v}_{\mathcal{S}^c}\|_1 \quad \forall\, \mathbf{v} \in \ker A \setminus \{0\}.$$

- The null space property of order s if

$$\|\mathbf{v}_{\mathcal{S}}\|_1 < \|\mathbf{v}_{\mathcal{S}^c}\|_1, \quad \forall\, \mathbf{v} \in \ker A \setminus \{0\},\ \forall \mathcal{S} \subset [n] \text{ with } |\mathcal{S}| \leq s.$$

Here $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

Null Space Property – Equivalent Conditions

Lemma (Lemma 1)

Let $A \in \mathbb{K}^{m \times n}$. The following statements are equivalent:

- *The matrix A satisfies the NSP of order s.*
- $\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{S^c}\|_1$ *for all* $\mathbf{v} \in \ker A \setminus \{0\},\ \forall S \subset [n]$ *with* $|S| \leq s$.
- $2\|\mathbf{v}_S\|_1 < \|\mathbf{v}\|_1$ *for all* $\mathbf{v} \in \ker A \setminus \{0\},\ \forall S \subset [n]$ *with* $|S| \leq s$.
- $2\|\mathbf{v}_S\|_1 < \|\mathbf{v}\|_1$ *for all* $\mathbf{v} \in \ker A \setminus \{0\}$ *and*
  $S = \{\textit{indices of } s \textit{ largest absolute entries of } \mathbf{v}\}$.
- $\|\mathbf{v}\|_1 < 2\|\mathbf{v}_{S^c}\|_1$ *for all* $\mathbf{v} \in \ker A \setminus \{0\},\ \forall S \subset [n]$ *with* $|S| \leq s$.
- $\|\mathbf{v}\|_1 < 2\sigma_s(\mathbf{v})_1,$ *for all* $\mathbf{v} \in \ker A \setminus \{0\}$.

Recall: The ℓ_p error of best s-term approximation to a vector $\mathbf{x}$ is given by
$\sigma_s(\mathbf{x})_p := \inf\{\|\mathbf{x} - \mathbf{z}\|_p : \quad \mathbf{z} \in \mathbb{C}^n \text{ is } s\text{-sparse}\}$.

Lemma (Lemma 2)

Let $A \in \mathbb{K}^{m \times n}$, $\mathcal{S} \subseteq [n]$, *and* $s \leq n$. *Consider two distinct vectors in* $\mathbb{K}^n$, $\mathbf{w} \neq \mathbf{z}$, *such that* $A\mathbf{w} = A\mathbf{z}$.

1. *Suppose* A *satisfies the NSP relative to the set* $\mathcal{S}$ *and* $\mathrm{supp}(\mathbf{w}) \subseteq \mathcal{S}$. *Then* $\mathrm{supp}(\mathbf{z}) \not\subseteq \mathcal{S}$ *and* $\|\mathbf{z}\|_1 > \|\mathbf{w}\|_1$.
2. *Suppose* A *satisfies the NSP of order* s *and* $\mathbf{w}$ *is* s*-sparse. Then* $\mathbf{z}$ *is not* s*-sparse and* $\|\mathbf{z}\|_1 > \|\mathbf{w}\|_1$.

Proof. 1. Let $\mathbf{v} := \mathbf{w} - \mathbf{z} \in \ker A - \{0\}$.

- Since $\mathrm{supp}(\mathbf{w}) \subseteq \mathcal{S}$, we have

$$\mathbf{v}_{\mathcal{S}} = \mathbf{w}_{\mathcal{S}} - \mathbf{z}_{\mathcal{S}} = \mathbf{w} - \mathbf{z}_{\mathcal{S}} \quad (Eq.1), \quad \mathbf{v}_{\mathcal{S}^c} = \mathbf{w}_{\mathcal{S}^c} - \mathbf{z}_{\mathcal{S}^c} = -\mathbf{z}_{\mathcal{S}^c} \quad (Eq.2).$$

- Since A satisfies the NSP relative to $\mathcal{S}$, $\|\mathbf{v}_{\mathcal{S}^c}\|_1 > \|\mathbf{v}_{\mathcal{S}}\|_1 \geq 0$. Therefore, $\|\mathbf{z}_{\mathcal{S}^c}\|_1 = \|\mathbf{v}_{\mathcal{S}^c}\|_1 > 0 \Rightarrow \mathrm{supp}(\mathbf{z}) \not\subseteq \mathcal{S}$.
- We also have

$$\begin{aligned}\|\mathbf{w}\|_1 \leq \|\mathbf{w} - \mathbf{z}_{\mathcal{S}}\|_1 + \|\mathbf{z}_{\mathcal{S}}\|_1 &\overset{Eq.1}{=} \|\mathbf{v}_{\mathcal{S}}\|_1 + \|\mathbf{z}_{\mathcal{S}}\|_1 \\ &\overset{\text{NSP of A}}{<} \|\mathbf{v}_{\mathcal{S}^c}\|_1 + \|\mathbf{z}_{\mathcal{S}}\|_1 \overset{Eq.2}{=} \|-\mathbf{z}_{\mathcal{S}^c}\|_1 + \|\mathbf{z}_{\mathcal{S}}\|_1 = \|\mathbf{z}\|_1,\end{aligned}$$

which completes the proof.

Proof. 2.

- Let $\mathcal{S} = \mathrm{supp}(\mathbf{w})$. Then $|\mathcal{S}| \leq s$. Since A satisfies the NSP of order s, A satisfies the NSP relative to the set $\mathcal{S}$. Applying part 1 for $(A, \mathcal{S})$, we have $\|\mathbf{z}\|_1 > \|\mathbf{w}\|_1$.
- Suppose $\mathbf{z}$ is s-sparse. Let $\mathcal{T} = \mathrm{supp}(\mathbf{z})$. Then since A satisfies the NSP of order s, A satisfies the NSP relative to the set $\mathcal{T}$. Applying part 1 for $(A, \mathcal{T})$, we have $\|\mathbf{w}\|_1 < \|\mathbf{z}\|_1$, a contradiction.
- Therefore, the assumption is wrong. That means $\mathbf{z}$ is not s-sparse.

Null Space Property - ℓ_0 and ℓ_1 Models

Using Lemma 2, we obtain the following results about the solutions of ℓ_0 and ℓ_1 Models.

Theorem

If $A \in \mathbb{K}^{m \times n}$ satisfies the NSP of order s, then for every $\mathbf{y} = A\mathbf{w}$ with s-sparse $\mathbf{w}$, the solution of the basis pursuit problem

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_1 \quad s.t. \quad \mathbf{y} = A\mathbf{z},$$

is the solution of the ℓ_0-minimization problem:

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_0 \quad s.t. \quad \mathbf{y} = A\mathbf{z}.$$

Proof. Let $\mathbf{y} = A\mathbf{w}$ where $\mathbf{w} \in \mathbb{K}^n$ is s-sparse.

- Suppose $\mathbf{x} \in \mathbb{K}^n$ is a solution of the ℓ_0-minimization problem:

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z}.$$

  Since $A\mathbf{x} = \mathbf{y} = A\mathbf{w}$, we have $\|\mathbf{w}\|_0 \geq \|\mathbf{x}\|_0$.

- Since $\mathbf{w}$ is s-sparse, $s \geq \|\mathbf{w}\|_0$. Therefore, $s \geq \|\mathbf{x}\|_0$. That is, $\mathbf{x}$ is also s-sparse.
- Since A satisfies the NSP of order s, $A\mathbf{x} = A\mathbf{w}$, and $\mathbf{x}$ and $\mathbf{w}$ are both s-sparse vectors, by Lemma 2 part 2, $\mathbf{x} = \mathbf{w}$.

Null Space Property - Exact Recovery Theorem

The reverse of Lemma 2 is also true. That is, we have the following results (Theorem 4.4 and Theorem 4.5).

Theorem 4.4. Given $A \in \mathbb{K}^{m\times n}$ and $\mathcal{S} \subseteq [n]$, every vector $\mathbf{w} \in \mathbb{K}^n$ supported on a set $\mathcal{S}$ is the unique solution of

$$\min_{\mathbf{z}\in\mathbb{K}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z},$$

where $\mathbf{y} = A\mathbf{w}$, if and only if A satisfies the null space property relative to $\mathcal{S}$.

Interpretation: The theorem says that if A satisfies the null space property relative to $\mathcal{S}$ and the output measurement $\mathbf{y}$ can be written as $\mathbf{y} = A\mathbf{w}$ for some $\mathbf{w} \in \{\mathbf{x} \in \mathbb{K}^n : \mathrm{supp}(\mathbf{x}) \subseteq \mathcal{S}\}$, then

$$\|\mathbf{w}\|_1 < \|\mathbf{z}\|_1, \quad \forall \mathbf{z} \in \mathbb{K}^n - \{\mathbf{w}\} \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z}.$$

That is, $\mathbf{w}$ is the unique solution of

$$\min_{\mathbf{z}\in\mathbb{K}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z}.$$

Null Space Property - Exact Recovery Theorem

Proof of Theorem 4.4. For any set $X \subseteq [n]$ and any $\mathbf{v} \in \mathbb{K}^n$, denote $\mathbf{v}_X$ be a vector in $\mathbb{K}^n$, whose entries with the indices in X are the corresponding entries from $\mathbf{v}$ and the remaining entries are zero.

($\Leftarrow$) Lemma 2 part 1.

($\Rightarrow$) Take $\mathbf{v} \in \ker A \setminus \{0\}$.

- Since $\mathbf{v} \in \ker A$, we have $A\mathbf{v}_{\mathcal{S}} = A(-\mathbf{v}_{\mathcal{S}^c})$.
- Since $\mathbf{v} \neq 0, \mathbf{v}_{\mathcal{S}} \neq (-\mathbf{v}_{\mathcal{S}^c})$.
- We have $\mathrm{supp}(\mathbf{v}_{\mathcal{S}}) \subseteq \mathcal{S}$. By the assumption, $\mathbf{v}_{\mathcal{S}}$ is the unique solution of the ℓ_1-optimization problem:

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad A\mathbf{v}_{\mathcal{S}} = A\mathbf{z}.$$

- Due to the uniqueness of the given ℓ_1-optimization problem, $\|\mathbf{v}_{\mathcal{S}^c}\|_1 > \|\mathbf{v}_{\mathcal{S}}\|_1$, which completes the proof.

Null Space Property versus s-Sparse Solution

Letting the index set $\mathcal{S}$ vary and applying Theorem 4.4, we obtain the following result:

Theorem (Theorem 4.5.)
Given $A \in \mathbb{K}^{m \times n}$, every s-sparse vector $\mathbf{w} \in \mathbb{K}^n$ is the unique solution of

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_1 \quad s.t. \quad \mathbf{y} = A\mathbf{z},$$

where $\mathbf{y} = A\mathbf{w}$, if and only if A satisfies the null space property of order s.

Null Space Property versus s-Sparse Solution

Interpretation: The theorem says that if A satisfies the null space property of order s and the output measurement $\mathbf{y}$ can be written as $\mathbf{y} = A\mathbf{w}$ for some s-sparse vector $\mathbf{w} \in \mathbb{K}^n$, then

$$\|\mathbf{w}\|_1 < \|\mathbf{z}\|_1, \quad \forall \mathbf{z} \in \mathbb{K}^n - \{\mathbf{w}\} \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z}.$$

That is, $\mathbf{w}$ is the unique solution of

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z}.$$

Moreover, by Lemma 2, part 2, if $A\mathbf{z} = A\mathbf{w}$ and $\mathbf{z} \neq \mathbf{w}$, then $\mathbf{z}$ is not s-sparse.

Null Space Property

Theorem

If $A \in \mathbb{K}^{m \times n}$ satisfies the NSP of order s, the following matrices also satisfy the NSP of order s:

$$\hat{A} := GA, \quad \textit{where } G \in \mathbb{K}^{m \times m} \textit{ is some invertible matrix},$$

$$\widetilde{A} := \begin{bmatrix} A \\ B \end{bmatrix}, \quad \textit{where} \;\; B \in \mathbb{K}^{m' \times n}.$$

Remark:

- If $A \in \mathbb{K}^{m \times n}$ satisfies the NSP of order s, there exists matrix $H \in \mathbb{K}^{n \times n}$ such that AH does not satisfy the NSP.
- The above theorem indicates that the sparse recovery property of basis pursuit is preserved if some measurements are rescaled, reshuffled, or added.

Recall: Coherence Condition for Basis Pursuit

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

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

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

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

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

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:

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

- 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

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

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

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

$$\Leftrightarrow 2\|v_S\|_1 < \|v\|_1, \ \forall v \in \ker A \setminus \{0\} \text{ and}$$

$$S = \{\text{indices of } s \text{ largest ab.entries of } v\}.$$

$$\Leftrightarrow \|v\|_1 < 2\sigma_s(v)_1, \quad \text{for all } v \in \ker A \setminus \{0\}.$$

Null Space Property - ℓ_0 and ℓ_1 Models

- **Theorem:** If $A \in \mathbb{K}^{m \times n}$ satisfies the NSP of order s, then for every $\mathbf{y} = A\mathbf{w}$ with s-sparse $\mathbf{w}$, the solution of the basis pursuit problem

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z},$$

is the solution of the ℓ_0-minimization problem:

$$\min_{\mathbf{z} \in \mathbb{K}^n} \|\mathbf{z}\|_0 \quad \text{s.t.} \quad \mathbf{y} = A\mathbf{z}.$$

- **Theorems 4.4** [9]. 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.

- **Question:** What happens if the output vector $\mathbf{y} = A\mathbf{w}$, but $\mathbf{w}$ is not sparse?

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

Stable Null Space Property

Definition (Definition 4.11.)

A matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy

- The stable null space property with constant $0 < \rho < 1$ relative to a set$S \subset [n]$ if

$$\|v_S\|_1 \leq \rho \|v_{S^c}\|_1 \quad \forall v \in \ker A.$$

- The stable null space property of order s with constant $0 < \rho < 1$ if

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

Stable Null Space Property - Verification Theorem

Theorem (Theorem 4.14.)

The matrix $A \in \mathbb{C}^{m \times n}$ *satisfies the stable null space property of order* s *with constant* $0 < \rho < 1$ *relative to a set* $S \subset [n]$ *if and only if*

$$\|z - x\|_1 \leq \frac{1+\rho}{1-\rho} \left(\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1 \right),$$

for all vectors $x, z \in \mathbb{C}^n$ *with* $Az = Ax$.

Stable Sparse Recovery

Theorem (Theorem 4.12.)

Suppose that $A \in \mathbb{C}^{m \times n}$ satisfies the stable null space property of order s with constant $0 < \rho < 1$. Then for any $w \in \mathbb{C}^n$, a solution $w^\#$ of the basis pursuit,

$$\min_z \|z\|_1 \quad s.t. \quad Az = Aw,$$

approximates the vector w with ℓ_1-error:

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

Remark: If $A \in \mathbb{C}^{m \times n}$ satisfies the stable null space property of order s with constant $0 < \rho < 1$, the basis pursuit may have more than one solution.

Question: What happen if the output measurement vector is noisy: $\mathbf{y} = A\mathbf{w} + \varepsilon$?

Robust Null Space Property

Definition (Definition 4.17.)

A matrix $A \in \mathbb{C}^{m \times n}$ is said to satisfy

- The robust null space property w.r.t. $\|\cdot\|$ with constants $0 < \rho < 1$ and $\tau > 0$ relative to a set$S \subset [n]$ if

$$\|v_S\|_1 \leq \rho \|v_{S^c}\|_1 + \tau \|Av\| \quad \forall v \in \mathbb{C}^n.$$

- The stable null space property of order s with constant $0 < \rho < 1$ if

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

Robust Sparse Recovery

Theorem (Theorem 4.19.)
Suppose a matrix $A \in \mathbb{C}^{m \times n}$ satisfies the robust null space property of order s with constant $0 < \rho < 1$ and $\tau > 0$. Then for any $w \in \mathbb{C}^n$, a solution $w^\#$ of the BPDN:

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

with $y = Aw + e$ and $\|e\| \leq \eta$ approximates the vector w with ℓ_1-error:

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

Robust Sparse Recovery

Theorem 4.19 is a special case of Theorem 4.20 below.

Theorem (Theorem 4.20.)

The matrix $A \in \mathbb{C}^{m \times n}$ *satisfies the robust null space property with constant* $0 < \rho < 1$ *and* $\tau > 0$ *relative to* $\mathcal{S}$ *if and only if*

$$\|\mathbf{z} - \mathbf{x}\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + 2\|\mathbf{x}_{\mathcal{S}^c}\|_1\right) + \frac{2\tau}{1-\rho}\|A(\mathbf{z} - \mathbf{x})\|,$$

for all vectors $\mathbf{x}, \mathbf{z} \in \mathbb{C}^n$.

ℓ_2-Robust Null Space Property

Definition (Definition 4.21.)

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

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

ℓ_2-Robust Null Space Property

Theorem (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:

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

with $y = Aw + e$ and $\|e\|_2 \leq \eta$ approximates the vector w with ℓ_p-error:

$$\|w - w^\#\|_p \leq \frac{C}{s^{1-1/p}}\sigma_s(w)_1 + Ds^{1/p-1/2}\eta, 1 \leq p \leq 2,$$

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

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.

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

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

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

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 (2) 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.

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

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

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,

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

Restricted Isometry Property

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,

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

RIP Theorems

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

$$\delta_{2s} < \frac{4}{\sqrt{41}}.$$

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

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

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

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

Proof sketch: A satisfies the RIP condition for BP, then A satisfies the ℓ_2-robust NSP of order s with constants $\rho \in (0,1)$ and $\tau > 0$ depending only on δ_{2s}.

RIP Theorems

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

$$\delta_{6s} < \frac{1}{\sqrt{3}}.$$

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

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

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

Summary: RIP Conditions

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 we have error estimations.

Reference

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