COMPRESSED SENSING

Hu Fu, Nov 2025

SHUFE

SPARSE SIGNALS

- High-dimensional but sparse signals arise in many applications
  - Image/media files are often sparse when expressed w.r.t. the right bases
    - E.g. wavelet transform
    - $\mathbf{x} \in \mathbb{R}^n$, with $\|\mathbf{x}\|_0 \leq s$. We say $\mathbf{x}$ is s-sparse
- We often have linear measurements of such signals
  - $\mathbf{y} = A\mathbf{x}$, where A is a matrix in $\mathbb{R}^{m \times n}$, with $m \ll n$
  - Given $\mathbf{y}$, can we recover $\mathbf{x}$? We can design both the measurements A and the recovery algorithm.

SPARSE RECOVERY WITH COUNT SKETCH

- Recall that with Count Sketch, we were able to recover sparse signals
- By taking the largest (in absolute value) s coordinates of the sketch, with high probability, we get an s-sparse $\tilde{\mathbf{x}}$ s.t. $\|\tilde{\mathbf{x}} - \mathbf{x}\|_2 \leq (1 + \epsilon)E_2^s(\mathbf{x})$
  - $E_2^s(\mathbf{x})$ is the ℓ_2-norm of $\mathbf{x}$ with its largest k coordinates zeroed out
  - If $\mathbf{x}$ is s-sparse, with high probability $\tilde{\mathbf{x}}$ is an exact recovery
  - Count Sketch consists of randomized linear measurements of $\mathbf{x}$. $\tilde{\mathbf{x}}$ is computed from them

NON-UNIFORM VS. UNIFORM RECOVERY

- Non-uniform schemes: $\forall \mathbf{x} \in \mathbb{R}^n$, $\Pr[\tilde{\mathbf{x}} \text{ recovers } \mathbf{x}] \geq 1 - \delta$
  - Count Sketch gives us such a guarantee
- Uniform schemes: $\Pr[\forall \mathbf{x} \in \mathbb{R}^n, \tilde{\mathbf{x}} \text{ recovers } \mathbf{x}] \geq 1 - \delta$
  - This is stronger
  - Today: compressed sensing
  - Pioneered by Candes & Tao

THE HALLMARK OF CS

- A natural way to frame the problem as an optimization: given $\mathbf{y} = A\mathbf{x}$, find $\tilde{\mathbf{x}}$, with $||\tilde{\mathbf{x}}||_0$ minimized, satisfying $A\tilde{\mathbf{x}} = \mathbf{y}$
  - This problem turns out NP-hard for an arbitrary A
- Compressed sensing solves the following linear program instead:
  - $\min ||\hat{\mathbf{x}}||_1$, s.t., $A\hat{\mathbf{x}} = \mathbf{y}$.   (*)
  - This is solvable in polynomial time
  - Intuitively, why is the solution to this LP a good recovery of $\mathbf{x}$?

RESTRICTED ISOMETRY PROPERTY

- In order for (*) to produce good recoveries, we need A to approximately preserve ℓ_2 norms for all sparse vectors.

Def. A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the *restricted isometry property* (RIP) with parameters α, β and s if the inequality

$$\alpha\|\mathbf{v}\|_2 \leq \|A\mathbf{v}\|_2 \leq \beta\|\mathbf{v}\|_2$$

holds for all vectors $\mathbf{v} \in \mathbb{R}^n$ such that $\|\mathbf{v}\|_0 \leq s$

Compare this with the distributional JL lemma

The difference in quantifier reflects the difference from Count Sketch

RANDOM MATRICES SATISFY RIP

Theorem. Consider an $m \times n$ matrix A whose entries are i.i.d. drawn from the standard Gaussian $N(0,1)$. There are constants C and $c > 0$ such that, if $m \geq Cs\log(en/s)$, then with probability at least $1 - 2\exp(-cm)$, the random matrix A satisfies the RIP with parameters $\alpha = 0.9\sqrt{m}, \beta = 1.1\sqrt{m}$ and s.

The proof is deferred to the end.

RIP IMPLIES EXACT RECOVERY

Theorem. Suppose that an $m \times n$ matrix A satisfies the RIP with some parameters α, β and $(1+\lambda)s$, where $\lambda \geq (\beta/\alpha)^2$. Then every s-sparse vector $\mathbf{x} \in \mathbb{R}^n$ is recovered exactly by solving the program (*), i..e, the solution satisfies $\hat{\mathbf{x}} = \mathbf{x}$.

Notations: Let $\mathbf{h} = \mathbf{x} - \hat{\mathbf{x}}$. We need to show $\mathbf{h} = \mathbf{0}$.

For any $T \subseteq [n]$, let $\mathbf{h}_T$ denote the vector $\mathbf{h}$ restricted to T, i.e., coordinates not in T are zeroed out

Let I_0 be the set of non-zero coordinates of $\mathbf{x}$, then $|I_0| \leq s$.

Lemma. $\|\mathbf{h}_{\overline{I_0}}\|_1 \leq \|\mathbf{h}_{I_0}\|_1$

Proof. $\|\mathbf{x}\|_1 \geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x}+\mathbf{h}\|_1 = \|\mathbf{x}_{I_0}+\mathbf{h}_{I_0}\|_1 + \|\mathbf{h}_{\overline{I_0}}\|_1 \geq \|\mathbf{x}\|_1 - \|\mathbf{h}_{I_0}\|_1 + \|\mathbf{h}_{\overline{I_0}}\|_1$

First "$\geq$": by the linear program

RIP IMPLIES EXACT RECOVERY

Theorem. Suppose that an $m \times n$ matrix A satisfies the RIP with some parameters α, β and $(1+\lambda)s$, where $\lambda \geq (\beta/\alpha)^2$. Then every s-sparse vector $\mathbf{x} \in \mathbb{R}^n$ is recovered exactly by solving the program (*), i..e, the solution satisfies $\hat{\mathbf{x}} = \mathbf{x}$.

Notations: Let $\mathbf{h} = \mathbf{x} - \hat{\mathbf{x}}$. We need to show $\mathbf{h} = \mathbf{0}$.

For any $T \subseteq [n]$, let $\mathbf{h}_T$ denote the vector $\mathbf{h}$ restricted to T, i.e., coordinates not in T are zeroed out

Let I_0 be the set of non-zero coordinates of $\mathbf{x}$, then $|I_0| \leq s$.

Sort coordinates of $\mathbf{h}$ in $\overline{I_0}$ in absolute values. Let I_1 be the set of next largest λs coordinates, I_2 the next λs coordinates, and so on. Let $I_{0,1}$ be $I_0 \cup I_1$

Theorem. Suppose that an $m \times n$ matrix A satisfies the RIP with some parameters α, β and $(1+\lambda)s$, where $\lambda \geq (\beta/\alpha)^2$. Then every s-sparse vector $\mathbf{x} \in \mathbb{R}^n$ is recovered exactly by solving the program (*), i..e, the solution satisfies $\hat{\mathbf{x}} = \mathbf{x}$.

Proof. We have $A\mathbf{h} = \mathbf{0}$, we have $A\mathbf{h}_{I_{0,1}} = -A\mathbf{h}_{\overline{I_{0,1}}}$, so $\|A\mathbf{h}_{I_{0,1}}\|_2 = \|A\mathbf{h}_{\overline{I_{0,1}}}\|_2$

Since $|I_{0,1}| \leq s + \lambda s$, RIP yields $\|A\mathbf{h}_{I_{0,1}}\|_2 \geq \alpha\|\mathbf{h}_{I_{0,1}}\|_2$

On the other side, $\|A\mathbf{h}_{\overline{I_{0,1}}}\|_2 \leq \sum_{i\geq 2} \|A\mathbf{h}_{I_i}\|_2 \leq \beta \sum_{i\geq 2} \|\mathbf{h}_{I_i}\|_2$

Crucial step: for $i \geq 2$, $\forall j \in I_i$, we have $|h_j| \leq \frac{1}{\lambda s}\|\mathbf{h}_{I_{i-1}}\|_1$, and so $\|\mathbf{h}_{I_i}\|_2 \leq \frac{1}{\sqrt{\lambda s}}\|\mathbf{h}_{I_{i-1}}\|_1$

Therefore $\sum_{i\geq 2} \|\mathbf{h}_{I_i}\|_2 \leq \frac{1}{\sqrt{\lambda s}} \sum_{i\geq 1} \|\mathbf{h}_{I_i}\|_1 = \frac{1}{\sqrt{\lambda s}}\|\mathbf{h}_{\overline{I_0}}\|_1 \leq \frac{1}{\sqrt{\lambda s}}\|\mathbf{h}_{I_0}\|_1 \leq \frac{1}{\sqrt{\lambda}}\|\mathbf{h}_{I_0}\|_2 \leq \frac{1}{\sqrt{\lambda}}\|\mathbf{h}_{I_{0,1}}\|_2$

Theorem. Suppose that an $m \times n$ matrix A satisfies the RIP with some parameters α, β and $(1+\lambda)s$, where $\lambda \geq (\beta/\alpha)^2$. Then every s-sparse vector $\mathbf{x} \in \mathbb{R}^n$ is recovered exactly by solving the program (*), i..e, the solution satisfies $\hat{\mathbf{x}} = \mathbf{x}$.

Proof. We have $A\mathbf{h} = \mathbf{0}$, we have $0 = \|A\mathbf{h}\|_2 \geq \|A\mathbf{h}_{I_{0,1}}\|_2 - \|A\mathbf{h}_{\overline{I_{0,1}}}\|_2$

Since $|I_{0,1}| \leq s + \lambda s$, RIP yields $\|A\mathbf{h}_{I_{0,1}}\|_2 \geq \alpha \|\mathbf{h}_{I_{0,1}}\|_2$

On the other side, $\|A\mathbf{h}_{\overline{I_{0,1}}}\|_2 \leq \sum_{i \geq 2} \|A\mathbf{h}_{I_i}\|_2 \leq \beta \sum_{i \geq 2} \|\mathbf{h}_{I_i}\|_2$

So we have $\dfrac{\beta}{\sqrt{\lambda}} \|\mathbf{h}_{I_{0,1}}\|_2 \geq \alpha \|\mathbf{h}_{I_{0,1}}\|_2$. But since $\lambda \geq (\beta/\alpha)^2$, this implies $\mathbf{h}_{I_{0,1}} = \mathbf{0}$, which by definition means $\mathbf{h} = \mathbf{0}$.

PROOF OF RIP

Theorem. Consider an $m \times n$ matrix A whose entries are i.i.d. drawn from the standard Gaussian $N(0, \frac{1}{\sqrt{m}})$. There are constants C and $c > 0$ such that, if $m \geq Cs\log(en/s)$, then with probability at least $1 - 2\exp(-cm)$, the random matrix A satisfies the RIP with parameters $\alpha = 0.9, \beta = 1.1$ and s.

Recall the JL distributional lemma (slightly modified):

Lemma. Fix $\epsilon \in (0, \frac{1}{2})$. Consider an $m \times n$ matrix A whose entries are i.i.d. drawn from the standard Gaussian $N(0,1)$. There is a constant $c > 0$ such that, for any $\mathbf{v} \in \mathbb{R}^n$ with $\|\mathbf{v}\| = 1$,

$$\mathbf{Pr}\left[1 - \epsilon \leq \frac{\|A\mathbf{v}\|}{\sqrt{m}} \leq 1 + \epsilon\right] \geq 1 - e^{-cm}$$

PROOF OF RIP

Idea: 1. Use an ϵ-net to discretize the ℓ_2-ball in $\mathbb{R}^s$. A preserves w.h.p. the norm of all points in this ϵ-net, and must also preserve w.h.p. the norm of all the points in the ℓ_2-ball

Def. Given a set K, a subset $L \subseteq K$ is an ϵ-net of K if $\forall x \in K, \exists y \in L$ s.t. $\|x - y\| \leq \epsilon$.

2. The argument above works for the subspace of vectors in $\mathbb{R}^n$ supported on any set T of coordinates with $|T| = s$. There are $\leq \binom{n}{s}$ such subspaces in $\mathbb{R}^n$. Apply union bound on all of them.

ϵ-NET OF ℓ_2-BALLS

Lemma. There is an ϵ-net for the unit ℓ_2-ball $B(0,1)$ in $\mathbb{R}^n$ of size $\leq \left(\frac{3}{\epsilon}\right)^n$.

Pf. First construct a set $L \subseteq B(1)$ as follows: add an arbitrary point to L; whenever there is a point $x' \in B(1)$ such that $\min_{y \in L} \|x' - y\| > \epsilon$, add x' to L.

When the process finishes, L must be an ϵ-net by definition.

Key observation: for each point $x \in L$, draw a ball of radius $\epsilon/2$ around it, then no two of these balls intersect with each other, and all of them are contained within the ball $B(0,1+\epsilon/2)$.

ϵ-NET OF ℓ_2-BALLS

Lemma. There is an ϵ-net for the unit ℓ_2-ball $B(0,1)$ in $\mathbb{R}^n$ of size $\leq \left(\frac{3}{\epsilon}\right)^n$.

Pf. The volume of an ℓ_2-ball of radius r in $\mathbb{R}^n$ is Cr^n for some constant C relying on n.

We therefore have $|T| \cdot C\left(\frac{\epsilon}{2}\right)^n \leq C\left(1+\frac{\epsilon}{2}\right)^n$

Therefore $|T| \leq \left(\frac{2}{\epsilon}+1\right)^n \leq \left(\frac{3}{\epsilon}\right)^n$.

USE ϵ-NET TO APPROXIMATE ℓ_2-BALLS

Let A' be the $m \times s$ matrix with each entry drawn i.i.d. from $\mathcal{N}\left(0, \frac{1}{\sqrt{m}}\right)$; let T be an ϵ-net so constructed for the ℓ_2-ball in $\mathbb{R}^s$. Then. **Pr** $\left[\forall \mathbf{v} \in T, (1-\epsilon)\|\mathbf{v}\| \leq \|A'\mathbf{v}\| \leq (1+\epsilon)\|\mathbf{v}\|\right] \geq 1 - e^{-cm} \cdot \left(\frac{3}{\epsilon}\right)^s$

Take a realization of A' that does preserve the norms of all vectors in T.

Let α be the smallest number such that $\|A'\mathbf{v}\| \leq (1+\alpha)\|\mathbf{v}\|$ for all $\mathbf{v} \in B(0,1)$.

Let $\mathbf{x}$ be the vector for which the equality attains. Then there exists $\mathbf{y} \in T$ with $\|\mathbf{y} - \mathbf{x}\| \leq \epsilon$.

So $1 + \alpha = \|A'\mathbf{x}\| \leq \|A'\mathbf{y}\| + \|A'(\mathbf{y} - \mathbf{x})\| \leq 1 + \epsilon + (1+\alpha)\epsilon \Rightarrow \alpha \leq \frac{2\epsilon}{1-\epsilon} \leq 4\epsilon$.

ENUMERATING SUBSPACES

Now for each $T \subseteq [n]$ with $|T| = s$, the argument above works for the matrix A applied to vectors in $\mathbb{R}^n$ supported on T.

There are $\binom{n}{s} \leq \left(\frac{en}{s}\right)^s$ such subspaces. The probability that A does not approximately preserve the norm of any s-sparse vector is therefore at most

$e^{-cm} \cdot \left(\frac{3}{\epsilon}\right)^s \cdot \left(\frac{en}{s}\right)^s$. Substitute $m \geq Cs \log(en/s)$ for appropriate C, we obtain the theorem.