AMATH 840, Lecture 08, Jan 31, 2022

RIP — $m \gtrsim O(s \ln(\frac{n}{s}))$

Today

1. Examples of matrices that satisfy the RIP cond.

   Coherence $m \gtrsim O(s^2)$

2. Sparse Optimization Methods

   i. Proximal Methods : nondiff $\|x\|_1$

   ii. Lagrangian Methods : constrained → unconstrained

   Image processing → $\|\nabla x\|_1$ → ADMM

References

1. A Mathematical Intro. to Compressive Sensing by Foucart and Rauhut, Chapters 9, 12, 14.
2. Convex Optimization, by Boyd & Vandenberghe
3. Slides from | ECE 236C, by Vandenberghe, UCLA
   | CMSC 764, by T. Goldstein, Uni. of Maryland

① Recall RIP condition

<u>Def</u> (i) The restricted isometry constant δ_s of a matrix $A \in \mathbb{C}^{m \times n}$ is the smallest $\delta \geqslant 0$ s.t

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2, \quad \forall x : s\text{-sparse}.$$

(ii) The matrix A is said to satisfy the RIP condition if $\delta_{ks} < \delta_*$, where constants k and δ_* depend on the algorithms/ models :

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

6.12

Theorem: If $A \in \mathbb{C}^{m \times n}$ satisfy the RIP condition for BP

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

① Every s-sparse vector w is exactly recovered from (BP):

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

② Moreover, the reconstruction is stable when sparsity is replaced by compressibility and the reconstruction is robust when measurement error occurs. More precisely, for any $w \in \mathbb{C}^n$ and $y \in \mathbb{C}^m$ with $\|y - Aw\|_2 \leq \eta$, a solution $w^\#$ of BPDN

$$\min \|z\|_1 \quad \text{s.t} \quad \|y - Az\|_2 \leq \eta$$

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

② Examples of matrices that satisfy RIP condition

Concentration Inequalities / Berstein Ineq / Covering number.

Example 1 : Gaussian Random Matrices: the entries of A are independent standard Gaussian random variables.

Example 2: Bernoulli Random Matrices: the entries of A are independent Rademacher variables (taking values ± 1 with prob. $\frac{1}{2}$)

Example 3: Subgaussian Random Matrices: the entries of A are independent mean-zero subgaussian random variables with variance 1 and $\mathbb{E}(e^{sA_{j,k}}) \leq e^{\sigma^2 s^2/2}$, $\forall s \in \mathbb{R}$, $j \in [m], k \in [n]$.

Here σ is a constant.

<u>Remark</u>: Gaussian and Bernoulli random matrices are subgaussian since:

* If X is $\mathcal{N}(0,\sigma^2)$, for any $t>0$, we have:

$$\mathbb{E}(e^{sX}) = e^{\sigma^2 s^2/2}, \quad \forall s \in \mathbb{R}$$

* If X is a Rademacher random variable, then

$$\mathbb{E}(e^{sX}) = \frac{1}{2}\left(e^{s} + e^{-s}\right)$$

$$= \frac{1}{2}\left(\sum_{k=0}^{\infty} \frac{s^k}{k!} + \sum_{k=0}^{\infty} \frac{(-s)^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{s^{2k}}{(2k)!}$$

$$\leq 1 + \sum_{k=1}^{\infty} \frac{s^{2k}}{2^k k!} = e^{s^2/2}$$

Theorem 9.2 Let $A \in \mathbb{C}^{m\times n}$ subgaussian random matrix

Then there exists a constant C (depending on the subgaussian parameter σ) s.t the restricted isometry constant of $\frac{1}{\sqrt{m}}A$ satisfies $\delta_s \leq \delta$ with probability at least $1-\varepsilon$, provided

$$m \geq C\,\delta^{-2}\left[s \ln\left(\frac{en}{s}\right) + \ln\left(2\varepsilon^{-1}\right)\right]$$

Letting $\varepsilon = 2\exp(-\delta^2 m/(2C))$ yields:

$$\boxed{m \geq 2C\,\delta^{-2} s \ln\left(\frac{en}{s}\right)}$$

$\mathbb{E}(\|\frac{1}{\sqrt{m}}Ax\|_2^2)$

$= \mathbb{E}(\|x\|_2^2)$

Example 4: Structured Random Matrices / Random Sampling Matrices

Def: Bounded Orthonormal System (BOS)

Let $\mathcal{D} \subset \mathbb{R}^d$ be endowed with a probability measure ν.

$\Phi := \{\phi_1, \dots, \phi_n\}$; where $\phi_k : \mathcal{D} \to \mathbb{C}$ $\forall k \in [n]$ and

① $$\int_{\mathcal{D}} \phi_j(t)\,\overline{\phi_k(t)}\,d\nu(t) = \delta_{jk}.$$

② $$\|\phi_j\|_\infty := \sup_{t \in \mathcal{D}} |\phi_j(t)| \leq K \qquad \forall j \in [n]$$

a constant.

Then Φ is called a bounded orthonormal system with constant K

Def (cont'd) Let $t_1, \dots, t_m \in \mathcal{D}$ be some sampling points.

The matrix $A = (A_{\ell,k} = \Phi_k(t_\ell)) \in \mathbb{C}^{m \times n}$

is called a sampling matrix.

$$A = \begin{pmatrix} \Phi_1(t_1) & \Phi_2 & \cdots & \Phi_n(t_1) \\ \downarrow & & & \vdots \\ t_m & & & \Phi_n(t_m) \end{pmatrix}$$

Examples of BOS

$k = \{-q, -q+1, \dots, 0, 1, \dots, q\}$

① Trigonometric polynomials: $\mathcal{D} = [0,1]$, $\Phi_k(t) = e^{2\pi i k t}$

② Real Trigonometric polynomials, $\mathcal{D} = [0,1]$

$\Phi_0(t) = 1$, $\Phi_{2k}(t) = \sqrt{2}\cos(2\pi k t)$, $\Phi_{2k-1}(t) = \sqrt{2}\sin 2\pi k t$

$k \geq 1$.

③ Partial discrete Fourier transform

$$F \in \mathbb{C}^{n \times n}, \quad F_{l,k} = \frac{1}{\sqrt{n}} e^{2\pi i (l-1)(k-1)/n}$$

$$l, k \in [n]$$

The normalized columns of F, $\sqrt{n}\, F_k \in \mathbb{C}^n$ form an orthonormal system w.r.t. the discrete uniform measure on $[n]$

④ Legendre polynomials $\quad \mathcal{D} = [-1, 1]$

$$\underbrace{P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1), \ \ldots}$$

↓ normalized

Setting

Consider functions of the form

$$f(t) = \sum_{k=1}^{n} w_k \phi_k(t) \quad , \quad f: D \subset \mathbb{R} \to \mathbb{C}$$

$$\{\phi_k : D \to \mathbb{C}\} \quad \text{BOS}$$

Suppose we are given sample values $(t_j, f(t_j))_{j=1}^{m} \subset D \times \mathbb{C}$.

$$y_\ell = f(t_\ell) = \sum_{k=1}^{n} w_k \phi_k(t_\ell)$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}, \quad \text{where } A \in \mathbb{C}^{m \times n}$$

$$A = \left(A_{\ell,k} = \phi_k(t_\ell)\right)_{\ell \in [m],\, k \in [n]}$$

$$\vec{y} = A \vec{w}$$

Goal: Given sample values $\vec{y}$ and sampling matrix A find $\vec{w} \in \mathbb{C}^n$. ($\Leftrightarrow$ find function f)

If f is s-sparse w.r.t $\{\phi_1, \dots, \phi_n\}$, find s-sparse vector $w \in \mathbb{C}^n$ s.t $\vec{y} = A\vec{w}$.

Remark: Assume the sampling points $t_1, \dots, t_m$ are selected independently at random according to the prob. measure ν. Then the sampling matrix A is called a random sampling matrix associated to a BOS with constant $K \geq 1$.

Theorem 12.22 Let $w \in \mathbb{C}^n$ and let $A \in \mathbb{C}^{m \times n}$ be the random sampling matrix associated to a BOS with constant $K \geq 1$. For $y = Aw + e$ with $\|e\|_2 \leq \sqrt{m}\,\eta$ for some $\eta \geq 0$, let $w^\#$ be a solution of

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

depend on the max degree of ~~max~~ legendre poly.

If $m \geq C K^2 s \ln(n) \ln(\varepsilon^{-1})$,

then with probability at least $1 - \varepsilon$, the reconstruction error satisfies

$$\|w - w^\#\|_2 \leq C_1\, \sigma_s(w)_1 + C_2 \sqrt{s}\,\eta$$

C, C_1, C_2 are global constants

Theorem 12.32: let $A \in \mathbb{C}^{m \times n}$ be the random sampling matrix associated to a BOS with constant $\tau \geq 1$. Under some conditions on m ($O(s\ln(s))$), the matrix A satisfies the RIP condition with high prob.

$\downarrow$ Robust

Stable

prob measure ν $\xrightarrow{\text{build}}$ $\{\Phi_k\}_{k=1}^{n}$ BOS

Drawback Depend ν

$\{t_k\}_{k=1}^{m} \sim \nu$ randomly

Input data

soln 1 Empirical Distribution

~~seen~~ 2 Transfer randomness of input data $\longrightarrow$ randomness of the Random Feature

Random weight controllable

$$A = \left(\underline{\sigma \langle t_k, \omega_k \rangle} \right)$$

A also satisfies RIP-like conditions