Computational Principles for High-dim Data Analysis (Lecture Six)

Yi Ma

EECS Department, UC Berkeley

September 16, 2021



EECS208, Fall 2021

Convex Methods for Sparse Signal Recovery (Matrices with Restricted Isometry Property)

The Johnson-Lindenstrauss Lemma

2 RIP of Gaussian Matrices

3 RIP of Non-Gaussian Matrices

"Algebra is but written geometry; geometry is but drawn algebra." – Sophie Germain

Restricted Isometry Property (Recap)

Definition (Restricted Isometry Property)

The matrix A satisfies the *restricted isometry property (RIP)* of order k, with constant $\delta \in [0, 1)$, if

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

The order-k restricted isometry constant $\delta_k(\mathbf{A})$ is the smallest number δ such that the above inequality holds.

Example of Gaussian Matrices: If A_{I} is a large $m \times k$ (k < m) matrix with entries independent $\mathcal{N}(0, 1/m)$,

$$\sigma_{min}(\boldsymbol{A}_{\mathsf{I}}^*\boldsymbol{A}_{\mathsf{I}}) \approx (\sqrt{1} - \sqrt{k/m})^2 \ge 1 - 2\sqrt{k/m},$$

$$\sigma_{max}(\boldsymbol{A}_{\mathsf{I}}^*\boldsymbol{A}_{\mathsf{I}}) \approx (\sqrt{1} + \sqrt{k/m})^2 \le 1 + 3\sqrt{k/m}.$$

イロト イポト イヨト イヨト 二日

Length Concentration of Gaussian Vectors

Lemma (Gaussian Vectors¹)

Let $g = [g_1, \ldots, g_m]^* \in \mathbb{R}^m$ be an *m*-dimensional random vector whose entries are iid $\mathcal{N}(0, 1/m)$. Then for any $t \in [0, 1]$,

$$\mathbb{P}\left[\left|\|\boldsymbol{g}\|_{2}^{2}-1\right| > t\right] \leq 2\exp\left(-\frac{t^{2}m}{8}\right).$$
(2)

Proof (sketch): The moment generating function of a standard normal random variable is

$$\mathbb{E}\left[e^{\lambda x^2}\right] = (1 - 2\lambda)^{-1/2}, \quad \lambda < 1/2.$$

Then we have $\mathbb{E}\left[e^{\lambda g_i^2}\right] = \left(1 - \frac{2\lambda}{m}\right)^{-1/2}, \quad \lambda < \frac{m}{2}.$

¹This result can be obtained via the Cramer-Chernoff exponential moment method (see Appendix E) or the book: *High-Dimensional Probability*, Roman Vershynin, 2018.

Length Concentration of Gaussian Vectors

Proof (continued):

$$\begin{split} \mathbb{P}\Big[\sum_{i=1}^{m} g_i^2 > t+1\Big] &= \mathbb{P}\Big[\exp\left(\lambda \sum_{i=1}^{m} g_i^2\right) > \exp\left(\lambda(t+1)\right)\Big] \\ &\leq e^{-\lambda(t+1)} \mathbb{E}\Big[\exp\left(\lambda \sum_{i=1}^{m} g_i^2\right)\Big] \\ &= e^{-\lambda(t+1)} \prod_{i=1}^{m} \mathbb{E}\Big[e^{\lambda g_i^2}\Big] = e^{-\lambda(t+1)} \Big(1 - \frac{2\lambda}{m}\Big)^{-m/2} \\ &\leq \exp(-\lambda t) \end{split}$$

since for sufficiently small $\lambda = \frac{tm}{C}$, we have $\left(1 - \frac{2\lambda}{m}\right)^{-m/2} \leq e^{\lambda}$.

- 3

イロト 不得下 イヨト イヨト

The JL Lemma: Distance Preserving Random Projections

Theorem (Johnson-Lindenstrauss Lemma)

Let $v_1, \ldots, v_n \in \mathbb{R}^D$ and $A \in \mathbb{R}^{m \times D}$ be a random matrix whose entries are *i.i.d.* $\mathcal{N}(0, 1/m)$. Then for any $\epsilon \in (0, 1)$, with probability at least $1 - 1/n^2$, the following holds:

 $\forall i \neq j, \ (1-\epsilon) \| \boldsymbol{v}_i - \boldsymbol{v}_j \|_2^2 \leq \| \boldsymbol{A} \boldsymbol{v}_i - \boldsymbol{A} \boldsymbol{v}_j \|_2^2 \leq (1+\epsilon) \| \boldsymbol{v}_i - \boldsymbol{v}_j \|_2^2, \ (3)$ provided $m > 32 \frac{\log n}{\epsilon^2}.$



The JL Lemma

Proof.

Finite cases: let $g_{ij} = A \frac{v_i - v_j}{\|v_i - v_j\|_2}$ for any $i \neq j \in \{1, \dots, n\}$. Tail bound: g_{ij} is distributed as an iid Gaussian vector, with entries $\mathcal{N}(0, 1/m)$. Applying Lemma:

$$\mathbb{P}\left[\left|\left\|\boldsymbol{g}_{ij}\right\|_{2}^{2}-1\right| > t\right] \leq 2\exp\left(-t^{2}m/8\right).$$
(4)

Union bound: Summing the probability of failure over all $i \neq j$, and then plugging in $t = \epsilon$ and $m \geq 32 \log n/\epsilon^2$, we get

$$\mathbb{P}\left[\exists (i,j) : \left| \|\boldsymbol{g}_{ij}\|_{2}^{2} - 1 \right| > t \right] \leq \frac{n(n-1)}{2} \times 2 \exp\left(-t^{2}m/8\right) \leq n^{-2}.$$
 (5)

Hence $\left| \|\boldsymbol{g}_{ij}\|_2^2 - 1 \right| \leq \epsilon$ with probability $1 - n^{-2}$.

イロト 不得下 イヨト イヨト 二日

The JL Lemma: Generalization to ℓ^p Norms

Locality-Sensitive Hashing²: for $p \in (0, 2]$, there exist the so-called *p-stable distributions* such that a random matrix A drawn from a *p*-stable distribution will preserve ℓ^p distance between vectors:

$$(1-\epsilon) \|\boldsymbol{v}_{i} - \boldsymbol{v}_{j}\|_{p}^{2} \leq \|\boldsymbol{A}\boldsymbol{v}_{i} - \boldsymbol{A}\boldsymbol{v}_{j}\|_{p}^{2} \leq (1+\epsilon) \|\boldsymbol{v}_{i} - \boldsymbol{v}_{j}\|_{p}^{2}.$$
 (6)

Example: For ℓ^1 norm, the corresponding distribution is the Cauchy distribution with density:

$$p(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}.$$

EECS208, Fall 2021

The JL Lemma: Fast Nearest Neighbors

Compact Code for Fast Nearest Neighbor³:

- 1: **Goal:** Generate compact binary code for efficient nearest neighbor search of high-dimensional data points.
- 2: Input: $x_1, \ldots, x_n \in \mathbb{R}^D$ and $m = O(\log n)$.
- 3: Generate a Gaussian matrix $\mathbf{R} \in \mathbb{R}^{m \times D}$ with entries i.i.d. $\mathcal{N}(0, 1)$.
- 4: for $i=1,\ldots,n$ do
- 5: Compute $oldsymbol{R}oldsymbol{x}_i$,
- 6: Set $y_i = \sigma(\mathbf{R} x_i)$ where $\sigma(\cdot)$ is the entry-wise binary thresholding.
- 7: end for
- 8: **Output:** $y_1, \ldots, y_n \in \{0, 1\}^m$.

Instead of $O(\log n)$ real numbers, one only needs $O(\log n)$ binary bits!

³Compact projection: Simple and efficient near neighbor search with practical memory requirements, K. Min, J. Wright, and Y. Ma, CVPR 2010.

RIP of Gaussian Matrices

Theorem (RIP of Gaussian Matrices)

There exists a numerical constant C > 0 such that if $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a random matrix with entries independent $\mathcal{N}\left(0, \frac{1}{m}\right)$ random variables, with high probability, $\delta_k(\mathbf{A}) < \delta$, provided

$$m \ge Ck \log(n/k)/\delta^2.$$
 (7)

Implications: ℓ^1 minimization can successfully recover k-sparse solutions x_o from about

$$m \ge Ck \log(n/k) \sim \Omega(k)$$

random measurements.

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

$$\delta_k({m A}) \ \le \ \delta$$
 if and only if $\sup_{{m x}\in {m \Sigma}_k} \left| \|{m A}{m x}\|_2^2 - 1
ight| \ \le \ \delta$ where

$$\Sigma_{k} = \{ \boldsymbol{x} \mid \| \boldsymbol{x} \|_{0} \le k, \| \boldsymbol{x} \|_{2} = 1 \}.$$
(8)

Construct a finite (minimal) ϵ -**net** for Σ_k .



э

Proof: Step 1. Discretization to Finite Cases An ϵ -net (or covering) N for a given set S if

 $\forall x \in \mathsf{S}, \quad \exists \, \bar{x} \in \mathsf{N} \quad \text{such that} \quad \|x - \bar{x}\|_2 \leq \epsilon.$ (9)

A set M is ϵ -separated if every pair of distinct points x, x' in M has distance at least ϵ :

$$\|\boldsymbol{x} - \boldsymbol{x}'\|_2 \ge \epsilon. \tag{10}$$

Fact: A maximal ϵ -separated subset M \subset S is a (minimal) ϵ -net of S.



EECS208, Fall 2021

Lemma (ϵ -Nets for the Unit Ball)

There exists an ϵ -net for the unit ball $\mathsf{B}(0,1) \subset \mathbb{R}^d$ of size at most $(3/\epsilon)^d$.

Proof: Let $N \subset B(0,1)$ be a *maximal* ϵ -separated set. The balls $B(x, \epsilon/2)$ with $x \in N$ are contained in $B(0, 1 + \epsilon/2)$. Thus,

$$|\mathsf{N}| \operatorname{vol}(\mathsf{B}(\mathbf{0}, \epsilon/2)) \leq \operatorname{vol}(\mathsf{B}(\mathbf{0}, 1+\epsilon/2)).$$
 (11)



Hence,

$$|\mathsf{N}| \leq \frac{\operatorname{vol}(\mathsf{B}(\mathbf{0}, 1 + \epsilon/2))}{\operatorname{vol}(\mathsf{B}(\mathbf{0}, \epsilon/2))}$$
(12)
$$= \left(\frac{1 + \epsilon/2}{\epsilon/2}\right)^{d} = (1 + 2/\epsilon)^{d}$$
(13)
$$\leq (3/\epsilon)^{d}$$
(14)

Lemma (Discretization)

Suppose we have a set $\bar{N} \subseteq \Sigma_k$ with the following property: for all $x \in \Sigma_k$, there exists $\bar{x} \in \bar{N}$ such that

 $\bullet \ |\textit{supp}(\bar{\bm{x}}) \cup \textit{supp}(\bm{x}|) \ \leq \ k$

•
$$\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2 \leq \epsilon.$$

set

$$\delta_{\bar{\mathsf{N}}} = \max_{\bar{\boldsymbol{x}} \in \bar{\mathsf{N}}} \left| \|\boldsymbol{A}\bar{\boldsymbol{x}}\|_2^2 - 1 \right|.$$
(15)

Then

$$\delta_k(\mathbf{A}) \leq \frac{\delta_{\bar{\mathbf{N}}} + 2\epsilon}{1 - 2\epsilon}.$$
 (16)

Implications: RIP constant δ does not change much if we restrict our calculation to a finite ϵ -covering set \overline{N} .

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

Lemma (ϵ -Nets for Σ_k)

There exists an ϵ -net \bar{N} for Σ_k satisfying the two properties required in Lemma 6, with

$$\bar{\mathsf{N}} \Big| \le \exp\Big(k\log(3/\epsilon) + k\log(n/k) + k\Big).$$
 (17)

Proof.

Constructing an ϵ -Net for each ball in Σ_k and take the union. Using the Stirling's formula,⁴ we can estimate

$$\left|\bar{\mathsf{N}}\right| \leq (3/\epsilon)^k \binom{n}{k} \leq (3/\epsilon)^k \left(\frac{ne}{k}\right)^k.$$
 (18)

⁴Stirling's formula gives the bounds for factorials: $\sqrt{2\pi k} \left(\frac{k}{e}\right)^k \leq k! \leq e\sqrt{k} \left(\frac{k}{e}\right)^k$

EECS208, Fall 2021

15 / 21

Proof: Steps 2 and 3

Step 2: Tail Bound for Probability of Each Failure Case: For each $x \in \overline{N}$, Ax is a random vector with entries independent $\mathcal{N}(0, 1/m)$. We have

$$\mathbb{P}\left[\left|\left\|\boldsymbol{A}\boldsymbol{x}\right\|_{2}^{2}-1\right| > t\right] \leq 2\exp(-mt^{2}/8).$$
(19)

Step 3: Union Bound for Probability of All Failure Cases: Summing over all elements of $\bar{N},$ we have

$$\mathbb{P}\left[\delta_{\bar{\mathsf{N}}} > t\right] \leq 2 \left|\bar{\mathsf{N}}\right| \exp\left(-mt^2/8\right) \tag{20}$$

$$\leq 2 \exp\left(-\frac{mt^2}{8} + k \log\left(\frac{n}{k}\right) + k \left(\log\left(\frac{3}{\epsilon}\right) + k\right)\right). \tag{21}$$

On the complement of the event $\delta_{\bar{\mathsf{N}}} > t$, we have

$$\delta_k(\mathbf{A}) \le \frac{2\epsilon + t}{1 - 2\epsilon}.$$
(22)

Setting $\epsilon = \delta/8$, $t = \delta/4$, and ensuring that $m \ge Ck \log(n/k)/\delta^2$ for sufficiently large numerical constant C, we obtain the result.

RIP of Order k for Gaussian Matrices $\boldsymbol{A} \in \mathbb{R}^{m imes n}$

From the above derivation, especially from equation (21), we see that a slight more tight bound for m is of the form

 $m \ge 128k \log(n/k)/\delta^2 + (\log(24/\delta) + 1)k/\delta^2 \doteq C_1 k \log(n/k) + C_2 k.$

For a small δ , the constants C_1 and C_2 can be rather large.

A much tighter bound (one of the best known) for m is given as⁵:

 $m \ge 8k \log(n/k) + 12k.$

A precise (phase transition) expression of m as function of k, n exists (Lecture Eight: Section 3.6 or Chapter 6).



⁵On sparse reconstruction from Fourier and Gaussian measurements, M. Rudelson and R. Vershynin. Comm. on Pure and Applied Mathematics, 2008.

EECS208, Fall 202

RIP of Random Unitary Matrices

Motivating example: recall the MRI sensing model:

 $oldsymbol{y} = oldsymbol{F}_\Omega oldsymbol{\Psi} oldsymbol{x}, \quad$ with $oldsymbol{F}$ Fourier and $oldsymbol{\Psi}$ wavelet.

Theorem

Let $U \in \mathbb{C}^{n \times n}$ be unitary $(U^*U = I)$ and Ω is a random subset of m elements from $\{1, \ldots, n\}$. Suppose that

$$\|\boldsymbol{U}\|_{\infty} \leq \zeta/\sqrt{n}. \tag{23}$$

lf

$$m \geq \frac{C\zeta^2}{\delta^2} k \log^4(n), \tag{24}$$

then with high probability, $\mathbf{A} = \sqrt{\frac{n}{m}} \mathbf{U}_{\Omega,\bullet}$ satisfies the RIP of order k, with constant $\delta_k(\mathbf{A}) \leq \delta$.

Circulant Convolution Matrices

A (random) circulant convolution:

$$\boldsymbol{r} * \boldsymbol{x} = \begin{bmatrix} r_0 & r_{n-1} & \dots & r_2 & r_1 \\ r_1 & r_0 & r_{n-1} & & r_2 \\ \vdots & r_1 & r_0 & \ddots & \vdots \\ r_{n-2} & \ddots & \ddots & r_{n-1} \\ r_{n-1} & r_{n-2} & \dots & r_1 & r_0 \end{bmatrix} \boldsymbol{x} \doteq \boldsymbol{R} \boldsymbol{x}. \quad (25)$$

Fact: any circulant matrix can be diagonalized by the discrete Fourier transform⁶:

$$R = FDF^*$$
.

Select a (random) subset of the measurements:

$$\boldsymbol{y} = \mathcal{P}_{\Omega}[\boldsymbol{r} \ast \boldsymbol{x}] = \boldsymbol{A}\boldsymbol{x}, \tag{26}$$

⁶See Appendix A.7.

Ma (EECS Department, UC Berkeley)

19/21

RIP of Random Circulant Convolution Matrices

Let r be a random vector with independent zero-mean, subgaussian random variables of variance one.

Theorem

Let $\Omega \subseteq \{1, \ldots, n\}$ be any fixed subset of size $|\Omega| = m$. Then if

$$m \ge \frac{Ck \log^2(k) \log^2(n)}{\delta^2},\tag{27}$$

then with high probability, A has RIP of order k with $\delta_k(A) \leq \delta$.

Approximate isometric property is the key to deep convolution neural networks!⁷

⁷Deep Isometric Learning for Visual Recognition, H. Qi, C. You, X. Wang, Yi Ma, and J. Malik, ICML 2020.

Ma (EECS Department, UC Berkeley)

EECS208, Fall 2021

September 16, 2021 20 / 21

Assignments

• Reading: Section 3.4 of Chapter 3.

э

Image: A matrix