Topic of the class:

Singular Value of Random Mat

# 1. Introduction

## 1.1 Why Singular Values?

① Linear equations $y = Ax$

are simplest approximation

for any $\underset{\text{smooth}}{\underline{\text{continuous}}}$ model

due 2 Taylor's Thm for small

variable range → good approximation.

(under some regularity conditions)

important aspect:

stability: : How do small pert in $x$ change $y$?

How do small permitations in $y$ change the reconstruction quality for $x$

suitable measure for the quality of A

condition number: $\frac{S_{max}}{S_{min}}$

How singular values are allowed to control stability?

When we have freedom to design A (possibly via model para)

we want condition number to be small

Ideal situation: Approximate isometry

(all singular values $\approx 1$)

1.2 Why random matrices?

(a) Compressed Sensing

- For square matrices identity is perfectly well-conditioned

- Flat rectangular matrix

nontrival kernel

→ worst conditioning (no unique solution)

$\boxed{\text{In signal processing}}$, signal of interest are often modeled as being approximately sparse

→ most entries are very small (but we do not know where the large entries are)

<u>Question:</u> Can we choose $A$ such that all $k$-column matrices are approximate isometries?

do random construction

How many rows do we need ?

sample $N$ rows identity

deterministic : $\sim k^{2/y}$ rows $/ y = 10^{-2}$

random : $\sim k \log N$

(b) $\boxed{\text{Dimension Reduction}}$

$p$ points $\{x_1, \cdots, x_p\}$ in $\mathbb{R}^N$

can we project $\mathbb{R}^N \to \mathbb{R}^n$ using a matrix $A$ such that <u>geometry is preserved</u> ?

$$(1-\varepsilon)\|x_i - x_j\|_2 \le \|Ax_i - Ax_j\|_2$$
$$\le (1+\varepsilon)\|x_i - x_j\|_2$$

- Deterministic method must adapt to $\{x_1, \dots, x_l\}$, no single method will work for all sets

(isometry Dimension reduction) bad, we don't know set

- Random methods work for every set with high probability

$\Rightarrow$ no adaption necessary

(just set of failure is different)

Strong Advantage for high-dimensional data processing

## 1.3 Asymptotic and non-asymptotic regimen (组织方法)

Rand mat theory studies

properties of

$N \times n$ matrices $A$ chosen from some distribution on the set of all matrices

Observation:

$N, n \to \infty \Rightarrow$ spectrum of $A$ stabilize

Math formulation:

$\boxed{\text{limit laws}}$ (rand mat version of CLT)

Example:

Bai-Yin Law:

$A \in \mathbb{R}^{n \times n}$, entries i.i.d. standard normal r.v.s

Then $\frac{S_{max}(A)}{2\sqrt{n}} \to 1$ a.s. as $n \to \infty$

This gives no info about the rate.

We prefer non-asy version:

$S_{max}(A) \leq C\sqrt{n}$ with prob at least $1 - e^{-n}$ for some absolute constant $C$.

o less precise absolute constant $C$ :

most quantative

(exp small prob of failure for fixed dimension)

## 1.4 Guiding paradigm

Tall random matrices should act as approximate isometries

$N \times n$ random matrix $A$ with $N >> n$ should satisfy

$$(1-\delta)\, k \|x\|_2 \le \|Ax\|_2 \le (1+\delta)\, k \|x\|_2 \quad \forall x \in \mathbb{R}^n$$

normalization

$\delta \ll 1$

Equiv, $(1-\delta)k \le s_{min}(A) \le (1+\delta)k$;

yet equiv, condition number

$$\frac{s_{max}(A)}{s_{min}(A)} \le \frac{1+\delta}{1-\delta} \approx 1$$

1.5 In this class

We study tall random matrices with <u>independent rows</u> or <u>independent columns</u>

either with <u>strong moment assumptions</u> (sub-Gaussian)

or <u>no moment assumptions</u> (except

finite variance) $\to$ heavy tailed

Four types of senerious:

1/2 Matrices with inde subgaussian rows/cols

3/4 ... inde heavy-tailed rows/cols

important examples: Gaussian, Bernoulli, random Fourier

Applications: compressed sensing
dim reduction

# 2. Preliminaries

## 2.1 Matrices and Singular values

We mostly study tall $A \in \mathbb{R}^{N \times n}$ or $\mathbb{C}^{N \times n} \to N \geq n > 1$

(for flat matrices consider adjecnt)

**Definition 2.1** (Singular Values)

The numbers

$$s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A) \geq 0$$

such that $s_i^2(A)$ are the eigen value of $A^* A$ are called

singular values of $A$

We also write the extreme singular values as

$$s_{max}(A) = s_1(A)$$

$$s_{min}(A) = s_m(A)$$