RECALL

- Sub-gaussian random vectors X: one-dim marginals are sub-gaussians.

$$\|X\|_{\psi_2} := \sup_{z \in S^{n-1}} \|\langle X, z\rangle\|_{\psi_2}$$

- Special case

Vectors X with independent sub-gaussian components $K := \max_i \|X_i\|_{\psi_2}$

(i) $\|X\|_{\psi_2} \le CK$

(ii) Norm Concentration (Annulus Thm.)

$$\| \|X\| - \sqrt{n} \|_{\psi_2} \le C K^2$$

- **Johnson - Lindenstrauss (JL) Lemma**

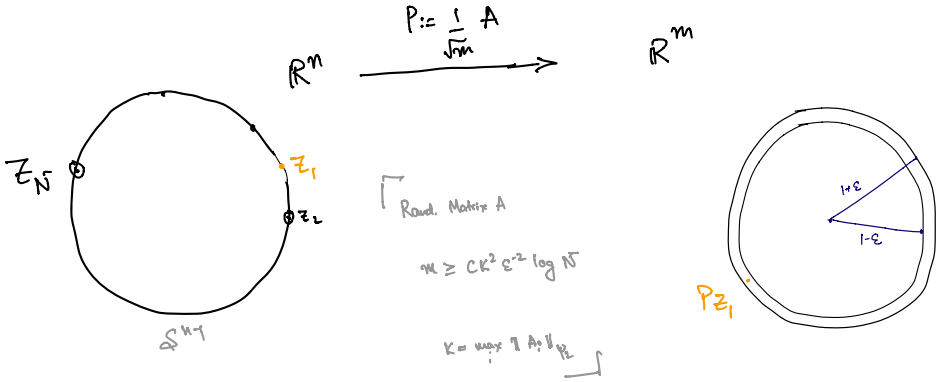
$X \subseteq \mathbb{R}^n \qquad |X| = N$

with high prob, for all $z \in X$

$$\|Pz\| = (1 \pm \varepsilon)\|z\|$$

$$\mathbb{E}\left[\max_z \left| \|Pz\| - \|z\| \right|\right] \le \varepsilon$$

for $m \gtrsim K^2 \varepsilon^{-2} \log N$



Sparse JL

C.f. [Achlioptas '93]

JL holds as long as the rows of rand. matrix A are independent, mean zero, isotropic, and sub-gaussian.

Hence, the following sparse A suffices

$$A_{ij} \sim (\text{ind}) \begin{cases} \sqrt{3} & \text{w.p. } 1/6 \\ -\sqrt{3} & \text{w.p. } 1/6 \\ 0 & \text{w.p. } 2/3 \end{cases}$$

[in expectation, $\frac{2}{3}$rd entries of A are zero.]

$P = \frac{1}{\sqrt{m}} A$

FAST JL [Ailon-Chazelle '06]

Standard JL transform requires $O\left(\frac{n \log N}{\varepsilon^2}\right)$ operations per vector x

Γ $A \in \mathbb{R}^{m \times n}$ $z \in X \subseteq \mathbb{R}^n$

$m = O\left(\frac{\log N}{\varepsilon^2}\right)$ $\rfloor$

A JL transform with running time $O(n \log n + m)$, per vector

[Use case $N \gg n$]

+ Fix unit vector $u \in S^{n-1}$ (wish to preserve norm of u)

+ Consider sparse matrix $S \in \mathbb{R}^{m \times n}$ with ind. rows $S_1, .. S_m$ with

$S_i \in_R \sqrt{n} \{ \pm e_1, \pm e_2, .. \pm e_n \}$ [e_i : standard basis vector $\rfloor$]

$\mathbb{E} S_i = 0$ (Mean zero) $\quad \mathbb{E} S_i S_i^T = I_n$ (Isotropic)

R.V. $S_i^T u$ is bounded

$\| S_i^T u \|_{\psi_2} \le C \| S_i^T u \|_\infty$

$= C \sqrt{n} \cdot \|u\|_\infty$

+ Each row is one sparse. Hence, computing Su requires $O(n)$ time (and space)

+ The vector $Su = \begin{pmatrix} S_1^T u \\ \vdots \\ S_m^T u \end{pmatrix}$ has ind, sub-g components with

$\mathbb{E} (S_i^T u)^2 = u^T (\mathbb{E} S_i S_i^T) u$

$= u^T I_n u = 1$

+ Norm Concentration

$$\left\| \|Su\| - \sqrt{m} \right\|_{\psi_2} \le C n \|u\|_\infty^2 \quad \text{— (1)}$$

Lemma Let $u \in S^{n-1}$ be a fixed unit vector, with $\|u\|_\infty \le \sqrt{\frac{\log n}{n}}$.

Also, let $S \in \mathbb{R}^{m \times n}$ be a rand. row sampling matrix with $m \gtrsim \frac{\log^2(n/\delta)}{\varepsilon^2}$.

Then, $\mathbb{P}_S \left\{ \left\| \frac{1}{\sqrt{m}} Su \right\| \notin (1 \pm \varepsilon) \cdot \|u\| \right\} \le \delta/2$.

Key Idea: • Randomly "rotate the basis" such that

$x \longmapsto u$

$\|u\|_\infty \le \sqrt{\frac{\log n}{n}}$ holds with high prob.

• Use a "structured" matrix M such that

(i) $\|Mx\|_\infty \le \sqrt{\frac{\log n}{n}}$

(ii) M admits fast matrix-vector mult.

Hadamard Matrices

Recursive Defn. $H_d := \frac{1}{\sqrt{2}} \begin{pmatrix} H_{d/2} & H_{d/2} \\ H_{d/2} & -H_{d/2} \end{pmatrix}$

$H_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

Write $H := H_n$

$H = \frac{1}{\sqrt{n}} \begin{pmatrix} \cdots \pm 1 \cdots \\ \vdots \end{pmatrix}$

(a) Rows of H are orthonormal

(b) Recursive defn. ensures Hx can be computed in $O(n \log n)$ time.

(c) Cannot expect $\|Hx\|_\infty \le \sqrt{\frac{\log n}{n}}$ for all x (e.g., x is a row of H).

Idea: Randomize over column signs

Let diagonal matrix D such that $D_{ii} \in_R \{\pm 1\}$

For fixed unit vector x, consider HDx

Lemma: Let $x \in S^{n-1}$ and $u := HDx$, where HD is a random Hadamard matrix. Then,

$\Pr_D \left\{ \|u\|_\infty > \sqrt{\frac{\log(n/\delta)}{n}} \right\} \le \delta/2$.

Pf: $H_{ij} \in_R \left\{ \frac{1}{\sqrt{n}}, \frac{-1}{\sqrt{n}} \right\}$ Ind. Bernoulli (columnwise ind.)

$\|H_{ij}\|_{\psi_2} \le \frac{C'}{\sqrt{n}}$

$\|H_i^T x\|_{\psi_2}^2 \le C' \cdot \|x\|^2 \cdot \frac{C'}{n}$ (Hoeffding's Inequ.)

$= \frac{C}{n}$.

$\|H_i^T x\|_{\psi_2} \le \frac{C}{\sqrt{n}}$.

$\mathbb{E} \max_i |H_i^T x| \le \frac{C}{\sqrt{n}} \cdot \sqrt{\log n}$

Using tail prob. we get the desired dependence on δ.

Thm (Fast JL): There exists a random matrix $M \in \mathbb{R}^{m \times n}$ with $m \gtrsim \varepsilon^{-2} \log\left(\frac{n}{\delta}\right)$ such that for any $x \in \mathbb{R}^n$

$$\|Mx\| = (1 \pm \varepsilon) \cdot \|x\|$$ holds w.p. at least $(1-\delta)$.

In addition, matrix-vector multiplication with M takes $O(n \log n + m)$ time.

Pf: Fix $x \in S^{n-1}$. Two "bad" events

$$B_1 := \left\{ \|HDx\|_\infty > \sqrt{\frac{\log(m/\delta)}{n}} \right\}$$

$$B_2 := \left\{ \|Mx\| \notin (1 \pm \varepsilon)\|x\| \right\}$$

$M = S \cdot HD$

To prove the thm, suffices to bound $\Pr\{B_2\} \le \delta$.

$\Pr\{B_1\} \le \delta/2$ (Lemma 2)

$\Pr\{B_2 \mid \neg B_1\} \le \delta/2$ (Lemma 1)

$$\Pr\{B_2\} = \Pr\{B_2 \mid B_1\} \cdot \Pr\{B_1\} + \Pr\{B_2 \mid \neg B_1\} \cdot \Pr\{\neg B_1\}$$

$$\le 1 \cdot \frac{\delta}{2} + \frac{\delta}{2} \cdot 1 = \delta$$

□

Applications of JL

① Pairwise Distances (Euclidean)

$X \subseteq \mathbb{R}^n$

- Naive, Exact Computation $O(N^2 n)$
- Approx computation JL

  JL Matrix $P: \mathbb{R}^n \to \mathbb{R}^m$

  Computation time

  $O(\text{JL time} + N^2 m)$

  Using Fast JL we get $m = O(\varepsilon^{-2} \log^2 n)$

  with JL time $(Nn \log n)$

  Overall running time

  $O(Nn \log n + N^2 \varepsilon^{-2} \log^2 N)$

② Streaming Algo. for F_2 estimation

[Space Constrained Environment]

Pioneering Work of: Alon, Matias, & Szegedy
* Streaming Model of Computation

- Given a stream of N indices
  $i_1, i_2, \ldots i_N$ $\quad i_t \in \{1, 2, \ldots, n\}$

- Maintain fq. vector
  $f_j := |\{t : i_t = j\}|$

Ex $n=3$ $N=6$
$[1, 1, 3, 3, 1, 2]$
$f = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$

[Trivial in $O(n \log n)$ space. Here, will use $O(\text{poly} \log n)$ space]

- Goal: Approximate the ℓ_2 norm $\|f\|_2$

Idea: Instead of storing f, maintain a projection of f.
$$y = P \cdot f$$
[$P \leftarrow$ Fast JL transform]

At time t, alg receives $i_t = j$

Update $y \leftarrow y + P \cdot e_j$

(add the jth column of P to y)

Analysis

Via JL we have that, whp,

$$(1-\varepsilon)\|f\| \le \|Pf\| \le (1+\varepsilon)\|f\|$$

Setting $m \gtrsim \frac{1}{\varepsilon^2}$, we have this bound.

(with $m \gtrsim \frac{1}{\varepsilon^2} \log N$ this guarantee is achieved for all N steps).