CIS 700: "algorithms for Big Data"

Lecture 5: Dimension Reduction

Slides at http://grigory.us/big-data-class.html

Grigory Yaroslavtsev

http://grigory.us

Penn
University of Pennsylvania

Today

- Dimensionality reduction
  - AMS as dimensionality reduction
  - Johnson-Lindenstrauss transform

L_p-norm Estimation

- Stream: $\boldsymbol{m}$ updates $(x_i, \Delta_i) \in [n] \times \mathbb{R}$ that define vector f where $f_j = \sum_{i:x_i=j} \Delta_i$.
- Example: For $n = 4$

$$\langle (1,3), (3, 0.5), (1,2), (2, -2), (2,1), (1, -1), (4,1) \rangle$$
$$f = (4, -1, 0.5, 1)$$

- L_p-norm: $||f||_p = \left(\sum_i |f|^p\right)^{\frac{1}{p}}$

L_p-norm Estimation

- L_p-norm: $||f||_p = (\sum_i |f|^p)^{\frac{1}{p}}$
- Two lectures ago:
  - $||f||_0 = F_0$-moment
  - $||f||_2^2 = F_2$-moment (via AMS sketching)
- Space: $O\left(\frac{\log n}{\epsilon^2} \log \frac{1}{\delta}\right)$
- Technique: linear sketches
  - $||f||_0$: $\sum_{i \in S} f_i$ for random sets S
  - $||f||_2^2$: $\sum_i \sigma_i f_i$ for random signs σ_i

AMS as dimensionality reduction

- Maintain a "linear sketch" vector

$$\boldsymbol{Z} = (Z_1, \dots, Z_k) = Rf$$

$$Z_i = \sum_{j \in [n]} \sigma_{ij} f_j, \quad \text{where } \sigma_{ij} \in_R \{-1,1\}$$

- Estimator $\boldsymbol{Y}$ for $||f||_2^2$:

$$\frac{1}{k} \sum_{i=1}^{k} Z_i^2 = \frac{||Rf||_2^2}{k}$$

- "Dimensionality reduction": $x \to Rx$, "heavy" tail

$$\Pr\left[\left|\boldsymbol{Y} - ||f||_2^2\right| \geq c \left(\frac{2}{k}\right)^{\frac{1}{2}} ||f||_2^2\right] \leq \frac{1}{c^2}$$

Normal Distribution

- Normal distribution $N(0,1)$
  - Range: $(-\infty, +\infty)$
  - Density: $\mu(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}$
  - Mean = 0, Variance = 1
- Basic facts:
  - If X and Y are independent r.v. with normal distribution then $X + Y$ has normal distribution
  - $Var[cX] = c^2 Var[X]$
  - If X, Y are independent, then $Var[X + Y] = Var[X] + Var[Y]$

Johnson-Lindenstrauss Transform

- Instead of ± 1 let σ_i be i.i.d. random variables from normal distribution $N(0,1)$

$$Z = \sum_i \sigma_i f_i$$

- We still have $\mathbb{E}[Z^2] = \sum_i f_i^2 = ||f||_2^2$ because:
  - $\mathbb{E}[\sigma_i]\mathbb{E}[\sigma_j] = 0$; $\mathbb{E}[\sigma_i^2] =$ "variance of σ_i" $= 1$
- Define $\boldsymbol{Z} = (Z_1, \ldots, Z_k)$ and define:

$$||\boldsymbol{Z}||_2^2 = \sum_j Z_j^2 \quad \left(\mathbb{E}\left[||\boldsymbol{Z}||_2^2\right] = k||f||_2^2\right)$$

- JL Lemma: There exists $C > 0$ s.t. for small enough $\epsilon > 0$:

$$\Pr\left[\left|||\boldsymbol{Z}||_2^2 - k\,||f||_2^2\right| > \epsilon k||f||_2^2\right] \le \exp(-C\epsilon^2 k)$$

Proof of JL Lemma

- JL Lemma: $\exists C > 0$ s.t. for small enough $\epsilon > 0$:

$$\Pr\left[\left|\left||\boldsymbol{Z}|\right|_2^2 - k\left||f|\right|_2^2\right| > \epsilon k\left||f|\right|_2^2\right] \le \exp(-C\epsilon^2 k)$$

- Assume $\left||f|\right|_2^2 = 1$.
- We have $\boldsymbol{Z}_i = \sum_j \sigma_{ij} f_i$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_k)$

$$\mathbb{E}\left[\left||\boldsymbol{Z}|\right|_2^2\right] = k\left||f|\right|_2^2 = k$$

- Alternative form of JL Lemma:

$$\Pr\left[\left||\boldsymbol{Z}|\right|_2^2 > k(1+\epsilon)^2\right] \le \exp(-\epsilon^2 k + O(k\,\epsilon^3))$$

Proof of JL Lemma

- Alternative form of JL Lemma:

$$\Pr\left[\left||\boldsymbol{Z}|\right|_2^2 > k(1+\epsilon)^2\right] \le \exp(-\epsilon^2 k + O(k\,\epsilon^3))$$

- Let $\boldsymbol{Y} = \left||\boldsymbol{Z}|\right|_2^2$ and $\alpha = k(1+\epsilon)^2$
- For every $\boldsymbol{s} > 0$ we have:

$$\Pr[\boldsymbol{Y} > \alpha] = \Pr[e^{\boldsymbol{s}\boldsymbol{Y}} > e^{\boldsymbol{s}\alpha}]$$

- By Markov and independence of $\boldsymbol{Z}_i's$:

$$\Pr[e^{\boldsymbol{s}\boldsymbol{Y}} > e^{\boldsymbol{s}\alpha}] \le \frac{\mathbb{E}[e^{\boldsymbol{s}\boldsymbol{Y}}]}{e^{\boldsymbol{s}\alpha}} = e^{-\boldsymbol{s}\alpha}\mathbb{E}\left[e^{\boldsymbol{s}\sum_i \boldsymbol{Z}_i^2}\right] = e^{-\boldsymbol{s}\alpha}\prod_{i=1}^{k}\mathbb{E}\left[e^{\boldsymbol{s}\boldsymbol{Z}_i^2}\right]$$

- We have $Z_i \in N(0,1)$, hence:

$$\mathbb{E}\left[e^{\boldsymbol{s}\boldsymbol{Z}_i^2}\right] = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty} e^{\boldsymbol{s}t^2} e^{-\frac{t^2}{2}}dt = \frac{1}{\sqrt{1-2\boldsymbol{s}}}$$

Proof of JL Lemma

- Alternative form of JL Lemma:

$$\Pr\left[\left||\boldsymbol{Z}|\right|_2^2 > k(1+\epsilon)^2\right] \le \exp(-\epsilon^2 k + O(k\,\epsilon^3))$$

- For every $\boldsymbol{s} > 0$ we have:

$$\Pr[\boldsymbol{Y} > \alpha] \le e^{-\boldsymbol{s}\alpha} \prod_{i=1}^{k} \mathbb{E}\left[e^{\boldsymbol{s}\boldsymbol{Z}_i^2}\right] = e^{-\boldsymbol{s}\alpha}(1 - 2\boldsymbol{s})^{-\frac{k}{2}}$$

- Let $\boldsymbol{s} = \frac{1}{2}\left(1 - \frac{k}{\alpha}\right)$ and recall that $\alpha = k(1+\epsilon)^2$
- A calculation finishes the proof:

$$\Pr[\boldsymbol{Y} > \alpha] \le \exp(-\epsilon^2 k + O(k\,\epsilon^3))$$

Johnson-Lindenstrauss Transform

- Single vector: $k = O\left(\frac{\log\frac{1}{\delta}}{\epsilon^2}\right)$
  - Tight: $k = \Omega\left(\frac{\log\frac{1}{\delta}}{\epsilon^2}\right)$ [Woodruff'10]
- $\boldsymbol{n}$ vectors simultaneously: $k = O\left(\frac{\log \boldsymbol{n} \log\frac{1}{\delta}}{\epsilon^2}\right)$
  - Tight: $k = \Omega\left(\frac{\log \boldsymbol{n} \log\frac{1}{\delta}}{\epsilon^2}\right)$ [Molinaro, Woodruff, Y. '13]
- Distances between $\boldsymbol{n}$ vectors = $O(\boldsymbol{n}^2)$ vectors:

$$k = O\left(\frac{\log \boldsymbol{n} \log\frac{1}{\delta}}{\epsilon^2}\right)$$

Random Variables and Norms

- For a random variable $\boldsymbol{X}$ and $\mathrm{p} \geq 1$ let:

$$||\boldsymbol{X}||_p = \mathbb{E}[X^p]^{1/p}$$

Facts:

- For any c: $||c\boldsymbol{X}||_p = c||\boldsymbol{X}||_p$
- $||\cdot||_p$ is a norm (Minkowski's inequality)
- $||\cdot||_p \leq ||\cdot||_q$ for $p \leq q$ (Monotonicity of norms)
- Jensen's inequality (used a lot for $F = |x|^p$):

If F is convex then $F(\mathbb{E}[\boldsymbol{X}]) \leq \mathbb{E}[F(\boldsymbol{X})]$

Khintchine Inequality

- [Khintchine]For $p \geq 1, x \in \mathbb{R}^n$ and σ_i i.i.d. Rademachers:

$$\left\|\sum_i \sigma_i x_i\right\|_p \lesssim \sqrt{p}\,\|x\|_2$$

- For r_i (either σ_i or $g_i \sim N(0,1)$) expand $\mathbb{E}[(\sum_i r_i x_i)^p]$
- All odd powers of r_i are zero
- All even moments for σ_i are 1, and for g_i are ≥ 1
- $\left\|\sum_i \sigma_i x_i\right\|_p \leq \left\|\sum_i g_i x_i\right\|_p$
- $\sum_i g_i x_i \sim N\left(0, \|x\|_2^2\right) \Rightarrow \left\|\sum_i g_i x_i\right\|_p \lesssim \sqrt{p}\|x\|_2$

Symmetrization

- [Symmetrization]: If $Z_1, \dots, Z_n$ are independent and σ_i are i.i.d. Rademachers:

$$\left\| \sum_i Z_i - \mathbb{E} \sum_i Z_i \right\|_p \le 2 \left\| \sum_i \sigma_i Z_i \right\|_p$$

- Let $Y_1 \dots Y_n$ be independent with the same distribution as Z_i
- $\left\|\sum_i Z_i - \mathbb{E} \sum_i Z_i\right\|_p = \left\|\sum_i Z_i - \mathbb{E}_Y \sum_i Y_i\right\|_p$

$\le \left\|\sum_i (Z_i - Y_i)\right\|_p$ (Jensen)

$= \left\|\sum_i \sigma_i (Z_i - Y_i)\right\|_p$ ($\mathrm{Z_i} - \mathrm{Y_i}$ are independent and symmetric)

$\le 2 \left\|\sum_i \sigma_i Z_i\right\|_p$ (triangle inequality)

Decoupling

- Let $x_1, \ldots x_n$ be independent with mean 0 and $x'_1, \ldots x'_n$ identically distributed as x_i and independent of them. For any a_{ij} and $p \geq 1$:

$$\left\|\sum_{i \neq j} a_{ij} x_i x_j\right\|_p \leq 4 \left\|\sum_{i,j} a_{ij} x_i x'_j\right\|_p$$

- Let $\eta_1, \ldots, \eta_n$ be i.i.d. Bernoullis (0/1 w.p. 1/2):

$$\left\|\sum_{i \neq j} a_{ij} x_i x_j\right\|_p = 4 \left\|\mathbb{E}_\eta \sum_{i \neq j} a_{ij} x_i x_j \eta_i (1 - \eta_j)\right\|_p$$

$$\leq 4 \left\|\sum_{i \neq j} a_{ij} x_i x_j \eta_i (1 - \eta_j)\right\|_p \text{ (Jensen)}$$

- There exists $\eta' \in \{0,1\}^n$ such that:

$$\left\|\sum_{i \neq j} a_{ij} x_i x_j \eta_i (1 - \eta_j)\right\|_p \leq \left\|\sum_{i \in S} \sum_{j \in \bar{S}} a_{ij} x_i x_j\right\|_p$$

where $S = \{i : \eta' = 1\}$.

Decoupling (continued)

Let x_S be an S-dimensional vector of x_i for $i \in S$.

- $\left\|\sum_{i\in S}\sum_{j\in\bar{S}} a_{ij}x_ix_j\right\|_p = \left\|\sum_{i\in S}\sum_{j\in\bar{S}} a_{ij}x_i{x'}_j\right\|_p$

$= \left\|\mathbb{E}_{x_{\bar{S}}}\mathbb{E}_{x'_S}\sum_{i,j} a_{ij}x_ix_j'\right\|_p$ $(\mathbb{E}[\mathrm{x_i}] = \mathbb{E}[\mathrm{x'_i}] = 0)$

$\leq ||\sum_{i,j} a_{ij}\, x_i\, x_j'\, ||_p$ (Jensen)

- Overall:

$$\left\|\sum_{i\neq j} a_{ij}x_ix_j\right\|_p \leq 4\left\|\sum_{i,j} a_{ij}x_ix_j'\right\|_p$$

Hanson-Wright Inequality

- For $\sigma_1, \dots, \sigma_n$ independent Rademachers and $A \in \mathbb{R}^{n \times n}$ real and symmetric for all $p \geq 1$:
$$\left\| \sigma^T A \sigma - \mathbb{E}[\sigma^T A \sigma] \right\|_p \lesssim \sqrt{p} \, \|A\|_F + p \, \|A\|$$
- Recall:
  - $\|A\|_F = \sqrt{\sum_{ij} a_{ij}^2} = \sqrt{Tr(A^T A)}$
  - $\|A\| = \sup_{\{v \neq 0\}} \frac{\|Av\|_2}{\|v\|_2}$

Hanson-Wright Inequality

- For $\sigma_1, \ldots, \sigma_n$ independent Rademachers and $A \in \mathbb{R}^{n \times n}$ real and symmetric for all $p \geq 1$:

$$\left|\left|\sigma^T A\sigma - \mathbb{E}[\sigma^T A\sigma]\right|\right|_p \preccurlyeq \sqrt{p}\left|\left|A\right|\right|_F + p\left|\left|A\right|\right|$$

$$\left|\left|\sigma^T A\sigma - \mathbb{E}[\sigma^T A\sigma]\right|\right|_p \preccurlyeq \left|\left|\sigma^T A\sigma'\right|\right|_p \text{ (decoupling)}$$

$$\preccurlyeq \sqrt{p}\left|\left|\left|\left|A\sigma\right|\right|_2\right|\right|_p \text{ (Khintchine)}$$

$$= \sqrt{p}\left|\left|\left|\left|A\sigma\right|\right|_2^2\right|\right|_{p/2}^{\frac{1}{2}}$$

$$\leq \sqrt{p}\left|\left|\left|\left|A\sigma\right|\right|_2^2\right|\right|_p^{\frac{1}{2}} \text{ (monotonicity of norms)}$$

Hanson-Wright (continued)

$$\sqrt{p}\ \left\|\left||A\sigma|\right|_2\right\|_p \leq \cdots \leq \sqrt{p}\ \left\|\left||A\sigma|\right|_2^2\right\|_p^{\frac{1}{2}}$$

$$\leq \sqrt{p}\left(\mathbb{E}\left[\left||A\sigma|\right|_2^2\right] + \left\|\left||A\sigma|\right|_2^2 - \ \mathbb{E}\left[\left||A\sigma|\right|_2^2\right]\right\|_p\right)^{\frac{1}{2}} \text{(triangle ineq.)}$$

$$= \sqrt{p}\left(\left||A|\right|_F^2 + \left\|\left||A\sigma|\right|_2^2 - \ \mathbb{E}\left[\left||A\sigma|\right|_2^2\right]\right\|_p\right)^{\frac{1}{2}}$$

$$\leq \sqrt{p}\left||A|\right|_F + \sqrt{p}\left\|\left||A\sigma|\right|_2^2 - \ \mathbb{E}\left[\left||A\sigma|\right|_2^2\right]\right\|_p^{\frac{1}{2}}$$

$$\preccurlyeq \sqrt{p}\left||A|\right|_F + \sqrt{p}\left|\left|\sigma^T A^T A \sigma'\right|\right|_p^{\frac{1}{2}} \ \text{(decoupling)}$$

$$\preccurlyeq \sqrt{p}\left||A|\right|_F + p^{\frac{3}{4}}\ \left|\left|\left|\left|A^T A\sigma\right|\right|_2\right|\right|_p^{1/2} \text{(Khintchine)}$$

$$\preccurlyeq \ \sqrt{p}\left||A|\right|_F + p^{\frac{3}{4}}\left||A|\right|^{\frac{1}{2}}\left|\left|\left||Ax|\right|_2\right|\right|_p^{\frac{1}{2}}$$

Hanson-Wright (continued)

$$\sqrt{p}\ \left\| \left| \|A\sigma\|_2 \right| \right\|_p \lesssim \sqrt{p}\|A\|_F + p^{\frac{3}{4}}\|A\|^{\frac{1}{2}} \left| \left| \|A\sigma\|_2 \right| \right|_p^{\frac{1}{2}}$$

Let E = $\left| \left| \|Ax\|_2 \right| \right|_p^{\frac{1}{2}}$ then $E^2 - Cp^{\frac{1}{4}}\ \|A\|^{\frac{1}{2}} E\ - C\|A\|_F \leq 0$

- E $\leq$ larger root of the quadratic equation above
- $E^2 \leq \sqrt{p}\ \|A\|_F + p\ \|A\|$
- (Hanson-Wright) For $\sigma_1, \ldots, \sigma_n$ independent Rademachers and $A \in\ \mathbb{R}^{n \times n}$ real and symmetric for all $p \geq 1$:

$$\left\| \sigma^T A\sigma\ -\ \mathbb{E}[\sigma^T A\sigma] \right\|_p \lesssim \sqrt{p}\ \|A\|_F + p\ \|A\|$$

Recap

- For a random variable $\boldsymbol{X}$ and $\mathrm{p} \geq 1$ let:
$$\left|\left|\boldsymbol{X}\right|\right|_p = \mathbb{E}[X^p]^{1/p}$$
- [Khintchine]For $p \geq 1, x \in \mathbb{R}^n$ and σ_i i.i.d. Rademachers:
$$\left|\left|\sum_i \sigma_i x_i\right|\right|_p \lesssim \sqrt{p}\left|\left|x\right|\right|_2$$
- [Symmetrization]: If $Z_1, \dots, Z_n$ are independent and σ_i are i.i.d. Rademachers:
$$\left\|\sum_i Z_i - \mathbb{E}\sum_i Z_i\right\|_p \leq 2\left\|\sum_i \sigma_i Z_i\right\|_p$$
- [Hanson-Wright]For $\sigma_1, \dots, \sigma_n$ independent Rademachers and $A \in \mathbb{R}^{n\times n}$ real and symmetric for all $p \geq 1$:
$$\left|\left|\sigma^T A\sigma - \mathbb{E}[\sigma^T A\sigma]\right|\right|_p \lesssim \sqrt{p}\left|\left|A\right|\right|_F + p\left|\left|A\right|\right|$$

Bernstein Inequality

- Let $X_1, \dots, X_n$ be indep. r.v's such that $|\mathrm{X_i}| \leq K$ almost surely and $\mathbb{E}\left[\sum_i X_i^2\right] \leq \sigma^2$. For all $p \geq 1$:

$$\left\| \sum_i X_i - \mathbb{E}[X_i] \right\|_p \lesssim \sigma\sqrt{p} + Kp$$

$\left||\sum_i X_i - \mathbb{E}[X_i]|\right|_p \leq 2\left||\sum_i \sigma_i X_i|\right|_p$ (symmetrization)

$\lesssim \sqrt{p}||(\sum_i X_i^2)^{\frac{1}{2}}||_p$ (Khintchine)

$= \sqrt{p}\,||\sum_i X_i^2\,||_{\frac{p}{2}}^{\frac{1}{2}}$

$\leq \sigma\sqrt{p} + \sqrt{p}\,\left||\sum_i X_i^2 - \mathbb{E}\left[X_i^2\right]|\right|_p^{1/2}$ (triangle inequality)

Bernstein Inequality (cont.)

$$\left\|\sum_i X_i - \mathbb{E}[X_i]\right\|_p \leq \ \ldots \preccurlyeq \sqrt{p}\,\|(\textstyle\sum_i X_i^2)^{\frac{1}{2}}\|_p$$

$$\leq \sigma\sqrt{p} + \sqrt{p}\ \left\|\sum_i X_i^2 - \mathbb{E}\left[X_i^2\right]\right\|_p^{\frac{1}{2}}$$

$$\leq \sigma\sqrt{p} + \sqrt{p}\ \left\|\textstyle\sum_i \sigma_i X_i^2\right\|_p^{\frac{1}{2}} \ \text{(symmetrization)}$$

$$\preccurlyeq \sigma\sqrt{p} + p^{\frac{3}{4}}\left\|\textstyle\sum_i \left(X_i^4\right)^{1/2}\right\|_p^{\frac{1}{2}} \ \text{(Khintchine)}$$

$$\leq \sigma\sqrt{p} + p^{\frac{3}{4}}\sqrt{K}\left\|\sum_i \left(X_i^2\right)^{1/2}\right\|_p^{\frac{1}{2}}$$

Bernstein Inequality (cont.)

- Let $E = ||(\sum_i X_i^2)^{\frac{1}{2}}||_p$ then for some $C > 0$:

$$E^2 - Cp^{\frac{1}{4}}\sqrt{K}E - C\sigma \leq 0$$

- $E \geq$ larger root of this quadratic equation
- $E \lesssim \sigma\sqrt{p} + Kp$
- [Bernstein] Let $X_1, \dots, X_n$ be indep. r.v's such that $|\mathrm{X_i}| \leq K$ almost surely and $\mathbb{E}\left[\sum_i \quad X_i^2\right] \leq \sigma^2$. For all $p \geq 1$:

$$\left\|\sum_i X_i - \mathbb{E}[X_i]\right\|_p \lesssim \sigma\sqrt{p} + Kp$$

Sparse Johnson-Lindenstrauss Transform

- Let $\Pi \in \mathbb{R}^{m \times n}$ be a JL-matrix where $\mathrm{m} = O\left(\frac{1}{\epsilon^2} \log \frac{1}{\delta}\right)$ which satisfies for $\left||x|\right|_2 = 1$:

$$\Pr_{\Pi}\left[\left|\left||\Pi x|\right|_2^2 - 1\right| \geq \epsilon\right] \leq \delta$$

- Takes $\mathrm{O}\left(m\left||x|\right|_0\right)$ time to compute JL
- Would be $\mathrm{O}\left(s\left||x|\right|_0\right)$ time Π only had s non-zero entries per column

Basic Sparse JL Transform

- Pick 2-wise indep. hash function $h : [n] \to [m]$
- Pick 4-wise indep. hash function $\sigma : [n] \to \{-1,1\}$
- For each $i \in [n]$ let $\Pi_{h(i),i} = \sigma(i)$, the rest are 0
- [Thorup, Zhang'12]: This is JL if $m \gtrsim \frac{1}{\epsilon^2 \delta}$
- Best possible since $s = 1$
- Analysis: standard expectation/variance using bounded independence + Chebyshev
- To improve m let's use Hanson-Wright (higher moment than Chebyshev's second)

Sparse JL Transform: Construction

- $\Pi_{\mathrm{r,i}} = \eta_{r,i}\sigma_{r,i}/\sqrt{s}$, where η_i are Bernoullis and σ_i are Rademachers
- For all r, i: $\mathbb{E}[\eta_{r,i}] = \frac{s}{m}$
- For all i: $\sum_i \eta_{r,i} = s$ (s non-zeros per column)
- $\eta_{r,i}$ are negatively correlated:

$$\mathbb{E}\left[\prod_{(r,i)\in S} \eta_{r,i}\right] \leq \prod_{(r,i)\in S} \mathbb{E}[\eta_{r,i}] = \left(\frac{s}{m}\right)^{|S|}$$

- Each column chosen uniformly from $\mathrm{Binom}(m\ ,s)$ columns of weight s works here

Sparse JL Transform: Analysis

Thm [KN'14]: If $m = O\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$ and $s \approx \epsilon m$:

$$\forall x: ||x||_2 = 1, \Pr_{\Pi}\left[\left|||\Pi x||_2^2 - 1\right| \geq \epsilon\right] \leq \delta$$

- $Z = ||\Pi x||_2^2 - 1 =$

$$\frac{1}{s}\sum_{r=1}^{m}\sum_{i \neq j} \eta_{r,i}\,\eta_{r,j}\sigma_{r,i}\sigma_{r,j}x_i x_j \equiv \sigma^T A_{x,\eta}\,\sigma$$

- $A_{x,\eta}$ is a block-diagonal matrix with m blocks where r-th block is $\frac{1}{s}\,x^{(r)}\left(x^{(r)}\right)^T$ but with zeros on the diagonal
- $x^{(r)}$ is a vector with entries $x_i^{(r)} = \eta_{r,i}x_i$
- By Hanson-Wright: $||Z||_p \leq \left\|\sqrt{p}\ ||A_{x,\eta}||_F + p\ ||A_{x,\eta}||\right\|_p$

$$\leq \sqrt{p}\ ||\ ||A_{x,\eta}||_F\ ||_p + p\ ||\ ||A_{x,\eta}||\ ||_p$$

Sparse JL Transform: Analysis

- (Operator norm) Since $A_{x,\eta}$ is block-diagonal $||A_{x,\eta}||$ is the largest norm of any block
- Eigenvalues in the r-th block are at most

$$\frac{2}{s}\max\left(\left|\left|x^{(r)}\right|\right|_2^2, \left|\left|x^{(r)}\right|\right|_\infty^2\right) \leq \frac{2}{s}$$

- $\left|\left|A_{x,\eta}\right|\right| \leq \frac{2}{s}$

Sparse JL Transform: Analysis

- Define $Q_{i,j} = \sum_{r=1}^{m} \eta_{r,i}\, \eta_{r,j}$ so that:

$$\left|\left|A_{x,\eta}\right|\right|_F^2 = 1/s^2 \sum_{i \neq j} x_i^2 x_j^2 Q_{i,j}$$

- Lemma: If $p \approx s^2/m$ then $\forall i, j$ $\left|\left|Q_{i,j}\right|\right|_p \preccurlyeq p$

- $|| \left|\left|A_{x,\eta}\right|\right|_F \Big|_p = || \left|\left|A_{x,\eta}\right|\right|_F^2 \Big| |^{\frac{1}{2}}{}_p$

$$\leq \left|\left|\frac{1}{s^2} \sum_{i \neq j} x_i^2 x_j^2\ Q_{i,j}\right|\right|_p^{\frac{1}{2}}$$

$\leq \frac{1}{s}\left(\sum_{i \neq j} x_i^2 x_j^2\ \left|\left|Q_{i,j}\right|\right|_p\right)^{1/2}$ (triangle ineq.)

$\leq 1/\sqrt{m}$

Sparse JL Transform: Analysis

- By Markov ($m = O\left(\frac{1}{\epsilon^2}\log 1/\delta\right), s \approx \epsilon m, p \approx \frac{s^2}{m}$):

$$\Pr\left[\left|\|\Pi x\|_2^2 - 1\right| > \epsilon\right] =$$

$$\Pr\left[\left|\sigma^T A_{x,\eta}\sigma\right|^p > \epsilon^p\right] \leq$$

$\epsilon^{-p}\mathbb{E}\left[\left|\sigma^T A_{x,\eta}\sigma\right|^p\right]$ (Markov)

$$\leq \epsilon^{-p} C^p \left(\frac{\sqrt{p}}{\sqrt{m}} + \frac{p}{s}\right)^p = \epsilon^{-p} C^p (\epsilon + \epsilon)^p \leq \delta$$

Sparse JL Transform: Analysis

- Lemma: If $p \approx s^2/m$ then $\forall i, j \left|\left|Q_{i,j}\right|\right|_p \lesssim p$
- Suppose $\eta_{a_1,i}, \ldots, \eta_{a_s,i}$ are all 1 where $a_1 < \cdots < a_s$.
- Note that $Q_{ij} = \sum_{t=1}^{s} Y_t$ where t is an indicator r.v. for the event $\eta_{a_t,j} = 1$.
- Y_t's are not indep. but negatively correlated ⇒ p-th moment at most p-th moments of i.i.d. Bernoullis with expectation $\frac{s}{m}$ (expand $(\sum_t Y_t)^p$ and compare term by term)
- By Bernstein inequality:

$$\left|\left|Q_{ij}\right|\right|_p = \left|\left|\sum_t Y_t\right|\right| \lesssim \sqrt{\frac{s^2}{m}}\sqrt{p} + p \approx p$$

FFT-based Fast JL-Transform

- [Ailon, Chazelle'09] Running time $O(n \log n)$
- Define $\Pi \in \mathbb{R}^{m \times n}$ as $\Pi = \frac{1}{\sqrt{m}} \mathrm{S\,H\,D}$
- $S = \mathrm{m} \times n$ sampling matrix (with replacement)
- $H =$ unnormalized bounded orthonormal system, i.e. $\mathrm{H} \in \mathbb{R}^{n \times n}; H^T H = I; \max_{i,j} \left|H_{i,j}\right| \leq 1$
- $D = diag(\alpha)$ for $(\alpha_1, \dots, \alpha_n)$ i.i.d. Rademachers
- If H = Hadamard matrix $\Rightarrow O(n \log n)$ time to compute $\Pi \mathrm{x}$

FFT-based Fast JL-Transform

- Change S to $S_\eta = diag(\eta_1, \ldots, \eta_n)$ where η_i are Bernoullis with expectation $\mathbb{E}[\eta_i] = m/n$
- [CNW'15] If $\Pi = \frac{1}{\sqrt{m}} S_\eta HD$, $m \succcurlyeq \epsilon^{-2} \log\frac{1}{\delta} \log\frac{1}{\epsilon\delta}$:

$$\forall x: \left|\left|x\right|\right|_2 = 1, \Pr_\Pi\left[\left|\left|\left|\Pi x\right|\right|_2^2 - 1\right| \geq \epsilon\right] \leq \delta$$

- Let $z = HDx$ so $\left|\left|\Pi x\right|\right|_2^2 = \frac{1}{m}\sum_i \eta_i z_i^2$
- Will show that $\left|\left|\frac{1}{m}\sum_{i=1}^n \eta_i z_i^2 - 1\right|\right|_p$ is small

FFT-based Fast JL-Transform

$$\left\|\frac{1}{m}\sum_{i=1}^{n}\eta_i z_i^2 - 1\right\|_p \leq \frac{2}{m}\left\|\sum_i \sigma_i \eta_i z_i^2\right\|_p \text{ (symmetrization)}$$

$$\lesssim \frac{\sqrt{p}}{m}\left\|\left(\sum_i \eta_i z_i^4\right)^{1/2}\right\|_p \text{ (Khintchine)}$$

$$\leq \frac{\sqrt{p}}{m}\left\|(\max_i \eta_i |z_i|)\left(\sum_i \eta_i z_i^2\right)^{\frac{1}{2}}\right\|_p$$

$$\leq \frac{\sqrt{p}}{m}\left\|\max_i \eta_i z_i^2\right\|_p^{\frac{1}{2}} \left\|\left(\sum_i \eta_i z_i^2\right)\right\|_p^{1/2}$$

$$\leq \frac{\sqrt{p}}{m}\left\|\max_i \eta_i z_i^2\right\|_p^{\frac{1}{2}} \left(\left\|\frac{1}{m}\sum_{i=1}^{n}\eta_i z_i^2 - 1\right\|_p^{\frac{1}{2}} + 1\right) \text{ (triangle inequality)}$$

FFT-based Fast JL-Transform

$$\left\| \max_i \eta_i z_i^2 \right\|_p^{\frac{1}{2}} \leq \left\| \max_i \eta_i z_i^2 \right\|_q^{\frac{1}{2}} \text{ for } q = \max(p, \log m)$$

$$\left\| \max_i \eta_i z_i^2 \right\|_q^{\frac{1}{2}} = \mathbb{E}_{\alpha,\eta}\left[\max_i \eta_i z_i^{2q}\right]^{1/q} \leq \mathbb{E}_{\alpha,\eta}\left[\sum_i \eta_i z_i^{2q}\right]^{\frac{1}{q}} =$$

$$\sum_i \mathbb{E}_{\alpha,\eta}[\eta_i z_i^{2q}]^{\frac{1}{q}} \leq \left(n \max_i \mathbb{E}_{\alpha,\eta}\left[\eta_i z_i^{2q}\right]\right)^{\frac{1}{q}}$$

$$= \left(n \max_i \mathbb{E}_{\eta}[\eta_i]\mathbb{E}_{\alpha}[z_i^{2q}]\right)^{\frac{1}{q}} = \left(m \max_i \mathbb{E}_{\alpha}[z_i^{2q}]\right)^{\frac{1}{q}}$$

$$\leq 2 \max_i \left\| z_i^2 \right\|_q \quad (m^{\frac{1}{q}} \leq 2 \text{ by choice of } q)$$

$$= 2 \max_i \left\| z_i \right\|_{2q}^2 \lesssim q \quad \text{(Khintchine)}$$

FFT-based Fast JL-Transform

- Let $E = \left\|\left\|\frac{1}{m}\sum_{i=1}^{n}\eta_i z_i^2 - 1\right\|\right|_p^{\frac{1}{2}}$
- $E^2 - C\sqrt{\frac{pq}{m}}E - C\sqrt{\frac{pq}{m}} \le 0$
- $E^2 \le \max\left(\sqrt{\frac{pq}{m}}, \frac{pq}{m}\right)$
- Markov: $\Pr_{\Pi}\left[\left|\left|\left|\Pi x\right|\right|_2^2 - 1\right| \ge \epsilon\right] \le \epsilon^{-p}E^{2p} \le \delta$
- $p = \log 1/\delta$ and $m \geqslant \frac{1}{\epsilon^2}\log\frac{1}{\delta}\log\frac{m}{\delta}$