Class 9

Compressed Sensing and the RIP

Warm-Up

Let $\varepsilon \in (0, 1/4)$. Suppose that $A \in \mathbb{R}^{m \times n}$ is a distribution on matrices so that, for some constant c:

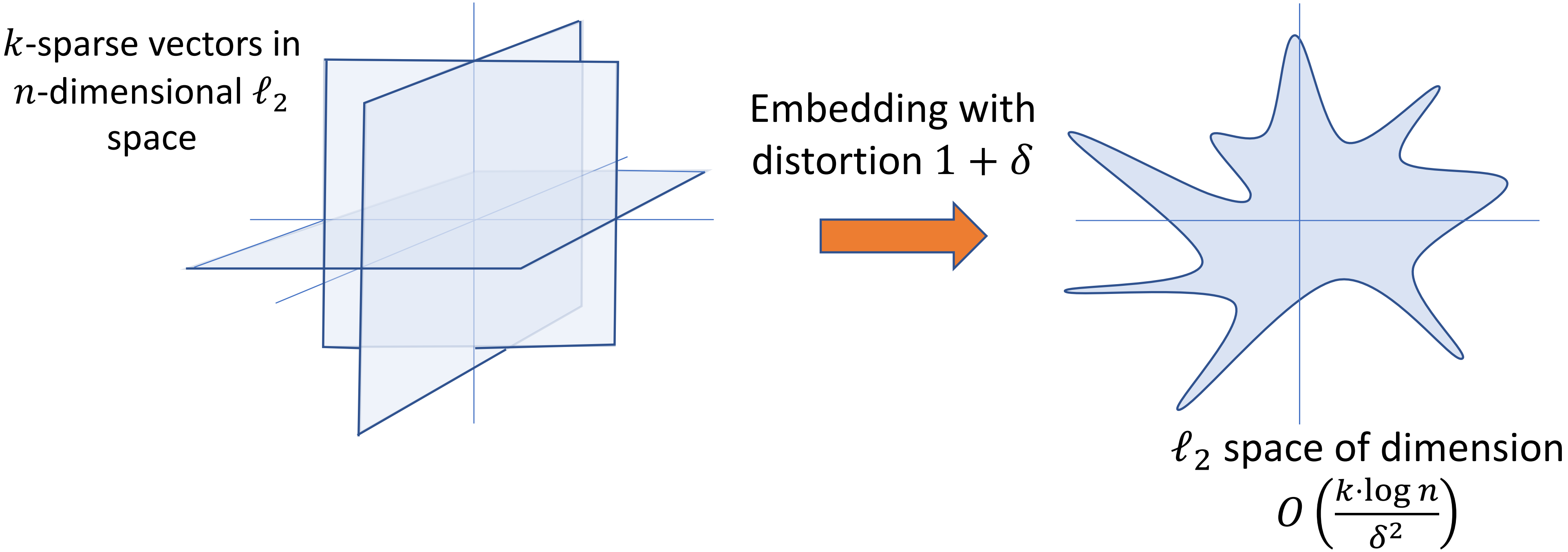
$$\forall x \in \mathbb{R}^n, \Pr\{|\|Ax\|_2 - \|x\|_2| \geq \varepsilon\|x\|_2\} \leq 2\exp(-cm\varepsilon^2). \tag{1}$$

1. Is it the case that A is a good JL transform (aka, for any set $S \subseteq \mathbb{R}^n$ of size N, $\|A(x-y)\|_2 = (1 \pm \varepsilon)\|x-y\|_2$ with high probability), with $m = O(\varepsilon^{-2}\log N)$?
2. Is it the case that, with high probability, A has the (k, ε)-RIP with $m = O(\varepsilon^{-2} k \log n)$?

Announcements

- HW3 due Friday!
- HW4 out now!
  - Problem 1: Sparse stuff (today)
  - Problems 2,3: Probabilistic Method (Monday)

Recap



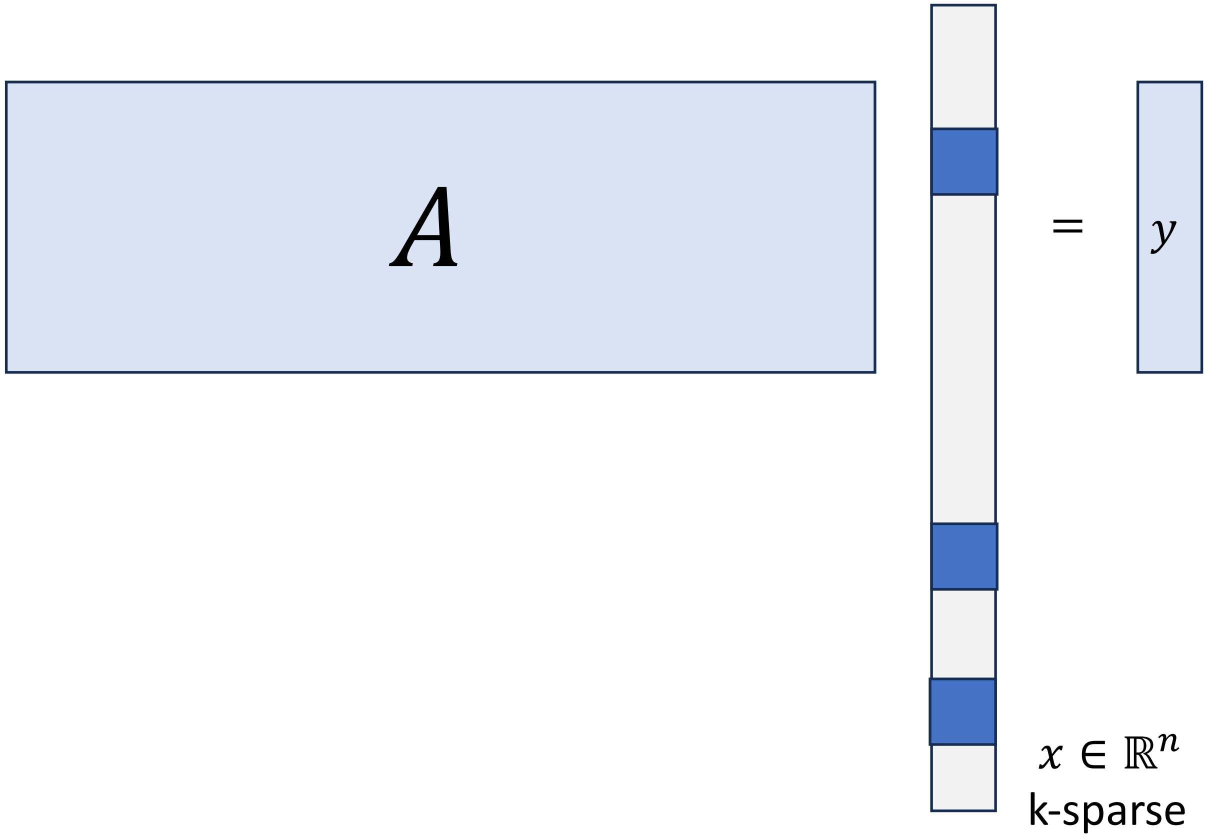
- Restricted Isometry Property: Embedding sparse vectors in low-dimensional space.
- A matrix $A \in \mathbb{R}^{m \times n}$ has the **Restricted Isometry Property** (RIP) with parameters k, δ if for any $x \in \mathbb{R}^n$ with $||x||_0 \leq k$,

$$(1 - \delta)||x||_2 \leq ||Ax||_2 \leq (1 + \delta)||x||_2$$

- Random Gaussian Matrices have the RIP whp!

Why do we care?

- RIP matrices allow us to (efficiently) solve this problem:



Goal:
Given y, find x

Questions?

Minilectures, warmup, quiz?

Warm-Up

Let $\varepsilon \in (0, 1/4)$. Suppose that $A \in \mathbb{R}^{m \times n}$ is a distribution on matrices so that, for some constant c:

$$\forall x \in \mathbb{R}^n, \Pr\{|\|Ax\|_2 - \|x\|_2| \geq \varepsilon\|x\|_2\} \leq 2\exp(-cm\varepsilon^2). \quad (1)$$

1. Is it the case that A is a good JL transform (aka, for any set $S \subseteq \mathbb{R}^n$ of size N, $\|A(x-y)\|_2 = (1 \pm \varepsilon)\|x-y\|_2$ with high probability), with $m = O(\varepsilon^{-2} \log N)$?
2. Is it the case that, with high probability, A has the (k, ε)-RIP with $m = O(\varepsilon^{-2} k \log n)$?

They are both true!

- When we proved JL / RIP for Gaussian matrices, this is the only property we used.

For JL

Say we have a set $S \subseteq \mathbb{R}^n$ of size N that we'd like to embed

- Suppose that for all $x \in \mathbb{R}^n$,

$$\Pr\left[\left|\|Ax\|_2 - \|x\|_2\right| \geq \epsilon \left|\left|x\right|\right|_2\right] \leq 2\exp(-cm\epsilon^2)$$

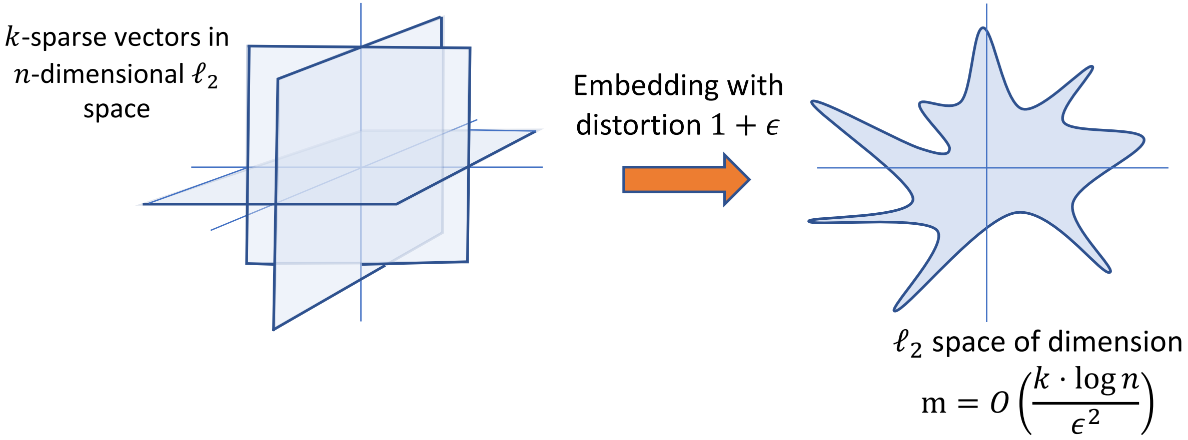
- Apply this to $x = y_1 - y_2$ for all $y_1, y_2 \in S$
- Union bound over all N^2 pairs from S:

$$\Pr\left[\exists y_1, y_2 : \left|\left|y_1 - y_2\right|\right|_2 \neq (1 \pm \epsilon)\left|\left|Ay_1 - Ay_2\right|\right|_2\right] \leq 2N^2\exp(-cm\epsilon^2)$$

- … which is tiny if $m \geq \frac{C \log N}{\epsilon^2}$ for an appropriate constant C.

For RIP

We want A to have the (k, ϵ)-RIP



- Suppose that for all $x \in \mathbb{R}^n$,

$$\Pr\left[\left|\|Ax\|_2 - \|x\|_2\right| \geq \epsilon \left|\left|x\right|\right|_2\right] \leq 2\exp(-cm\epsilon^2)$$

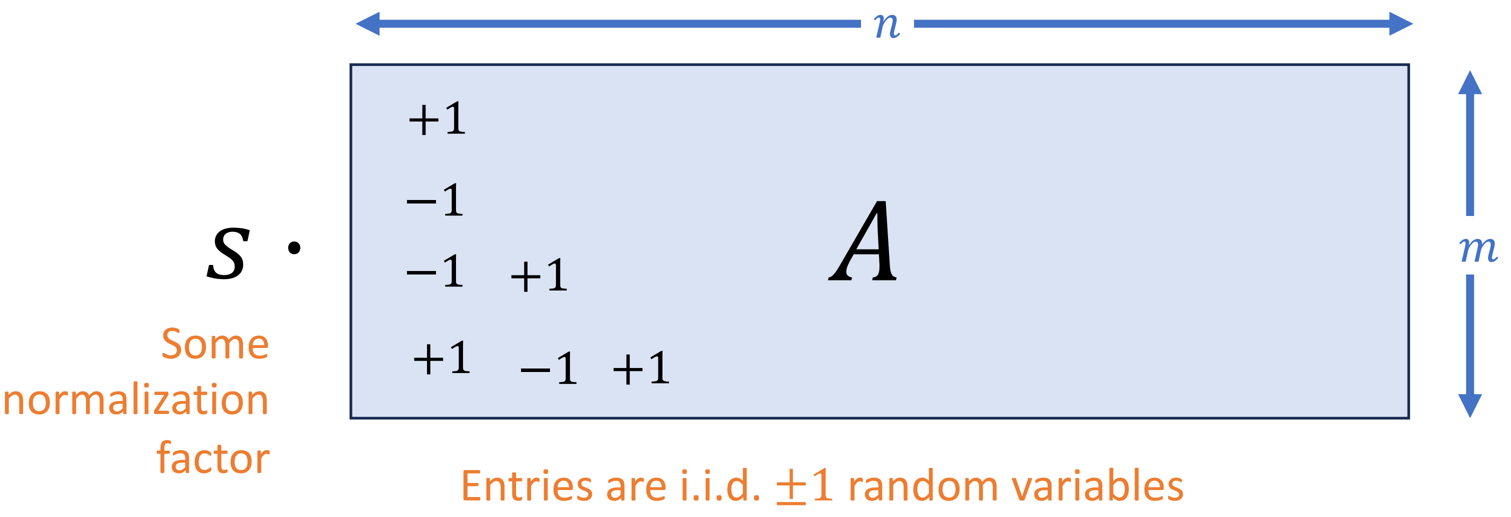
- Apply this to $x \in \mathcal{N}$ for an ϵ-net $\mathcal{N}$ of [all k-sparse vectors of norm 1]
  - Such a net has size $\binom{n}{k}\left(\frac{3}{\epsilon}\right)^k = \exp(k\log(n/\epsilon))$
- Union bound over all $x \in \mathcal{N}$

$$\Pr\left[\exists x \in \Sigma_k : \left|\|Ax\|_2 - \|x\|_2\right| \geq \epsilon \left|\left|x\right|\right|_2\right]$$
$$\leq \exp(k\log(n/\epsilon)) \cdot 2\exp(-cm\epsilon^2)$$

- ... which is tiny if $m = C \cdot \frac{k\log(n)}{\epsilon^2}$ for a large enough constant C.

Today: Other matrices with the RIP

- Random Gaussian Matrices have the RIP whp…
  - But those are really obnoxious to store/compute with!
- Random ± 1 random variables would be much nicer…



Group Work 1

- (a) What should s be (as a function of m, n) so that for any $x \in \mathbb{R}^n$,
$$\mathbb{E}\|sAx\|_2^2 = \|x\|_2^2?$$

- (b) For $x \in \mathbb{R}^n$ with norm 1, show that $\Pr[\|sAx\|^2 \geq 1 + \epsilon] \leq e^{-\epsilon^2 m/100}$

- (c) Are we done? If not, what's left?

(a) How to choose s so that $\mathbb{E}\|sAx\|_2^2 = \|x\|_2^2$

- $s = 1/\sqrt{m}$
  - This makes each column of sA have norm 1
  - So $||sAe_i||_2 = 1$ for all i … so if there is such an s, this better be it.

- $\mathbb{E}||sAx||_2^2 = s^2 m\, \mathbb{E}(\sum_{i=1}^n Y_i x_i)^2 \quad = s^2 m\, \mathbb{E} \sum_{i,j=1}^n Y_i Y_j x_i x_j$

Y_i are iid ± 1

$$= s^2 m \sum_{i,j=1}^n \mathbb{E}[Y_i Y_j] x_i x_j$$

$$= s^2 m \sum_{i=1}^n x_i^2$$

$\mathbb{E}[Y_i Y_j]$ is 1 iff i=j, else 0

$$= s^2 m \sum_{i=1}^n x_i^2$$

$$= s^2 m ||x||_2^2$$

Choose $s = 1/\sqrt{m}$

(b) For $x \in \mathbb{R}^n$ with norm 1, $\Pr[\|sAx\|^2 \geq 1+\epsilon] \leq e^{-\epsilon^2 m/100}$

- Let $Z = \sum_{i=1}^{n} Y_i x_i$. FACT: $\mathbb{E}\left[e^{tZ^2}\right] \leq 1 + t + 12t^2$ for $t \in \left(0, \frac{1}{3}\right)$
- $\Pr[\|sAx\|^2 \geq 1+\epsilon] = \Pr\left[\sum_{i=1}^{m} Z_i^2 \geq (1+\epsilon)m\right]$

$$= \Pr\left[e^{t\sum_{i=1}^{m} Z_i^2} \geq e^{t(1+\epsilon)m}\right]$$

$$\leq \frac{\mathbb{E}\Pi_i e^{tZ_i^2}}{e^{t(1+\epsilon)m}} \leq \frac{(1+t+12t^2)^m}{e^{t(1+\epsilon)m}}$$

$$\leq \frac{e^{(t+12t^2)m}}{e^{t(1+\epsilon)m}} = e^{m(12t^2 - \epsilon t)}$$

$$= e^{m(-\epsilon^2/48)}$$

Pick $t = \frac{\epsilon}{24}$

(c) Are we done?

- Almost!
- We showed that $\Pr[\|sAx\|^2 \geq 1+\epsilon] \leq e^{-\epsilon^2 m/100}$
- We also need to show that $\Pr[\|sAx\|^2 \leq 1-\epsilon] \leq e^{-\epsilon^2 m/100}$
  - Basically the same argument
- Now, as per the Warm-Up, we can use A as a JL transform or RIP matrix!

Let $\varepsilon \in (0, 1/4)$. Suppose that $A \in \mathbb{R}^{m \times n}$ is a distribution on matrices so that, for some constant c:

$$\forall x \in \mathbb{R}^n, \Pr\{|\|Ax\|_2 - \|x\|_2| \geq \varepsilon \|x\|_2\} \leq 2\exp(-cm\varepsilon^2).$$

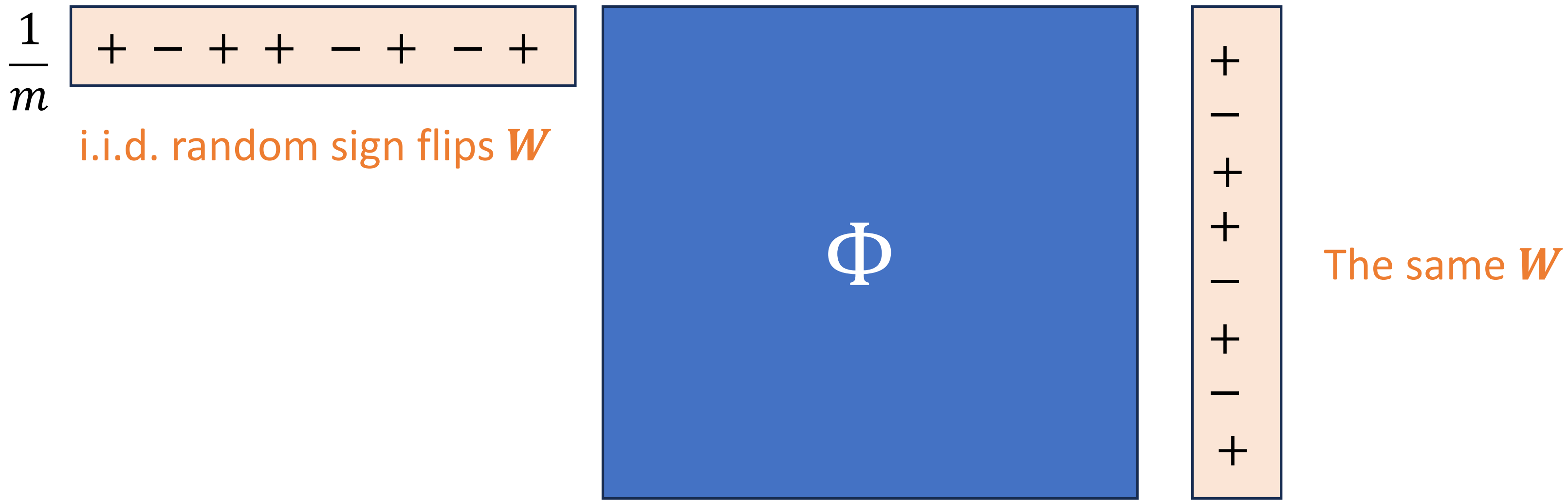
Have this with $c = \frac{1}{100}$

1. Is it the case that A is a good JL transform (aka, for any set $S \subseteq \mathbb{R}^n$ of size N, $\|A(x-y)\|_2 = (1 \pm \varepsilon)\|x-y\|_2$ with high probability), with $m = O(\varepsilon^{-2} \log N)$?
2. Is it the case that, with high probability, A has the (k, ε)-RIP with $m = O(\varepsilon^{-2} k \log n)$?

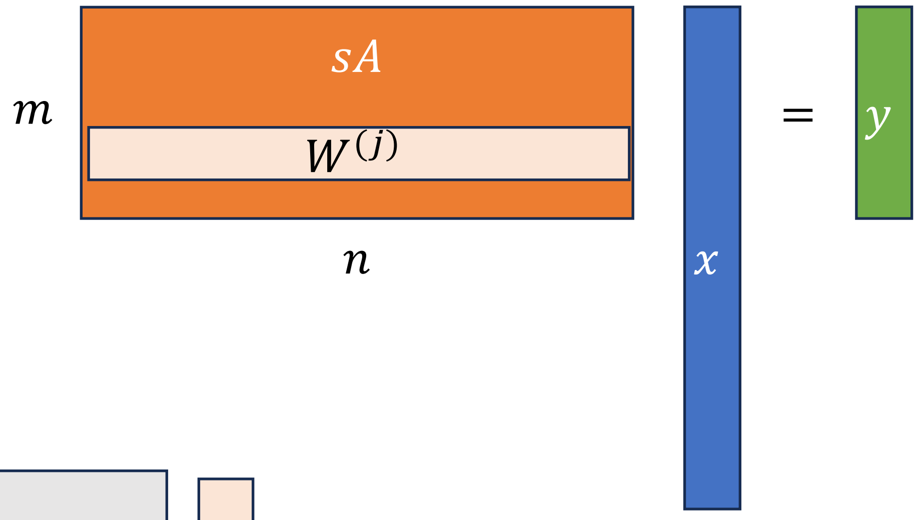
2: Hanson-Wright Inequality!

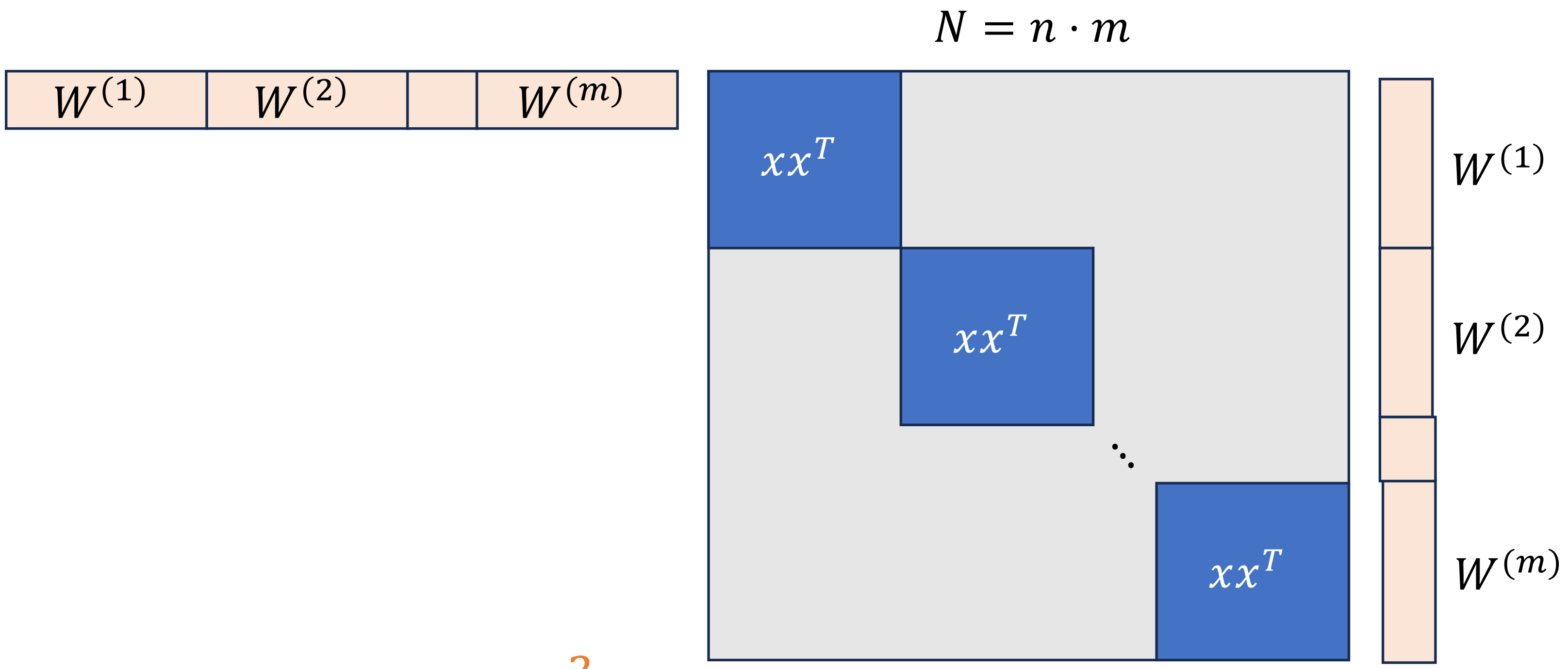
- Let $W_1, W_2, \ldots, W_N$ be i.i.d $\pm\frac{1}{\sqrt{m}}$ valued random variables (mean 0).
- Let $\Phi \in \mathbb{R}^{N\times N}$ be any matrix.
- For any $\epsilon > 0$,

$$\Pr[\,|\boldsymbol{W}^T\Phi\boldsymbol{W} \;-\; \mathbb{E}\boldsymbol{W}^T\Phi\boldsymbol{W}| > \epsilon] \le 2\exp\left(-c\min\left(\frac{\epsilon^2 m^2}{||\Phi||_F^2}, \frac{\epsilon m}{||\Phi||}\right)\right)$$



Pick a special Φ ...





$$= ||sAx||_2^2$$

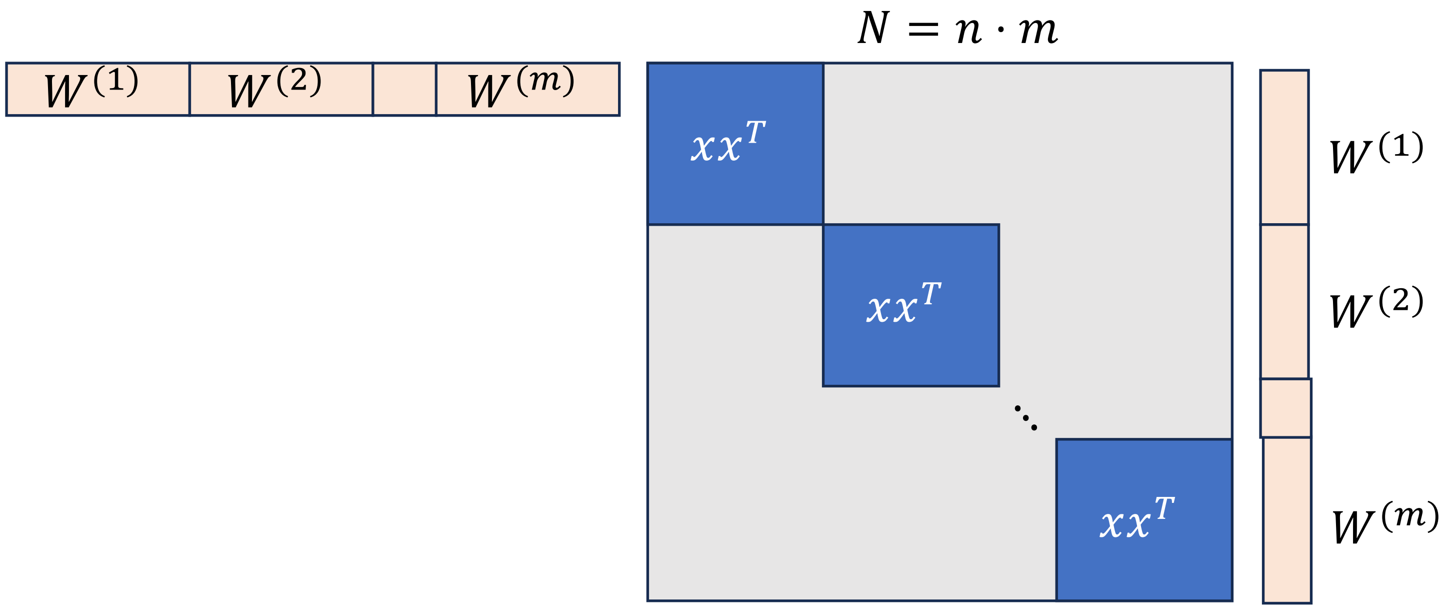
$$||sAx||_2^2$$

$$\Pr[\,|\boldsymbol{W}^T\Phi\boldsymbol{W} - \mathbb{E}\boldsymbol{W}^T\Phi\boldsymbol{W}| > t] \leq 2\exp\left(-c\min\left(\frac{t^2m^2}{||\Phi||_F^2}, \frac{tm}{||\Phi||}\right)\right)$$

$$\Pr\left[\left|\,||sAx||_2^2 - \mathbb{E}\boldsymbol{W}^T\Phi\boldsymbol{W}\right| > \epsilon\right] \leq 2\exp\left(-c\min\left(\frac{\epsilon^2 m^2}{||\Phi||_F^2}, \frac{\epsilon m}{||\Phi||}\right)\right)$$

1

- $\mathbb{E}\boldsymbol{W}^T\Phi\boldsymbol{W}$
- $= \mathbb{E}\sum_{\ell,r=1}^{N}\Phi_{\ell,r}W_\ell W_r$
- $= \sum_{\ell,r=1}^{N}\Phi_{\ell,r}\mathbb{E}\, W_\ell W_r$
- $= \frac{1}{\mathrm{m}}\sum_{\ell=1}^{N}\Phi_{\ell,\ell}$
- $= \sum_{i=1}^{n}(xx^T)_{i,i}$
- $= \sum_{i=1}^{n}x_i^2 = 1$

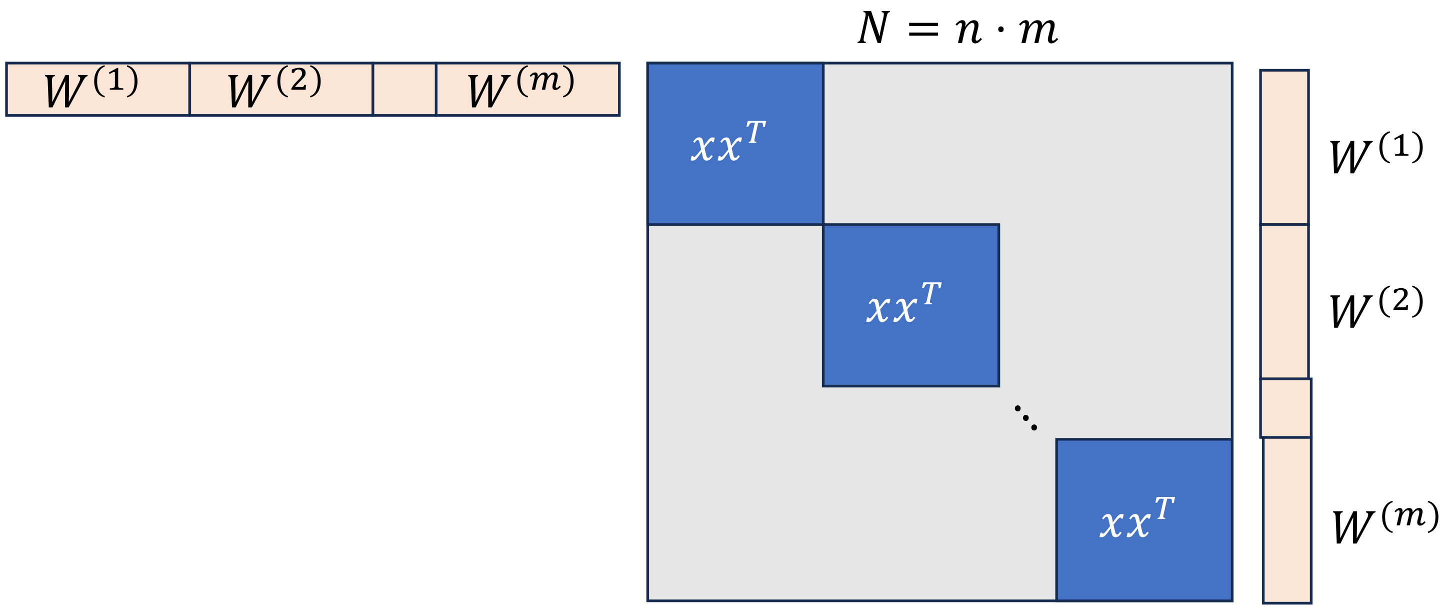


$$\Pr\left[\left|\|sAx\|_2^2 - 1\right| > \epsilon\right] \leq 2\exp\left(-c\min\left(\frac{\epsilon^2 m^2}{\|\Phi\|_F^2}, \frac{\epsilon m}{\|\Phi\|}\right)\right)$$

m 1

- $\|\Phi\|_F^2 = m\|xx^T\|_F^2$
- $\qquad = m\|x\|_2^4 = m$

- $\|\Phi\| = \|xx^T\|$
- $\qquad = \|x\|_2^2 = 1$



Putting it together

$$\Pr\left[\left|\|sAx\|_2^2 - 1\right| > \epsilon\right] \leq 2\exp\left(-c\min\left(\frac{\epsilon^2 m^2}{m}, \frac{\epsilon m}{1}\right)\right)$$

$$= 2\exp(-c\epsilon^2 m)$$

Yay!

Recap

- We can use ± 1 values entries instead of Gaussians and still get JL/RIP.

- Two proofs!
  - One Chernoff-like
  - One via Hanson-Wright Inequality

Next time

- The Probabilistic Method!