CS265, Winter 2025

Class 9: Agenda, Questions, and Links

1 Warm-Up

Group Work

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

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

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

Group Work: Solutions

Yes, they are both true. Going back to the lecture notes when we proved that random Gaussian matrices satisfy both of these, (1) is the only property that we used.

2 Announcements

- HW3 due Friday!
- HW4 out now! (Or soon...)

3 Lecture Recap and Questions?

Questions from minilectures and pre-class quiz? (Compressed sensing; RIP; Gaussian matrices have the RIP with high probability.)

4 More matrices with the RIP whp

Group Work

1. Let $A \in (\pm 1)^{m \times n}$ be a matrix where every entry is independently selected to be either $+1$ or -1. In this question, you'll show that for a cleverly chosen constant s, the matrix $\Psi = sA$ satisfies (1) from the Warm-Up. (Notice that sA is much easier

to generate than a random Gaussian matrix, and is also nicer to compute with).

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

(b) For a vector $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$, let Z denote the random variable representing the inner product of x a row of matrix A. Namely $Z = \sum_{i=1}^n Y_i x_i$ where Y_i is independently chosen to be ± 1 with probability $1/2$ each, and x_i denotes the ith coordinate of x. The following bound on the moment generating function of Z^2 is not too hard (but a bit tedious) to prove: for any $t \in (0, 1/3)$,

$$\mathbb{E}[e^{tZ^2}] \leq 1 + t + 12t^2.$$

Using this bound on the moment generating function of Z^2, prove that

$$\Pr[\|sAx\|_2^2 \geq (1+\varepsilon)] = \Pr[Z_1^2 + Z_2^2 + \ldots + Z_m^2 \geq (1+\varepsilon)m] \leq e^{-\varepsilon^2 m/100},$$

where the Z_i's represent independent realizations of the random variable Z.
Hint*: Proceed as in the proof of Chernoff bounds...*
Hint*: In the final step, you may want to plug in an "optimal" value of t (the parameter in the proof of the Chernoff bounds). Try something like $t = \varepsilon/24$ to get the math to work out cleanly.*

(c) Conclude that $\Psi = sA$ satisfies property (1) from the warm-up. ***Hint****: We're not* ***quite*** *done...*

2. Here's another way to show that a random ± 1 matrix (normalized appropriately) achieves (1) from the Warm-Up. This way also has the advantage that we get to learn a new tail bound, called the *Hanson-Wright inequality!* Here's a statement of one form of this inequality:

   Theorem. Let $W_1, \ldots, W_N$ be $\pm\frac{1}{\sqrt{m}}$-valued independent mean-zero random variables. Let $\Phi \in \mathbb{R}^{N \times N}$ be any matrix. Then for any $t \geq 0$,

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

   where above $\vec{W} = (W_1, \ldots, W_N)$ is the length-N vector with the random variables Z_i in it, $\|\Phi\|_F^2 = \sum_{i,j} \Phi_{i,j}^2$ denotes the Frobenius norm, and $\|\Phi\| = \sup_{v \in \mathbb{R}^N \setminus \{0\}} \frac{\|\Phi v\|_2}{\|v\|_2}$ is the operator norm.

   Use the Hanson-Wright inequality to show that (1) holds for the matrix sA (with the same s that you found in 1(a)).

   Hint*: Let $N = nm$, and write $\|sAx\|_2^2$ as $\vec{W}^T \Phi \vec{W}$ for some matrix Φ, where the elements of $\vec{W}$ are the entries of sA.*

Hint: *A further hint for how to do the above: Let* a_i *be the* i*'th row of* A*. Then* $\|Ax\|_2^2 = \sum_i a_i^T(xx^T)a_i$ *(why?). Consider a matrix* Φ *that is block-diagonal where each block is equal to the matrix* xx^T*:*

$$\Phi = \begin{pmatrix} xx^T & & & & \\ & xx^T & & & \\ & & xx^T & & \\ & & & \ddots & \\ & & & & xx^T \end{pmatrix}$$

Hint: It might be useful that (a) for a vector x, we have $\|xx^T\|_F^2 = \|x\|_2^4$ and $\|xx^T\| = \|x\|_2^2$, and (b) for a block-diagonal matrix Φ with blocks $\Phi_1, \Phi_2, \ldots$ on the diagonal, $\|\Phi\| = \max_i \|\Phi_i\|$. (These facts are not too hard to derive, but you can take them as given if you like).

Group Work: Solutions

1. (a) We should set $s = 1/\sqrt{m}$. By linearity of expectation, $\mathbb{E}\left[\|Ax\|_2^2\right] = m\mathbb{E}\left[(\sum_{i=1}^n Y_i x_i)^2\right]$ where the Y_i's are independent ± 1 random variables that are $+1$ and -1 with probability 1/2 each. Expanding out the terms in the expression being squared, we have

$$\mathbb{E}\left[(\sum_{i=1}^n Y_i x_i)^2\right] = \mathbb{E}\left[\sum_{i,j} x_i x_j Y_i Y_j\right] = \sum_{i,j} x_i x_j \mathbb{E}\left[Y_i Y_j\right] = \sum_i x_i^2 = \|x\|_2^2,$$

since for $i \neq j, \mathbb{E}\left[Y_j Y_i\right] = 0$. Hence $\mathbb{E}\left[\|Ax\|_2^2\right] = m\|x\|_2^2$, so if we multiply A by $1/\sqrt{m}$, we will cancel this factor of m.

(b) For any $t > 0$,

$$\Pr[\|sAx\|_2^2 \geq (1+\varepsilon)] = \Pr[\sum_i Z_i^2 \geq (1+\varepsilon)m] = \Pr[e^{t\sum Z_i^2} \geq e^{tm(1+\varepsilon)}] \leq \frac{\prod_i \mathbb{E}\left[e^{tZ_i^2}\right]}{e^{tm(1+\varepsilon)}}.$$

Plugging in the fact that $\mathbb{E}\left[e^{tZ_i^2}\right] \leq 1 + t + 12t^2$, yields the following:

$$\frac{\prod_i \mathbb{E}\left[e^{tZ_i^2}\right]}{e^{tm(1+\varepsilon)}} \leq \frac{(1+t+12t^2)^m}{e^{tm(1+\varepsilon)}} \leq \frac{e^{m(t+12t^2)}}{e^{tm(1+\varepsilon)}} = e^{m(12t^2 - \varepsilon \cdot t)}.$$

Plugging in $t = \varepsilon/24$ and simplifying yields a bound of $e^{-m\varepsilon^2 \frac{1}{48}}$, as desired.

(c) The main thing we are missing is a bound on the probability that $\|sAx\|_2^2 \leq (1-\varepsilon)\|x\|$, though this can be proved analogously to the upper bound, in the same way that we proved the lower Chernoff bounds.

2. Define Φ as in the hint. Break up $\vec{W} \in \{+1/\sqrt{m}, -1/\sqrt{m}\}^{nm}$ into chunks $\vec{W} = \vec{W}^{(1)}, \ldots, \vec{W}^{(m)}$, so that each chunk $\vec{W}^{(j)}$ has length n, and corresponds to the j'th row of sA. Then with Φ as in the hint, we have

$$\vec{W}^T \Phi \vec{W} = \sum_{j=1}^{m} (\vec{W}^{(j)})^T x x^T \vec{W}^{(j)} = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} W_i^{(j)} x_i \right)^2 .$$

On the other hand, if we write down what $\|sAx\|_2^2$ is, we get the same thing! Indeed, the j'th coordinate of sAx is $\sum_{i=1}^n W_i^{(j)} x_i$, and then we sum up the squares of those to get the ℓ_2 norm.

Now that we have this, we need to figure out $\mathbb{E}\vec{W}^T \Phi \vec{W}$. This is

$$\mathbb{E}\vec{W}^T \Phi \vec{W} = \sum_{i,j=1}^{N} \mathbb{E} W_i W_j \Phi_{i,j} = \frac{1}{m} \sum_i \Phi_{i,i}.$$

(So, it's $1/m$ times the sum of the diagonal entries). In our case, $\sum_i \Phi_{i,i} = m \sum_j x_j^2$, so we have $\mathbb{E}\vec{W}^T \Phi \vec{W} = \|x\|_2^2 = 1$.

Finally, we have to bound $\|\Phi\|_F^2$ and $\|\Phi\|$. We have

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

and

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

Plugging all this in, the Hanson-Wright inequality tells us that

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

which is what we wanted. Yay!

Group Work

Here are a few "challenge" questions to think about:

1. What other distributions on a matrix A can you come up with (other than i.i.d. Gaussians and i.i.d. ± 1 entries) that are (a) natural and (b) seem like they'd satisfy (1)? For example, what about any matrix with i.i.d. mean-zero entries?

What about any matrix with i.i.d. mean-zero *bounded* entries? (e.g., the entries should never be larger than 100).

2. We saw in the warm-up that (1) implies that Φ is a good JL transform. Is the converse true? Must any JL transform for sets of N points also satisfy (1) for a single point? In particular, could we get a *better* JL transform for N points than we could by using (1) and a union bound?

3. Suppose that A has the RIP. Consider a matrix $A \cdot D$, where D is a diagonal matrix with i.i.d. mean-zero ± 1 entries on the diagonal. Show that AD satisfies (1), up to log factors.

   Hint*: This is pretty tricky to do quantitatively, but you may be able to come up with some intuition for why it should be true qualitatively.*

   Hint*: Write $\|ADx\|_2^2 = Z^T \Phi Z$ where Z is a vector of independent sign flips, and apply the Hanson-Wright inequality above...if A has the RIP, what can you say about every $k \times k$ block of Φ?*

Group Work: Solutions

1. If the entries are *sub-gaussian*, then everything we did today goes through.
2. Unfortunately, we can't beat the union bound here. See `https://arxiv.org/abs/1609.02094` for a proof.
3. See this paper: `https://arxiv.org/pdf/1009.0744.pdf`