

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 :

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

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 $m = O(\varepsilon^{-2} k \log n)$?

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 represent-

ing 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 i th 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

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

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, \dots, 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 $\vec{W} = (W_1, \dots, W_N)$ 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 Φ_1, Φ_2, \dots 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

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 ?*