Class 9: Agenda, Questions, and Links

1 Announcements

- HW5 is out, due next Wednesday.

2 Lecture Recap and Questions?

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

3 More matrices with the RIP whp

Group Work

1. 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:

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

   (a) 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 $m = O(\varepsilon^{-2}\log N)$?

   (b) Is it the case that, with high probability, A has the (k, ε)-RIP with $m = O(\varepsilon^{-2} k \log n)$?

2. 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 sA satisfies (1). (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 $t \in (0, 1/3)$,

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

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 by multiplying both sides by t, exponentiating, applying Markov's inequality, then using the given bounds on the moment-generating function of each Z_i.*
Hint: *In the final step, you'll want to plug in an 'optimal' value of t. Try something like $t = \varepsilon/24$ to get the math to work out cleanly.*

(c) Have we proved that sA satisfies property (1)? If not, what is missing?

3. Here are a few "challenge" questions to think about:
   (a) What other distributions on a matrix A can you come up with (other than i.i.d. Gaussians and i.i.d. $\pm 1/\sqrt{m}$ 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? (i.e., the entries should never be larger than 100).
   (b) 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: *For a complete solution, check out this paper: https://arxiv.org/pdf/1009.0744.pdf*

Group Work: Solutions

1. Yes, both of these are true. If you go back to our proof that Gaussian matrices are good JL transforms and that they have the RIP, this is the only property we use.
2. (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:

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

Plugging in $t = \varepsilon/24$ and simplifying yields a bound of $e^{-m\varepsilon^2 \frac{3}{4\cdot 24}}$, 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.

4 Connected Components in Random Graphs

As one final super-cool application of Chernoff bounds, in this problem we will prove a really cool property of the sizes of the connected components in a natural random graph model.

Let $G_{n,p}$ denote the Erdos-Renyi random graph model, where each edge exists (independently) with probability $p = c/n$ for some constant c that does not vary with n.

Theorem 1. *Let G be drawn from $G_{n,p}$, with $p = c/n$ for some constant c:*

- *If $c < 1$, with probability tending to 1 as $n \to \infty$, the largest connected component of G has size $O(\log n)$.*
- *If $c > 1$, with probability tending to 1 as $n \to \infty$, the largest connected component of G has size $f(c)n \pm o(n)$, where $f(c)$ is a constant, independent of n, satisfying $f(c) \in (0,1)$ for all $c > 1$.[1], and the second-largest connected component of G has size $O(\log n)$.*

Group Work

1. Spend one minute pondering why we shouldn't expect any "medium" sized connected components. What is the intuition behind the above theorem?
2. In this problem we prove the $c < 1$ case of the above theorem.

[1] $f(c)$ can actually be defined fairly cleanly: Suppose on day 1, we start with 1 rabbit. On each day, each existing rabbit will have a number of offspring drawn from (independent) Poisson random variables of expectation c and then the original rabbits die, leaving only the offspring. $f(c)$ is the probability that this process never dies out (ie one minus the probability that there is some day with no more rabbits).

(a) For a given vertex v, prove that

$$\Pr[v \text{ in connected component of size } \geq k] \leq \Pr\left[X \geq k-1\right],$$

where X is distributed according to $Binomial[k \cdot n, c/n]$. [Hint: consider doing a breadth-first search of the neighborhood of v in the graph.]

(b) Assuming the above, using a union bound over Chernoff bounds, prove that

$$\Pr[\text{there is a connected component of size } \geq \frac{10 \log n}{(1-c)^2}] \leq 1/n.$$

3. In this problem, we prove the $c > 1$ case of the above theorem.

(a) Given a random node v in the graph, prove that for any k satisfying $\frac{100c \log n}{(c-1)^2} \leq k \leq n^{3/4}$, the probability that the connected component of v has size k is no more than n^{-10}.

Hint*: Consider a sort of breadth-first search that starts with a set that contains only v, then "marks" v and adds all the neighbors of v to the set , and then iteratively continues by "marking" an unmarked node of the set and adding all its neighbors to the set. Suppose we have "marked" k nodes, what is the chance that there are no more "unmarked" nodes in our set? Based on this, prove that, with high probability, if the connected component of v has size at least k, it will have size at least $k+1$. Be mindful of the way you condition events.*

(b) Prove that we do not expect any connected components to have size in the interval $[\frac{100c \log n}{(c-1)^2}, n^{3/4}]$.

(c) Prove that with probability tending to 1 as $n \to \infty$, there is at most one connected component of size $\geq n^{3/4}$. [Hint: conditioned on the neighborhood of both v and u having size at least $n^{3/4}$, show that the probability that they are not connected is tiny, then union bound over the at most n such neighborhoods.]

(d) Challenge: Show that the size of the large component is within $o(n)$ of its expectation with probability tending to 1 as $n \to \infty$. [Hint: bound the variance of the number of nodes that are in "small" components of size at most $\frac{100c \log n}{(c-1)^2}$, then use Chebyshev's inequality.]

Group Work: Solutions

1. The lack of "medium-sized" components, for both $c < 1$ and $c > 1$ might seem intuitive, though a nice challenge is to try to figure out what happens in the regime where $c \approx 1$ (e.g. where $c = 1 + 1/\sqrt{n}$, for example. . . .

2. The next three parts will all be based on the following careful way of revealing

the connected component of a single vertex, v. We will explore v's connected component in a way that carefully ensures that we are aware of which edges in the graph have been "explored", and which edges we haven't yet looked at. [This will avoid the issue that arises if we sloppily try to condition on properties of the size of v's connected component, after which it will no longer be true that each edge is independently present with probability p.]

We do this exploration as follows: We will proceed iteratively, and step t, we will have two sets, A_t and B_t, where A_t corresponds to nodes in v's connected component for which at time t we've already checked all $n-1$ potential edges between them and the other nodes, and B_t corresponds to nodes that we have found to be in v's connected component by time t, but for which we haven't yet checked to see which neighbors they have. At time $t=0$, we start with $B_0 = \{v\}$, and $A_0 = \emptyset$. At $t=1$, we set $A_1 = \{v\}$ and B_1 to be all the neighbors of v. In general, at step t, we select an arbitrary node $w_t \in B_{t-1}$ (for example, the node of smallest index, if we number the nodes 1 through n), and add all its neighbor that aren't already in A_{t-1} or B_{t-1} to set B_{t-1} to form B_t. Namely $B_t = B_{t-1} \cup neighbors(w_t) \setminus A_{t-1}$. We then remove w_t from B_t and set $A_t = A_{t-1} \cup \{w_t\}$. We continue this process until the set B_{t-1} is empty, in which case A_{t-1} must be v's entire connected component.

To analyze this process, let X_i denote the number of nodes added to B_{i-1} to form B_i. Because we haven't looked at any potential edges from node w_i except those going to a node in set A_{i-1}, and we are only adding neighbors of w_i that aren't already in sets A_{i-1} or B_{i-1}, it is the case that X_i is distributed as a binomial consisting of $n - \sum_{j=1}^{i-1} X_j$ tosses of a p-biased coin: $Bin(n - \sum_{j=1}^{i-1} X_i, p)$.

We will now solve this part of the problem. Let $c < 1$, and note that if v's connected component has size greater than k, then it must be the case that $X_1 > 0$, and $X_1 + X_2 > 1$, $\ldots$, and $\sum_{i=1}^{k} X_i > k$. (If this wasn't the case—for example, if $X_1 = 0$, then the connected component has size 1. If $X_1 > 0$ but $X_1 + X_2 = 1$ then the connected component has size 1, etc.) Now, note that the probability that these things all happen is at most the probability that $\sum_{i=1}^{k} Y_i > k$, where each Y_i is an independent binomial random variable corresponding to n tosses of a p-biased coin. This is true because each X_i consists of a binomial of at *most* n tosses of a p-biased coin, and so flipping a few extra coins can only help the probability of getting more "heads". To wrap things up, note that $Y = \sum_{i=1}^{k} Y_i$ is distributed exactly like $Bin(kn, p)$, and if $p = c/n$ then its expectation is ck. So for $c < 1$, we have $\Pr[Y > k] = \Pr[Y > (1/c)\mathbb{E}[Y]]$, and we can bound this probability via a standard Chernoff bound. In the case that $c < 1$ is a constant, the Chernoff bound will give us a probability of the form $e^(- f(c)k)$ for some function f of c, in which case for k some large constant time $\log n$, will give us a probability less than, say, $1/n^2$, in which case we can do a union bound over the at most n connected components to prove that with probability tending to 1 as $n \to \infty$, the largest connected component has size $O(\log n)$. [Exercise: pin down the precise Chernoff bound, to get the right dependence on c...]

3. For the next part, the analysis of $c > 1$, for simplicity we'll give the argument in the case that $c = 2$. We leverage the same construction of set A_t, B_t as described above. Assuming that v's connected component has size at least k, it follows that B_{k-1} cannot be empty, in which case $B_k = \sum_{i=1}^{k} X_i$. If this sum is less than $n^{3/4}$, then each term in it has the form $Bin(m, p)$, where $m > n - n^{3/4}$, which is greater than $0.9n$ for large n. Additionally, the assumption on the partial sums of these binomials, ie $X_1 > 0$, $X_1 + X_2 > 1$ etc., can only increase the probability that this sum is large. Hence the probability that $|B_k| = 0$ is at most the probability that a sum of k independent binomials $Bin(0.9n, 2/n)$ is at most k. [Recall we are assuming $c = 2$ for simplicity.]$So, this probability is bounded by$ $\Pr[Bin[0.9kn, 2/n] \leq k] < \Pr[W < (1-0.4)\mathbb{E}[W]]$, where W is distributed according to $Bin[0.9kn, 2/n]$. By a standard Chernoff bound, this probability is inverse exponential in $\mathbb{E}[W] = 1.8k$, and hence as long as k is a large constant times $\log n$, this probability is much less than $1/n^3$, and we can union bound over the $O(n)$ choices of such k, showing that for a single connected component, the probability it has size between $O(\log n)$ and $n^{3/4}$ is at most $O(1/n^2)$, and thenalso union bound over the $O(n)$ connected components. [Note that we assumed that none of the partial sums of X_i's exceeded $n^{3/4}$, which is a fine assumption because if that isn't the case, the connected component already has size at least $n^{3/4}$, which is what we want!!]

4. To show that any two large connected components (ie components of size $\geq n^{3/4}$ must actually be connected, the following does NOT work: consider two such connected components, and now observe that there are $\approx n^{3/2}$ possible edges connecting them, and hence the probability none of these exist is $(1 - c/n)^{n^{3/2}} = o(1)$. That approach is fatally flawed because by assuming you start with two large connected components, you are implicitly already implying that you've looked to see which edges are present, and hence those potential $n^{3/2}$ edges connecting the two components have already been examined, and don't actually have any randomness left!! For example, with probability > 0.9999 there DO exist two sets of $\Theta(n)$ nodes that are disconnected from each other—just take the $\Theta(n)$ nodes of degree 0 and divide this into two equal-sized sets. Despite there being $\approx n^2$ potential edges between these two sets, the sets are disconnected.

   The moral of the above is that we need to be careful to make sure that if we make an argument of the form "surely some of these potential edges exist", we need to explicitly construct things in such a way that we haven't peeked at the potential edges before making that claim.

   One nice way to do this follows the same strategy of forming sets A_t, B_t. Imagine doing this starting from two nodes, v and w, where A_t^v, B_t^v denote the sets corresponding to v and A_t^w, B_t^w correspond to the sets spawned by w. Consider iteratively building these sets for $t = 1, 2, \ldots$ and stopping the first time one of the following things happens: 1) $t = n^{3/4}$, 2) the neighborhoods have already been found to intersect, namely $(A_t^v \cup B_t^v) \cap (A_t^w \cup B_t^w) \neq \emptyset$, or 3) one of the neighborhoods is fully discovered, e.g. $B_t^v = \emptyset$. The only stopping condition we need to worry about is 1). If that occurs, then, with high probability (as argued above) for $t = n^{3/4}$, we have $|B_t^v| = \Theta(n^{3/4})$ and $|B_t^w| = \Theta(n^{3/4})$. Now, if B_t^v and B_t^w are disjoint, then we haven't yet looked at any of the $\Theta(n^{3/2})$ potential edges between these two sets (and if they do intersect, then the v and w must be in the same connected component anyway...)!! So now we can safely argue that the probability that none of these $m = \Theta(n^{3/2})$ edges exist is actually $(1 - c/n)^m = o(1/n^3)$, and hence we can union bound over the at most $O(n^2)$ pairs of connected components.