PRINCETON COS 521: ADVANCED ALGORITHM DESIGN

Matrix Concentration Inequalities and the Planted Clique Problem

In this lecture, our goal is to investigate methods to prove the spectral norm bounds on random matrices that we saw in the last lecture.

Let's focus on the following theorem that is a special case of what we used in the last lecture.

Theorem 1. *Let A be a $n \times n$ symmetric random matrix with each entry being an independent random variable that takes the value $1-p$ with probability p and $-p$ with probability $1-p$. Then, with probability $1-o_n(1)$, whenever $p \gg \log^4 n/n$, we have:*

$$\|A\|_2 \leq O(\sqrt{pn})$$

We will introduce a general-purpose tool to prove such results that will give a weaker version of the theorem. We will discuss methods to prove the bounds of the above form in the next lecture.

The following key tool should be thought of as a generalization of Chernoff-Hoeffding bounds:

Theorem 2 (Matrix Bernstein Inequality)**.** *Let $A_1, A_2, \ldots, A_N$ be $n \times n$ symmetric, real-entry, independent random matrices such that for each $1 \leq i \leq N$, $\|A_i\|_2 \leq 1$ with probability 1. Let $\sigma^2 = \left\|\mathbb{E} \sum_i A_i^2\right\|_2$ be the "variance" parameter. Then,*

$$\Pr[\left\|\sum_i A_i\right\|_2 \geq \tau] \leq 2n \exp\left(-\frac{\tau^2}{\sigma^2 + \tau/3}\right).$$

Notice that the expectation is taken before taking the norm in computing the variance term.

It is instructive to apply the theorem to $n = 1$ in which case it gives a scalar inequality that is quite similar to the Chernoff-Hoeffding bounds we studied in this course. The key difference in the case of matrices is the dimension dependence in the bound on the probability of deviation (i.e. the factor n multiplying the exponential tail).

Here's a corollary that is useful (it is an important exercise to derive it yourself):

Corollary 1. *Under the hypothesis of the theorem above, we have that with probability at least $1 - 1/n^{100}$,*

$$\left\|\sum_i A_i\right\|_2 \leq O(\sigma \log n).$$

That is, the typical/high probability value of the spectral norm of the sum $\sum_i A_i$ is about $\log n$ larger than the variance term σ.

Let us apply the Matrix Bernstein inequality to prove a weaker version of the main theorem we are after in this lecture. Notice that the weaker bound requires no assumptions on p unlike the main theorem above.

Corollary 2. *Let A be a $n \times n$ symmetric random matrix with each entry being an independent random variable that takes the value $1-p$ with probability p and $-p$ with probability $1-p$. Then, with probability at least $1-1/n^{100}$,*

$$\|A\|_2 \leq O(\sqrt{pn} logn)$$

Proof. We want to write A as a sum of independent random matrices and apply the Matrix Bernstein inequality. Towards this, let $B_{i,j}$ be the matrix with (i,j) and (j,i) entry equal to $1-p$ with probability p and $-p$ with probability $1-p$. All other $B_{i,j}$ entries are 0. Then, clearly, $B_{i,j}$s are independent random matrices and further $A = \sum_{i<j} B_{i,j}$.

Next, notice that $\|B_{i,j}\|_2 \leq \max_a \|B_{i,j}(a;)\|_1 \leq 1$ where $\|B_{i,j}(a;)\|_1$ is the maximum ℓ_1 norm of any row of $B_{i,j}$.

Next, let's compute the variance parameter. We first observe that $B_{i,j}^2$ is a diagonal matrix with exactly two non-zero entries – the (i,i) and the (j,j) entries. The entries themselves are $(1-p)^2$ with probability p and p^2 with probability $1-p$. The expectation of the entry is thus $p(1-p)^2 + (1-p)p^2 = p(1-p)$.

Since $B_{i,j}^2$ is diagonal, so is $\sum_{i<j} B_{i,j}^2$. Further, each entry on the diagonal is exactly equal to $(n-1)p(1-p)$. The spectral norm of a diagonal matrix is the maximum entry so is thus exactly $(n-1)p(1-p)$. Thus, $\sigma^2 = (n-1)p(1-p)$.

Thus, by Matrix Bernstein inequality, with probability at least $1-1/n^{100}$, the spectral norm of A is at most $O(\sqrt{np}\log n)$. □

The Planted Clique Model

Back in 1976, only <5 years after the discovery of the proof of the Cook-Levin theorem, Karp proposed studying the task of finding the maximum clique in a random graph. Precisely speaking, let $G(n,1/2)$ be the distribution on graphs on n vertices where every edge is included independently in the graph with probability $1/2$. Such a graph is dense — it has $\sim n^2/4$ edges in expectation.

Properties of such random graphs are extremely well-studied. In particular, we know that the maximum clique in such graphs is of size $\sim 2\log_2 n$ with all but a negligible probability as $n \to \infty$. Obtaining an estimate of $\Theta(\log n)$ on the size of the maximum clique is not so hard and is an application of the first and second-moment method. First, calculate the expected number of k-cliques for a parameter k by linearity of expectation. Notice that this turns out to be $2^{-\binom{k}{2}}\binom{n}{k}$ and is $\ll 1$ if $k \gg 2\log_2 n$ which already implies by Markov's inequality that the maximum clique cannot be more than $O(\log n)$ in size. To show that there is a clique of size $\sim 2\log_2 n$, we need to compute the variance of the random variable that counts the number of k-cliques. This takes some more effort but is still elementary.

It turns out that the concentration of the maximum clique in $G(n, 1/2)$ is much stronger than the above argument might suggest. Indeed, this random variable exhibits what is called a *two-point* concentration inequality that says that the maximum clique takes one of two possible values $\lfloor 2\log_2 n \lfloor$ and $\lfloor 2\log_2 n \lfloor +1$ with $1 - o_n(1)$ probability.

So we know that $G \sim G(n, 1/2)$ has a maximum clique of size $2\log_2 n$. Can we find it?

There's a simple greedy algorithm that finds a clique of size $\log_2 n$ (i.e. about half the size of the maximum clique): repeat until no longer possible: 1) Let $S = \emptyset$, 2) take any vertex in the common neighborhood of S and add it to S.

Exercise 1. *Analyze the above algorithm and argue that it finds a clique of size* $\geq log_2 n - o(\log_2 n)$ *with probability at least* 0.99.

The idea of the analysis is easy: each time we add a vertex, the size of the common neighborhood slashes by $\sim 1/2$ and the edges between the vertices in the common neighborhood are independent of all the "randomness" we have seen so far.

Karp asked if we could improve on this algorithm. No improved algorithm is known so far. There are lower bounds on restricted class of methods that at least justify our failure in finding better algorithms somewhat [1].

[1]

The Planted Clique Model Planted clique model $G(n, 1/2, k)$ is a variant of the above question of Karp. In this model, the input graph is generated by taking a $G \sim G(n, 1/2)$, taking a fixed set of k vertices and adding edges to create a k-clique on it. We choose $k \gg 2\log_2 n$ so the maximum clique in the new graph must be the added or planted clique since there was no clique larger than $2\log_2 n$ in the graph before. The goal of the algorithm is to find the vertices in the planted k-clique.

There is a simple quasipolynomial algorithm for the problem that searches for $3\log_2 n$ size cliques iteratively and observes that each such clique must be a subclique of the planted k-clique.

Exercise 2. *Flesh out the details of this simple method to recover the planted k-clique*

The key algorithmic question is to find the planted k-clique (with high probability over the draw of the graph) in polynomial time.

We will solve a related but easier question in this lecture. Given a graph G that is either generated by taking a sample from $G(n, 1/2)$ ("random") or from $G(n, 1/2, k)$ ("planted"), decide correctly which model was used to generate it.

The following is an easy by key observation that relates the spectral norm of a (variant of) the adjacency matrix of a graph and the size of its largest clique.

Lemma 1. *Let G be a graph with a clique of size ω. Let A be the ± 1-adjacency matrix of G. That is, $A(i,j) = +1$ if $\{i,j\}$ is an edge in G and -1 otherwise. Then,*

$$\|A\|_2 + 1 \geq \omega$$

Proof. Let $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ be the eigenvalues of A. By the Courant-Fisher theorem, we know that

$$\|A\|_2 \geq \lambda_1(A) = \max_{x \neq 0} \frac{x^\top A x}{\|x\|_2^2}$$

Thus, to prove the lemma, it is enough to find a vector x such that the ratio in the RHS is at least $\omega - 1$. We take the 0-1 indicator x vector of the ω-clique, say S, in G. Note that $\|x\|_2^2 = \omega$.

Further, $x^\top A x = \sum_{i,j} x_i x_j A(i,j) = \omega(\omega - 1)$ since only the $\{i,j\}$ such that $i, j \in S$ contribute a non-zero value and in that case, in fact each term contributes a $+1$ whenever $i \neq j$.

Thus, $\frac{x^\top A x}{\|x\|_2^2} = \omega(\omega - 1)/\omega = \omega - 1$ as desired.

□

We can now describe and analyze a distinguishing algorithm.

1. Construct the ± 1 adjacency matrix of the input graph.
2. Compute $\|A\|_2$.
3. If $\|A\|_2 \leq C\sqrt{n}$, output "random". Otherwise output "planted".

Lemma 2. *There is a constant $C > 0$ such that the above distinguishing algorithm succeeds correctly with high probability whenever $k \geq 2C\sqrt{n}$.*

Proof. The analysis is simple. Observe that A is a random matrix with independent entries up to symmetry when $G \sim G(n, 1/2)$. So by the theorem above (for $p = q = 1/2$), $\|A\|_2 \leq C\sqrt{n}$ for some constant $C > 0$.

On the other hand, by the lemma above, in the case when $k \geq 2C\sqrt{n}$, the spectral norm of $\|A\|_2 \geq 2C\sqrt{n}$. □

Bibliography

[1] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. ITCS 2014