1 The operator norm of random matrices

1.1 Concentration inequalities

Last time we proved the following lemmas.

Lemma 1.1. *Suppose that $\xi_{i,j}$ are independent, $\mathbf{E}\xi_{i,j} = 0$ and $|\xi_{i,j}| \leq 1$. Let $x \in \mathbb{C}^n$ be a unit vector. Then for sufficiently large A, we have*

$$\mathbf{P}(\|Mx\|_2 \geq A\sqrt{n}) \leq C\exp(-cAn)$$

for some absolute constants $C, c > 0$.

Lemma 1.2. *Let $\varepsilon \in (0,1)$, and let Σ be an ε-net of $\{x \in \mathbb{C}^n : \|x\|_2 = 1\}$. Then there exists $C > 0$ such that $\#\Sigma \leq (C/\varepsilon)^n$.*

Lemma 1.3. *Let Σ be a maximal $1/2$-net of $\{x \in \mathbb{C}^n : \|x\|_2 = 1\}$. Then for any $\lambda > 0$, we have*

$$\mathbf{P}(\|M\| > \lambda) \leq \mathbf{P}\left(\bigcup_{y\in\Sigma}\{\|My\|_2 > \lambda/2\}\right).$$

Applying these lemmas, we obtain the following upper tail estimate for bounded i.i.d. matrix ensembles.

Corollary 1.4. *Suppose that $\xi_{i,j}$ are independent, $\mathbf{E}\xi_{i,j} = 0$ and $|\xi_{i,j}| \leq 1$. Then for sufficiently large A, we have*

$$\mathbf{P}(\|M\| \geq A\sqrt{n}) \leq C\exp(-cAn)$$

for some absolute constants $C, c > 0$.

We can actually deduce an upper tail estimate for random Hermitian matrices from this.

Corollary 1.5. *Suppose that M is a random Hermitian matrix, with $|\xi_{i,j}| \leq 1$. Then for sufficiently large A, we have*

$$\mathbf{P}(\|M\| \geq A\sqrt{n}) \leq C\exp(-cAn)$$

for some absolute constants $C, c > 0$.

Proof. Write $M = L + U$, where L is lower triangular and U is strictly upper triangular. Then apply previous Corollary to L and U. □

Remark. *The upper tail estimates still hold if we assume $\xi_{i,j}$ are (sub)Gaussian random variables instead of $|\xi_{i,j}| \leq 1$. In particular, the estimates also hold for GOE and GUE.*

We will skip Section 2.3.2.

Other than $\mathbf{P}(\|M\| > \lambda)$, we are also interested in how $\|M\|$ deviates from its mean, like the Chebyshev inequality. More precisely, we would like to bound $\mathbf{P}(|\|M\| - \mathbf{E}\|M\|| > \lambda)$. We will use Talagrand's inequality, which is very powerful.

Theorem 1.6. *Let $K > 0$, and let $X_1, \ldots, X_n$ be independent complex random variables with $|X_i| \leq K$ for all $1 \leq i \leq n$. Let $F : \mathbb{C}^n \to \mathbb{R}$ be a 1-Lipschitz convex function, where we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ for the purposes of defining Lipschitz and convex. Then for any $\lambda \geq 0$, we have*

$$\mathbf{P}(|F(X) - \mathbf{E}F(X)| \geq \lambda K) \leq C \exp(-c^2\lambda)$$

and

$$\mathbf{P}(|F(X) - \mathbf{M}F(X)| \geq \lambda K) \leq C \exp(-c^2\lambda)$$

for some absolute constants $C, c > 0$, where $\mathbf{M}F(X)$ is a median of $F(X)$.

The following is a quick application of Talagrand's inequality.

Proposition 1.7. *Suppose that $\xi_{i,j}$ are independent, $\mathbf{E}\xi_{i,j} = 0$ and $|\xi_{i,j}| \leq 1$. Then for any $\lambda \geq 0$, we have*

$$\mathbf{P}(|\|M\| - \mathbf{E}\|M\|| \geq \lambda) \leq C \exp(-c\lambda^2)$$

and

$$\mathbf{P}(|\|M\| - \mathbf{M}\|M\|| \geq \lambda) \leq C \exp(-c\lambda^2)$$

for some absolute constants $C, c > 0$.

Proof. We can view $\|M\|$ as a function $F : \mathbb{C}^{n^2} \to \mathbb{R}$ defined by $F((\xi_{i,j})) = \|M\|$. Write $N = (\eta_{i,j})$. F is convex because

$$F(t(\xi_{i,j})+(1-t)(\eta_{i,j})) = \|tM+(1-t)N\| \leq t\|M\|+(1-t)\|N\| = tF((\xi_{i,j}))+(1-t)F((\eta_{i,j})).$$

Also, writing $Q = M - N$,

$$|F((\xi_{i,j})) - F((\eta_{i,j}))| \leq \|Q\| = \sqrt{\lambda_{\max}(Q^*Q)} \leq \sqrt{\operatorname{tr}(Q^*Q)} = \sqrt{\sum_{i,j=1}^{n} |\xi_{i,j} - \eta_{i,j}|^2},$$

and so F is 1-Lipschitz. Therefore we can apply Talagrand's inequality. □

Similarly, one can show that this inequality holds for bounded Hermitian matrix ensembles. Although from the above Proposition we know that $\|M\|$ is close to its mean, we don't know what the value of the mean is (we only have some rough estimates). Therefore, we will need some other methods, such as the moment method.

1.2 Moment method

We will first assume M is symmetric or Hermitian. In this case,

$$\|M\| = \max_{1 \leq i \leq n} |\lambda_i|.$$

Also, we have

$$\operatorname{tr}(M^k) = \sum_{i=1}^{n} \lambda_i^k.$$

If $k \in 2\mathbb{N}$, then we have

$$\|M\|^k \le \mathrm{tr}(M^k) \le n\|M\|^k,$$

or equivalently,

$$\|M\| \le \mathrm{tr}(M^k)^{1/k} \le n^{1/k}\|M\|. \tag{1.1}$$

If we have knowledge on the k-th moment $\mathrm{tr}(M^k)$, then we will have control on $\|M\|$, up to a factor $n^{1/k}$. When k gets larger, we will expect that we can get better controls.

Let's first consider the case $k = 2$. In this case,

$$\mathrm{tr}(M^2) = \sum_{i,j=1}^{n} |\xi_{i,j}|^2.$$

Note that $|\xi_{i,j}|^2$ are independent. Therefore, by the law of large numbers, we obtain that $\mathrm{tr}(M^2) \sim n^2$. From (1.1), we see that

$$\sqrt{n} \lesssim \|M\| \lesssim n.$$

Recall that from the concentration inequality, we have $\|M\| \lesssim \sqrt{n}$ with high probability, and so we obtain a nice lower bound, but the upper bound should be far from optimal.

Let's see what happens if we try $k = 4$. For simplicity, we will assume that $\mathbf{E}\xi_{i,j} = 0$ and $\mathrm{Var}(\xi_{i,j}) = 1$. We will also assume that $|\xi_{i,j}| \le K$. Note that

$$\mathrm{tr}(M^4) = \sum_{1\le i_1,i_2,i_3,i_4\le n} \xi_{i_1,i_2}\xi_{i_2,i_3}\xi_{i_3,i_4}\xi_{i_4,i_1}.$$

Taking expectation, we have

$$\mathbf{E}\mathrm{tr}(M^4) = \sum_{1\le i_1,i_2,i_3,i_4\le n} \mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_3}\xi_{i_3,i_4}\xi_{i_4,i_1}.$$

One can view this sum graphically, as a sum over length four cycles in the vertex set $\{1,\ldots,n\}$. If all the edges are distinct, then by independence, $\mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_3}\xi_{i_3,i_4}\xi_{i_4,i_1} = 0$. There are many such terms will vanish. One can show that up to cyclic permutations of i_1, i_2, i_3, i_4, there are only a few types of cycles in which $\mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_3}\xi_{i_3,i_4}\xi_{i_4,i_1}$ does not vanish, namely

1. $i_1 = i_3$, but i_2, i_4 are distinct from each other and from i_1. In this case,

   $$\mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_3}\xi_{i_3,i_4}\xi_{i_4,i_1} = \mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_1}\xi_{i_1,i_4}\xi_{i_4,i_1} = \mathbf{E}\xi_{i_1,i_2}^2\xi_{i_1,i_4}^2 = 1.$$

   There are $O(n^3)$ such terms, so the total contribution to $\mathbf{E}\mathrm{tr}(M^4)$ is at most $O(n^3)$.

2. $i_1 = i_3$, $i_2 = i_4$ but $i_1 \ne i_2$. In this case,

   $$\mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_3}\xi_{i_3,i_4}\xi_{i_4,i_1} = \mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_1}\xi_{i_1,i_2}\xi_{i_2,i_1} = \mathbf{E}\xi_{i_1,i_2}^4 \le K^2\mathbf{E}\xi_{i_1,i_2}^2 = K^2.$$

   There are $O(n^2)$ such terms, so the total contribution to $\mathbf{E}\mathrm{tr}(M^4)$ is at most $O(n^2K^2)$.

3. $i_1 = i_2 = i_3$, but $i_4 \neq i_1$. In this case,

$$\mathbf{E}\xi_{i_1,i_2}\xi_{i_2,i_3}\xi_{i_3,i_4}\xi_{i_4,i_1} = \mathbf{E}\xi_{i_1,i_1}\xi_{i_1,i_1}\xi_{i_1,i_4}\xi_{i_4,i_1} = \mathbf{E}\xi^2_{i_1,i_1}\xi^2_{i_1,i_4} = 1.$$

   The total contribution to $\mathbf{E}\mathrm{tr}(M^4)$ is at most $O(n^2)$.

4. $i_1 = i_2 = i_3 = i_4$. The total contribution is $O(nK^2)$.

Combining all the cases, we see that

$$\mathbf{E}\mathrm{tr}(M^4) \leq O(n^3) + O(n^2K^2).$$

If $K = O(\sqrt{n})$, then $\mathbf{E}\mathrm{tr}(M^4) \leq O(n^3)$. By Markov inequality, this gives $\mathrm{tr}(M^4) \leq O(n^3)$ with high probability. In particular, (1.1) gives

$$\|M\| \lesssim n^{3/4}$$

with high probability. The bound has been improved!

We will see what will happen when k gets larger next time.

References

[1] Tao, Terence. Topics in random matrix theory. Graduate Studies in Mathematics, 132. American Mathematical Society, Providence, RI, 2012. x+282 pp. ISBN: 978-0-8218-7430-1