SEEM 5380: Optim. Methods for High-Dim. Statistics **2021–22 Second Term**

Handout 1: Spectrum of a Gaussian Random Matrix

Instructor: Anthony Man–Cho So January 21, 2022

Our goal in this handout is to give upper and lower bounds on the singular values of a Gaussian random matrix. Before we proceed, let us recall some basic definitions and results concerning the singular values of a matrix.

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a given matrix. The i-th singular value of $\boldsymbol{A}$ can be computed as $s_i(\boldsymbol{A}) = \sqrt{\lambda_i(\boldsymbol{A}^T\boldsymbol{A})}$, where $\lambda_i(\boldsymbol{A}^T\boldsymbol{A})$ is the i-th eigenvalue of $\boldsymbol{A}^T\boldsymbol{A}$. The spectral norm (i.e., the largest singular value) of $\boldsymbol{A}$ is given by

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{v} \in \mathbb{S}^{n-1}} \|\boldsymbol{A}\boldsymbol{v}\|_2 = \max_{\boldsymbol{u} \in \mathbb{S}^{m-1},\, \boldsymbol{v} \in \mathbb{S}^{n-1}} \boldsymbol{u}^T\boldsymbol{A}\boldsymbol{v}, \tag{1}$$

where $\mathbb{S}^{n-1} = \{\boldsymbol{v} \in \mathbb{R}^n : \|\boldsymbol{v}\|_2 = 1\}$. Observe that when $m = n$ and $\boldsymbol{A}$ is symmetric, we have

$$\|\boldsymbol{A}\| = \max_{\boldsymbol{u} \in \mathbb{S}^{n-1}} |\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u}|.$$

This can be established by considering the spectral decomposition of $\boldsymbol{A}$.

Suppose that $m \geq n$. It is easy to show that for any $t \in (0,1)$, if $\|\boldsymbol{A}^T\boldsymbol{A} - \boldsymbol{I}_n\| \leq t$, then $1 - t \leq s_i(\boldsymbol{A}) \leq 1 + t$ for $i \in \{1, \dots, n\}$. This gives us a handle to bound all the singular values of $\boldsymbol{A}$.

Now, let $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ be a matrix whose entries are independent and identically distributed (iid) standard Gaussian random variables; i.e., $X_{ij} \sim \mathcal{N}(0,1)$ for $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$. Observe that $\mathbb{E}[\boldsymbol{X}^T\boldsymbol{X}] = m\boldsymbol{I}_n$. Hence, our goal of bounding the singular values of $\boldsymbol{X}$ can be achieved by bounding

$$\left\| \frac{1}{m}\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{I}_n \right\|.$$

1 Computing Spectral Norm on a Net

Recall from (1) that the spectral norm of a matrix involves the maximization of a certain quantity over the sphere $\mathbb{S}^{n-1}$. In the study of random matrices, it is often useful to perform the maximization over a *finite* subset $\mathcal{N}$ of $\mathbb{S}^{n-1}$. To get a small approximation error, the set $\mathcal{N}$ should cover $\mathbb{S}^{n-1}$ sufficiently well. This motivates the following definition:

Definition 1 (ϵ-Net) *We say that $\mathcal{N}$ is an ϵ-net of $\mathbb{S}^{n-1}$ if (i) $\mathcal{N} \subseteq \mathbb{S}^{n-1}$ and (ii) for every $\boldsymbol{v} \in \mathbb{S}^{n-1}$, there exists a $\boldsymbol{v}_0 \in \mathcal{N}$ such that $\|\boldsymbol{v} - \boldsymbol{v}_0\|_2 \leq \epsilon$.*

By considering suitable nets on $\mathbb{S}^{m-1}$ and $\mathbb{S}^{n-1}$, we have the following result:

Proposition 1 (Bilinear Form on a net) *Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a given matrix. Let $\mathcal{M}$ be a δ-net of $\mathcal{S}^{m-1}$ and $\mathcal{N}$ be an ϵ-net of $\mathbb{S}^{n-1}$, where $\delta, \epsilon \in [0,1)$. Then,*

$$\|\boldsymbol{A}\| \leq \frac{1}{(1-\delta)(1-\epsilon)} \max_{\boldsymbol{u} \in \mathcal{M},\, \boldsymbol{v} \in \mathcal{N}} \boldsymbol{u}^T\boldsymbol{A}\boldsymbol{v}.$$

Proof Let $\bar{\boldsymbol{v}} \in \mathbb{S}^{n-1}$ be such that $\|\boldsymbol{A}\| = \|\boldsymbol{A}\bar{\boldsymbol{v}}\|_2$. By definition of a net, we can find a $\bar{\boldsymbol{v}}_0 \in \mathcal{N}$ such that $\|\bar{\boldsymbol{v}} - \bar{\boldsymbol{v}}_0\|_2 \leq \epsilon$. Hence, we have

$$\|\boldsymbol{A}\| = \|\boldsymbol{A}(\bar{\boldsymbol{v}}_0 + (\bar{\boldsymbol{v}} - \bar{\boldsymbol{v}}_0)\|_2 \leq \|\boldsymbol{A}\bar{\boldsymbol{v}}_0\|_2 + \|\boldsymbol{A}\| \cdot \|\bar{\boldsymbol{v}} - \bar{\boldsymbol{v}}_0\|_2 = \|\boldsymbol{A}\bar{\boldsymbol{v}}_0\|_2 + \epsilon\|\boldsymbol{A}\|,$$

which implies that

$$\|\boldsymbol{A}\| \leq \frac{1}{1-\epsilon}\|\boldsymbol{A}\bar{\boldsymbol{v}}_0\|_2. \tag{2}$$

Now, let $\bar{\boldsymbol{u}} \in \mathbb{S}^{m-1}$ be such that $\|\boldsymbol{A}\bar{\boldsymbol{v}}_0\|_2 = \bar{\boldsymbol{u}}^T\boldsymbol{A}\bar{\boldsymbol{v}}_0$. Again, by definition of a net, we can find a $\bar{\boldsymbol{u}}_0 \in \mathcal{M}$ such that $\|\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}_0\|_2 \leq \delta$. Hence, we have

$$\|\boldsymbol{A}\bar{\boldsymbol{v}}_0\|_2 = \bar{\boldsymbol{u}}_0^T\boldsymbol{A}\bar{\boldsymbol{v}}_0 + (\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}_0)^T\boldsymbol{A}\bar{\boldsymbol{v}}_0 \leq \max_{\boldsymbol{u}\in\mathcal{M},\, \boldsymbol{v}\in\mathcal{N}} \boldsymbol{u}^T\boldsymbol{A}\boldsymbol{v} + \delta\|\boldsymbol{A}\bar{\boldsymbol{v}}_0\|_2. \tag{3}$$

Putting (2) and (3) together yields the desired result. □

In the case where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is symmetric, we have the following result:

Proposition 2 (Quadratic Form on a net) *Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a given symmetric matrix. Let $\mathcal{N}$ be an ϵ-net of $\mathbb{S}^{n-1}$, where $\epsilon \in [0, 1/2)$. Then,*

$$\|\boldsymbol{A}\| \leq \frac{1}{1-2\epsilon}\max_{\boldsymbol{u}\in\mathcal{N}} |\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u}|.$$

Proof Let $\bar{\boldsymbol{u}} \in \mathbb{S}^{n-1}$ be such that $\|\boldsymbol{A}\| = |\bar{\boldsymbol{u}}^T\boldsymbol{A}\bar{\boldsymbol{u}}|$ and $\bar{\boldsymbol{u}}_0 \in \mathcal{N}$ be such that $\|\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}_0\|_2 \leq \epsilon$. Noting that $\|\boldsymbol{A}\| = \max_{\boldsymbol{u}\in\mathbb{S}^{n-1}} \|\boldsymbol{A}\boldsymbol{u}\|_2$ and $\|\bar{\boldsymbol{u}}_0\|_2 = 1$ by definition, we compute

$$\begin{aligned}
\|\boldsymbol{A}\| = |\bar{\boldsymbol{u}}^T\boldsymbol{A}\bar{\boldsymbol{u}}| &\leq |\bar{\boldsymbol{u}}^T\boldsymbol{A}(\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}_0)| + |\bar{\boldsymbol{u}}^T\boldsymbol{A}\bar{\boldsymbol{u}}_0| \\
&\leq \epsilon\|\boldsymbol{A}\bar{\boldsymbol{u}}\|_2 + |(\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}_0)^T\boldsymbol{A}\bar{\boldsymbol{u}}_0| + |\bar{\boldsymbol{u}}_0^T\boldsymbol{A}\bar{\boldsymbol{u}}_0| \\
&\leq \epsilon\|\boldsymbol{A}\| + \|\boldsymbol{A}(\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}_0)\|_2 + \max_{\boldsymbol{u}\in\mathcal{N}} |\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u}| \\
&\leq 2\epsilon\|\boldsymbol{A}\| + \max_{\boldsymbol{u}\in\mathcal{N}} |\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u}|.
\end{aligned}$$

This completes the proof. □

Next, let us bound the size of an ϵ-net of $\mathbb{S}^{n-1}$ using a volume argument.

Proposition 3 (Size of a Net) *Let $\epsilon \in (0, 1)$ be given. Then, there exists an ϵ-net $\mathcal{N}$ of $\mathbb{S}^{n-1}$ with*

$$|\mathcal{N}| \leq \left(\frac{2}{\epsilon} + 1\right)^n - \left(\frac{2}{\epsilon} - 1\right)^n.$$

Proof Let $\mathcal{N}$ be a maximal cardinality subset of $\mathbb{S}^{n-1}$ such that for any distinct $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{N}$, we have $\|\boldsymbol{u} - \boldsymbol{v}\|_2 > \epsilon$. By the maximality of $\mathcal{N}$, we see that $\mathcal{N}$ is an ϵ-net of $\mathbb{S}^{n-1}$. Now, observe that $B(\boldsymbol{u}, \epsilon/2) \cap B(\boldsymbol{v}, \epsilon/2) = \emptyset$ for every distinct $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{N}$, where $B(\boldsymbol{x}, \epsilon)$ denotes the Euclidean ball centered at $\boldsymbol{x} \in \mathbb{R}^n$ with radius ϵ. Moreover, it is clear that

$$\bigcup_{\boldsymbol{u}\in\mathcal{N}} B(\boldsymbol{u}, \epsilon/2) \subseteq B(\boldsymbol{0}, 1 + \epsilon/2) \backslash B(\boldsymbol{0}, 1 - \epsilon/2).$$

Hence, by comparing the volumes of the balls, we have

$$|\mathcal{N}| \cdot \mathrm{vol}(B(\boldsymbol{0}, \epsilon/2)) \leq \mathrm{vol}(B(\boldsymbol{0}, 1 + \epsilon/2)) - \mathrm{vol}(B(\boldsymbol{0}, 1 - \epsilon/2)). \tag{4}$$

Now, recall that

$$\text{vol}(B(\boldsymbol{x},\epsilon)) = \epsilon^n \text{vol}(B(\boldsymbol{0},1)) \quad \text{for any } \boldsymbol{x} \in \mathbb{R}^n.$$

Hence, upon dividing both sides of (4) by $\text{vol}(B(\boldsymbol{0},1))$, we obtain

$$|\mathcal{N}| \cdot \left(\frac{\epsilon}{2}\right)^n \leq \left(1+\frac{\epsilon}{2}\right)^n - \left(1-\frac{\epsilon}{2}\right)^n.$$

This implies the desired result. $\square$

2 Concentration Inequality for χ^2 Random Variable

Let $\mathcal{N}$ be an ϵ-net of $\mathbb{S}^{n-1}$. Using Proposition 2, we have

$$\left\|\frac{1}{m}\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{I}_n\right\| \leq \frac{1}{1-2\epsilon}\max_{\boldsymbol{u}\in\mathcal{N}}\left|\boldsymbol{u}^T\left(\frac{1}{m}\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{I}_n\right)\boldsymbol{u}\right| = \frac{1}{1-2\epsilon}\max_{\boldsymbol{u}\in\mathcal{N}}\left|\frac{1}{m}\|\boldsymbol{X}\boldsymbol{u}\|_2^2 - 1\right|. \tag{5}$$

Now, using the fact that X_{ij} are iid according to $\mathscr{N}(0,1)$ for $i \in \{1,\ldots,m\}$, $j \in \{1,\ldots,n\}$ and $\|\boldsymbol{u}\|_2 = 1$ for any $\boldsymbol{u} \in \mathcal{N}$, we have

$$\frac{1}{m}\|\boldsymbol{X}\boldsymbol{u}\|_2^2 - 1 = \frac{1}{m}\sum_{i=1}^m (g_i^2 - 1),$$

where $g_1,\ldots,g_m$ are iid according to $\mathscr{N}(0,1)$. As an aside, the random variable $\sum_{i=1}^m g_i^2$ follows the so-called χ^2-distribution with m degrees of freedom. The above development motivates us to estabish the following result:

Proposition 4 (Concentration Inequality for χ^2 Random Variable) *Let $g_1,\ldots,g_m$ be iid according to $\mathscr{N}(0,1)$. Then, for any $t \in [0,1]$,*

$$\Pr\left(\left|\frac{1}{m}\sum_{i=1}^m (g_i^2-1)\right| \geq t\right) \leq 2\exp(-mt^2/8).$$

Proof Note that

$$\Pr\left(\left|\frac{1}{m}\sum_{i=1}^m (g_i^2-1)\right| \geq t\right) \leq \Pr\left(\frac{1}{m}\sum_{i=1}^m (g_i^2-1) \geq t\right) + \Pr\left(\frac{1}{m}\sum_{i=1}^m (g_i^2-1) \leq -t\right). \tag{6}$$

Hence, it suffices to bound the two terms on the right-hand side of (6) separately. To bound the first term in (6), observe that by Markov's inequality, for any $\lambda \geq 0$, we have

$$\begin{aligned}\Pr\left(\frac{1}{m}\sum_{i=1}^m (g_i^2-1) \geq t\right) &= \Pr\left[\exp\left(\lambda\sum_{i=1}^m (g_i^2-1)\right) \geq \exp(\lambda m t)\right] \\ &\leq \exp(-\lambda m t)\cdot\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^m (g_i^2-1)\right)\right]. \end{aligned}\tag{7}$$

Since $g_1,\ldots,g_m$ are iid, we have

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^m (g_i^2-1)\right)\right] = \prod_{i=1}^m \mathbb{E}\left[\exp\left(\lambda(g_i^2-1)\right)\right] = \left(\mathbb{E}\left[\exp\left(\lambda(g_1^2-1)\right)\right]\right)^m.$$

Using the density function of a standard Gaussian random variable, for any $\lambda < 1/2$, we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda(g_1^2-1)\right)\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(\lambda(x^2-1))\cdot\exp(-x^2/2)\,dx \\ &= \sigma\cdot\exp(-\lambda)\cdot\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp(-x^2/2\sigma^2)\,dx,\end{aligned}$$

where $\sigma = (1-2\lambda)^{-1/2}$. Upon noting that

$$x \mapsto \frac{1}{\sqrt{2\pi}\sigma}\exp(-x^2/2\sigma^2)$$

is the density function of a Gaussian random variable with mean 0 and variance σ^2 and using the inequality $-x-x^2 \le \ln(1-x)$, which is valid for any $x \in [0,1/2]$, we obtain, for any $\lambda \in [0,1/4]$,

$$\mathbb{E}\left[\exp\left(\lambda(g_1^2-1)\right)\right] = \frac{\exp(-\lambda)}{\sqrt{1-2\lambda}} \le \exp(2\lambda^2).$$

Substituting the above into (7), we get, for any $\lambda \in [0,1/4]$,

$$\Pr\left(\frac{1}{m}\sum_{i=1}^{m}(g_i^2-1)\ge t\right) \le \exp(m(2\lambda^2-\lambda t)).$$

Since the above bound holds for any $\lambda \in [0,1/4]$, we can get the best bound by minimizing the right-hand side over λ. In particular, by setting $\lambda = t/4$ and noting that $\lambda \in [0,1/4]$ whenever $t \in [0,1]$, we obtain

$$\Pr\left(\frac{1}{m}\sum_{i=1}^{m}(g_i^2-1)\ge t\right) \le \exp(-mt^2/8).$$

To bound the second term in (6), we proceed in a similar manner. Specifically, by Markov's inequality, for any $\lambda \ge 0$, we have

$$\Pr\left(\frac{1}{m}\sum_{i=1}^{m}(g_i^2-1)\le -t\right) \le \exp(-\lambda m t)\cdot\left(\mathbb{E}\left[\exp\left(\lambda(1-g_1^2)\right)\right]\right)^m.$$

Since the inequality $(1+2x)^{-1/2} \le \exp(2x^2-x)$ holds for all $x \ge 0$, a simple calculation yields

$$\mathbb{E}\left[\exp\left(\lambda(1-g_1^2)\right)\right] = \frac{\exp(\lambda)}{\sqrt{1+2\lambda}} \le \exp(2\lambda^2).$$

Using the same argument as before, we obtain, for any $t \in [0,1]$,

$$\Pr\left(\frac{1}{m}\sum_{i=1}^{m}(g_i^2-1)\le -t\right) \le \exp(-mt^2/8).$$

This completes the proof. $\square$

3 Bounding the Spectrum of Gaussian Random Matrix

With the results in the previous sections, we are now ready to give upper and lower bounds on the singular values of a Gaussian random matrix. Let $\mathcal{N}$ be a (1/4)-net of $\mathbb{S}^{n-1}$. By (5) and Propositions 3 and 4, we have

$$\begin{aligned}\Pr\left(\left\|\frac{1}{m}\boldsymbol{X}^T\boldsymbol{X}-\boldsymbol{I}_n\right\|\geq 2t\right) &\leq \Pr\left(\max_{\boldsymbol{u}\in\mathcal{N}}\left|\frac{1}{m}\|\boldsymbol{X}\boldsymbol{u}\|_2^2-1\right|\geq t\right)\leq\sum_{\boldsymbol{u}\in\mathcal{N}}\Pr\left(\left|\frac{1}{m}\|\boldsymbol{X}\boldsymbol{u}\|_2^2-1\right|\geq t\right)\\ &\leq 2\cdot 9^n\cdot\exp(-mt^2/8).\end{aligned}$$

By setting $t^2=\frac{8\ln 9}{m}(n+\eta^2)$ for some $\eta\geq 0$ and noting that $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$ for any $a,b\geq 0$, we get

$$\Pr\left(\left\|\frac{1}{m}\boldsymbol{X}^T\boldsymbol{X}-\boldsymbol{I}_n\right\|\geq 2\cdot\sqrt{\frac{8\ln 9}{m}}\cdot(\sqrt{n}+\eta)\right)\leq 2\exp(-(\ln 9)\eta^2).$$

It follows that when $\eta>0$ is sufficiently large, we have

$$\sqrt{m}-c(\sqrt{n}+\eta)\leq s_i(\boldsymbol{X})\leq\sqrt{m}+c(\sqrt{n}+\eta) \tag{8}$$

for $i\in\{1,\ldots,n\}$ with high probability, where $c=2\sqrt{8\ln 9}$.

4 Remarks

1. It is possible to establish upper and lower bounds on the singular values of more general classes of random matrices; see [3, Chapter 4] for a development in this direction.

2. Note that the lower bound in (8) becomes less useful as m tends to n; i.e., the matrix $\boldsymbol{X}$ becomes more square. To estimate the least singular value of an almost-square Gaussian random matrix, one needs more sophisticated machinery. We refer the interested reader to the paper [1] and survey [2] for further reading.

References

[1] M. Rudelson and R. Vershynin. Smallest Singular Value of a Random Rectangular Matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009.

[2] M. Rudelson and R. Vershynin. Non–Asymptotic Theory of Random Matrices: Extreme Singular Values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010)*, pages 1576–1602, 2011.

[3] R. Vershynin. *High–Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018.