

Improved bounds for random projection

Miaoyan Wang, Oct 15, 2019

Notation. We say that an event A occurs “with high probability” if $\mathbb{P}(A)$ tends to 1 as the dimension $d_{\min} = \min\{d_1, d_2\}$ tends to infinity. We say that A occurs “with very high probability” if $\mathbb{P}(A)$ tends to 1 faster than any polynomial of d_{\min} .

Let $\mathbf{E} \in \mathbb{R}^{d_1 \times d_2}$ be a real matrix. The matrix spectral norm is defined as

$$\|\mathbf{E}\|_\sigma = \max_{(\mathbf{a}, \mathbf{b}) \in \mathbb{S}^{d_1-1} \times \mathbb{S}^{d_2-1}} \mathbf{a}^T \mathbf{E} \mathbf{b},$$

where $\mathbb{S}^{d-1} = \{\mathbf{a} \in \mathbb{R}^d : \|\mathbf{a}\|_\sigma = 1\}$ denotes the unit sphere in \mathbb{R}^d .

Theorem 0.1. Consider a noisy rank-1 matrix model $\mathbf{D} = \lambda \mathbf{a} \otimes \mathbf{b} + \mathbf{E}$, where $\lambda \in \mathbb{R}_+$ is a scalar, $\mathbf{a} \in \mathbb{R}^{d_1}$, $\mathbf{b} \in \mathbb{R}^{d_2}$ are unit-1 vectors, and $\mathbf{E} \in \mathbb{R}^{d_1 \times d_2}$ is a Gaussian matrix with i.i.d. $N(0, \sigma^2)$ entries. Define a random projection

$$\hat{\mathbf{a}} = \mathbf{D} \mathbf{\Omega}, \text{ where } \mathbf{\Omega} = (z_1, \dots, z_{d_2})^T \stackrel{i.i.d.}{\sim} N(0, 1).$$

Suppose $\lambda \gg \sigma \sqrt{d_2 \max(d_1, d_2)}$ as $d_1, d_2 \rightarrow \infty$. Then, $\cos \Theta(\mathbf{a}, \hat{\mathbf{a}}) \rightarrow 1$ in probability.

Proof. The perturbed rank-1 model $\mathbf{D} = \lambda \mathbf{a} \otimes \mathbf{b} + \mathbf{E}$ implies that

$$\hat{\mathbf{a}} = \mathbf{D} \mathbf{\Omega} = \lambda Z \mathbf{a} + \mathbf{E} \mathbf{\Omega}, \tag{1}$$

where $Z \stackrel{\text{def}}{=} \langle \mathbf{b}, \mathbf{\Omega} \rangle$ is an $N(0, 1)$ random variable. Note that $\mathbf{E} \mathbf{\Omega} \in \mathbb{R}^{d_1}$ is a random vector with length $\|\mathbf{E} \mathbf{\Omega}\|_2 \leq \|\mathbf{E}\|_\sigma \|\mathbf{\Omega}\|_2$.

To show $\cos(\mathbf{a}, \hat{\mathbf{a}}) \rightarrow 1$, it suffices to show that $\cot(\mathbf{a}, \hat{\mathbf{a}}) \rightarrow \infty$. Based on (1), we have

$$\cot \Theta(\mathbf{a}, \hat{\mathbf{a}}) \geq \frac{\|\lambda Z \mathbf{a}\|_2}{\|\mathbf{E} \mathbf{\Omega}\|_2} \geq \frac{\lambda |Z|}{\|\mathbf{E}\|_\sigma \|\mathbf{\Omega}\|_2}. \tag{2}$$

Now consider the asymptotical property of (2) as $d_1, d_2 \rightarrow \infty$. The fact $\|\mathbf{\Omega}\|_2^2 \sim \chi^2(d_2)$ implies that $\|\mathbf{\Omega}\|_2 \asymp (1 + o(1))\sqrt{d_2}$ with very high probability, for some constant $C > 0$. Furthermore, we have that $\|\mathbf{E}\|_\sigma \asymp (2 + o(1))\sigma \sqrt{\max(d_1, d_2)}$ by Lemma 1. Therefore, for any fixed $L > 0$,

$$\begin{aligned} \mathbb{P}(\cot(\mathbf{a}, \hat{\mathbf{a}}) \geq L) &\geq \mathbb{P}\left(\frac{\lambda |Z|}{\|\mathbf{E}\|_\sigma \|\mathbf{\Omega}\|_2} \geq L\right) \\ &\geq \mathbb{P}\left(|Z| \geq \frac{2L\sigma \sqrt{d_2 \max(d_1, d_2)}}{\lambda}\right) \\ &\geq 1 - \frac{4\lambda_*}{\sqrt{2\pi}}, \end{aligned} \tag{3}$$

where $\lambda_* \stackrel{\text{def}}{=} L\sigma \sqrt{d_2 \max(d_1, d_2)}/\lambda$. Here the last line of (3) uses the fact that $\mathbb{P}(|N(0, 1)| \geq t) \geq$

$1 - 2t\phi(0)$ for any $t \geq 0$, where $\phi(\cdot)$ is the pdf for standard normal. Based on assumption $\lambda_* \rightarrow 0$ we conclude that $\cot(\mathbf{a}, \hat{\mathbf{a}}) \geq L$ with high probability. Sending $L \rightarrow \infty$ gives the desired result. \square

Lemma 1 (Spectral norm of Gaussian matrix [1]). *Let $\mathbf{E} \in \mathbb{R}^{d_1 \times d_2}$ be a random matrix with i.i.d. $N(0, 1)$ entries. Then, we have, with very high probability,*

$$\|\mathbf{E}\|_\sigma \asymp (2 + o(1))\sqrt{\max(d_1, d_2)}.$$

References

- [1] Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 1576–1602. World Scientific, 2010.