Upper Bounds on the Spectral Norm of the Pseudo-Inverse of Non-Standard Gaussian Matrices

Arvind Rathnashyam
RPI CS and Math
rathna@rpi.edu

October 3, 2023

Abstract

In this paper we explore upper bounds on the spectral norm for Gaussian Matrices with columns standard from Central Correlated Multivariate Normal Distributions. We utilize a lemma from [Chi17, CWS09] and extend the analysis from [CD05]. These bounds find applications in the generalization of the randomized SVD given in [BT22] and wireless network science.

1 Introduction

The study of the expectation of the norms of the pseudoinverse of standard normal gaussian matrices first appeared in [HMT11] when analyzing the error bounds for the Randomized SVD algorithm. The bounds developed in [HMT11] used theory developed in analyzing the condition numbers of standard normal matrices in [CD05]. In a generalization of the Randomized SVD, the need for bounds on the expectation of the spectral norm for correlated Gaussian matrices appears in [BT22].

2 Relevant Work in Standard Uncorrelated Matrices

In this section we will briefly discuss bounds developed for the inequalities of standard normal matrices.

Proposition 1. *(HMT Proposition 10.2). Draw a $k \times (k+p)$ standard Gaussian matrix $\mathbf{G}$ with $k \geq 2$ and $p \geq 2$. Then*

$$\mathbb{E}\left\|\mathbf{G}^{\dagger}\right\| \leq \frac{e\sqrt{k+p}}{p} \tag{1}$$

From our search in the literature, there is no bound on equation 1 when the columns are not sampled from a multiple of the identity.

3 Theory

We will first introduce the necessary lemmas needed to prove our main results.

3.1 Necessary Lemmas

Lemma 2. *[Jam64, Eq. (58,59)]. If $\lambda_1 \geq \ldots \geq \lambda_m$ are the eigenvalues of $\mathbf{W}$ s.t. $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{C})$ s.t. $n > m-1$, then the joint PDF of eigenvalues is*

$$f(\lambda_1, \ldots, \lambda_m) = K_{m,n} \left(\det \mathbf{C}\right)^{-n/2} \exp\left(-\frac{1}{2}\operatorname{Tr}\left(\mathbf{C}^{-1}\mathbf{W}\right)\right) \prod_{i=1}^{m} \lambda_i^{(n-m-1)/2} \prod_{i<j} (\lambda_i - \lambda_j) \tag{2}$$

where

$$K_{m,n} = \frac{\pi^{m^2/2}}{\Gamma_m\left(\frac{1}{2}m\right)\Gamma_m\left(\frac{1}{2}n\right)} \tag{3}$$

Lemma 3. *[WLRT08, Lemma 3.6]. Let $m, n \in \mathbb{N}$ s.t. $n \geq m$. Suppose $\mathbf{A} \in \mathbb{R}^{n\times m}$, then if $(\mathbf{A}^\top \mathbf{A})$ is invertible*

$$\left\| \left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \right\| = \frac{1}{\sigma_m(\mathbf{A})} \tag{4}$$

Lemma 4. *[Chi17, Lemma 1]. Draw a $m \times n$ matrix $\mathbf{G}$ s.t. the columns of $\mathbf{G}$ are sampled from $\mathcal{N}_m(\mathbf{0}, \mathbf{C})$ where the eigenvalues of $\mathbf{C}$ are represented as $\sigma_1 > \sigma_2 > \cdots > \sigma_m$. Let $\mathbf{W} \sim \mathcal{W}_m(n, \mathbf{C})$. The eigenvalue distribution is given as*

$$f(x_1, \ldots, x_n) = K_{\mathbf{C}} \left|\mathbf{E}(\mathbf{x}, \boldsymbol{\sigma})\right| \cdot \prod_{i=1}^{m-1} \prod_{j=i+1}^{m} (x_i - x_j) \prod_{i=1}^{n} x_i^{n-m} \tag{5}$$

where $\mathbf{E}(\mathbf{x}, \boldsymbol{\sigma}) = \left\{e^{-x_i/\sigma_j}\right\}_{i,j=1}^{m} = \begin{bmatrix} e^{-\frac{x_1}{\sigma_1}} & \ldots & e^{-\frac{x_1}{\sigma_m}} \\ \vdots & \ddots & \vdots \\ e^{-\frac{x_m}{\sigma_1}} & \ldots & e^{-\frac{x_m}{\sigma_m}} \end{bmatrix}$ *and*

$$K_{\mathbf{C}}^{-1} = \prod_{i=1}^{m-1} \prod_{j=i+1}^{m} (\sigma_i - \sigma_j) \prod_{i=1}^{m} \sigma_i^{n-m+1} (n-i)! \tag{6}$$

Theorem 5. *Consider a sequence of independent random matrices, $\mathbf{X}_k \in \mathbb{R}^{n\times m}$, such that*

$$\mathbb{E}\left[\mathbf{X}_k\right] = \mathbf{0} \tag{7}$$

$$\|\mathbf{X}_k\| \leq R \quad \forall k \tag{8}$$

$$\nu \triangleq \max\left\{ \left\| \mathbb{E} \sum_k \mathbf{X}_k \mathbf{X}_k^\top \right\|, \left\| \mathbb{E} \sum_k \mathbf{X}_k^\top \mathbf{X}_k \right\| \right\} \tag{9}$$

Then we have for all $t \geq 0$,

$$\mathbb{P}\left\{ \left\| \sum_k \mathbf{X}_k \right\| \geq t \right\} \leq (m+) \exp\left(\frac{-t^2/2}{\nu + Rt/3} \right) \tag{10}$$

With these lemmas we will go to proving the main results.

3.2 Main Results

Theorem 6. *Draw a $m \times m$ matrix $\mathbf{G}$ s.t. the columns of $\mathbf{G}$ are sampled from $\mathcal{N}_m(\mathbf{0}, \mathbf{C})$ where the eigenvalues of $\mathbf{C}$ are represented as $\sigma_1 > \sigma_2 > \cdots > \sigma_m$. Then*

$$\mathbb{E}\left\|\mathbf{G}^\dagger\right\| \leq \sqrt{\pi \sum_{k=1}^{m} \frac{1}{\sigma_k}} \tag{11}$$

Proof. We will first note

$$\left\|\mathbf{G}^\dagger\right\| \overset{\text{lem. 3}}{=} \frac{1}{\sigma_m(\mathbf{G})} = \frac{1}{\sqrt{\lambda_{\min}(\mathbf{G}\mathbf{G}^\top)}} \tag{12}$$

For $\mathbf{W}$ sampled from $\mathcal{W}_m(m, \mathbf{C})$. We will now derive the distribution for minimum eigenvalue of $\mathbf{W}$ similar to [NZYY08].

$$f_{\lambda_{\min}}(x_m) = \int_{x_2}^{\infty} \cdots \int_{x_{m-1}}^{\infty} K_{\mathbf{C}} \left|\mathbf{E}(\mathbf{x}, \boldsymbol{\sigma})\right| \cdot \prod_{i<j}^{m} (x_i - x_j) \prod_{i=1}^{m} x_j^{m-m} \prod_{i=1}^{m-1} dx_i \tag{13}$$

$$= K_{\mathbf{C}} \int_{x_2}^{\infty} \cdots \int_{x_{m-1}}^{\infty} \left|\mathbf{E}(\mathbf{x}, \boldsymbol{\sigma})\right| \cdot \prod_{i=1}^{m-2} \prod_{j=i+1}^{m-1} (x_i - x_j) \prod_{i=1}^{m-1} (x_i - x_m) \prod_{i=1}^{m-1} dx_i \tag{14}$$

$$\overset{\zeta_1}{=} e^{-\sum_{i=1}^{m} \frac{x_m}{\sigma_i}} \left(\int_{y_2}^{\infty} \cdots \int_{y_{m-1}}^{\infty} \sum_{i=1}^{m} (-1)^{i+m} K_{\mathbf{C}} \left|\mathbf{E}_i\left(\mathbf{x} - \mathbf{x}_m, \boldsymbol{\sigma}\right)\right| \prod_{i=1}^{m-2} \prod_{j=i+1}^{m-1} (y_i - y_j) \prod_{i-1}^{m-1} dy_i \right) \tag{15}$$

$$\overset{\zeta_2}{=} \Xi e^{-\sum_{i=1}^{m} \frac{x_m}{\sigma_i}} \tag{16}$$

(ζ_1) follows due to the properties of the determinant. (ζ_2) follows as the intergral expresion in Equation (15) no longer integrates over x_m and thus integrates to some constant we define as Ξ. Since the PDF must integrate to 1, we thus have,

$$f_{\lambda_{\min}}(x) = \left(\sum_{k=1}^{m} \frac{1}{\sigma_k} \right) e^{-x \sum_{k=1}^{m} \frac{1}{\sigma_k}} \tag{17}$$

The Expected Value follows from a simple integration.

$$\mathbb{E}\left\|\mathbf{G}^{\dagger}\right\| = \int_{0}^{\infty} \frac{1}{\sqrt{x}} e^{-x \sum_{k=1}^{m} \sigma_k^{-1}} dx \tag{18}$$

$$= \sqrt{\pi \sum_{k=1}^{m} \frac{1}{\sigma_k}} \operatorname{erf}\left(\sqrt{\pi \sum_{k=1}^{m} \frac{1}{\sigma_k}} \right) \leq \sqrt{\pi \sum_{k=1}^{m} \frac{1}{\sigma_k}} \tag{19}$$

In probability

$$\mathbb{P}\left\{ \mathbf{G}^{\dagger} > t \right\} = \int_{t}^{\infty} \frac{1}{\sqrt{t}} e^{-t \sum_{k=1}^{m} \sigma_k^{-1}} dt \tag{20}$$

The proof is complete. ■

In our next theorem, we will consider the matrix is rectangle and all the singular values of the covariance matrix are distinct.

Theorem 7. *Draw a $m \times n$ matrix $\mathbf{G}$ s.t. the columns of $\mathbf{G}$ are sampled from $\mathcal{N}_m(\mathbf{0}, \boldsymbol{\Sigma})$ where the eigenvalues of $\boldsymbol{\Sigma}$ are represented as $\sigma_1 > \sigma_2 > \cdots > \sigma_m$. Let $\mathbf{W} \sim \mathcal{W}_m(n, \boldsymbol{\Sigma})$. Then,*

$$\mathbb{E}\left\|\mathbf{G}^{\dagger}\right\| \leq n\sigma_{\min}\left(\boldsymbol{\Sigma}\right) - \sigma_{\max}\left(\boldsymbol{\Sigma}\right) \left(\sqrt{2\left(mK - 1\right) n \log(2m)} + \frac{1}{3} mK \log\left(2m\right) \right) \tag{21}$$

Proof. First, let us represent $\mathbf{W} = \sum_{i=1}^{n} \mathbf{x}\mathbf{x}^\top$ where $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Then we can lower $\sigma_{\min}(\mathbf{W})$,

$$\sigma_{\min}(\mathbf{W}) = \sigma_{\min}\left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^\top\right) \tag{22}$$

$$= \sigma_{\min}\left(n\mathbb{E}\left[\mathbf{x}\mathbf{x}^\top\right] + \sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^\top - \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i^\top\right]\right) \tag{23}$$

$$\geq n\sigma_{\min}\left(\mathbf{\Sigma}\right) - \sigma_{\max}\left(\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma}\right) \tag{24}$$

$$= n\sigma_{\min}\left(\mathbf{\Sigma}\right) - \sigma_{\max}\left(\sum_{i=1}^{n} \left(\mathbf{\Sigma}^{1/2}\mathbf{v}_i\right)\left(\mathbf{\Sigma}^{1/2}\mathbf{v}_i\right)^\top - \mathbf{\Sigma}\right) \tag{25}$$

$$= n\sigma_{\min}\left(\mathbf{\Sigma}\right) - \sigma_{\max}\left(\mathbf{\Sigma}^{1/2}\left(\sum_{i=1}^{n} \mathbf{v}_i\mathbf{v}_i^\top - \mathbf{I}\right)\mathbf{\Sigma}^{1/2}\right) \tag{26}$$

$$\geq n\sigma_{\min}\left(\mathbf{\Sigma}\right) - \sigma_{\max}\left(\mathbf{\Sigma}\right)\underbrace{\sigma_{\max}\left(\sum_{i=1}^{n} \mathbf{v}_i\mathbf{v}_i^\top - \mathbf{I}\right)}_{\text{A}} \tag{27}$$

$$\tag{28}$$

There has been significant theory in Random Matrix Theory and High Dimensional Probability analyzing Covariance Estimation, especially in the standard normal case, [T+15, Ver20, Rig15]. We will utilize the Matrix Bernstein inequality to upper bound A in probability and in expectation. To use Matrix Bernstein we need to upper bound $\mathbb{V}(\mathbf{v}_i\mathbf{v}_i^\top - \mathbf{I})$.

$$\mathbb{V}\left(\mathbf{v}\mathbf{v}^\top - \mathbf{I}\right) = \mathbb{E}\left[\left(\mathbf{v}\mathbf{v}^\top - \mathbf{I}\right)^\top\left(\mathbf{v}\mathbf{v}^\top - \mathbf{I}\right)\right] \tag{29}$$

$$= \mathbb{E}\left[\mathbf{v}\mathbf{v}^\top\mathbf{v}\mathbf{v}^\top - 2\mathbf{v}^\top + \mathbf{I}\right] \tag{30}$$

$$= \mathbb{E}\left[\mathbf{v}\mathbf{v}^\top\mathbf{v}\mathbf{v}^\top\right] - \mathbf{I} \tag{31}$$

$$= \mathbb{E}\left[\left\|\mathbf{v}\right\|^2\mathbf{v}\mathbf{v}^\top\right] - \mathbf{I} \tag{32}$$

$$= (mK - 1) \tag{33}$$

Then from Theorem 5, we have

$$\mathbb{E}\left[\text{A}\right] \leq \sqrt{2\left(mK - 1\right)n\log(2m)} + \frac{1}{3}mK\log\left(2d\right) \tag{34}$$

Therefore we have in expectation,

$$\mathbb{E}\left\|\mathbf{G}^\dagger\right\| \leq n\sigma_{\min}\left(\mathbf{\Sigma}\right) - \sigma_{\max}\left(\mathbf{\Sigma}\right)\left(\sqrt{2\left(mK - 1\right)n\log(2m)} + \frac{1}{3}mK\log\left(2m\right)\right) \tag{35}$$

4 Numerical Experiments

In Figure 1, we verify the results given in Theorem 6.
In Figure 2, we verify the results given in **??**.

5 Conclusions

In this paper, we derive novel upper bounds for the spectral norm of Gaussian matrices with columns sampled from a central correlated multivariate normal distribution with various distributions of the singular values of the covariance matrix.

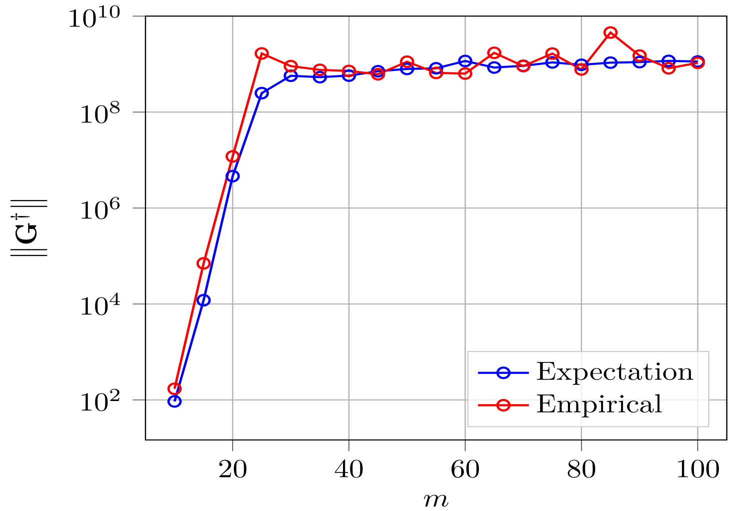

Figure 1: Comparing the expected norm upper bound on $\|\mathbf{G}^{\dagger}\|$ where $\mathbf{G} \in \mathbb{R}^{m \times m}$ and the columns of $\mathbf{G}$ are sampled from $\mathcal{N}_m(\mathbf{0}, \mathbf{K})$ with the average norm of $\mathbf{G}^{\dagger}$ over 100 samples. The expected norm is calculated with Proposition 6.

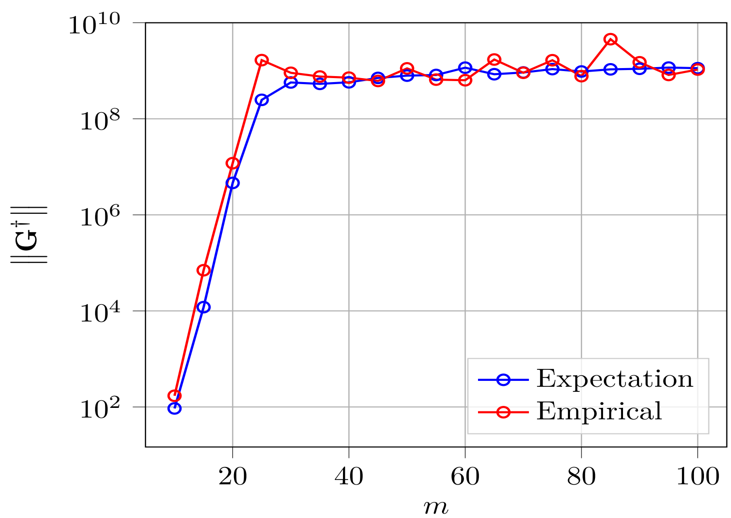

Figure 2: Comparing the expected norm upper bound on $\|\mathbf{G}^{\dagger}\|$ where $\mathbf{G} \in \mathbb{R}^{m \times m}$ and the columns of $\mathbf{G}$ are sampled from $\mathcal{N}_m(\mathbf{0}, \mathbf{K})$ with the average norm of $\mathbf{G}^{\dagger}$ over 100 samples. The expected norm is calculated with Proposition 6.

References

[BT22] Nicolas Boulle and Alex Townsend. A generalization of the randomized singular value decomposition. In *International Conference on Learning Representations*, 2022.

[CD05] Zizhong Chen and Jack J. Dongarra. Condition numbers of gaussian random matrices. *SIAM Journal on Matrix Analysis and Applications*, 27(3):603–620, 2005.

[Chi17] Marco Chiani. On the probability that all eigenvalues of gaussian, wishart, and double wishart random matrices lie within an interval, 2017.

[CWS09] Marco Chiani, Moe Z Win, and Hyundong Shin. Mimo networks: The effects of interference. *IEEE Transactions on information theory*, 56(1):336–349, 2009.

[HMT11] N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

[Jam64] Alan T. James. Distributions of matrix variates and latent roots derived from normal samples. *The Annals of Mathematical Statistics*, 35(2):475–501, 1964.

[NZYY08] Fangfang Niu, Haochuan Zhang, Hongwen Yang, and Dacheng Yang. Distribution of the smallest eigenvalue of complex central semi-correlated wishart matrices. In *2008 IEEE International Symposium on Information Theory*, pages 1788–1792, 2008.

[Rig15] Philippe Rigollet. 18. s997: High dimensional statistics. *Lecture Notes), Cambridge, MA, USA: MIT Open-CourseWare*, 2015.

[Rob55] Herbert Robbins. A remark on stirling's formula. *The American mathematical monthly*, 62(1):26–29, 1955.

[T+15] Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[Ver20] Roman Vershynin. High-dimensional probability. *University of California, Irvine*, 2020.

[WLRT08] Franco Woolfe, Edo Liberty, Vladimir Rokhlin, and Mark Tygert. A fast randomized algorithm for the approximation of matrices. *Applied and Computational Harmonic Analysis*, 25(3):335–366, 2008.