| GAFA Geometric And Functional Analysis

NON-ASYMPTOTIC RESULTS FOR SINGULAR VALUES OF GAUSSIAN MATRIX PRODUCTS

BORIS HANIN AND GRIGORIS PAOURIS

Abstract. This article provides a non-asymptotic analysis of the singular values (and Lyapunov exponents) of Gaussian matrix products in the regime where N, the number of terms in the product, is large and n, the size of the matrices, may be large or small and may depend on N. We obtain concentration estimates for sums of Lyapunov exponents, a quantitative rate for convergence of the empirical measure of the squared singular values to the uniform distribution on $[0, 1]$, and results on the joint normality of Lyapunov exponents when N is sufficiently large as a function of n. Our technique consists of non-asymptotic versions of the ergodic theory approach at $N = \infty$ due originally to Furstenberg and Kesten (Ann Math Stat 31(2):457–469, 1960) in the 1960s, which were then further developed by Newman (Commun Math Phys 103(1):121–126, 1986) and Isopi and Newman (Commun Math Phys 143(3):591–598, 1992) as well as by a number of other authors in the 1980s. Our key technical idea is that small ball probabilities for volumes of random projections gives a way to quantify convergence in the multiplicative ergodic theorem for random matrices.

1 Introduction

This article is about the spectral theory of random matrix products

$$X_{N,n} := A_N \cdots A_1, \tag{1.1}$$

where A_i are independent $n \times n$ matrices with independent real Gaussian entries $(A_i)_{\alpha\beta} \sim \mathcal{N}(0, 1/n)$ of mean zero and variance $1/n$. We are primarily interested in the situation when N is large and finite, while n may depend on N and may be either small or large. Our results concern the singular values of $X_{N,n}$:

$$s_1(X_{N,n}) \geq \cdots \geq s_n(X_{N,n}), \tag{1.2}$$

and can be summarized informally as follows:

1. We prove that as N, n tend to infinity *at any relative rate* the global distribution of the normalized squared singular values $\left\{ s_i(X_{N,n})^{2/N}, i = 1, \ldots, n \right\}$ converges

to the uniform distribution on $[0, 1]$ (see Section 1.2 and Theorem 1.2). Unlike previous results, we obtain quantitative concentration estimates valid for all N, n larger than a fixed constant. See also Section 1.3 for a heuristic explanation of why the uniform distribution appears in this context.

2. We prove that as long as N is sufficiently large as a function of n, the Lyapunov exponents

$$\lambda_i = \lambda_i(X_{N,n}) := \frac{1}{N} \log s_i(X_{N,n}) \tag{1.3}$$

of $X_{N,n}$ are approximately independent and Gaussian (see Theorem 1.3 in Section 1.4). Unlike previous results, our estimates simultaneously treat all the Lyapunov exponents and provide quantitative concentration estimates when N is large but finite even when n grows with N.

3. The statements listed above derived from our main technical result, Theorem 1.1, which gives quantitative deviation estimates on sums of Lyapunov exponents of $X_{N,n}$:

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=m}^{k} (\lambda_i - \mu_{n,i}) \right| \geq s \right) \leq c_1 e^{-c_2 nNs \min\{1, ng_{n,k}(s)\}}, \quad s \geq \frac{k}{nN} \log\left(\frac{en}{k} \right),$$

where $\mu_{n,i}$ is defined in (1.5) and $g_{n,k}(s)$ is a function defined in (1.4). It is known that (e.g. equations (1) and (7) in [New86]) $\mu_{n,i}$ is the almost sure limit of λ_i when $N \to \infty$.

In this article, we exclusively treat the case of A_i having iid real Gaussian entries. This simplifies a number of arguments, but we conjecture that similar results hold if we assume only that the distribution of the entries of A_i have finite fourth moments and bounded density. We leave this for future work.

1.1 Main Technical Result. Let us set some notation. Denote as in (1.3) by $\lambda_i = \lambda_i(X_{N,n})$ the Lyapunov exponents of $X_{N,n}$. Further, define for any $s > 0$

$$g_{n,k}(s) = \begin{cases} \min\left\{ 1, \frac{ns}{k} \right\}, & k \leq \frac{n}{2}, \\ \min\left\{ \delta_{n,k}, \frac{s}{\log 1/\delta_{n,k}} \right\}, & \frac{n}{2} < k \leq n, \end{cases} \tag{1.4}$$

where for $k \geq n/2$ we've set

$$\delta_{n,k} := \frac{n - k + 1}{n} \in \left[\frac{1}{n}, \frac{n-1}{n} \right].$$

Finally, write

$$\mu_{n,k} := \mathbb{E}\left[\frac{1}{2} \log\left(\frac{1}{n} \chi^2_{n-k+1} \right) \right] = \frac{1}{2}\left(\log\left(\frac{2}{n} \right) + \psi\left(\frac{n-k+1}{2} \right) \right), \tag{1.5}$$

where $\psi(z) = \frac{d}{dz} \log \Gamma(z)$ is the digamma function and χ^2_m is a chi-squared random variable with m degrees of freedom. Our main technical result is the following

Theorem 1.1 (Deviation Estimates for Sums of Lyapunov Exponents). *There exist universal constants $c_1, c_2, c_3 > 0$ with the following property. Fix $1 \leq m \leq k \leq n$ as well as $N \geq 1$. Then,*

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=m}^{k} (\lambda_i - \mu_{n,i}) \right| \geq s \right) \leq c_2 \exp\left(-c_3 n N s \min\{1, n g_{n,k}(s)\} \right), \qquad (1.6)$$

provided $s \geq c_1 \frac{k}{nN} \log(en/k)$.

Theorem 1.1 holds for every $n, N \geq 1$ and reveals a great deal about the singular values and Lyapunov exponents of $X_{N,n}$. For instance, in the bulk (i.e. when k is comparable to n), the restriction on s in (1.6) reduces simply to $s > C/N$, giving information about $X_{N,n}$ as soon as N is large, regardless of n. This turns out to be enough to prove Theorem 1.2, given in Section 1.2 below, which states that the squared singular values of $X_{N,n}$ approximate the uniform distribution on $[0, 1]$ when N, n tend to infinity at any relative rate.

Theorem 1.1 also gives precise information about the top Lyapunov exponents of $X_{N,n}$. Indeed, taking k to be fixed in (1.6) gives non-trivial information on $\lambda_1, \ldots, \lambda_k$ as soon as $N \gg \log(n)$. Further, note that standard estimates for the digamma function ψ yield

$$\mu_{n,k} = \log\left(1 - \frac{k-1}{n} \right) - \frac{1}{n-k+1} + O\left(\frac{1}{(n-k+1)^2} \right). \qquad (1.7)$$

This shows that the difference between the means $\mu_{n,1}$ and $\mu_{n,2}$ of λ_1 and λ_2 is on the order of $1/n$. As soon as $N \gg n \log(n)$, we may apply (1.6) with $s \ll \lambda_1 - \lambda_2$ to conclude that

$$\frac{s_1(X_{N,n})}{s_2(X_{N,n})} = e^{N(\lambda_1 - \lambda_2)} \geq e^{cN/n} \quad \text{with high probability.}$$

Hence, we find that $X_{N,n}$ begins to have a large spectral gap in the "near ergodic" regime $N \gg n \log(n)$. In fact, in Theorem 1.3, we prove that in this regime $\lambda_1, \ldots, \lambda_k$ are also approximately independent Gaussians. We refer the reader to Section 1.4 for the details.

A notable aspect of Theorem 1.1 is that it applies to any finite $n, N \geq 1$, allowing us to "interpolate" between the ergodic $N \gg n$ and free $n \gg N$ regimes. To explain this point, note that matrix products of the form (1.1) have been studied primarily in two setting. The first, which we refer to as the free probability regime occurs when N is fixed and $n \to \infty$. This is a kind of maximum entropy regime in which the global distribution of singular values can be characterized in terms of maximizing the non-commutative entropy (cf eg [AG97, BBCC11]). The second, which we call the ergodic regime, occurs when n is fixed and $N \to \infty$. This is a kind of minimal entropy regime in which the Lyapunov exponents (and singular values of $X_{N,n}$) tend to almost sure limits.

In both the ergodic and the free regimes, it is often difficult to obtain finite size corrections. Theorem 1.1 supplies such information. Moreover, since the ergodic and free regimes are usually treated by rather different means, it is unclear which techniques can give information that can interpolate between them. Our approach extends the ergodic techniques pioneered by Furstenburg–Kesten [FK60], further developed in connection to random Schrödinger operators by Carmona [Car82] and Le Page [LP82] (cf also [BLR85]), and applied in a very similar context as ours by Newman [New86] and Isopi–Newman [IN92]. It is therefore not surprising that in all of our results, we need N to be in some sense large.

Although we do not take this approach in the present article, it is also natural to study spectra of random matrix products by adapting techniques originally developed to treat the case when $N = 1$. Indeed, in this setting, there has been considerable effort to obtain non-asymptotic analogs of classical random matrix theory results when $n = \infty$ [Rud14, Ver12, Rud17], culminating in the resolution of a number of long-standing open problems [RV08, RV09, ALPTJ10, Tik20]. More recently, several groups of authors [HW20, HNWTW20, KMS20] have started to extend techniques for obtaining concentration for random matrices tailored (see [Tro15]) to the small N regime for understanding the kinds of matrix products considered in this article. From this point of view, our article takes a complementary approach, finding extensions of techniques originally coming from the ergodic theory used to analyze the case when $N = \infty$.

1.2 Convergence of Squared Singular Values to the Uniform Distribution.
Prior work [IN92, Kar08, Tuc10, GS18, LWW18, Ahn19] shows that in a variety of settings where $n, N \to \infty$, the global distribution of singular values of $X_{N,n}^{1/N}$ converges to the so-called triangle law after proper normalization. Informally, this means

$$\lim_{N,n\to\infty} \frac{1}{n}\#\left\{ j \le n \mid s_i^{1/N}(X_{N,n}) \le t \right\} = \int_{-\infty}^{t} 2s\mathbf{1}_{\{s\in[0,1]\}}ds =: TL(t). \quad (1.8)$$

The graph of the density $2s\mathbf{1}_{\{s\in[0,1]\}}$ of TL has the shape of a triangle, giving the distribution its name. With the exception of the articles [LWW18, GS18], which obtain much more precise information for products of *complex* Gaussian matrices and the article [Ahn19] concerning β-Jacobi products as well as the real Gaussian case, the majority of prior results about (1.8) (e.g. [IN92, Kar08, Tuc10]) do not allow n, N to tend to infinity simultaneously. Moreover, all prior results we are aware of do not give quantitative rates of convergence. Theorem 1.2 provides both for the real Gaussian case we consider here. To state it, we note that if a random variable T is distributed according to the triangle law, then T^2 is uniformly distributed on $[0, 1]$, i.e. has the following cumulative distribution function:

$$\mathcal{U}(t) := \int_{-\infty}^{t} \mathbf{1}_{[0,1]}(t)dt.$$

Theorem 1.2 (Global Convergence to Triangle Law). *There exist universal constants* $c_1, c_2, c_3, c_4 > 0$ *with the following property. For all* $\varepsilon \in (0, c_1)$, *if* $N > c_2/\varepsilon^2$ *and* $n > c_3 \log(1/\varepsilon)/\varepsilon$, *then the probability*

$$\mathbb{P}\left(\sup_{t \in \mathbb{R}} \left| \frac{1}{n} \# \left\{ 1 \le i \le n \mid s_i^{2/N}(X_{N,n}) \le t \right\} - \mathcal{U}(t) \right| \ge \varepsilon \right)$$

that the cumulative distribution for the squared singular values of $X_{N,n}$ *deviates from the uniform distribution by more than* ε *is bounded above by*

$$4 \exp\left[-c_4 n N \varepsilon^2 \min\left\{ 1, n g_{n,k}(\varepsilon^2) \right\} \right].$$

In the next section we use the circular law (1.12) for the (complex) *eigenvalues* of $X_{N,n}^{1/N}$ to give an intuitive but heuristic explanation for why the uniform distribution (or equivalently the triangle law) should appear as the limiting distribution of singular values on $X_{N,n}$. Before doing so, we briefly discuss the dependence of Theorem 1.2 on N, n, starting with the former. For fixed N, consider an iid random sequence $\{X_{N,n}\}_{n=1}^{\infty}$ with the product measure. Taking $\varepsilon = 2(c_2/N)^{1/2} =: CN^{-1/2}$, Theorem 1.2 shows that

$$\mathbb{P}\left(\sup_{t \in \mathbb{R}} \left| \frac{1}{n} \# \left\{ 1 \le i \le n \mid s_i^{2/N}(X_{N,n}) \le t \right\} - \mathcal{U}(t) \right| \ge \frac{C}{\sqrt{N}} \right) \le 4 e^{-cn}, \quad c > 0.$$

Thus, by Borel–Cantelli, we find that

$$\sup_{t \in \mathbb{R}} \left| \lim_{n \to \infty} \frac{1}{n} \# \left\{ 1 \le i \le n \mid s_i^{2/N}(X_{N,n}) \le t \right\} - \mathcal{U}(t) \right| \le \frac{C}{\sqrt{N}} \quad \text{with probability 1,}$$

$$(1.9)$$

where by $\lim_{n \to \infty} a_n$ we mean any limit point of the sequence a_n. This $1/\sqrt{N}$ can be seen as a Berry–Esseen-type estimate. To make this precise, consider

$$\rho_{N,\infty} := \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{s_i(X_{N,n})^{1/N}},$$

the large matrix limit for the empirical distribution of normalized singular values for $X_{N,n}$. It is known [BBCC11, Thm 6.1] that

$$\rho_{N,\infty} = \mathrm{qc}^{\boxtimes N}, \qquad \mathrm{qc}(x) := \frac{1}{2\pi} \sqrt{x(2-x)} \mathbf{1}_{\{[0,2]\}}(x),$$

where qc is the quarter circle law and $\boxtimes$ is the multiplicative free convolution. Kargin [Kar08] and Tucci [Tuc10] show that, consistent with Theorem 1.2,

$$\lim_{N \to \infty} \rho_{N,\infty} = \mathrm{TL}.$$

As far as we know, the optimal rate of convergence for such repeated multiplicative free convolution is unknown. However, from this point of view, (1.9) shows that the rate of convergence is at least as fast as in the usual central limit theorem.

To understand the dependence of Theorem 1.2 on n, we send N to infinity in Theorem 1.2 to obtain as before that there is $C > 0$ so that

$$\sup_{t \in \mathbb{R}} \left| \lim_{N \to \infty} \frac{1}{n} \# \left\{ 1 \le i \le n \mid s_i^{2/N}(X_{N,n}) \le t \right\} - \mathcal{U}(t) \right| \le \frac{C \log(n)}{n} \qquad \text{with probability 1,}$$

(1.10)

Apart from the $\log(n)$, this estimate is sharp. Indeed, the empirical distribution

$$\rho_{\infty,n} := \lim_{N \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{s_i(X_{N,n})^{1/N}}$$

of singular values in the large number of matrices limit exists almost surely and is deterministic by the Multiplicative Ergodic Theorem. Among other things, Theorem 1.3 below computes, in agreement with the early work of Newman [New86], this limit in our Gaussian case. The subsequent work of Isopi–Newman [IN92] showed that, under minimal assumptions,

$$\lim_{n \to \infty} \rho_{\infty,n} = \text{TL}.$$

This of course agrees with Theorem 1.2, which via (1.10) provides a natural rate of convergence. This rate is optimal, perhaps up to the $\log(n)$, because the spacing of the atoms in $\rho_{\infty,n}$ is approximately $1/n$. Hence, the distance between $\rho_{\infty,n}$ and triangle law TL, which is a continuous distribution, is bounded below by a constant times $1/n$.

1.3 Why the Uniform Distribution in Theorem 1.2?   A number of articles [New86, IN92, Kar08, Tuc10, LWW18, GS18] show as in (1.8) that in the limit where n, N tend to infinity, the singular values $s_i(X_{N,n})^{1/N}$ (or for similar matrix products) converge to the triangle law (and hence their squares converge to the uniform distribution on $[0,1]$). These articles use a variety of techniques ranging from free probability to ergodic theory and special functions. Why does the uniform distribution appear? The purpose of this section to give an intuitive explanation for this phenomenon. After writing an initial draft of this article, we learned from G. Akemann that an explanation similar to the one below can be found on pages 3, 4 in [ABK14]. We also refer the reader to the work of Kieberg–Kösters [KK16] about an exact relation between eigenvalues and singular values for products of complex Ginibre matrices.

Since $X_{N,n}$ is not normal with probability 1, its spectral properties are captured not only its singular values but also by its eigenvalues

$$|\zeta_1(X_{N,n})| \ge \cdots \ge |\zeta_n(X_{N,n})|, \qquad \zeta_i(X_{N,n}) \in \mathbb{C}. \tag{1.11}$$

Our argument for why the triangle law appears in Theorem 1.2 relates the singular values and eigenvalues of $X_{N,n}$ and consists of two observations. First, consider

the (complex) eigenvalues of $X_{N,n}^{1/N}$ as defined in (1.11). It is shown in [GT10, OS11] that for each fixed N the empirical distribution of the eigenvalues of $X_{N,n}^{1/N}$ converges weakly almost surely to the uniform measure on the unit disk in $\mathbb{C}$. This result is often called the circular law. Informally, it reads

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{\zeta_i^{1/N}}(z) = \frac{1}{\pi}\mathbf{1}_{\{|z|\leq 1\}}, \qquad z \in \mathbb{C}. \tag{1.12}$$

Precise results on the rate of convergence can be found in [GJ18, Jal19] and local limit theorems are obtained in [Nem17]. Since in polar coordinates (r,θ) the radial part of the uniform measure on the unit disk is $2rdr$, a corollary of the circular law is that

For N fixed, as $n \to \infty$, squared eigenvalue moduli $|\zeta_i|^{2/N}$ of $X_{N,n}^{1/N}$ converge to $\mathcal{U}$. (1.13)

Thus, the uniform distribution $\mathcal{U}$ appears naturally as the distribution of the squared moduli of eigenvalues of $X_{N,n}^{1/N}$ for every N! On the other hand, it has been proved that for any fixed finite n [Red16, Red19] that when N is large

$$\forall i = 1,\ldots,n \qquad |\zeta_i|^{1/N} \approx s_i^{1/N}.$$

Thus, we extract another piece of intuition:

For n fixed, as $N \to \infty$, eigenvalue moduli and singular values of $X_{N,n}^{1/N}$ coincide. (1.14)

Putting together (1.13) and (1.14), we conclude heuristically that if both n, N tend to infinity then the distribution of the singular values $s_i^{1/N}$ should converge to the triangle law. This is precisely the content of Theorem 1.2. While the heuristic for (1.14) was previously established only when n is fixed, we believe it can also be proved in the regime where n is allowed to grow with N but leave this for future work.

1.4 Distribution of Lyapunov Exponents in the Near Ergodic Regime.
In addition to studying the global distribution of singular values of $X_{N,n}$, we also obtain in Theorem 1.3 precise estimates for the joint distribution of the Lyapunov exponents

$$\lambda_i = \lambda_i(X_{N,n}) = \frac{1}{N}\log s_i(X_{N,n}) \tag{1.15}$$

of $X_{N,n}$ in the regime when $N \gg n\log^2(n)$. To state it, we need some notation. Recall first that for each $1 \leq k \leq n$ we had set

$$\mu_{n,k} = \mathbb{E}\left[\frac{1}{2}\log\left(\frac{1}{n}\chi_{n-k+1}^2\right)\right] = \frac{1}{2}\left(\log\left(\frac{2}{n}\right) + \psi\left(\frac{n-k+1}{2}\right)\right), \tag{1.16}$$

where $\psi(z) = \frac{d}{dz} \log \Gamma(z)$ is the digamma function and χ_m^2 is a chi-squared random variable with m degrees of freedom. We also recall the estimate (1.7):

$$\mu_{n,k} = \log\left(1 - \frac{k-1}{n}\right) - \frac{1}{n-k+1} + O\left(\frac{1}{(n-k+1)^2}\right).$$

The quantity $\mu_{n,k}$ already appears in [New86, IN92] as the mean of λ_k when $N \to \infty$. We futher define

$$\sigma_{n,k}^2 := \mathrm{Var}\left[\frac{1}{2}\log\left(\frac{1}{n}\chi_{n-k+1}^2\right)\right] = \psi'\left(\frac{n-k+1}{2}\right) = \frac{1}{2(n-k+1)} + O\left(\frac{1}{(n-k+1)^2}\right),$$
(1.17)

and set

$$\mu_{n,\leq k} := (\mu_{n,1}, \ldots, \mu_{n,k}), \qquad \sigma_{n,\leq k}^2 := (\sigma_{n,1}^2, \ldots, \sigma_{n,k}^2).$$
(1.18)

Finally, we will consider for two $\mathbb{R}^k$-valued random variables X, Y the following high-dimensional generalization of the usual Kolmogorov–Smirnov distance:

$$d(X,Y) := \sup_{C \in \mathcal{C}_k} |\mathbb{P}(X \in C) - \mathbb{P}(Y \in C)|,$$
(1.19)

where $\mathcal{C}_k$ is the collection of all convex subsets of $\mathbb{R}^k$.

Theorem 1.3 (Asymptotic Normality of Lyapunov Exponents). *There exist constants $C_1, C_2 > 0$ with the following property. Suppose $X_{N,n}$ is as in (1.1), fix $1 \leq k \leq n$, and write*

$$\Lambda_k = (\lambda_1, \ldots, \lambda_k)$$

for the vector of the top k Lyapunov exponents of $X_{N,n}$. Then, $\lambda_1, \ldots, \lambda_k$ are approximately independent and Gaussian when N is sufficiently large as a function of k, n:

$$d\left(\Lambda_k, \mathcal{N}\left(\mu_{n,\leq k}, \frac{1}{N}\mathrm{Diag}\left(\sigma_{n,\leq k}^2\right)\right)\right) \leq C_2 \left(\frac{k^{7/2} n \log^2(n) \log^2(N/n)}{N}\right)^{1/2}.$$
(1.20)

Here $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian with mean μ and co-variance Σ and for any $v = (v_1, \ldots, v_k) \in \mathbb{R}^k$ we have written $\mathrm{Diag}(v)$ for the diagonal matrix with $\mathrm{Diag}(v)_{ii} = v_i$.

REMARK 1.4. The arguments in [AB12, AB12, ABK14, ABK19] strongly suggest (see Section 2.2) that for k fixed and independent of n, a necessarily and sufficient condition for $\lambda_1, \ldots, \lambda_k$ to be close to independent and Gaussian is $N \gg n$. Thus, the $\log^2(n) \log^2(N/n)$ in (1.20) is likely sub-optimal. It is not clear whether the power $k^{7/2}$ can be improved.

For $k \geq 1$ fixed independent of n, N, Theorem 1.3 shows that the top k Lyapunov exponents of $X_{N,n}$ are close to independent Gaussian as soon as $N \gg n \log^2(n) \log^2(N/n)$. This is a significant refinement of the result in [Car82] (see also Theorem 5.4 in [BLR85]), which states that when n is fixed λ_1 is asymptotically normal. It also refines the recent result of Reddy [Red19, Theorem 11], which holds only for fixed finite n and does not give estimates at finite N. The advantage of Theorem 1.3 is that it treats simultaneously any number of Lyapunov exponents and gives a rate of convergence. For example, taking $k = n$, we find that if $N \gg n^{9/2} \log^2(n) \log^2(N/n)$, then *all* Lyapunov exponents of $X_{N,n}$ are approximately independent Gaussians. However, results in articles such as [Car82] are for matrix products $A_N \cdots A_1$ in which the entries of A_i have mean zero, variance $1/n$ and satisfy some mild regularity assumptions, whereas our results hold only for the Gaussian case. We conjecture that Theorem 1.3 holds in this more general setting as well but leave this to future work.

2 Prior Work and Intuitions

The purpose of this section is to give an exposition of prior work and provide several intuitions for thinking about the matrix products $X_{N,n}$, especially about the differences between the near-ergodic $N \gg n$ and the near-free $n \gg N$ regimes. We do this by first giving in Section 2.1 a basic intuition from dynamical systems, which suggests that one can think of N as a time variable and n as a system size. This intuition dovetails with the multiplicative ergodic theorem. We proceed in Section 2.2 to explain an exact correspondence derived in [AB12, ABKN14, ABK14, ABK19] at a physical level of rigor in which n/N plays the role of a time parameter for the evolution of the n singular values of $X_{N,n}$. This helps to explain why even simple linear statistics behave differently depending on the relative size of n, N.

2.1 $X_{N,n}$ at Fixed n as a Dynamical System.   One way to intuitively think of $X_{N,n} = A_N \cdots A_1$ is as defining the time 0 to time N map for a dynamical system in which the time one dynamics are very chaotic and are modelled as multiplication by an iid random matrix. In this analogy, N takes on the role of a time parameter, whereas n denotes the system size. Since large systems take longer to come to equilibrium, we should expect that N and n are "in tension." If we fix n and let N tend to infinity, then the size of the long time image $||X_{N,n}u||$ of an unit length input $u \in \mathbb{R}^n$ satisfies a pointwise ergodic theorem:

$$\lim_{N \to \infty} \frac{1}{N} \log ||X_{N,n}u|| = E, \tag{2.1}$$

where E is a constant (independent of u) depending on the measure μ according to which the entries of the matrices A_i making up the matrix product $X_{N,n}$ are distributed. This can be proved in a variety of ways (e.g. Corollary 3.2 in [CN84]).

In fact, much more is true. It was shown by Kesten–Furstenberg in [FK60], that this statement tolerates taking a supremum over u:

$$\lim_{N \to \infty} \lambda_1(X_{N,n}) = E \qquad \text{almost surely.}$$

Later, in his seminal work [Ose68] Oseledets proved the multiplicative ergodic theorem. In the context of iid products of N matrices of size $n \times n$, it says that under some mild conditions on μ if n is fixed, then the full list of Lyapunov exponents $\lambda_1(X_{N,n}), \ldots, \lambda_n(X_{N,n})$ converges almost surely to a deterministic limit. We refer the reader to [Fil19] for a review of the vast literature on this subject and to [BLR85] for an exposition specifically about matrix products.

Determining the values of the limiting Lyapunov exponents in the multiplicative ergodic theorem is in general quite difficult and has applications to Anderson localization for random Schrödinger operators [BLR85, Dam11].

Moreover, the work of LePage [LP82] as well as subsequent analysis [Car82, BLR85] showed that the top Lyapunov exponent of matrix products such as $X_{N,n}$ (not necessarily Gaussian) is asymptotically normal in the sense that there exist $a_n, b_{N,n} \in \mathbb{R}$ so that

$$b_{N,n} \left(\lambda_1(X_{N,n}) - a_n \right) \quad \overset{d}{\longrightarrow} \quad \mathcal{N}(0,1),$$

where the d indicates that the convergence is in the sense of distribution. As far as we are aware, all known mathematical proofs of asymptotic normality results hold only for finite fixed n, for the top Lyapunov exponent λ_1 and do not include quantitative rates of convergence. For the real Gaussian case we study, our Theorem 1.3 overcomes these deficiencies. However, at the physical level of rigor, we refer the reader to the excellent articles [AB12, ABKN14, ABK14, AI15, ABK19] that derive in the case of complex Gaussian matrix products asymptotic normality and much more for the top Lyapunov exponents (cf Section 2.2).

While the preceding discussion concerned matrix products with any entry distribution μ with mean 0 and variance $1/n$, the Gaussian $\mu = \mathcal{N}(0, 1/n)$ considered in this article leads to some significant simplifications. For instance, Newman [New86] computed the exact expression, which can be written in terms of the digamma function, for the limiting Lyapunov exponents. Similarly, (2.1) is a simple fact in this case since $\frac{1}{N} \log ||X_{N,n}u||$ turns out to be sum of iid random variables (see Lemma 9.5). These simplifications stem from the fact that the distribution of each matrix A_i is left and right-invariant under multiplication by an orthogonal matrix.

2.2 n/N as a Time Parameter in an Interacting Particle System.   In the regime where n/N is bounded away from 0 and ∞ as $n, N \to \infty$, even the behavior of an innocuous seeming log-linear statistic depends very much on the ratio of n and N. Informally,

$$\log ||X_{N,n}u|| \ \approx \ \mathcal{N} \left(-\frac{N}{4n}, \frac{N}{4n} \right) + O \left(\frac{N}{n^2} \right), \qquad u \in \mathbb{R}^n, \ ||u|| = 1, \quad (2.2)$$

where $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian with mean μ and covariance Σ. As mentioned above, in the Gaussian case we consider in this article, this approximate normality is easy to see since $\log ||X_{N,n}u||$ is a sum of iid variables (see Lemma 9.5). Precise versions of (2.2) also hold true when the matrices A_i in the definition (1.1) of $X_{N,n}$ have symmetric but non-Gaussian entries (see Theorem 1 in [HN19]). It is interesting to compare the almost sure convergence to a constant in (2.1) (note that $1/N$ normalization) with the asymptotic normality in (2.2).

The relation (2.2) already suggests that $t = n/N$ is an important parameter for interpolating between the ergodic regime, defined by $t = 0$ and the asymptotically free regime, in which $t = \infty$. A number of remarkable articles [AB12, ABKN14, ABK14] and especially [ABK19] establish a correspondence between t and the time parameter in the stochastic evolution of an interacting particle system. This correspondence between singular values for products of complex Ginibre matrices and DBM appears to be initially due to Maurice Duits.

The particles in question are the limiting Lyapunov exponents λ_i of $X_{N,n}$. When $t = 0$, they are approximately uniformly spaced (see Theorem 1.3) and are interpreted as an initial condition for Dyson Brownian motion (DBM)

$$d\lambda_i = dB_i + \sum_{j \neq i} \frac{dt}{\lambda_i - \lambda_j}, \qquad i = 1, \ldots, n \tag{2.3}$$

the dynamics induced on the spectrum of a matrix by allowing each entry to evolve for time t under and independent Brownian motion [Dys62]. The surprising observation is that, at least in the bulk of the spectrum (i.e. λ_k with k proportional to n) the joint distribution of the Lyapunov exponents of $X_{N,n}$ satisfies (2.3) at time t with an equally spaced initial condition in the limit when $n/N = t$ and $n, N \to \infty$.

The idea of the derivations in [AB12, ABKN14, ABK14, ABK19] is to use that when $X_{N,n}$ is a product of complex Ginibre matrices, the joint distribution of all of its singular values, at any finite n, N, is given by a determinantal point process. One may then study the scaling limit of the corresponding determinantal kernel at any fixed $t = n/N$. This kernel coincides with the solution to DBM from equally spaced initial conditions, which is also determinantal [Joh04].

A rigorous analysis of the determinantal kernel for the joint distribution of singular values for products of complex Gaussian matrices was undertaken in a variety of articles [For13, For14, FL16, LWZ16, LWW18]. In particular, [LWZ16] shows that when N is arbitrary but fixed and $n \to \infty$, the determinantal kernel for singular values in products of N iid complex Gaussian matrices of size $n \times n$ converges to the familiar sine and Airy kernels that arise in the local spectral statistics of large GUE matrices in the bulk and edge, respectively. This agrees with the prediction from [ABK19]. Indeed, in this regime, the time parameter $t = n/N$ is infinite and the limiting distribution of DBM is that of the eigenvalues for a large GUE matrix. Moreover, [LWW18] rigorously obtained an expression for the limiting determinantal kernel when $t = n/N$ is arbitrary in the context of products of complex Ginibre

matrices.We refer the reader also to the subsequent article of Liu–Wang [LW19] that performs a similar analysis for the *eigenvalues* in the same setting.

Also in the regime where n/N is fixed while $n, N \to \infty$, we refer the reader to Gorin–Sun [GS18]. This article shows that the fluctuations of the singular values of $X_{N,n}$ around the triangle law always converge to a Gaussian field. We also refer the reader to [Ahn19], which obtains a CLT for linear statistics of top singular values when n/N is fixed and finite.

3 Idea of Proof: Reduction to Small Ball Estimates

Before turning to the formal proofs of Theorems 1.2 and 1.3, we give a brief overview of our approach, which begins with the following representation (cf e.g. [New86, IN92]) for sums of Lyapunov exponents from Lemma 5.1 (see Section 5):

$$\lambda_1 + \cdots + \lambda_k = \sup_{\Theta \in \mathrm{Fr}_{n,k}} \frac{1}{N} \log \|X_{N,n}(\Theta)\|. \tag{3.1}$$

In the previous line, we've denoted by $\mathrm{Fr}_{n,k}$ the collection of all orthonormal k-frames in $\mathbb{R}^n$ (i.e. collections of k orthonormal vectors $v_1, \ldots, v_k$). We have also set

$$X_{N,n}(\Theta) = X_{N,n}\theta_1 \wedge \cdots \wedge X_{N,n}\theta_k, \qquad \Theta = (\theta_1, \ldots, \theta_k) \in \mathrm{Fr}_{n,k},$$

where we recall that for $a, b \in \mathbb{R}^n$ $a_1 \wedge \cdots \wedge a_k$ is the anti-symmetrization of $a_1 \otimes \cdots \otimes a_k$. We refer the reader to Section 5 for more background and to the start of Section 7 in [Spi70] and to Section 2.6.1 in [Tao10] for more on wedge products. As pointed out in [IN92], information about the sums $\lambda_1 + \cdots + \lambda_k$ can easily be translated into the information about their cumulative distribution function, ultimately resulting in Theorem 1.2. Similarly, the vector of the top k Lyapunov exponents considered in Theorem 1.3 can be obtained by an affine transformation of the vector of partial sums $\lambda_1, \lambda_1 + \lambda_2, \ldots, \lambda_1 + \cdots + \lambda_k$. Thus, the focus of our proofs is to obtain precise concentration estimates for the expression on the right hand side of (3.1). An important idea for analyzing (3.1), which goes back to the work of Furstenburg–Kesten [FK60] is that when N is large, one can almost drop the supremum:

$$\lim_{N \to \infty} \left| \frac{1}{N} \log \|X_{N,n}(\Theta)\| - \sup_{\Theta' \in \mathrm{Fr}_{n,k}} \frac{1}{N} \log \left| \|X_{N,n}(\Theta')\| \right| \right| = 0 \tag{3.2}$$

for any fixed $\Theta \in \mathrm{Fr}_{n,k}$. As explained below this is plausible since the ratio $s_k(X_{N,n})/s_{k+1}(X_{N,n})$ of the kth and $(k + 1)^{st}$ singular values grows exponentially with N, causing the wedge product $X_{N,n}(\Theta)$ to align almost entirely with the wedge product of the top k singular vectors of $X_{N,n}$ for almost every Θ.

The "pointwise" quantity $\frac{1}{N} \log \|X_{N,n}(\Theta)\|$ is a sum of iid random variables (Lemma 9.5) and can be analyzed using a result of Łatała [Lat97] (see Theorem 6.1). It then remains to obtain quantitative versions of (3.1) valid for large but finite

N, n. One possible approach is via energy-entropy estimates using ε-nets on $\mathrm{Fr}_{n,k}$. However, while this gives some results, this approach is suboptimal for large N. The reason that ε-nets fail is that, due to the N^{-1} normalization,

$$N \ \text{large} \quad \Rightarrow \quad \mathrm{Var}_\Theta \left[\frac{1}{N} \log ||X_{N,n}(\Theta)|| \right] \ll \mathrm{Var}_{X_{N,n}} \left[\frac{1}{N} \log ||X_{N,n}(\Theta)|| \right]$$

by which we mean that the variance of $\frac{1}{N} \log ||X_{N,n}(\Theta)||$ over $\Theta \in \mathrm{Fr}_{n,k}$ for a typical realization of $X_{N,n}$ is much smaller than its variance over the randomness in $X_{N,n}$ for any fixed Θ, causing the optimal net to have constant cardinality.

   The main technical novelty of our proofs is that we quantify (3.2) not through net arguments but rather via small ball probabilities for volumes of random projections, which are already known (cf Proposition 8.3). The key result is the following:

PROPOSITION 3.1. *For any $\varepsilon \in (0,1)$ and any $\Theta \in \mathrm{Fr}_{n,k}$ we have*

$$\mathbb{P}\left(\left| \frac{1}{N} \log ||X_{N,n}\Theta|| - \sup_{\Theta' \in \mathrm{Fr}_{n,k}} \frac{1}{N} \log ||X_{N,n}\Theta'|| \right| \geq \frac{1}{N} \log \left(\frac{1}{\varepsilon} \right) \right) \leq \mathbb{P}\left(||P_F(\Theta)|| \leq \varepsilon \right),$$

where F is a Haar distributed k-dimensional subspace of $\mathbb{R}^n$ and

$$P_F(\Theta) = P_F\theta_1 \wedge \cdots \wedge P_F\theta_k$$

with P_F denoting the orthogonal onto F.

For the proof of Proposition 3.1 see Lemma 8.2. The appearance of small ball probabilities is natural, although perhaps somewhat unexpected. Let us briefly describe why in the simplest case of $k = 1$. Denote by $v^{(i)}$ the right eigenvector of $X_{N,n}$ corresponding to the singular value $s_i(X_{N,n})$. For any $\theta \in S^{n-1}$, we may write

$$||X_{N,n}\theta||^2 = \sum_{i=1}^n \left| \left\langle X_{N,n}\theta, v^{(i)} \right\rangle \right|^2.$$

When $N \gg n$, the matrix $X_{N,n}$ is highly degenerate in the sense that there exists a universal constant $C > 0$ so that

$$\frac{s_1(X_{N,n})}{s_2(X_{N,n})} \geq e^{CN/n}.$$

This is easy to see intuitively since in this regime $\lambda_1 - \lambda_2 \approx \frac{1}{n}$ (cf Theorem 1.3). Hence,

$$||X_{N,n}\theta||^2 \approx \left| \left\langle X_{N,n}\theta, v^{(1)} \right\rangle \right|^2 \tag{3.3}$$

unless $\left| \langle \theta, v^{(1)} \rangle \right|$ is unusually small. In fact, for all θ

$$0 \geq \frac{1}{N} \log ||X_{N,n}\theta|| - \lambda_1 = \frac{1}{2N} \log \left(\frac{||X_{N,n}\theta||^2}{s_1^2(X_{N,n})} \right) \geq \frac{1}{N} \log \left| \langle \theta, v^{(1)} \rangle \right|.$$

This lower bound is essentially sharp by (3.3) unless θ has small overlap with $v^{(1)}$, an event whose probability is controlled precisely by a small ball estimate.

4 Organization of the Rest of the Article

The rest of this article is structured as follows. First, in Section 5 we collect some well-known results on the relation between the exterior algebra of $\mathbb{R}^n$ and the singular values of any linear map $A : \mathbb{R}^n \to \mathbb{R}^n$. We also record several elementary observations (Lemmas 5.3 and 5.4) about polar decompositions and Haar measures on orthonormal frames and their flags. We will use this formalism throughout our proofs.

Next, in Section 6 recalls two kinds of results. The first, Theorem 6.1, is a result of Łatała [Lat97] that gives precise information on moments (and hence tail behavior) for sums of independent centered random variables. The second is a set of results related to the multivariate central limit theorem (Theorem 6.4) and the Gaussian content of boundaries of convex sets (Theorem 6.5). The latter allows us to prove Proposition 6.3, a stability result for the Kolmogorov–Smirnov-type distance function d used in the statement of Theorem 1.3. Section 7 follows, containing a brief road map to the proofs of Theorems 1.1, 1.2 and 1.3. Then, Section 8 is devoted to explaining how to use small ball estimates on volumes of random projections to formalize the ergodicity (3.2).

Further, the results in Section 9 are used in all our proofs. The main result there is Proposition 9.1, which together with Proposition 8.1 and Lemma 9.5 explains the appearance of chi-squared random variables in the statement of Theorem 1.3. The proof of Proposition 9.6 is the most technical part of our arguments. Next, we complete the proof of Theorem 1.1 in Section 10. We then use Theorem 1.1 to complete in Sections 11 and 12 the proofs of Theorems 1.2 and 1.3, respectively.

5 Singular Values via Wedge Products

In this section, we recall some background on wedge products and refer the reader to the start of Section 7 in [Spi70] and to Section 2.6.1 in [Tao10] for more details. The usual ℓ_2-structure on $\mathbb{R}^n$ gives rise in a functorial way to an ℓ_2 structure on the exterior powers $\Lambda^k \mathbb{R}^n$. If $x_1, \cdots, x_k$ are in $\mathbb{R}^n$ (e.g. are a frame for an element of $G_{n,k}$) we denote the resulting norm by

$$\|x_1 \wedge \cdots \wedge x_k\|.$$

If we denote by X^* the $n \times k$ matrix $(x_1, \cdots, x_k)$, the Gram identity reads

$$\|x_1 \wedge \cdots \wedge x_k\| = \sqrt{\det(XX^*)} = \mathrm{vol}_k \left(\mathcal{P}\left(x_1, \ldots, x_k\right)\right), \tag{5.1}$$

where $\mathcal{P}\left(x_1, \ldots, x_k\right)$ is the parallelopiped spanned by $x_1, \ldots, x_k$. The following Lemma gives a well-known characterization of products of singular values in terms of norms of wedge products, which we will use repeatedly in the proofs of our results.

LEMMA 5.1. *Let A be an $n \times n$ real matrix with singular values $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A)$. If $\theta_1, \cdots, \theta_k$ are unit vectors in $\mathbb{R}^n$, then*

$$\|A\theta_1 \wedge \cdots \wedge A\theta_k\| \leq \sup_{\theta'_1, \ldots, \theta'_k \in S^{n-1}} \|A\theta'_1 \wedge \cdots \wedge A\theta'_k\| = \prod_{i=1}^{k} s_i(A), \qquad (5.2)$$

with equality if and only if θ_i are orthonormal and $\text{span}\{\theta_i, i \leq k\}$ is the subspace spanned by the eigenvectors of AA^* that correspond to the largest singular values of A.

Proof. The inequality on the left is clear. To derive the equality, note that, for any $\theta_1,', \ldots, \theta'_k \in S^{n-1}$,

$$\theta'_1 \wedge \cdots \wedge \theta'_k = \theta''_1 \wedge \cdots \wedge \theta''_k,$$

where

$$\theta''_j = \Pi^{\perp}_{\leq j-1} \theta'_j, \qquad \theta''_1 = \theta_1$$

and $\Pi^{\perp}_{\leq j-1}$ is the projection onto the orthogonal complement of the span of $\theta'_1, \ldots, \theta'_{j-1}$. This follows immediately from the fact that $a_1 \wedge \cdots \wedge a_k$ is zero if $\{a_j\}$ is linearly dependent. Thus, the supremum in (5.2) can be taken over $\theta'_1, \ldots, \theta'_k$ that are orthogonal. Over such collections, the supremum is obtained by letting $\theta'_1, \ldots, \theta'_k$ be any permutation of the k right singular vectors of A. $\qquad\square$

Next, we record in Lemma 5.2 some basic properties of this norm of wedge products that we will use.

LEMMA 5.2. *Let $x, x_1, \cdots, x_k$ be vectors in $\mathbb{R}^n$. Then we have the following basic properties:*

1. **Homogeneity**: *If $\lambda_i > 0$,*

$$\|\lambda x_1 \wedge \cdots \wedge \lambda_k x_k\| = \left(\prod_{i=1}^{n} \lambda_i\right) \|x_1 \wedge \cdots \wedge x_k\|. \qquad (5.3)$$

2. **Projection formula**: *Let $P_{V_i^{\perp}}$ be the orthogonal projection onto the orthogonal complement of $V_i := \text{span}\{x_1, \cdots, x_k\}, V_0 = \{0\}, 1 \leq k \leq n-1$. We have*

$$\|x_1 \wedge \cdots \wedge x_k\| = \prod_{i=1}^{k} \|P_{V_{i-1}^{\perp}} x_i\|_2. \qquad (5.4)$$

3. **Pythagorean Theorem**: *Let $e_1, \ldots, e_n$ be any orthonormal basis of $\mathbb{R}^n$, and define for each multi-index $I = (i_1, \ldots, i_k)$*

$$e_I := e_{i_1} \wedge \cdots \wedge e_{i_k}.$$

Then,

$$||x_1 \wedge \cdots \wedge x_k||^2 = \sum_{\substack{I=(i_1,\ldots,i_k) \\ 1 \le i_1 < \cdots < i_k \le n}} \langle x_1 \wedge \cdots \wedge x_k, e_I \rangle^2. \qquad (5.5)$$

4. **Generalized Gram Identity***: Let $\Theta = (\theta_1, \ldots, \theta_k)$ be an orthonormal system of k vectors in $\mathbb{R}^n$ and write P_Θ for the orthogonal projection onto the span of the θ_i. Consider arbitrary linearly independent vectors $v_1, \ldots, v_k$ in $\mathbb{R}^n$, and denote by V the $n \times k$ matrix whose columns are v_i. Then*

$$\langle v_1 \wedge \cdots \wedge v_k, \theta_1 \wedge \cdots \wedge \theta_k \rangle^2 = \det(P_\Theta V V^* P_\Theta) = ||P_\Theta v_1 \wedge \cdots \wedge P_\Theta v_k||^2. \qquad (5.6)$$

Proof. Homogeneity is immediate from the multi-linearity of the determinant (5.1). The projection formula (5.4) follows from (5.1) and the fact that $\sqrt{\det(XX^*)}$ is the volume of the parallelopiped spanned by $x_1, \ldots, x_k$.. Next, the Pythagorean theorem follows from the fact that in the definition of the ℓ_2 structure on $\Lambda^k \mathbb{R}^n$,

$$\{e_I, \ I = (i_1, \ldots, i_k), \ 1 \le i_1 < \cdots < i_k \le n\}$$

is an orthonormal basis. Finally, to show (5.6), assume first that

$$\theta_j = e_j, \qquad j = 1, \ldots, k$$

are the first k standard unit vectors. Then the right equality follows immediately from the Gram identity (5.1). To see the left equality, write

$$v_j = \sum_{i=1}^n v_{j,i} e_i.$$

We have

$$v_1 \wedge \cdots \wedge v_k = \sum_{i_1,\ldots,i_k} \prod_{j=1}^k v_{j,i_j} e_{i_1} \wedge \cdots \wedge e_{i_k} = \sum_{\substack{I=(i_1,\ldots,i_k) \\ 1 \le i_1 < \cdots < i_k \le n}} \left(\sum_{\sigma \in S_k} (-1)^{\mathrm{sgn}(\sigma)} \prod_{j=1}^k v_{j,\sigma(j)} \right) e_I.$$

Hence, writing V_k for the matrix obtained from V by keeping only the first k rows, we find from the Pythagorean theorem (5.5),

$$\langle v_1 \wedge \cdots \wedge v_k, \theta_1 \wedge \cdots \wedge \theta_k \rangle^2 = \det(V_k)^2 = \det(V_k V_k^*).$$

The case of general θ_i follows by considering any orthogonal matrix U satisfying

$$\theta_i = U e_i, \qquad i = 1, \ldots, k.$$

Then

$$\langle v_1 \wedge \cdots \wedge v_k, \theta_1 \wedge \cdots \wedge \theta_k \rangle^2 = \det((U^T V)_k (U^T V)_k^*) = \det(P_\Theta V (P_\Theta V)^*). \qquad \square$$

5.1 Haar Measure on Frames. It well-know that if ξ is a standard Gaussian on $\mathbb{R}^n$, then $\widehat{\xi} = \xi/\|\xi\|$ is independent of $\|\xi\|$ and that $\widehat{\xi}$ is uniform on the unit sphere. We will need natural generalizations of these facts to orthonormal frames, Lemmas 5.3 and 5.4.

LEMMA 5.3. (Polar Decomposition for Haar Measure on Flags) *Fix integers* $n \geq k \geq 1$. *Let* $\xi_1, \ldots, \xi_k \in \mathbb{R}^n$ *be independent standard Gaussian random vectors. The following collections of random variables are independent:*

$$\{\|\xi_1\|, \|\xi_1 \wedge \xi_2\|, \ldots, \|\xi_1 \wedge \cdots \wedge \xi_k\|\}, \qquad \left\{ \frac{\xi_1}{\|\xi_1\|}, \ldots, \frac{\xi_1 \wedge \cdots \wedge \xi_k}{\|\xi_1 \wedge \cdots \wedge \xi_k\|} \right\}. \quad (5.7)$$

Moreover, denote by $P_{\leq i}$ *the orthogonal projection onto the complement of the span of* $\xi_1, \ldots, \xi_i$. *Then, the random variables terms* $\|P_{\leq i-1}\xi_i\|$ *are joiltly independent.*

Proof. We begin by recalling a fact from elementary probability. Namely, let X, Y, Z be any random variables defined on the same probability space. Then,

$$X \perp Y \qquad \text{and} \qquad X \perp Z | Y \qquad \Rightarrow \qquad X \perp (Y, Z). \qquad (5.8)$$

In words, if X is independent of Y and Z is independent of X given Y, then, (Y, Z) is independent of X.

We proceed by induction on k. The case $k = 1$ follows from the fact that the radial and angular parts of a standard Gaussian are independent. Suppose now $k \geq 2$ and we have proved the statement for $k - 1$. For any $\xi \in \mathbb{R}^k \setminus \{0\}$, let us write $P_{\xi^\perp}$ for the orthogonal projection onto the orthogonal complement to the line spanned by ξ. Define for $\ell = 2, \ldots, k$

$$\xi'_\ell := P_{\xi_1^\perp} \xi_\ell.$$

Note that

$$\|\xi_1 \wedge \cdots \wedge \xi_\ell\| = \|\xi_1\| \|\xi'_2 \wedge \cdots \wedge \xi'_\ell\|$$

and that

$$\xi_1 \wedge \cdots \wedge \xi_\ell = \xi_1 \wedge \xi'_2 \wedge \cdots \wedge \xi'_\ell.$$

With this notation, it is enough to show that the collections

$$\{\|\xi_1\|, \|\xi'_2\|, \ldots, \|\xi'_2 \wedge \cdots \wedge \xi'_k\|\}, \qquad \left\{ \frac{\xi_1}{\|\xi_1\|}, \frac{\xi'_2}{\|\xi'_2\|}, \ldots, \frac{\xi'_2 \wedge \cdots \wedge \xi'_k}{\|\xi'_2 \wedge \cdots \wedge \xi'_k\|} \right\}$$

are independent since if X, Y are independent, then so are $f(X), g(Y)$ for any measurable functions f, g. To see this, first observe that, aside from $\|\xi_1\|$, all other random variables in all both collections are measurable functions of $\xi_1/\|\xi_1\|, \xi_2, \ldots, \xi_k$.

Hence, $||\xi_1||$ is independent of all other variables in both collections. It therefore remains to check only that

$$\mathcal{A} = \left\{ ||\xi_2'||, \ldots, ||\xi_2' \wedge \cdots \wedge \xi_k'|| \right\}, \qquad \mathcal{B} = \left\{ \frac{\xi_1}{||\xi_1||}, \frac{\xi_2'}{||\xi_2'||}, \ldots, \frac{\xi_2' \wedge \cdots \wedge \xi_k'}{||\xi_2' \wedge \cdots \wedge \xi_k'||} \right\}$$

are independent. Observe that, given, $\xi_1/||\xi_1||$, the random variables ξ_ℓ', $\ell = 2, \ldots, k$ are projections of iid standard Gaussians onto a fixed dimension k subspace and hence are themselves iid standard Gaussians. By the inductive hypothesis, conditional on $\xi_1/||\xi_1||$, the collection $\mathcal{A}$ is independent of the collection $\mathcal{B} \backslash \{\xi_1/||\xi_1||\}$. Moreover, the random variables in $\mathcal{A}$ are independent of ξ_1. Invoking (5.8) therefore completes the proof of the statement that the two collections in (5.7) are independent. To finish the proof of this Lemma, let us check that the terms

$$||\xi_1||, ||P_{\leq 1}\xi_2||, \ldots, ||P_{\leq k-1}\xi_k|| \tag{5.9}$$

are independent by induction on k. When $k = 1$ the claim is trivial. For the inductive step, use the projection formula (5.4) to we write

$$||P_{\leq k-1}\xi_k|| = \left|\left| \frac{\xi_1 \wedge \cdots \wedge \xi_{k-1}}{||\xi_1 \wedge \cdots \wedge \xi_{k-1}||} \wedge \xi_k \right|\right|.$$

By the first part of this Lemma, we know that

$$\frac{\xi_1 \wedge \cdots \wedge \xi_{k-1}}{||\xi_1 \wedge \cdots \wedge \xi_{k-1}||}, \qquad \{||\xi_1 \wedge \cdots \wedge \xi_j||, \qquad j \leq k-1\}$$

are independent. Hence, since ξ_k is independent of $\xi_1, \ldots, \xi_{k-1}$, we conclude that the terms in (5.9) are independent as well.   □

We will also use this following result:

LEMMA 5.4 (Haar Measure on Flags). *Fix integers $n \geq k \geq 1$. Let $\xi_1, \ldots, \xi_k \in \mathbb{R}^n$ be independent standard Gaussian random vectors. For each $i = 1, \ldots, k$ define*

$$\xi_i' = \begin{cases} \xi_1, & i = 1, \\ P_{V_{i-1}}\xi_i, & i > 1, \end{cases}$$

where $V_{i-1} = \text{Span}\{\xi_j, \ 1 \leq j < i\}^{\perp}$ and $P_{V_{i-1}}$ is the orthogonal projection onto V_{i-1}. Then $\{\xi_i'/||\xi_i'||, \ i = 1, \ldots, k\}$ is distributed according to Haar measure on the space of such flags of orthonormal frames. In particular, if $v_1, \ldots, v_k$ is Haar-distributed in the space of k-frames in $\mathbb{R}^n$, then

$$v_1 \wedge \cdots \wedge v_k \stackrel{d}{=} \frac{\xi_1 \wedge \cdots \wedge \xi_k}{||\xi_1 \wedge \cdots \wedge \xi_k||}. \tag{5.10}$$

Proof. The random variable $(\xi_i', \, i = 1, \dots, k)$ clearly takes values in the set of k-frames in $\mathbb{R}^n$. Moreover, it is invariant under the action of the orthogonal group on such frames and hence must be distributed according to the Haar measure. Indeed, since the angular part of a standard Gaussian is uniform on the sphere, $\xi_1'/\|\xi_i'\|$ is uniform on S^{n-1}. Similarly, ξ_2' is a standard Gaussian in the orthogonal complement $\xi_1'^\perp$ to ξ_i'. Hence, $\xi_2'/\|\xi_2'\|$ is uniform on the unit sphere in $\xi_1'^\perp$ and so on. Finally, to derive (5.10), note that

$$\xi_1 \wedge \cdots \wedge \xi_k = \xi_1' \wedge \cdots \xi_k'$$

since the wedge product of any linearly dependent set of vectors vanishes. Combining this with the projection formula (5.4), we conclude that

$$\frac{\xi_1 \wedge \cdots \wedge \xi_k}{\|\xi_1 \wedge \cdots \wedge \xi_k\|} = \frac{\xi_1'}{\|\xi_1'\|} \wedge \cdots \wedge \frac{\xi_k'}{\|\xi_k'\|}$$

and (5.10) follows from the first part of this Lemma. $\qquad\square$

6 Background on Sums of Independent Random Variables

6.1 A Result of Łatała: Precise Behavior for Moments of Sums. In the proof of our pointwise esimate Proposition 9.1, we will use the following result of R. Latała [Lat97, Thm. 2, Cor. 2, Rmk. 2]:

Theorem 6.1. *Let* $X_1, \cdots, X_N$ *be mean zero, independent r.v. and* $p \geq 1$. *Then*

$$\left(\mathbb{E} \left[\left| \sum_{j=1}^N X_j \right|^p \right] \right)^{\frac{1}{p}} \simeq \inf \left\{ t > 0 : \sum_{j=1}^N \log \left[\mathbb{E} |1 + \frac{X_j}{t}|^p \right] \leq p \right\}, \qquad (6.1)$$

where $a \simeq b$ *means there exist universal constants* c_1, c_2 *so that* $c_1 a \leq b \leq a c_2$. *Moreover, if* X_i *are also identically distributed then*

$$\left(\mathbb{E} \left[\left| \sum_{j=1}^N X_j \right|^p \right] \right)^{\frac{1}{p}} \simeq \sup_{\max\{2, \frac{p}{N}\} \leq s \leq p} \frac{p}{s} \left(\frac{N}{p} \right)^{\frac{1}{s}} \|X_i\|_s. \qquad (6.2)$$

6.2 Small Ball Estimates for Sums of iid Random Variables with Bounded Density. We will have occasion to use the following standard result (e.g. Lemma 2.2 in [RV08]).

Theorem 6.2. *Let* $\zeta_k, \, k = 1, \dots, n$ *be iid non-negative random variables and suppose there exists* K *such that* $\mathbb{P}(\zeta_k < \varepsilon) \leq K\varepsilon$ *for all* $\varepsilon \geq 0$, *then for all* $\varepsilon \geq 0$

$$\mathbb{P} \left(\sum_{k=1}^n \zeta_k^2 < n\varepsilon^2 \right) \leq (CK\varepsilon)^n,$$

where C *is an absolute constant.*

6.3 Quantitative Multivariate CLT. One of our goals in this article is to measure the approximate normality for Lyapunov exponents of $X_{N,n}$ when $N \gg n$ (see Theorem 1.3). As explained in the introduction, we will measure normality using a distance function that is a natural high-dimensional generalization of the usual Kolmogorov–Smirnov distance:

$$d(X, Y) := \sup_{C \in \mathcal{C}_k} |\mathbb{P}(X \in C) - \mathbb{P}(Y \in C)|, \tag{6.3}$$

where $\mathcal{C}_k$ is the collection of all convex subsets of $\mathbb{R}^k$. The distance function d has three desirable properties. First, it is affine invariant in sense that if T is any invertible affine transformation and A is any convex set, then $T^{-1}A$ is also convex and hence for any random variables X, Y on the same probability space

$$d(TX, TY) = d(X, Y). \tag{6.4}$$

The second desirable property of d is that it is stable to small ℓ_2 perturbations, as explained in Proposition 6.3 below. To state this result, we write

$$\mathcal{N}(\mu, \Sigma), \qquad \mu \in \mathbb{R}^k, \quad \Sigma \in \mathrm{Sym}_k^+$$

for a Gaussian with mean μ and invertible covariance Σ.

PROPOSITION 6.3. *There exists $c > 0$ with the following property. Suppose that X, Y are $\mathbb{R}^k$-valued random variables defined on the same probability space. For all $\mu \in \mathbb{R}^k$, invertible symmetric matrices $\Sigma \in \mathrm{Sym}_k^+$, and $\delta > 0$ we have*

$$d(X + Y, \mathcal{N}(\mu, \Sigma)) \le 3d(X, \mathcal{N}(\mu, \Sigma)) + c\delta\sqrt{||\Sigma^{-1}||_{HS}} + 2\mathbb{P}(||Y||_2 > \delta). \tag{6.5}$$

We prove Proposition 6.3 in Section 6.4 below. Before doing so, however, we state the third desirable property of the distance d, namely, that it measures convergence in the multivariate CLT. We follow the notation in Bentkus [Ben05] and define

$$S := S_N = X_1 + \cdots + X_N,$$

where $X_1, \ldots, X_N$ are independent random vectors in $\mathbb{R}^k$ with common mean $\mathbb{E}X_j = 0$. We set

$$C := \mathrm{cov}(S)$$

to be the covariance matrix of S, which we assume is invertible. With the definition

$$\beta_j := \mathbb{E}||C^{-\frac{1}{2}}X_j||_2^3, \ \beta := \sum_{j=1}^{N} \beta_j, \tag{6.6}$$

we have the following [Ben05]:

Theorem 6.4 (Multivariate CLT with Rate). *There exists an absolute constant $c > 0$ such that*

$$d(S, C^{\frac{1}{2}}Z) \le ck^{\frac{1}{4}}\beta, \tag{6.7}$$

where $Z \sim \mathcal{N}(0, \mathrm{Id}_k)$ denotes a standard Gaussian on $\mathbb{R}^k$.

6.4   Proof of Proposition 6.3.   We rely on the following result (Theorem 6.5) of Nazarov [Naz03], which was an extension of a Theorem of Ball [Bal93].
Let Z_k be the standard Gaussian in $\mathbb{R}^k$. Let C be a positive definite symmetric matrix and let γ_C be the density of the $C^{-\frac{1}{2}}Z_k$, i.e.

$$d\gamma_C(y) := \frac{\sqrt{|\det C|}}{(2\pi)^{\frac{k}{2}}} e^{-\frac{\langle Cy, y \rangle}{2}} dy.$$

Let

$$K_\epsilon := \{x \in \mathbb{R}^k : \exists y \in K, \|x - y\|_2 < \epsilon\} \text{ and } K_{-\epsilon} : \{x \in K := x + \epsilon B_2^k \subseteq K\} \quad (6.8)$$

We define

$$\Gamma(C) := \sup_{K \in \mathcal{C}} \left\{ \lim_{\epsilon \to 0_+} \frac{\gamma_C(K_\epsilon \backslash K)}{\epsilon} \right\}. \quad (6.9)$$

Our proof of Proposition 6.3 crucially relies on the following result of Nazarov.

Theorem 6.5 ([Naz03]). *There exists absolute constants $c_1, c_2 > 0$ such that*

$$c_1 \sqrt{\|C\|_{HS}} \leq \Gamma(C) \leq c_2 \sqrt{\|C\|_{HS}}. \quad (6.10)$$

We will need an elementary corollary of (6.10).

COROLLARY 6.6. *For any $\varepsilon \in (0, 1)$ and $K \in \mathcal{D}_k$, we have*

$$\left| \mathbb{P}(C^{-\frac{1}{2}}Z_k \in K_\epsilon) - \mathbb{P}(C^{-\frac{1}{2}}Z_k \in K_{-\epsilon}) \right| \leq c_2 \epsilon \sqrt{\|C\|_{HS}}. \quad (6.11)$$

Proof. Our argument is along the lines of the proof in [Ben05] of equations (1.3), (1.4) or [Ben03] equation (1.2), which obtain similar statements in the special case where $C = \text{Id}$. We will use a standard estimate (see e.g. Lemma 10.5 in [Kal02]) that if K is any convex set in $\mathbb{R}^k$ and $(\partial K)_\epsilon$ is the ϵ-neighborhood of the boundary of K, then

$$\text{vol}((\partial K)_\epsilon) \leq 2 \left[\left(1 + \frac{\epsilon}{r(K)}\right)^k - 1 \right] \text{vol}(B_{r(K)}), \quad (6.12)$$

where $r(K)$ is the radius of the smallest ball containing K, vol is the Euclidean volume, and B_r is a ball of radius r. The crucial point is that the right hand side is bounded for all $\epsilon \in (0, 1)$ by $c\epsilon$ with c depending only on $r(K)$ and the ambient dimension k.

To derive from this estimate (6.11), observe that for every $\epsilon > 0$, if $K \in \mathcal{C}_k$ then also K_ϵ and $K_{-\epsilon}$ are in $\mathcal{C}_k$. Note also that the difference of probabilities on the left-hand side of (6.11) can be bounded above by $\mathbb{P}(W \in K_\epsilon \backslash K) + \mathbb{P}(W \in K \backslash K_{-\epsilon})$ where $W := C^{-\frac{1}{2}}Z_k$. It is therefore enough to check that each of these probabilities is in

turn bounded above by the right hand side of (6.11). To see this for $\mathbb{P}(W \in K_\epsilon \backslash K)$, denote for K convex and $t > 0$,

$$\omega_K(t) := \mathbb{P}\left(W \in K_t \backslash K\right).$$

Since $K_{t+\epsilon} \backslash K = (K_{t+\epsilon} \backslash K_t) \cup (K_t \backslash K)$ we have that

$$\omega_K(t+\epsilon) - \omega_K(t) = \omega_{K_t}(\epsilon). \tag{6.13}$$

This relation, together with (6.12), implies that $\omega_K(t)$ is absolutely continuous. Indeed, we may write

$$\omega_{K_t}(\epsilon) = \mathbb{P}\left(Z_k \in C^{1/2}K_{t+\epsilon} \backslash C^{1/2}K_t\right) \leq \frac{\text{vol}\left(C^{1/2}K_{t+\epsilon} \backslash C^{1/2}K_t\right)}{(2\pi)^{k/2}}.$$

Denoting by $\lambda_{\max}$ the maximal eigenvalue of $C^{1/2}$, we may write

$$C^{1/2}K_{t+\epsilon} = C^{1/2}\left(K_t\right)_\epsilon \subseteq (C^{1/2}K_t)_{\lambda_{\max}\epsilon}.$$

Thus,

$$\omega_{K_t}(\epsilon) \leq \frac{\text{vol}\left((C^{1/2}K_t)_{\lambda_{\max}\epsilon} \backslash C^{1/2}K_t\right)}{(2\pi)^{k/2}}.$$

Denoting by R the radius of the smallest ball containing $(C^{1/2}K_1)_{\lambda_{\max}}$, we obtain from (6.12) that there exists a constant $c > 0$ depending on C, k and R so that for any $0 \leq t, \epsilon \leq 1$

$$\omega_{K_t}(\epsilon) \leq c\epsilon.$$

Thus, using (6.13), we indeed see that $\omega_K(t)$ is absolutely continuous on $[0, 1]$. Hence, its derivative $\omega'_K(t)$ exists almost everywhere and

$$\omega_K(t) = \int_0^t \omega'_K(s)ds.$$

Moreover, combining (6.13) with (6.10), we find for any $t \in (0, 1)$ that

$$0 \leq \omega'_K(t) = \lim_{\epsilon \to 0_+} \frac{\omega_K(t+\epsilon) - \omega_K(t)}{\epsilon} = \lim_{\epsilon \to 0_+} \frac{\omega_{K_t}(\epsilon)}{\epsilon} \leq c\sqrt{||C||_{HS}}$$

Hence, since $\omega_K(0) = 0$, we find that $\omega_K(t) \leq ct\sqrt{||C||_{HS}}$, for all $t \leq 1$. Using a similar argument for $\bar{\omega}_K(t) := \mathbb{P}(W \in K \backslash K_{-t})$, we conclude that (6.11) indeed holds. $\qquad\square$

We are now ready to prove Proposition 6.3. Note that if S, T are any events on the same probability space, then

$$\left|\mathbb{P}\left(S\right) - \mathbb{P}\left(T\right)\right| \leq \mathbb{P}\left(S \Delta T\right),$$

where $S \Delta T$ is the symmetric difference. For any convex set $A \subseteq \mathbb{R}^k$, we have

$$\left|\mathbb{P}\left(X + Y \in A\right) - \mathbb{P}\left(X \in A\right)\right|$$

is bounded above by

$$\left|\mathbb{P}\left(X + Y \in A, \, ||Y||_2 \leq \delta\right) - \mathbb{P}\left(X \in A, \, ||Y||_2 \leq \delta\right)\right| + 2\mathbb{P}\left(||Y||_2 > \delta\right).$$

Note that

$$\left\{X + Y \in A, \, ||Y||_2 \leq \delta\right\} \Delta \left\{X \in A, \, ||Y||_2 \leq \delta\right\} \ \subseteq \ \left\{X \in A_\delta \backslash A_{-\delta}\right\}.$$

Thus, we find

$$\begin{aligned}
\left|\mathbb{P}\left(X + Y \in A\right) - \mathbb{P}\left(X \in A\right)\right| &\leq \mathbb{P}\left(X \in A_\delta\right) - \mathbb{P}\left(X \in A_{-\delta}\right) + 2\mathbb{P}\left(||Y||_2 > \delta\right) \\
&\leq \left|\mathbb{P}\left(X \in A_\delta\right) - \mathbb{P}\left(\mathcal{N}\left(\mu, \Sigma\right) \in A_\delta\right)\right| \\
&\quad + \left|\mathbb{P}\left(X \in A_{-\delta}\right) - \mathbb{P}\left(\mathcal{N}\left(\mu, \Sigma\right) \in A_{-\delta}\right)\right| \\
&\quad + \mathbb{P}\left(\mathcal{N}\left(\mu, \Sigma\right) \in A_\delta\right) - \mathbb{P}\left(\mathcal{N}\left(\mu, \Sigma\right) \in A_{-\delta}\right) + 2\mathbb{P}\left(||Y||_2 > \delta\right) \\
&\leq 2d(X, \mathcal{N}\left(\mu, \Sigma\right)) + c_0 \delta \sqrt{||\Sigma^{-1}||_{HS}} + 2\mathbb{P}\left(||Y||_2 > \delta\right)
\end{aligned}$$

where we have used (6.11). Putting this all together, we find

$$\begin{aligned}
d(X + Y, \mathcal{N}\left(\mu, \Sigma\right)) &\leq d(X, \mathcal{N}\left(\mu, \Sigma\right)) + d(X + Y, X) \\
&\leq 3d(X, \mathcal{N}\left(\mu, \Sigma\right)) + c_0 \delta \sqrt{||\Sigma^{-1}||_{HS}} + 2\mathbb{P}\left(||Y||_2 > \delta\right). \qquad \square
\end{aligned}$$

7   Roadmap for Proofs of Theorems 1.2 and 1.3

In this section, we explain the organization of the proofs of Theorems 1.2 and 1.3. Our starting point is in Section 8. There, in Proposition 8.1 we explain how to provide surprisingly useful bounds on the size of the difference

$$\frac{1}{N} \log ||X_{N,n}(\Theta)|| - (\lambda_1 + \cdots + \lambda_k) \tag{7.1}$$

using small ball estimates on determinants of volumes of random projections. This makes precise (3.2). We remind the reader that $\lambda_1, \ldots, \lambda_k$ are the top k Lyapunov exponents for $X_{N,n}$ and that

$$X_{N,n}(\Theta) = X_{N,n}\theta_1 \wedge \cdots X_{N,n}\theta_k,$$

where θ_j are is a fixed orthonormal k-frame in $\mathbb{R}^n$. We think of $X_{N,n}(\Theta)$ as a pointwise analog of $\lambda_1 + \cdots + \lambda_k$ since by Lemma 5.1 the supremum over Θ of $\frac{1}{N} \log \|X_{N,n}(\Theta)\|$ equals $\lambda_1 + \cdots + \lambda_k$.

Using Proposition 8.1, we analyze in Section 9 the concentration properties of $\frac{1}{N} \log \|X_{N,n}(\Theta)\|$. By Lemma 9.5 it is a sum of independent random variables, allowing us to apply Theorem 6.1 several times. The main result is Proposition 9.1, whose proof is the most technical part of this article.

Combining these concentration estimates for $\frac{1}{N} \log \|X_{N,n}(\Theta)\|$ with the bounds on (7.1) derived in Proposition 8.1, we derive Theorem 1.2 in Section 11, giving quantitative estimates about convergence of the global distribution of singular values of $X_{N,n}$ to the Triangle Law.

Finally, in Section 12, we combine Theorem 1.1 with Proposition 9.1 and the multivariate CLT (see Section 6.3) to prove the approximate normality of Lyapunov exponents stated in Theorem 1.3.

8 Lyapunov Sums Via Small Ball Estimates

The purpose of this section is to explain how to use small ball estimates on volumes of random projections to obtain concentration estimates on the difference

$$\frac{1}{N} \log \|X_{N,n}(\Theta)\| - \sup_{\Theta' \in \mathrm{Fr}_{n,k}} \frac{1}{N} \log \|X_{N,n}(\Theta')\| = \frac{1}{N} \log \|X_{N,n}(\Theta)\| - (\lambda_1 + \cdots + \lambda_k)$$

between the sum of the first k Lyapunov exponents of $X_{N,n}$ and the "pointwise" analog of this quantity evaluated at any fixed orthonormal system $\Theta = (\theta_1, \ldots, \theta_k)$ of k vectors in $\mathbb{R}^n$. Our main result is the following

PROPOSITION 8.1. *There exists $C > 0$ with the following property. For any $\varepsilon \in (0, 1)$ and any $\Theta \in \mathrm{Fr}_{n,k}$ we have*

$$\mathbb{P}\left(\left| \frac{1}{N} \log \|X_{N,n}(\Theta)\| - \sup_{\Theta' \in \mathrm{Fr}_{n,k}} \frac{1}{N} \log \left| \|X_{N,n}(\Theta')\| \right| \right| \geq \frac{k}{2N} \log \left(\frac{n}{k\varepsilon^2} \right) \right) \leq (C\varepsilon)^{k/2}.$$

Proof. The key observation is the following:

LEMMA 8.2. *For any $\varepsilon \in (0, 1)$ and any $\Theta \in \mathrm{Fr}_{n,k}$ we have*

$$\mathbb{P}\left(\left| \frac{1}{N} \log \|X_{N,n}(\Theta)\| - \sup_{\Theta' \in \mathrm{Fr}_{n,k}} \frac{1}{N} \log \|X_{N,n}(\Theta')\| \right| \geq \frac{1}{N} \log \left(\frac{1}{\varepsilon} \right) \right) \leq \mathbb{P}\left(\|P_F(\Theta)\| \leq \varepsilon \right),$$

where F is a Haar distributed k-dimensional subspace of $\mathbb{R}^n$ and

$$P_F(\Theta) = P_F \theta_1 \wedge \cdots \wedge P_F \theta_k$$

with P_F denoting the orthogonal onto F.

Proof of Lemma 8.2. Denote by $v^{(1)}, \ldots, v^{(n)}$ the right singular vectors of $X_{N,n}$ corresponding to its singular values $s_1 \geq \cdots \geq s_n$. By abuse of notation, we will write $X_{N,n} : \Lambda^k \mathbb{R}^n \to \Lambda^k \mathbb{R}^n$ for the linear transformation given by

$$X_{N,n}(x_1 \wedge \cdots \wedge x_k) = X_{N,n} x_1 \wedge \cdots \wedge X_{N,n} x_k.$$

The right singular vectors of $X_{N,n}$ acting on $\Lambda^k \mathbb{R}^n$ are

$$v^{(I)} := v^{(i_1)} \wedge \cdots \wedge v^{(i_k)}, \qquad I = (i_1, \ldots, i_k), \, i_1 < \cdots < i_k$$

and the corresponding singular values are

$$s_I := \prod_{i \in I} s_i.$$

Hence, the Pythagorean theorem for wedge products and the generalized Gram identity (see Lemma 5.2) yield

$$||X_{N,n}(\Theta)||^2 = \sum_{\substack{I=(i_1,\ldots,i_k) \\ 1 \leq i_1 < \cdots < i_k \leq n}} s_I^2 \left\langle v^{(I)}, \Theta \right\rangle^2 \geq \left(\prod_{i=1}^k s_i^2 \right) \left\langle v^{(1,\ldots,k)}, \Theta \right\rangle^2 = \left(\prod_{i=1}^k s_i^2 \right) ||P_k(\Theta)||^2,$$

where in the last equality we've denote by P_k the orthogonal projection into the span of the top k right singular vectors of $X_{N,n}$. We therefore obtain, using Lemma 5.1:

$$0 \geq \frac{1}{N} \log ||X_{N,n}(\Theta)|| - \sup_{\Theta' \in \mathrm{Fr}_{n,k}} \frac{1}{N} \log ||X_{N,n}\Theta'||$$

$$= \frac{1}{2N} \log \left(\frac{||X_{N,n}(\Theta)||^2}{\prod_{i=1}^k s_i^2} \right)$$

$$= \frac{1}{N} \log ||P_k(\Theta)||.$$

Since $X_{N,n}$ is invariant under right multiplication by a Haar orthogonal matrix, we see that P_k is equal in distribution to the orthogonal projection onto a Haar distributed k-dimensional subspace of $\mathbb{R}^n$. $\qquad\square$

In order to apply Lemma 8.2 we need small ball estimates on $||P_F(\Theta)||$. Gaussian analogs of such estimates are essentially available in the literature, but are phrased in the language of determinants of random matrices. To reduce to this case, note that if F is Haar distributed among k-dimensional subspaces of $\mathbb{R}^n$, an orthonormal basis $v_1, \ldots, v_k$ for F is Haar distributed on the space of such k-frames in $\mathbb{R}^n$. Thus,

by (5.10) from Lemma 5.4, we find

$$
\begin{aligned}
||P_F(\Theta)|| &= ||P_F\theta_1 \wedge \cdots \wedge P_F\theta_k|| \\
&= |\langle v_1 \wedge \cdots \wedge v_k, \theta_1 \wedge \cdots \wedge \theta_k\rangle| \\
&\overset{d}{=} \left|\left\langle \frac{g_1 \wedge \cdots \wedge g_k}{||g_1 \wedge \cdots \wedge g_k||}, \theta_1 \wedge \cdots \wedge \theta_k\right\rangle\right| \\
&\overset{d}{=} \frac{||G\theta_1 \wedge \cdots \wedge G\theta_k||}{\det(GG^*)^{\frac{1}{2}}} \\
&= \left(\frac{\det(G_kG_k^*)}{\det(GG^*)}\right)^{\frac{1}{2}}
\end{aligned}
\tag{8.1}
$$

where $\overset{d}{=}$ denotes equality in distribution, G is a $k \times n$ matrix with iid standard Gaussian columns g_i, G_k is the obtained from G by keeping only the first k columns, and we have used the Gram identity (5.1) and (5.6) in the last two lines. The relation (8.1), combined in Lemma 8.2, therefore gives that

$$
\mathbb{P}\left(\left|\frac{1}{N}\log||X_{N,n}(\Theta)|| - \sup_{\Theta' \in \mathrm{Fr}_{n,k}} \frac{1}{N}\log||X_{N,n}(\Theta')||\right| \geq \frac{1}{N}\log\left(\frac{1}{\varepsilon}\right)\right)
$$

is bounded above by

$$
\mathbb{P}\left(\left(\frac{\det(G_kG_k^*)}{\det(GG^*)}\right)^{\frac{1}{2}} \leq \varepsilon\right),
\tag{8.2}
$$

To complete the proof, we recall the following result.

PROPOSITION 8.3 (Lemma 4.2 in [PP13]).
There exist universal constants $c, C > 0$ with the following property. Let G be a $k \times n$ matrix with iid standard $\mathcal{N}(0,1)$ Gaussian entries. Then

$$
\left(\mathbb{E}\left[\det(GG^*)^{\frac{p}{2k}}\right]\right)^{\frac{1}{p}} \leq C\sqrt{n}, \ 0 < p \leq kn
\tag{8.3}
$$

and

$$
\left(\mathbb{E}\left[\det(GG^*)^{-\frac{p}{2k}}\right]\right)^{-\frac{1}{p}} \geq c\sqrt{n}, \ 0 < p \leq k(n-k+1-e^{-k(n-k+1)})
\tag{8.4}
$$

This allows use to estimate the probability in (8.2) via the following corollary, which in combination with (8.2) and Lemma 8.2 completes the proof of Proposition 8.1.

COROLLARY 8.4. *There exists a universal constant $c > 0$ with the following property. Let G be a $k \times n$ matrix with iid Standard Gaussian entries, and denote by G_k the matrix obtained from G by keeping only the first k columns.*

$$
\mathbb{P}\left(\left(\frac{\det(G_kG_k^*)}{\det(GG^*)}\right)^{\frac{1}{2k}} \leq \varepsilon\sqrt{\frac{k}{n}}\right) \leq (c\varepsilon)^{\frac{k}{2}}, \varepsilon > 0.
\tag{8.5}
$$

Proof. Using (8.3) and (8.4) and Markov's inequality, we have that for $t \geq 1$,

$$\mathbb{P}\left(\det(GG^*)^{\frac{1}{2k}} \geq tC\sqrt{n}\right) \leq \frac{1}{t^{nk}} \text{ and } \mathbb{P}\left(\det(GG^*)^{\frac{1}{2k}} \leq \varepsilon c\sqrt{n}\right) \leq (c\varepsilon)^{k(n-k+1-e^{-k(n-k+1)})}.$$

Note that G_k has the same distribution as a $k \times k$ matrix with iid standard $\mathcal{N}(0,1)$ Gaussian entries. So, we have that

$$\begin{aligned}
\mathbb{P}\left(\left(\frac{\det(G_k G_k^*)}{\det(GG^*)}\right)^{\frac{1}{2k}} \leq \varepsilon\sqrt{\frac{k}{n}}\right) &\leq \mathbb{P}\left(\frac{\det(G_k G_k^*)^{\frac{1}{2k}}}{\det(GG^*)^{\frac{1}{2k}}} \leq \varepsilon\sqrt{\frac{k}{n}} \text{ and } \det(GG^*)^{\frac{1}{2k}} \leq tC\sqrt{n}\right) \\
&\quad + \mathbb{P}\left(\det(GG^*)^{\frac{1}{2k}} \geq tC\sqrt{n}\right) \\
&\leq \mathbb{P}\left(\det(G_k G_k^*)^{\frac{1}{2k}} \leq \varepsilon t C\sqrt{k}\right) + \mathbb{P}\left(\det(GG^*)^{\frac{1}{2k}} \geq tC\sqrt{n}\right) \\
&\leq (c't\varepsilon)^{k(1-e^{-k})} + \frac{1}{t^{kn}} \leq (c\varepsilon)^{k/2},
\end{aligned}$$

where in the last line we've taken $t = \varepsilon^{-1/2n}$. □

9 Concentration for $\frac{1}{N}\log\|X_{N,n}(\Theta)\|$

As mentioned above, a key step towards proving Theorems 1.2 and 1.3 is to obtain precise concentration estimates for

$$\frac{1}{N}\log\|X_{N,n}(\Theta)\| = \frac{1}{N}\log\|X_{N,n}\theta_1 \wedge \cdots \wedge X_{N,n}\theta_k\|, \tag{9.1}$$

where $\Theta = (\theta_1, \ldots, \theta_k)$ is a fixed orthonormal system in $\mathbb{R}^n$. Define

$$M_j := n - j + 1, \qquad \xi_{n,k} = \frac{1}{n}\sum_{j=1}^{k}\frac{1}{M_j} \tag{9.2}$$

and as in (1.5) set

$$\mu_{n,j} := \frac{1}{2}\mathbb{E}\left[\log\left(\frac{1}{n}\chi^2_{n-j+1}\right)\right].$$

Our main result about the concentration for $\log\|X_{N,n}(\Theta)\|$ is the following.

PROPOSITION 9.1. *There exists a universal constant $c > 0$ with the following property. Fix any orthonormal system Θ of k vectors in $\mathbb{R}^n$. With $X_{N,n}(\Theta)$ as in (9.1), we have*

$$\mathbb{P}\left(\left|\frac{1}{nN}\log\|X_{N,n}(\Theta)\| - \frac{1}{n}\sum_{j=1}^{k}\mu_{n,j}\right| \geq s\right) \leq 2\exp\left\{-cnN\min\left\{M_k s, \frac{s^2}{\xi_{n,k}}\right\}\right\}, \qquad s > 0. \tag{9.3}$$

REMARK 9.2. The double behavior in the exponent of the estimates (9.3) is of Bernstein-type. We do not use any off-the-shelf Bernstein estimates for deriving it however, relying instead on Theorem 6.1 of Łatała [Lat97]. One advantage of our approach is that Łatała's estimates are all reversible (i.e. have matching upper and lower bounds). Hence, with a bit more work it is possible to show that the estimate (9.3) is sharp. We will not need this fact, however, and will provide only a proof of the upper bound.

REMARK 9.3. Although we focus in this article on the Gaussian case, we believe it is possible to prove that Proposition 9.1 holds under minimal assumptions on the distribution of the entries of A_i. Somewhat weaker results in this directions are proved in [LP82, Thms. 7,8] and [BLR85, Thm. 5.1].

REMARK 9.4. By Lemma 9.5 below, we have

$$\mathbb{E}\left[\frac{1}{N}\log||X_{N,n}(\Theta)||\right] = \sum_{j=1}^{k}\mu_{n,j}.$$

The proof of Proposition 9.1 proceeds from the observation that for the Gaussian case we consider here, $\log X_{N,n}(\Theta)$ is a sum of independent random variables.

LEMMA 9.5. Fix $n, N \geq 1$ and $1 \leq k \leq n$ as well as a collection Θ of k orthonormal vectors in $\mathbb{R}^n$. We have

$$\log||X_{N,n}(\Theta)|| \stackrel{d}{=} \sum_{i=1}^{N}\sum_{j=1}^{k}Y_{i,j},$$

where $\stackrel{d}{=}$ denotes equality in distribution, $Y_{i,j}$ are independent, and for each i, j the random variable $Y_{i,j}$ is distributed like the logarithm $\frac{1}{2}\log(\frac{1}{n}\chi^2_{n-j+1})$ of a normalized chi-squared random variable with $n - j + 1$ degrees of freedom.

Proof. We have

$$\log||X_{N,n}(\Theta)|| = \log||A_N\cdots A_1(\Theta)||$$
$$= \log\left|\left|A_N\cdots A_2\frac{A_1(\Theta)}{||A_1(\Theta)||}\right|\right| + \log||A_1(\Theta)||. \qquad (9.4)$$

Note that by Lemma 5.3, we have that

$$||A_1(\Theta)|| = ||A_1\theta_1 \wedge \cdots \wedge A_1\theta_k||$$

is independent of

$$\frac{A_1(\Theta)}{||A_1(\Theta)||} = \frac{A_1\theta_1 \wedge \cdots \wedge A_1\theta_k}{||A_1\theta_1 \wedge \cdots \wedge A_1\theta_k||}.$$

Hence the two terms in (9.4) are independent. Moreover, $A_2 \left(\frac{A_1(\Theta)}{||A_1(\Theta)||} \right)$ is independent of $A_3, \ldots, A_N$ and, in distribution, we have

$$A_2 \left(\frac{A_1(\Theta)}{||A_1(\Theta)||} \right) \overset{d}{=} A_2(\Theta). \tag{9.5}$$

Indeed, we may write

$$\frac{A_1(\Theta)}{||A_1(\Theta)||} = \frac{A_1\theta_1 \wedge \cdots \wedge A_k\theta_k}{||A_1\theta_1 \wedge \cdots \wedge A_k\theta_k||} = \frac{A_1\theta_1}{||A_1\theta_1||} \wedge \frac{\Pi_{\leq 1}A_1\theta_1}{||\Pi_{\leq 1}A_1\theta_1||} \wedge \cdots \wedge \frac{\Pi_{\leq k-1}A_1\theta_k}{||\Pi_{\leq k-1}A_1\theta_k||},$$

where we've written $\Pi_{\leq i}$ for the projection onto the complement of the span of $A_1\theta_1, \ldots, A_1\theta_i$. Next, we may choose an orthogonal matrix M so that

$$\Pi_{\leq i-1}A_1\theta_i = Me_i,$$

where e_i is the ith standard basis vector. For this choice of M, we find

$$\frac{A_1(\Theta)}{||A_1(\Theta)||} = Me_1 \wedge \cdots \wedge Me_k = M(e_1 \wedge \cdots \wedge e_k).$$

Since $A_2 \overset{d}{=} A_2M$, we conclude that (9.5) holds. In particular, we find that, in distribution,

$$\log ||X_{N,n}(\Theta)|| \overset{d}{=} \sum_{i=1}^{N} \log ||A_i(\Theta)||$$

is a sum of iid terms. Finally, for any fixed $i = 1, \ldots, N$

$$||A_i(\Theta)|| \overset{d}{=} ||\xi_1 \wedge \cdots \wedge \xi_k||,$$

where ξ_i are iid n-dimensional standard Gaussians. Hence, by the projection formula (5.4), we find that

$$||A_i(\Theta)|| \overset{d}{=} \prod_{i=1}^{k} ||P_{\leq i-1}\xi_i||,$$

where we've denoted by $P_{\leq j}$ the projection onto the orthogonal complement of the span of $\xi_1, \ldots, \xi_j$. The terms in the product are independent by Lemma 5.3, and the distribution of the ith term is precisely the same as that of $\sqrt{\frac{1}{n}\chi^2_{n-i+1}}$, completing the proof. $\square$

Lemma 9.5 allows us to obtain precise estimates on the rate of growth of moments of $\log X_{N,n}(\Theta)$ using the result of Łatała [Lat97] (Theorem 6.1 above). These moment estimates, in turn, yield Proposition 9.1 via Markov's inequality applied to the optimal power of $\log ||X_{N,n}(\Theta)||$. We carry out these details in Section 9.1.

9.1 Details for Proof of Proposition 9.1. The purpose of this section is to prove Proposition 9.1. Throughout this section C, C', c, c' etc will be universal constants that may change from line to line. Recalling from (9.2) the notation

$$M_j = n - j + 1, \quad \xi_{n,k} = \frac{1}{n} \sum_{j=1}^{k} \frac{1}{M_j}, \tag{9.6}$$

we seek to show that for $s > 0$

$$\mathbb{P}\left(\left| \frac{1}{nN} \log \|X_{N,n}(\Theta)\| - \frac{1}{n} \sum_{j=1}^{k} \mu_{n,j} \right| \geq s \right) \leq 2 \exp\left(-cnN \min\left\{ M_k s, \frac{s^2}{\xi_{n,k}} \right\} \right). \tag{9.7}$$

where we remind the reader that as in (1.5), we've set

$$\mu_{n,j} := \frac{1}{2} \mathbb{E}\left[\log\left(\frac{1}{n} \chi^2_{n-j+1} \right) \right].$$

According to Lemma 9.5, we have in distribution that

$$\frac{2}{N} \log \|X_{N,n}(\Theta)\| = \frac{1}{N} \sum_{i=1}^{N} T_i,$$

where T_i are independent and

$$T_i = \sum_{j=1}^{k} t_{i,j}, \qquad t_{i,j} \sim \log\left(\frac{1}{n} \chi^2_{n-j+1} \right) \quad iid. \tag{9.8}$$

Hence, we find in particular that

$$\mathbb{E}\left[\frac{1}{N} \log \|X_{N,n}(\Theta)\| \right] = \sum_{j=1}^{k} \mu_{n,j} \tag{9.9}$$

and see that Proposition 9.1 is equivalent to showing that for any $s > 0$

$$\mathbb{P}\left(\left| \frac{1}{nN} \sum_{i=1}^{N} \overline{T}_i \right| \geq s \right) \leq 2 \exp\left(-cnN \min\left\{ M_k s, \frac{s^2}{\xi_{n,k}} \right\} \right), \tag{9.10}$$

where for any random variable Y we will use the shorthand

$$\overline{Y} := Y - \mathbb{E}\left[Y \right].$$

We will obtain (9.10) by Markov's inequality applied to certain moments of the sum of the T_i's. Specifically, we will prove the following estimate.

PROPOSITION 9.6. *There exists a universal constant C so that for any n, N, k and $p \geq 1$*

$$\left(\mathbb{E}\left[\left| \sum_{i=1}^{N} \overline{T}_i \right|^p \right] \right)^{1/p} \leq C \left(\sqrt{pN \sum_{j=1}^{k} \frac{1}{M_j}} + \frac{p}{M_k} \right).$$

The proof of Proposition 9.6, which is straightforward but tedious, is given in Section 9.2 below. We assume it for now and complete the proof of (9.10). Write

$$p_0 := M_k^2 \sum_{j=1}^{k} \frac{1}{M_j} \tag{9.11}$$

and note that

$$p \leq Np_0 \quad \Longleftrightarrow \quad \frac{p}{M_k} \leq \sqrt{pN \sum_{j=1}^{k} \frac{1}{M_j}}.$$

Thus, applying Markov's inequality to Proposition 9.6 shows that there exists $C > 0$ so that for $1 \leq p \leq Np_0$

$$\mathbb{P}\left(\left| \frac{1}{nN} \sum_{i=1}^{N} \overline{T}_i \right| \geq \frac{C}{n\sqrt{N}} \sqrt{p \sum_{j=1}^{k} \frac{1}{M_j}} \right) \leq e^{-p}.$$

Equivalently, recalling that

$$\xi_{n,k} = \frac{1}{n} \sum_{j=1}^{k} \frac{1}{M_j},$$

we see that there exists $c > 0$ so that

$$\mathbb{P}\left(\left| \frac{1}{nN} \sum_{i=1}^{N} \overline{T}_i \right| \geq s \right) \leq 2e^{-cnN \frac{s^2}{\xi_{n,k}}}, \qquad 0 \leq s \leq C\xi_{n,k}M_k.$$

This establishes (9.10) in this range of s. To treat $s \geq CM_k \xi_{n,k}$, we again apply Markov's inequality to Proposition 9.6 to see that there exists $C > 0$ so that

$$p \geq Np_0 \quad \Longrightarrow \quad \mathbb{P}\left(\left| \sum_{i=1}^{N} \overline{T}_i \right| > C\frac{p}{M_k} \right) \leq e^{-p}.$$

Hence, there exists $c > 0$ so that

$$\mathbb{P}\left(\left| \frac{1}{nN} \sum_{i=1}^{N} \overline{T}_i \right| \geq s \right) \leq e^{-cnNM_k s}, \quad s \geq CM_k \xi_{n,k},$$

completing the proof of (9.10).

9.2 Proof of Proposition 9.6.   We seek to estimate the moments of

$$\sum_{i=1}^{N} \overline{T}_i = \sum_{i=1}^{N} \sum_{j=1}^{k} \bar{t}_{i,j}, \qquad t_{i,j} \sim \log(\frac{1}{n}\chi^2_{n-j+1}).$$

By Theorem 6.1, we have

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{N}\overline{T}_i\right|^p\right]\right)^{1/p} \simeq \sup_{\max\{2,\frac{p}{N}\}\leq s\leq p} \frac{p}{s}\left(\frac{N}{p}\right)^{\frac{1}{s}} \mathbb{E}\left[\left|\overline{T}_i\right|^s\right]^{\frac{1}{s}}. \tag{9.12}$$

Our strategy is therefore to estimate $\mathbb{E}\left[\left|\overline{T}_i\right|^s\right]^{1/s}$ and optimize over s. In particular, we seek to show that there exists $C > 0$ so that for every $p \geq 1$, we have

$$\left(\mathbb{E}\left[\left|\overline{T}_i\right|^p\right]\right)^{1/p} \leq C\left(\sqrt{p\sum_{j=1}^{k} M_j^{-1}} + \frac{p}{M_k}\right). \tag{9.13}$$

Our first step towards showing (9.13) is to obtain the following estimates on the moments of $\bar{t}_{i,j}$.

LEMMA 9.7. *There exist $C_1 > 0$ such that*

$$\left(\mathbb{E}\left[\left|\bar{t}_{i,j}\right|^p\right]\right)^{1/p} \leq C_1 \max\left\{\sqrt{\frac{p}{M_j}}, \frac{p}{M_j}\right\}.$$

Proof. We first make a reduction. Namely, let us check that the estimates in Lemma 9.7 for $\bar{t}_{i,j} = t_{i,j} - \mathbb{E}\left[\log(n^{-1}\chi^2_{M_j})\right]$ follow from the same estimates for

$$\widehat{t}_{i,j} := t_{i,j} - \log(n^{-1}M_j).$$

To see this, recall that by (1.16) and (1.7), we have

$$\mathbb{E}\left[\log\left(\frac{1}{n}\chi^2_{M_j}\right)\right] = \log\left(\frac{2}{n}\right) + \psi\left(\frac{M_j}{2}\right) = \log\left(\frac{M_j}{n}\right) + \varepsilon_j, \qquad \varepsilon_j = O(M_j^{-1})$$

where ψ is the digamma function, and we have used its asymptotic expansion $\psi(z) \sim \log(z) + O(z^{-1})$ for large arguments. Thus, we have for each i that

$$\mathbb{E}\left[\left|\bar{t}_{i,j}\right|^p\right] = \mathbb{E}\left[\left|\widehat{t}_{i,j} + \varepsilon_j\right|^p\right] \leq \sum_{k=0}^{p}\binom{p}{k}\mathbb{E}\left[\left|\widehat{t}_{i,j}\right|^k\right]|\varepsilon_j|^{p-k}.$$

So assuming that $\widehat{t}_{i,j}$ satisfy the conclusion of Lemma 9.7, we find

$$\mathbb{E}\left[\left|\bar{t}_{i,j}\right|^p\right] \leq \sum_{k=0}^{p}\binom{p}{k}\zeta_{k,j}^k|\varepsilon_j|^{p-k}, \qquad \zeta_{k,j} := C\max\left\{\sqrt{\frac{k}{M_j}}, \frac{k}{M_j}\right\}.$$

Since for $0 \leq k \leq p$ we have $\zeta_{k,j} \leq \zeta_{p,j}$, we see that

$$\mathbb{E}\left[|\bar{t}_{i,j}|^p\right] \leq \sum_{k=0}^{p} \binom{p}{k} \zeta_{p,j}^k |\varepsilon_j|^{p-k} \leq (\zeta_{p,j} + |\varepsilon_j|)^p.$$

Finally, since $\varepsilon_j = O(M_j^{-1}) = O(\zeta_{p,j})$, we find that there exists $C > 0$ so that

$$\left(\mathbb{E}\left[|\bar{t}_{i,j}|^p\right]\right)^{1/p} \leq C\zeta_{p,j} \leq C \max\left\{\sqrt{\frac{p}{M_j}}, \frac{p}{M_j}\right\},$$

as desired. It therefore remains to show that $\widehat{t}_{i,j} = t_{i,j} - \log\left(n^{-1}M_j\right)$ satisfies the conclusion of Lemma 9.7. To do this, we begin by checking that there exists $c_1 > 0$ so that, with $M_j = n - j + 1$, for all $s \geq 0$

$$\mathbb{P}\left(\left|t_{i,j} - \log\left(\frac{M_j}{n}\right)\right| \geq s\right) \leq 4e^{-c_1 M_j \min\{s,s^2\}}. \tag{9.14}$$

We have

$$\mathbb{P}\left(\left|t_{i,j} - \log\left(\frac{M_j}{n}\right)\right| \geq s\right) = \mathbb{P}\left(\left|\log\left(\frac{1}{n}\chi_{M_j}^2\right) - \log\left(\frac{M_j}{n}\right)\right| \geq s\right)$$

$$= \mathbb{P}\left(\left|\log\left(\frac{1}{M_j}\chi_{M_j}^2\right)\right| \geq s\right)$$

$$= \mathbb{P}\left(\chi_{M_j}^2 \geq M_j e^s\right) + \mathbb{P}\left(\chi_{M_j}^2 \leq M_j e^{-s}\right).$$

Let us first bound $\mathbb{P}\left(\chi_{M_j}^2 \geq M_j e^s\right)$. Note that the mean of $\chi_{M_j}^2$ is M_j and that $\chi_{M_j}^2 - M_j$ is a sub-exponential random variable with parameters $(4M_j, 4)$. Thus, Bernstein's tail estimates for sub-exponential random variables yield the existence of $c > 0$ such that for all $t \geq 0$

$$\mathbb{P}\left(\left|M_j^{-1}\chi_{M_j}^2 - 1\right| \geq t\right) \leq 2e^{-cM_j \min\{t,t^2\}}. \tag{9.15}$$

In particular,

$$\mathbb{P}\left(\chi_{M_j}^2 \geq M_j e^s\right) = \mathbb{P}\left(M_j^{-1}\chi_{M_j}^2 - 1 \geq e^s - 1\right) \leq 2e^{-cM_j \min\{e^s-1,(e^s-1)^2\}} \leq 2e^{-cM_j \min\{s,s^2\}},$$

where in the last inequality we used that $e^s - 1 \geq s$. This gives the first half of (9.14).

Let us now obtain a similar estimate on $\mathbb{P}\left(\chi_{M_j}^2 \leq M_j e^{-s}\right)$, which is a small ball estimate for a sum of iid random variables with bounded density. We need to consider two cases. Theorem 6.2 on small ball estimates for sums of iid random variables shows that there exists a universal constant $C > 0$ so that

$$\mathbb{P}\left(\chi_{M_j}^2 \leq M_j e^{-s}\right) \leq (Ce^{-s/2})^{M_j}.$$

Hence,

$$s > s_* := 2\log(C) \quad \Rightarrow \quad \exists c > 0 \text{ s.t. } \mathbb{P}\left(\chi^2_{M_j} \leq M_j e^{-s}\right) \leq e^{-csM_j},$$

giving the desired estimate in this range. Finally, suppose $s \leq s_*$. Then,

$$e^{-s} \leq 1 - e^{-s_*} s$$

since both sides equal 1 at $s = 0$ and the derivative $-e^{-s}$ of the left hand side is more negative than the derivative $-e^{-s_*}$ of the right hand side for all $s \in (0, s_*)$. Thus, we find for $s \in (0, s_*)$ that

$$\mathbb{P}\left(\chi^2_{M_j} \leq M_j e^{-s}\right) \leq \mathbb{P}\left(\chi^2_{M_j} \leq M_j(1 - e^{-s_*} s)\right) \leq \mathbb{P}\left(\left|M_j^{-1}\chi^2_{M_j} - 1\right| > e^{-s_*} s\right).$$

Using the Bernstein inequality (9.15), we obtain that there exists $c > 0$ depending only on s_* such that for $s \in (0, s_*)$,

$$\mathbb{P}\left(\chi^2_{M_j} \leq M_j e^{-s}\right) \leq 2e^{-cM_j \min\{s, s^2\}},$$

giving the desired estimate in this range as well. This establishes (9.14). To complete the proof that $\widehat{t}_{i,j}$ satisfy Lemma 9.7, we use (9.14) to write

$$\mathbb{E}\left[|\widehat{t}_{i,j}|^p\right] = \int_0^\infty \mathbb{P}\left(|\widehat{t}_{i,j}| > x\right) px^{p-1} dx$$

$$\leq p\left[\int_0^1 e^{-cM_j x^2} x^{p-1} dx + \int_1^\infty e^{-cM_j x} x^{p-1} dx\right].$$

The first term can be estimated by comparing to the moments of a Gaussian as follows:

$$p\int_0^1 e^{-cM_j x^2} x^{p-1} dx = p(2cM_j)^{-p/2}\int_0^{(2cM_j)^{1/2}} e^{-x^2} x^{p-1} dx$$

$$\leq p(2cM_j)^{-p/2}\int_0^\infty e^{-x^2} x^{p-1} dx$$

$$\leq p(2cM_j)^{-p/2} 2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)$$

$$\leq p\left(\frac{p}{2cM_j}\right)^{p/2},$$

where we used that for $z > 0$ we have $\Gamma(z) \leq z^z$. The second term can similarly be estimated by comparing the moments of an exponential:

$$p\int_1^\infty e^{-cM_j x} x^{p-1} dx = p(cM_j)^{-p}\int_{cM_j}^\infty e^{-x} x^{p-1} dx$$

$$= p(cM_j)^{-p}(p-1)!$$

$$\leq p\left(\frac{p}{cM_j}\right)^p.$$

Putting these two estimates together and taking $1/p$ powers, we find that there exists $C > 0$ so that

$$\left(\mathbb{E}\left[|\widehat{t}_{i,j}|^p\right]\right)^{1/p} \leq C \max\left\{\sqrt{\frac{p}{M_j}}, \frac{p}{M_j}\right\}$$

for all $p \geq 1$. This completes the proof.   □

With Lemma 9.7 in hand, we are now in a position to show (9.13). Since

$$\overline{T}_i = \sum_{j=1}^{k} \overline{t}_{i,j},$$

we have by Theorem 6.1 that

$$\left(\mathbb{E}|\overline{T}_i|^p\right)^{\frac{1}{p}} \simeq \inf\left\{t > 0 : \sum_{j=1}^{k} \log\left[\mathbb{E}|1 + \frac{\overline{t}_{i,j}}{t}|^p\right] \leq p\right\}, \tag{9.16}$$

where $\simeq$ means bounded above and below by absolute constants. We will use the notation from (9.11):

$$p_0 = M_k^2 \sum_{j=1}^{k} M_j^{-1}.$$

Since

$$\sqrt{p \sum_{j=1}^{k} M_j^{-1}} \leq \frac{p}{M_k} \qquad \Longleftrightarrow \qquad p \geq p_0,$$

we will show (9.13) by breaking into two cases. Namely, we will show that there exists $C > 0$ so that

$$p \leq p_0 \quad \Longrightarrow \quad \left(\mathbb{E}\left|\overline{T}_i\right|^p\right)^{\frac{1}{p}} \leq C\sqrt{p \sum_{j=1}^{k} M_j^{-1}} \tag{9.17}$$

as well as

$$p \geq p_0 \quad \Longrightarrow \quad \left(\mathbb{E}\left|\overline{T}_i\right|^p\right)^{\frac{1}{p}} \leq C\frac{p}{M_k}. \tag{9.18}$$

We may assume without loss of generality that p is an even integer. Indeed, once we've show (9.17) and (9.18) for even integers p (and a uniform constant C), we may use that

$$\left(\mathbb{E}\left[|\overline{T}_i|^p\right]\right)^{1/p} \leq \left(\mathbb{E}\left[|\overline{T}_i|^{p'}\right]\right)^{1/p'},$$

where p' is the smallest even integer greater than or equal to p. This yields (9.17) and (9.18) for any p with C replaced by $2C$.

We now turn to showing that (9.17) holds with p an even integer. To do this, we will need to evaluate the expectation $\mathbb{E}\left[|1 + t^{-1}t_{i,j}|^p\right]$ appearing in (9.16). A key point is to use that $\bar{t}_{i,j}$ are centered. Since p is even, we may bring this to bear most directly by noting that the absolute value in $\mathbb{E}\left[|1 + t^{-1}t_{i,j}|^p\right]$ is unnecessary and using that $\mathbb{E}\left[\bar{t}_{i,j}\right] = 0$. Lemma 9.7 and the well-known estimate

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{n}{k}\right)^k e^k, \qquad k \geq 1 \tag{9.19}$$

yield that for all i, j we have:

$$
\begin{aligned}
\mathbb{E}\left[\left(1 + \frac{\bar{t}_{i,j}}{t}\right)^p\right] &= 1 + \sum_{\ell=2}^{p}\binom{p}{\ell}\frac{\mathbb{E}\left[\bar{t}_{i,j}^\ell\right]}{t^\ell} \\
&\leq 1 + \sum_{\ell=2}^{\min\{p,M_j\}}\binom{p}{\ell}\left(\frac{C\ell}{t^2 M_j}\right)^{\ell/2} + \sum_{\ell=M_j+1}^{p}\binom{p}{\ell}\left(\frac{C\ell}{t M_j}\right)^\ell \\
&\leq 1 + \sum_{\ell=2}^{\min\{p,M_j\}}\left(\frac{p}{\ell}\right)^\ell\left(\frac{C\ell}{t^2 M_j}\right)^{\ell/2} + \sum_{\ell=M_j+1}^{p}\left(\frac{Cp}{t M_j}\right)^\ell \\
&\leq 1 + \sum_{\ell=2}^{\min\{p,M_j\}}\left(\frac{p}{\ell}\right)^{\ell/2}\left(\frac{Cp}{t^2 M_j}\right)^{\ell/2} + \sum_{\ell=M_j+1}^{p}\left(\frac{Cp}{t M_j}\right)^\ell. \tag{9.20}
\end{aligned}
$$

We now bound the first two terms in the previous line by breaking into the terms where ℓ is even and odd. When ℓ is even the terms in (9.20) are bounded above by

$$1 + \sum_{\substack{\ell=2 \\ \ell \text{ even}}}^{\min\{p,M_j\}}\left(\frac{p}{\ell}\right)^{\ell/2}\left(\frac{C'p}{t^2 M_j}\right)^{\ell/2} \leq 1 + \sum_{\substack{\ell=2 \\ \ell \text{ even}}}^{\min\{p,M_j\}}\binom{p/2}{\ell/2}\left(\frac{C'p}{t^2 M_j}\right)^{\ell/2} \leq \left(1 + \frac{C'p}{t^2 M_j}\right)^{p/2}. \tag{9.21}$$

To bound the odd terms in (9.20), let us first note that for any $0 \leq m \leq \ell \leq p$, we have

$$\left(\frac{p}{\ell}\right)^\ell \leq p^m\binom{p}{\ell - m}.$$

Indeed, when $\ell = m$, this equality simply reads

$$\left(\frac{p}{\ell}\right)^\ell \leq p^{\ell+1},$$

while if $m < \ell$, we note that the inequality is equivalent to

$$\frac{p^{\ell-m}}{\ell^\ell} \leq \binom{p}{\ell - m},$$

which follows by estimating the expression on the left hand side using (9.19) as follows:

$$\left(\frac{p}{\ell}\right)^{\ell-m}\frac{1}{\ell^m} = \left(\frac{p}{\ell-m}\right)^{\ell-m}\left(\frac{\ell-m}{\ell}\right)^{\ell-m}\frac{1}{\ell^m} \leq \binom{p}{\ell-m}.$$

Thus, the odd terms in (9.20) are bounded above by

$$\sum_{\substack{\ell=3 \\ \ell \text{ odd}}}^{\min\{p,M_j\}}\left(\frac{p}{\ell}\right)^{\ell/2}\left(\frac{Cp}{t^2M_j}\right)^{\ell/2} \leq \min_{m=1,3}\left\{\left(\frac{Cp^2}{t^2M_j}\right)^{\frac{m}{2}}\sum_{\substack{\ell=3 \\ \ell \text{ odd}}}^{p-1}\binom{p}{\ell-m}^{\frac{1}{2}}\left(\frac{Cp}{t^2M_j}\right)^{\frac{\ell-m}{2}}\right\}.$$

To proceed, note that for any $0 \leq b \leq a$

$$\binom{2a}{2b} \leq 2^b\binom{a}{b}^2.$$

This inequality follows by observing that for any $j = 0,\ldots,b-1$, we have

$$\frac{(2a-2j)(2a-2j-1)}{(2b-2j-1)} = \frac{(a-j)(a-j-1/2)}{(b-j)(b-j-1/2)} \leq 2\left(\frac{a-j}{b-j}\right)^2$$

and repeatedly applying this estimate to the terms in $\binom{2a}{2b}$. Thus, we obtain

$$\sum_{\substack{\ell=3 \\ \ell \text{ odd}}}^{\min\{p,M_j\}}\left(\frac{p}{\ell}\right)^{\ell/2}\left(\frac{Cp}{t^2M_j}\right)^{\ell/2} \leq \min_{m=1,3}\left\{\left(\frac{Cp^2}{t^2M_j}\right)^{\frac{m}{2}}\right\}\sum_{\ell=0}^{p/2}\binom{p/2}{\ell}\left(\frac{Cp}{t^2M_j}\right)^{\ell}$$

$$= \min_{m=1,3}\left\{\left(\frac{Cp^2}{t^2M_j}\right)^{\frac{m}{2}}\right\}\left(1+\frac{Cp}{t^2M_j}\right)^{p/2}$$

$$\leq \left(\frac{Cp^2}{t^2M_j}\right)\left(1+\frac{Cp}{t^2M_j}\right)^{p/2}, \qquad (9.22)$$

where in the last inequality we've used that $\min\left\{x^{1/2}, x^{3/2}\right\} \leq x$ for all $x \geq 0$. Putting together (9.21) and (9.22) we see that there exists $C > 0$ so that

$$\mathbb{E}\left[\left(1+\frac{\bar{t}_{i,j}}{t}\right)^p\right] \leq \left(1+\frac{Cp^2}{t^2M_j}\right)\left(1+\frac{Cp}{t^2M_j}\right)^{p/2} + \sum_{\ell=M_j+1}^{p}\left(\frac{Cp}{tM_j}\right)^{\ell}. \quad (9.23)$$

Let us now verify (9.17). Recall that for $j \leq k$ we have $M_j \geq M_k$ and that $p \leq p_0 = M_k^2\sum_{j=1}^{k}\frac{1}{M_j}$. We set

$$t = \sqrt{C'p\sum_{i=1}^{k}M_i^{-1}} \qquad (9.24)$$

where

$$C' = \max\left\{(16C)^2, \, 2C^{1/2}\right\}$$

is an absolute constant depending only on the constant C appearing in (9.23). For this choice of C', we have

$$\frac{Cp}{tM_j} = C\sqrt{\frac{p^2}{C'pM_j^2 \sum_{i=1}^k M_i^{-1}}} \leq C\sqrt{\frac{p}{C'M_k^2 \sum_{i=1}^k M_i^{-1}}} = C\sqrt{\frac{p}{C'p_0}} \leq \frac{1}{16}, \forall j \leq k$$

and

$$\sum_{j=1}^k \frac{Cp}{t^2 M_j} \leq \frac{C}{(C')^2} \leq \frac{1}{4}.$$

In particular for $j \leq k$ such that $M_j \leq p$,

$$\sum_{\ell=M_j+1}^p \left(\frac{Cp}{tM_j}\right)^\ell \leq \sum_{\ell=M_j+1}^p \frac{1}{16^\ell} \leq \frac{1}{4^{M_j}}$$

Hence, since $\log(a+b) \leq (\log a) + b$ for $a \geq 1$ and $b > 0$,

$$\sum_{j=1}^k \log \mathbb{E}\left[\left(1 + \left(\frac{\bar{t}_{i,j}}{t}\right)^p\right)\right] \leq \frac{p}{2}\sum_{j=1}^k \log\left(1 + \frac{Cp}{t^2 M_j}\right) + \sum_{j=1}^k \log\left(1 + \frac{Cp^2}{t^2 M_j}\right) + \sum_{j=1}^k \left(\frac{1}{4}\right)^{M_j}$$

$$\leq \frac{p}{2}\sum_{j=1}^k \frac{Cp}{t^2 M_j} + \sum_{j=1}^k \frac{Cp^2}{t^2 M_j} + \sum_{s=n-k+1}^n \frac{1}{4^s}$$

$$= \frac{3p}{8} + \frac{1}{2}$$

$$\leq p.$$

Hence, (9.17) follows from (9.16). We now turn to the case when $p \geq p_0$ and seek to show (9.18). Rather than (9.24), to show (9.18), we set

$$t = \frac{C'p}{M_k} \tag{9.25}$$

with

$$C' = \max\left\{4C, \, 2C^{1/2}\right\}.$$

Then,

$$\frac{Cp}{tM_j} = \frac{CM_k}{C'M_j} < \frac{C}{C'} \leq \frac{1}{4}$$

and

$$\sum_{j=1}^{k} \frac{Cp}{t^2 M_j} = \sum_{j=1}^{k} \frac{CM_k^2}{(C')^2 p M_j} = \frac{C}{(C')^2} \frac{p_0}{p} \le \frac{C}{(C')^2} \le \frac{1}{4}.$$

Hence from (9.23), we find

$$\sum_{j=1}^{k} \log \mathbb{E}\left[\left(1 + \left(\frac{\overline{t}_{i,j}}{t}\right)^p\right)\right] \le \frac{p}{2} \sum_{j=1}^{k} \log\left(1 + \frac{Cp}{t^2 M_j}\right) + \sum_{j=1}^{k} \log\left(1 + \frac{Cp^2}{t^2 M_j}\right) + \sum_{j=1}^{k} \left(\frac{1}{4}\right)^{M_j}$$

$$\le \frac{p}{2} \sum_{j=1}^{k} \frac{Cp}{t^2 M_j} + \sum_{j=1}^{k} \frac{Cp^2}{t^2 M_j} + \frac{1}{2}$$

$$\le \frac{3Cp}{2(C')^2} + \frac{1}{2}$$

$$\le p.$$

Thus, we see that relation (9.18) also follows from (9.16). We are now in a position to finish the proof of Proposition 9.1 by combining (9.12) with (9.17) and (9.18). We find from (9.12) that

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \overline{T}_i\right|^p\right]\right)^{1/p} \le \sup_{\max\{2, \frac{p}{N}\} \le s \le p} \frac{p}{s} \left(\frac{N}{p}\right)^{1/s} \mathbb{E}\left[|\overline{T}_i|^s\right]^{1/s}.$$

If $p \le p_0$, then (9.17) implies

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \overline{T}_i\right|^p\right]\right)^{1/p} \le C \sup_{\max\{2, \frac{p}{N}\} \le s \le p} \frac{p}{s} \left(\frac{N}{p}\right)^{1/s} \sqrt{s \sum_{j=1}^{k} M_j^{-1}}. \qquad (9.26)$$

Define

$$f(s) = \log\left(\frac{p}{s} \left(\frac{N}{p}\right)^{1/s} \sqrt{s}\right) = \log(p) - \frac{1}{2} \log(s) + \frac{1}{s} \log\left(\frac{N}{p}\right).$$

Note that

$$f'(s) = -\frac{1}{2s} - \frac{1}{s^2} \log\left(\frac{N}{p}\right).$$

When $p \le N$ the function f is manifestly monotone decreasing in $s > 0$. Therefore, taking $s = 2$, we find

$$p \le N \quad \Rightarrow \quad \left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \overline{T}_i\right|^p\right]\right)^{1/p} \le C \sqrt{pN \sum_{j=1}^{k} M_j^{-1}}.$$

On the other hand, when $p \geq N$, we have

$$\frac{p}{s} \left(\frac{N}{p}\right)^{1/s} \sqrt{s \sum_{j=1}^{k} M_j^{-1}} \leq p s^{-1/2} \sqrt{\sum_{j=1}^{k} M_j^{-1}},$$

which is strictly decreasing in $s > 0$. Hence, taking $s = p/N$, we again find

$$p \geq N \quad \Rightarrow \quad \left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \overline{T}_i\right|^p\right]\right)^{1/p} \leq C \sqrt{pN \sum_{j=1}^{k} M_j^{-1}}.$$

Hence, we find that if $p \leq p_0$, then

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \overline{T}_i\right|^p\right]\right)^{1/p} \leq C \sqrt{pN \sum_{j=1}^{k} M_j^{-1}},$$

as desired. Finally, if $p \geq p_0$, then (9.18) implies

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \overline{T}_i\right|^p\right]\right)^{1/p} \leq C \sup_{\max\{2,\frac{p}{N}\} \leq s \leq p} \frac{p}{s} \left(\frac{N}{p}\right)^{1/s} \frac{s}{M_k} = \frac{Cp}{M_k} \sup_{\max\{2,\frac{p}{N}\} \leq s \leq p} \left(\frac{N}{p}\right)^{1/s}. \tag{9.27}$$

Note that if $p/N \geq 1$, then for every $s \geq 2$ we have $(N/p)^{1/s} \leq 1$ and hence

$$\frac{Cp}{M_k} \sup_{\max\{2,\frac{p}{N}\} \leq s \leq p} \left(\frac{N}{p}\right)^{1/s} \leq \frac{Cp}{M_k}.$$

Further, if $p/N \leq 1$, then $(N/p)^{1/s}$ is monotonically decreasing with s and hence

$$\frac{Cp}{M_k} \sup_{\max\{2,\frac{p}{N}\} \leq s \leq p} \left(\frac{N}{p}\right)^{1/s} \leq \frac{Cp}{M_k} \left(\frac{N}{p}\right)^{1/2} = C\sqrt{pN M_k^{-2}} \leq C \sqrt{pN \sum_{j=1}^{k} M_j^{-1}}.$$

Thus, in all cases we find

$$\left(\mathbb{E}\left[\left|\sum_{i=1}^{N} \overline{T}_i\right|^p\right]\right)^{1/p} \leq C \left(\sqrt{pN \sum_{j=1}^{k} \frac{1}{M_j}} + \frac{p}{M_k}\right),$$

which is precisely the statement of Proposition 9.6.   □

10   Lyapunov Sums: Proof of Theorem 1.1

To prove Theorem 1.1, we start by combining previous estimates for $\frac{1}{N} \log \|X_{N,n}(\Theta)\|$ from Proposition 9.1 with the deviation estimates in Proposition 8.1. Let recall the definition of the function g (cf (1.4))

$$g_{n,k}(s) = \begin{cases} \min\left\{1, \frac{ns}{k}\right\}, & k \leq \frac{n}{2}, \\ \min\left\{\delta_{n,k}, \frac{s}{\log 1/\delta_{n,k}}\right\}, & k > \frac{n}{2}, \end{cases} \tag{10.1}$$

where we recall that for $k \geq \frac{n}{2}$

$$\delta_{n,k} := \frac{n-k+1}{n}.$$

Let $s > 0$ and $1 \leq k < m \leq n$. Writing

$$p_{s,k,m} := \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=m}^{k}(\lambda_i - \mu_{n,i})\right| \geq s\right),$$

we seek to show that there exist universal constants $c_1, c_2, c_3 > 0$ such that for any $1 \leq m \leq k \leq n$ and every $s \geq c_1\frac{k}{nN}\log\frac{en}{k}$ we have

$$p_{s,k,m} \leq c_2 \exp\left\{-c_3 \min\left\{nNs, n^2Nsg_{n,k}(s)\right\}\right\}. \tag{10.2}$$

To see this, note that the triangle inequality yields

$$p_{s,k,m} \leq p_{s/2,k,1} + p_{s/2,m-1,1}.$$

Hence, it suffices to prove (10.2) with $m = 1$. To do this, fix $\Theta \in \mathrm{Fr}_{n,k}$, an orthonormal k-frame in $\mathbb{R}^n$. We may write for any $s > 0$

$$p_{s,k,1} \leq \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{k}\lambda_i - \frac{1}{nN}\log\|X_{N,n}(\Theta)\|\right| \geq \frac{s}{2}\right) \tag{10.3}$$

$$+ \mathbb{P}\left(\left|\frac{1}{nN}\log\|X_{N,n}(\Theta)\| - \frac{1}{n}\sum_{i=1}^{k}\mu_{n,i}\right| \geq \frac{s}{2}\right). \tag{10.4}$$

We will show separately that the probabilities in (10.3) and (10.4) are both bounded above by the right hand side of (10.2). To check this for (10.4), note that Proposition 9.1 guarantees

$$\mathbb{P}\left(\left|\frac{\log\|X_{N,n}(\Theta)\|}{nN} - \frac{1}{n}\sum_{j=1}^{k}\mu_{n,j}\right| \geq s\right) \leq 2\exp\left\{-cnN\min\{M_ks, \frac{s^2}{\xi_{n,k}}\}\right\}, s > 0,$$

$$\tag{10.5}$$

where we remind the reader that

$$M_j := n - j + 1, \ \xi_{n,k} := \frac{1}{n} \sum_{j=1}^{k} \frac{1}{M_j}, \ \mu_{n,k} := \frac{1}{2} \mathbb{E} \left[\log \left(\frac{1}{n} \chi^2_{n-j+1} \right) \right].$$

Some routine algebra reveals

$$k \leq \frac{n}{2} \qquad \Rightarrow \qquad n\xi_{n,k} \simeq \frac{k}{n}, \qquad M_k \simeq n \qquad (10.6)$$

and

$$k \geq \frac{n}{2} \qquad \Rightarrow \qquad n\xi_{n,k} \simeq \log \left(\frac{1}{\delta_{n,k}} \right), \qquad M_k = \delta_{n,k} n, \qquad (10.7)$$

where $a \simeq b$ means that there exists $c_1, c_2 > 0$ so that $c_1 a \leq b \leq c_2 a$. Hence,

$$k \leq \frac{n}{2} \quad \Rightarrow \quad \min \left\{ M_k s, \frac{s^2}{\xi_{n,k}} \right\} \simeq ns \min \left\{ 1, \frac{ns}{k} \right\}$$

and similarly

$$k \geq \frac{n}{2} \quad \Rightarrow \quad \min \left\{ M_k s, \frac{s^2}{\xi_{n,k}} \right\} \simeq ns \left\{ \delta_{n,k}, \frac{s}{\log(1/\delta_{n,k})} \right\}.$$

Putting these two estimates together, we find that (10.5) yields for any $s > 0$

$$\mathbb{P} \left(\left| \frac{\log ||X_{N,n}(\Theta)||}{nN} - \frac{1}{n} \sum_{j=1}^{k} \mu_{n,j} \right| \geq s \right) \leq 2 \exp \left\{ -cn^2 N s g_{n,k}(s) \right\}, \qquad (10.8)$$

as desired. Turning to the probability in (10.3), recall that in Proposition 8.1, we have shown that for every $\varepsilon \in (0,1)$,

$$\mathbb{P} \left(\left| \frac{1}{nN} \log ||X_{N,n}(\Theta)|| - \frac{1}{n} \sum_{i=1}^{k} \lambda_i \right| \geq \frac{k}{2Nn} \log \left(\frac{n}{k\varepsilon^2} \right) \right) \leq (C\varepsilon)^{\frac{k}{2}}.$$

If we set $s := \frac{k}{nN} \log \frac{en}{k\varepsilon^2}$, then

$$(C\varepsilon)^{k/2} = \exp \left[-\frac{1}{4} snN + \frac{k}{4} \log \left(\frac{en}{k} \right) + \frac{k}{2} \log(C) \right].$$

Hence, assuming that

$$s \geq C' \frac{k}{nN} \log \left(\frac{en}{k} \right)$$

for C' sufficiently large, we arrive to the following expression:

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^{k} \lambda_i - \frac{\log ||X_{N,n}(\Theta)||}{nN} \right| \geq s \right) \leq e^{-\frac{snN}{4}}, \quad s \geq C' \frac{k}{nN} \log \left(\frac{en}{k} \right). \ (10.9)$$

Thus, putting together (10.8) and (10.9), we find that

$$p_{s,k,1} \leq c_2 \exp \left\{ -c_3 \min \left\{ nNs, n^2 N s g_{n,k}(s) \right\} \right\},$$

completing the proof. $\qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \square$

11   Convergence to the Triangle Law: Proof of Theorem 1.2

In this section, we derive Theorem 1.2 from Theorem 1.1. We will need the following elementary result.

LEMMA 11.1. *Fix positive integers* n, q, m *satisfying* $4 \leq m \leq q \leq n$. *Then*

$$\frac{m}{2n} \log\left(q/n\right) - \frac{1}{n} \sum_{j=n-q+1}^{n-q+m} \mu_{n,j} \geq \frac{(m-1)^2}{4nq}.$$

Further, assuming that $n - q - m \geq 0$, *we also have*

$$\frac{m}{2n} \log(q/n) - \frac{1}{n} \sum_{j=n-q-m+1}^{n-q} \mu_{n,j} \leq -\frac{m(m-3)}{3nq}.$$

Proof. Let us first prove the lower bound. Recall that

$$\mu_{n,j} = \frac{1}{2}\left(\log\left(\frac{2}{n}\right) + \psi\left(\frac{n-j+1}{2}\right)\right). \tag{11.1}$$

Moreover, a well-known estimate [AS64, Equation 6.3.18, p.259] for the digamma function is:

$$\psi(x) < \log(x).$$

Using this we obtain

$$\mu_{n,j} < \frac{1}{2} \log\left(1 - \frac{j-1}{n}\right),$$

which allows us to write

$$\frac{m}{2n} \log\left(q/n\right) - \frac{1}{n} \sum_{j=n-q+1}^{n-q+m} \mu_{n,j} \geq \frac{1}{2n} \sum_{j=n-q+1}^{n-q+m} \log\left(\frac{q}{n-j+1}\right) = \frac{1}{2n} \sum_{j=1}^{m-1} \log\left(\frac{1}{1-j/q}\right).$$

Since q, n are fixed, let us temporarily introduce

$$\xi := \frac{q}{n}.$$

With this notation, because $\log(1/(1-t))$ is monotonically increasing for $t \in [0, 1)$, we have

$$\frac{1}{2n} \sum_{j=1}^{m-1} \log\left(\frac{1}{1-j/q}\right) = \frac{1}{2} \sum_{j=1}^{m-1} \frac{1}{n} \log\left(\frac{1}{1-(j/n)/\xi}\right) \geq \frac{1}{2} \int_0^{\frac{m-1}{n}} \log\left(\frac{1}{1-t/\xi}\right) dt.$$

Some routine calculus therefore reveals that

$$\frac{m}{2n} \log\left(q/n\right) - \frac{1}{n} \sum_{j=n-q+1}^{n-q+m} \mu_{n,j} \geq \frac{\xi}{2}\left[(1-\varepsilon/\xi) \log\left(1-\varepsilon/\xi\right) + \varepsilon/\xi\right],$$

where we've set $\varepsilon := (m-1)/n$. Finally, note that for $x > 0$

$$(1-x)\log(1-x) + x = \sum_{k \geq 2} \frac{x^k}{k(k-1)} \geq x^2/2.$$

Hence, we obtain

$$\frac{m}{2n}\log\left(q/n\right) - \frac{1}{n}\sum_{j=n-q+1}^{n-q+m} \mu_{n,j} \geq \frac{\varepsilon^2}{4\xi} = \frac{(m-1)^2}{4nq},$$

as claimed. Let us now derive the upper bound. Using (11.1), we get

$$\frac{m}{2n}\log\left(q/n\right) - \frac{1}{n}\sum_{j=n-q-m+1}^{n-q} \mu_{n,j} = \frac{1}{2n}\sum_{j=n-q-m+1}^{n-q} \left\{\log\left(\frac{q}{2}\right) - \psi\left(\frac{n-j+1}{2}\right)\right\}$$

$$= \frac{1}{2n}\sum_{j=1}^{m}\left\{\log\left(\frac{q}{2}\right) - \psi\left(\frac{q+j}{2}\right)\right\}.$$

Using the inequality (see again [AS64, Equation 6.3.18, p.259])

$$\psi(x) > \log(x) - 1/x,$$

we arrive at the estimate

$$\frac{m}{2n}\log\left(q/n\right) - \frac{1}{n}\sum_{j=n-q-m+1}^{n-q} \mu_{n,j} \leq \frac{1}{2n}\sum_{j=1}^{m}\left\{\log\left(\frac{q}{q+j}\right) + \frac{2}{q+j}\right\}.$$

As before, we will estimate this sum above by an integral. Still writing $\xi = q/n$, we have as before

$$\frac{1}{2n}\sum_{j=1}^{m}\log\left(\frac{1}{1+j/q}\right) \leq \frac{\xi}{2}\int_0^{\varepsilon/\xi}\log\left(\frac{1}{1+t}\right)dt = \frac{\xi}{2}\left[-(1+\varepsilon/\xi)\log\left(1+\varepsilon/\xi\right) + \varepsilon/\xi\right]$$

where we've now set $\varepsilon = m/n$ (which is slightly different than above). For $x \in (0,1)$, we have

$$-(1+x)\log(1+x) + x = \sum_{k \geq 2}(-1)^{k+1}\frac{x^k}{k(k-1)} \leq -\frac{x^2}{2} + \frac{x^3}{6} \leq -\frac{x^2}{3},$$

we therefore find

$$\frac{1}{2n}\sum_{j=1}^{m}\log\left(\frac{1}{1+j/q}\right) \leq -\frac{\varepsilon^2}{3\xi}.$$

Next,

$$\frac{1}{q}\sum_{j=1}^{m}\frac{1}{n}\frac{1}{1+\xi^{-1}(j/n)} \leq \frac{1}{q}\int_0^{\varepsilon}\frac{dt}{1+t/\xi} = \frac{\xi}{q}\log\left(1+\varepsilon/\xi\right) \leq \frac{\varepsilon}{q}.$$

So all together we find the upper bound

$$\frac{m}{2n} \log{(q/n)} - \frac{1}{n} \sum_{j=n-q-m+1}^{n-q} \mu_{n,j} \leq -\frac{m^2}{3nq} + \frac{m}{nq} = -\frac{m(m-3)}{3nq}.$$

This completes the proof.   □

We now conclude the proof of Theorem 1.2. To do this, fix $\varepsilon > 0$ and assume that

$$n > \frac{c_1 \sqrt{\log(1/\varepsilon)}}{\varepsilon}, \qquad N > \frac{c_2}{\varepsilon^2}$$

for some constants $c_1, c_2 > 1$ that we will fix later. To prove Theorem 1.2 note that the bound above on n guarantees that $\varepsilon > c_1/n$. Hence, we need only consider such ε. Moreover, we may always assume that

$$\varepsilon = \frac{m}{n}$$

for some integer $5 \leq m \leq n$ since $\mathcal{U}(t)$ is 1-lipschitz, and will use ε and m/n interchangeably. Next, recall the following notation for the cumulative distributions

$$\mathcal{H}_{N,n}(s) := \frac{1}{n} \# \left\{ j \leq n \mid s_j(X_{N,n})^{2/N} \leq s \right\}, \qquad \mathcal{U}(s) := \begin{cases} 0, & s < 0, \\ s, & 0 \leq s \leq 1, \\ 1, & s > 1, \end{cases}$$

of the squared singular values of $X_{N,n}$ and the uniform distribution on $[0,1]$. Let us define the event

$$S_{n,m} := \left\{ \forall q \in \{m+1, \ldots, n\} \ \ |\mathcal{H}_{N,n}(q/n) - \mathcal{U}(q/n)| \leq \varepsilon \right\}.$$

On this event, since $\mathcal{H}_{N,n}$ and $\mathcal{U}$ are both monotone we have for $t \leq (m+1)/n$ that

$$\mathcal{H}_{N,n}(t) \leq \mathcal{H}_{N,n}((m+1)/n) \leq \varepsilon + \mathcal{U}((m+1)/n) = \varepsilon + (m+1)/n \leq 3\varepsilon.$$

Similarly if $t > 1$,

$$1 - \mathcal{H}_{N,n}(t) \leq 1 - \mathcal{H}_{N,n}(1) \leq 1 - \mathcal{U}(1) + \varepsilon = \varepsilon$$

Using the same idea we may write for any $t \in [(m+1)/n, 1]$

$$\begin{aligned} \mathcal{H}_{N,n}(t) - \mathcal{U}(t) &\leq \mathcal{H}_{N,n}((j+1)/n) - \mathcal{U}((j+1)/n) + \mathcal{U}(t) - \mathcal{U}((j+1)/n) \\ &\leq \varepsilon + 1/n \\ &\leq 2\varepsilon, \end{aligned}$$

where $m + 1 \leq j \leq n$ is the unique integer for which

$$\frac{j}{n} \leq t < \frac{j+1}{n}.$$

Hence,

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}}|\mathcal{H}_{N,n}(t)-\mathcal{U}(t)|>3\varepsilon\right)\leq\mathbb{P}\left(S_{n,m}^{c}\right)$$

and we must therefore bound $\mathbb{P}\left(S_{n,m}^{c}\right)$ from above. We will do this by rewriting all events involving the singular values s_i in terms of the Lyapunov exponents λ_j. Moving forward, let us agree that any event that involves λ_{-s} or λ_{n+s} for $s>0$ is by definition empty. Since

$$s_j(X_{N,n})^{2/N}\leq\frac{q}{n}\quad\Leftrightarrow\quad\lambda_j\leq\frac{1}{2}\log\left(\frac{q}{n}\right),$$

we find

$$|\mathcal{H}_{N,n}\left(q/n\right)-\mathcal{U}\left(q/n\right)|=\left|\frac{1}{n}\#\left\{j\leq n:\lambda_j\leq\frac{1}{2}\log\left(q/n\right)\right\}-q/n\right|$$

$$=\left|\frac{1}{n}\#\left\{j\leq n:\lambda_j>\frac{1}{2}\log\left(q/n\right)\right\}-(n-q)/n\right|.$$

For any positive integer $m+1\leq q\leq n$, define

$$p:=p(n,m,q,N)=\mathbb{P}\left(\left|\frac{1}{n}\#\left\{j\leq n:\lambda_j>\frac{1}{2}\log\left(q/n\right)\right\}-(n-q)/n\right|\geq\frac{m}{n}\right).$$

$$(11.2)$$

We have

$$\mathbb{P}\left(S_{n,m}^{c}\right)\leq\sum_{q=m+1}^{n}p(n,m,q,N),$$

and the proof of Theorem 1.2 therefore reduces to estimating the probabilities in this sum. To do this, we fix n,m,q,N and observe that since λ_j are decreasing, the event whose probability we've denoted by $p(n,m,q,N)$ is equal to

$$\left\{\lambda_{n-q+m}>\frac{1}{2}\log\left(q/n\right)\right\}\cup\left\{\lambda_{n-q-m+1}\leq\frac{1}{2}\log\left(q/n\right)\right\},$$

where we remind the reader that the second event is empty if $q\geq n-m+1$. Again using the monotonicity of λ_j, this implies

$$\left\{\frac{1}{n}\sum_{j=n-q+1}^{n-q+m}\lambda_j>\frac{m}{2n}\log\left(q/n\right)\right\}\cup\left\{\frac{1}{n}\sum_{j=n-q-m+1}^{n-q}\lambda_j\leq\frac{m}{2n}\log\left(q/n\right)\right\}.$$

So, the probability $p(n,m,q,N)$ we seek to bound is itself bounded above by

$$p_1+p_2:=\mathbb{P}\left(\frac{1}{n}\sum_{j=n-q+1}^{n-q+m}\lambda_j>\frac{m}{2n}\log\left(q/n\right)\right)+\mathbb{P}\left(\frac{1}{n}\sum_{j=n-q-m+1}^{n-q}\lambda_j\leq\frac{m}{2n}\log\left(q/n\right)\right).$$

To estimate p_1 note that by Lemma 11.1,

$$\frac{m}{2n} \log \left(q/n \right) - \frac{1}{n} \sum_{j=n-q+1}^{n-q+m} \mu_{n,j} \geq \frac{(m-1)^2}{4nq}.$$

Hence, we obtain

$$p_1 \leq \mathbb{P} \left(\left| \frac{1}{n} \sum_{j=n-q+1}^{n-q+m} \{\lambda_j - \mu_{n,j}\} \right| \geq \frac{(m-1)^2}{4nq} \right).$$

We will bound the right hand side by using Theorem 1.1. To do this, we must ensure that for c_2 sufficiently large, our assumption $N > c_2/\varepsilon^2$ implies that for the constant c_1 in Theorem 1.1, we have

$$\frac{(m-1)^2}{4nq} \geq c_1 \frac{n-q+m}{nN} \log \left(\frac{en}{n-q+m} \right). \tag{11.3}$$

To check this, note that since $x \log(e/x) \leq 1$ for $x \in [0,1]$ this estimate holds as soon as

$$N \geq c_1 \frac{4nq}{(m-1)^2}. \tag{11.4}$$

Recall that by construction, we have

$$q \leq n, \qquad (m-1)^2 \geq \frac{1}{2} m^2 = \frac{1}{2} \varepsilon^2 n^2.$$

Hence, (11.4) is satisfied once

$$N \geq 8 c_1 \varepsilon^{-2},$$

as claimed. Thus, we may apply Theorem 1.1 to conclude that

$$p_1(n,m,q,N) \leq c_3 \exp \left(c_4 \min \left\{ nNs, n^2 N s g_{n,k}(s) \right\} \right), \quad s = s(n,m,q) = \frac{(m-1)^2}{4nq}.$$

Since

$$\inf_{q=m+1,\dots,n} s(n,m,q) \geq \frac{\varepsilon^2}{8},$$

we find that

$$\sum_{q=m+1}^{n} p_1(m,n,q,N) \leq c_3 \exp \left(-c_4 \min \left\{ nN\varepsilon^2, n^2 N \varepsilon^2 g_{n,k}(\varepsilon^2) \right\} + \log(n) \right)$$

for some universal constants $c_3, c_4 > 0$. Further, note that by assumption,

$$nN\varepsilon^2 > c_2 n > \log(n)$$

as soon as $c_2 > 1$. Hence, at the cost of replacing c_4 by a slightly larger constant c_4', we find that

$$\sum_{q=m+1}^{n} p_1(m, n, q, N) \leq c_3 \exp\left(-c_4' \min\left\{nN\varepsilon^2, n^2 N\varepsilon^2 g_{n,k}(\varepsilon^2)\right\}\right).$$

Essentially the identical argument (but this time the upper bound from Lemma 11.1) implies that this same upper bound holds for p_2 as well, completing the proof of Theorem 1.2. □

12 Asymptotic Normality: Proof for Theorem 1.3

Theorem 1.3 concerns

$$\Lambda_k = (\lambda_1, \ldots, \lambda_k) = (\lambda_1(X_{N,n}), \ldots, \lambda_k(X_{N,n})),$$

the vector of the first k Lyapunov exponents of $X_{N,n}$. Our aim is to prove that there exist universal constants $C, C' > 0$ so that once $N \geq Cn\log(n)$ we have

$$d\left(\Lambda_k, \mathcal{N}\left(\mu_{n,\leq k}, \Sigma_{n,k,N}\right)\right) \leq C'\left(\frac{k^{7/2}n\log^2(n)\log(N/n)}{N}\right)^{1/2}, \qquad \Sigma_{n,k,N} := \frac{1}{N}\mathrm{Diag}(\sigma_{n,\leq k}^2) \tag{12.1}$$

where $\mu_{n,\leq k}, \sigma_{n,\leq k}^2$ are the vectors of means and variances of

$$\left(\frac{1}{2}\log\left(\frac{1}{n}\chi_{n-m+1}^2\right), m = 1, \ldots, k\right)$$

(see (1.18)) and d is the distance function defined in (6.3). To prove (12.1), we introduce

$$S_k = (\lambda_1, \lambda_1 + \lambda_2, \ldots, \lambda_1 + \cdots + \lambda_k)^*$$

and note that

$$S_k = T\Lambda_k, \tag{12.2}$$

where T is a lower triangular matrix all of whose lower-triangular entries are equal to 1. The explain the strategy for proving Theorem 1.3, let us fix $\Theta = \theta_1 \wedge \cdots \wedge \theta_k$, where $\{\theta_j\}$ form an orthonormal k-frame in $\mathbb{R}^n$. For $1 \leq m \leq k$, we will abbreviate

$$\Theta_{\leq m} = \theta_1 \wedge \cdots \wedge \theta_m.$$

The idea of the proof is to compare S_k, Λ_k to their "pointwise" analogs

$$\widehat{S}_k := \frac{1}{N}\left(\log\|X_{N,n}(\Theta_{\leq 1})\|, \ldots, \log\|X_{N,n}(\Theta_{\leq k})\|\right)^*$$

and

$$\widehat{\Lambda}_k := T^{-1}\widehat{S}_k, \tag{12.3}$$

where $\Theta = (\theta_1, \ldots, \theta_k)$ is any fixed collection of k orthonormal vectors in $\mathbb{R}^n$. Specifically, by Proposition 6.3 and the affine invariance (6.4) of d, we find that there exists $c_0 > 0$ so that for all $\delta > 0$

$$
\begin{aligned}
d\left(\Lambda_k, \mathcal{N}\left(\mu_{n,k}, \Sigma_{n,k}\right)\right) &= d\left(S_k, \mathcal{N}\left(T\mu_{n,k}, T\Sigma_{n,k,N}T^*\right)\right) \\
&\leq 3d\left(\widehat{S}_k, \mathcal{N}\left(T\mu_{n,k}, T\Sigma_{n,k,N}T^*\right)\right) + c_0\delta\left|\left|(T\Sigma_{n,k}T^*)^{-1}\right|\right|_{HS}^{1/2} \\
&\quad + 2\mathbb{P}\left(\left|\left|S_k - \widehat{S}_k\right|\right| > \delta\right) \\
&= 3d\left(\widehat{\Lambda}_k, \mathcal{N}\left(\mu_{n,k}, \Sigma_{n,k,N}\right)\right) + c_0\delta\left|\left|(T\Sigma_{n,k,N}T^*)^{-1}\right|\right|_{HS}^{1/2} \\
&\quad + 2\mathbb{P}\left(\left|\left|S_k - \widehat{S}_k\right|\right| > \delta\right). \tag{12.4}
\end{aligned}
$$

The remainder of the proof consists of bounding each of these three terms and then optimizing over δ. To start, let us check that the first term in (12.4) is small:

LEMMA 12.1. *In distribution,*

$$\widehat{\Lambda}_k = \frac{1}{N}\sum_{i=1}^{N}(Y_{i,1}, \ldots, Y_{i,k}), \tag{12.5}$$

where $\{Y_{i,j},\ 1 \leq i \leq N,\ 1 \leq j \leq k\}$ *are independent with*

$$Y_{i,j} \sim \frac{1}{2}\log\left(\frac{1}{n}\chi^2_{n-j+1}\right).$$

Consequently, by the multivariate central limit theorem, there exists $C > 0$ *so that*

$$d\left(\widehat{\Lambda}_k, \mathcal{N}\left(\mu_{n,\leq k}, \Sigma_{n,k,N}\right)\right) \leq \frac{Ck^{7/4}}{N^{1/2}} \tag{12.6}$$

where $\Sigma_{n,k,N} = \frac{1}{N}\mathrm{Diag}(\sigma^2_{n,\leq k})$, $\sigma^2_{n,j} := \mathrm{Var}\left[\frac{1}{2}\log\left(\frac{1}{n}\chi^2_{n-j+1}\right)\right]$.

Proof. Fix integers $N, n \geq 1$ and $1 \leq k \leq n$ and recall that $X_{N,n} = A_N \cdots A_1$ with A_i iid $n \times n$ Gaussian matrices. Note that for each $1 \leq m \leq k$, we have

$$\log||X_{N,n}(\Theta_{\leq m})|| = \sum_{i=1}^{N}\log\left|\left|A_i(\Theta^{(i)}_{\leq m})\right|\right|, \tag{12.7}$$

where

$$\Theta^{(1)}_{\leq m} = \Theta_{\leq m}, \qquad \Theta^{(i+1)}_{\leq m} = \frac{A_i\left(\Theta^{(i)}_{\leq m}\right)}{\left|\left|A_i\left(\Theta^{(i)}_{\leq m}\right)\right|\right|}.$$

Repeatedly applying Lemma 5.4, we therefore conclude that in distribution

$$\widehat{S}_k = \frac{1}{N} \sum_{i=1}^{N} \left(\log ||A_i(\Theta_{\leq 1})||, \ldots, \log ||A_i(\Theta_{\leq k})|| \right)^*$$

is equal to a sum of iid random vectors. Thus, using the definition (12.3) of $\widehat{\Lambda}_k$, we find that in distribution

$$\widehat{\Lambda}_k = \frac{1}{N} \sum_{i=1}^{N} \widehat{\Lambda}_{k,i}, \qquad \widehat{\Lambda}_{k,i} := T^{-1} \left(\log ||A_i(\Theta_{\leq 1})||, \ldots, \log ||A_i(\Theta_{\leq k})|| \right)^*,$$

where we recall that T is a lower triangular matrix with all lower triangular entries equal to 1. Namely,

$$T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \ldots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}, \qquad T^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

Note that $\left\{ \widehat{\Lambda}_{k,i}, \, m = 1, \ldots, k \right\}$ are independent collection for different i. Next, the mth component of $\widehat{\Lambda}_{k,i}$ is

$$\left(\widehat{\Lambda}_{k,i} \right)_m = \log ||A_i(\Theta_{\leq m})|| - \log ||A_i(\Theta_{\leq m-1})|| = \log \left\| \frac{A_i(\Theta_{\leq m-1})}{||A_i(\Theta_{\leq m-1})||} \wedge A\theta_m \right\|,$$
(12.8)

Since $\{\theta_i\}$ are orthonormal, the collection $\{A\theta_i\}$ are iid Gaussians. In particular, we see that $A\theta_m$ is independent of $\{A(\Theta_{\leq j}), 1 \leq j \leq m-1\}$. Also, by Lemma 5.3, the following collections of random variables are independent:

$$\{||A(\Theta_{\leq 1})||, \ldots, ||A(\Theta_{\leq m-1})||\}, \qquad \left\{ \frac{A(\Theta_{\leq 1})}{||A(\Theta_{\leq 1})||}, \ldots, \frac{A(\Theta_{\leq m-1})}{||A(\Theta_{\leq m-1})||} \right\}.$$

The left hand side of relation (12.8) shows that the $1, \ldots, m-1^{st}$ components of $\Lambda_{k,i}$ depend only $\{||A(\Theta_{\leq j})||, j = 1, \ldots, m-1\}$, whereas the right hand side of (12.8) shows that the mth component of $\Lambda_{k,i}$ depends only on $A(\Theta_{\leq m-1})/||A(\Theta_{\leq m-1})||$ and on $A\theta_m$. Therefore, the mth component of $\widehat{\Lambda}_{k,i}$ is independent of all the previous components. Proceeding in this way for $m = k, k-1, \ldots, 1$, we find that the components of $\widehat{\Lambda}_{k,i}$ are independent. Finally, let us denote by $\Pi_{\leq m-1}$ the orthogonal projection onto the orthogonal complement of the span of $\{\theta_1, \ldots, \theta_{m-1}\}$. We have by Lemma 5.2 that in distribution

$$\left(\widehat{\Lambda}_{k,i} \right)_m = \log ||\Pi_{\leq m-1}(A\theta_m)||.$$

Note that $A\theta_m$ is independent of $\Pi_{\leq m-1}$ since the latter depends only on $A\theta_1, \ldots, A\theta_{m-1}$. Hence, we have the following equality in distribution:

$$\left(\widehat{\Lambda}_{k,i}\right)_m = \frac{1}{2}\log\left(\frac{1}{n}\chi^2_{n-m+1}\right).$$

This completes the proof of (12.5). To conclude (12.6), we apply the multivariate CLT (Theorem 6.4) to

$$\widehat{\Lambda}_k - \mathbb{E}\left[\widehat{\Lambda}_k\right] = \sum_{i=1}^N \frac{1}{N}\left(\widehat{\Lambda}_{k,i} - \mu_{n,\leq k}\right).$$

Since the covariance matrix of $(Y_{i,1}, \cdots, Y_{i,k})$ is $\mathrm{Diag}(\sigma^2_{n,\leq k})$ by independence we have that $C := \mathrm{Cov}(\widehat{\Lambda}_K) := \frac{1}{N}\mathrm{Diag}(\sigma^2_{n,1}, \cdots, \sigma^2_{n,k})$. Recall that $\beta_i := \mathbb{E}\|C^{-\frac{1}{2}}(\overline{Y}_{i,1}, \cdots, \overline{Y}_{i,k})\|_2^3$. It is not difficult to check that $\log\chi^2_m$ is a log-concave random variable (i.e. its density is a log-concave function). Moreover, since $\sigma^{-1}_{n,j}\overline{Y}_{i,j}$ have mean zero and variance 1, $D := (\sigma^{-1}_{i,1}\overline{Y}_{i,1}, \cdots, \sigma^{-1}_{i,k}\overline{Y}_{i,k})$ is a log-concave random vector in $\mathbb{R}^k$ with covariance matrix equals to the identity. Therefore $\mathbb{E}\|D\|_2^2 = k$. It is known that the Euclidean norm of such vectors satisfies a reverse Hölder inequality with a universal constant, and in particular (see e.g. [Pao06] or [[AAGM15] Theorem 10.4.6] for a stronger result) that

$$\left(\mathbb{E}\|D\|_2^3\right)^{\frac{1}{3}} \leq C\left(\mathbb{E}\|D\|_2^2\right)^{\frac{1}{2}} = C\sqrt{k},$$

where $C > 0$ is an absolute constant. So,

$$\beta_i = \frac{1}{N^{\frac{3}{2}}}\mathbb{E}\|D\|_2^3 \leq \frac{C^3 k^{\frac{3}{2}}}{N^{\frac{3}{2}}}\ 1 \leq i \leq N.$$

Therefore,

$$\beta := \sum_{j=1}^N \beta_j \leq \frac{C^3 k^{\frac{3}{2}}}{N^{\frac{1}{2}}}$$

and we conclude that there exists an absolute constant $c > 0$ so that

$$d(\widehat{\Lambda}_k, \mathcal{N}(\mu_{n,\leq k}, \Sigma_{n,k})) \leq ck^{7/4}N^{-1/2}. \qquad \qquad \square$$

Having bounded the first term in (12.4), we write

$$(T\Sigma_{n,k,N}T^*)^{-1} = (T^*)^{-1}\Sigma^{-1}_{n,k,N}T^{-1}$$

and bound the second term using that the matrix Σ is diagonal and that T^{-1} a bi-diagonal:

LEMMA 12.2. *There exists $C > 0$ so that*

$$\left\|\left(T\Sigma_{n,k,N}T^*\right)^{-1}\right\|_{HS}^{1/2} \le Ck^{1/4}(nN)^{1/2}. \tag{12.9}$$

Proof. We have

$$T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \ldots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}, \qquad T^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

Thus, recalling that

$$\Sigma_{n,k,N} = \frac{1}{N}\mathrm{Diag}\left(\sigma_{n,\le k}\right) = \frac{1}{N}\left(\sigma_{n,1}^2, \ldots, \sigma_{n,k}^2\right),$$

we find

$$(T^*)^{-1}\Sigma_{n,k,N}^{-1}T^{-1} = N \begin{pmatrix} \sigma_{n,1}^{-2} + \sigma_{n,2}^{-2} & -\sigma_{n,2}^{-2} & 0 & \cdots & 0 \\ -\sigma_{n,2}^{-2} & \sigma_{n,2}^{-2} + \sigma_{n,3}^{-2} & -\sigma_{n,3}^{-2} & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\sigma_{n,k-1}^{-2} & \sigma_{n,k-1}^{-2} + \sigma_{n,k}^{-2} & -\sigma_{n,k}^{-2} \\ 0 & \cdots & 0 & -\sigma_{n,k}^{-2} & \sigma_{n,k}^{-2} \end{pmatrix}$$

Hence, using (1.17), we find that for some $C > 0$

$$\left\|(T^*)^{-1}\Sigma_{n,k,N}^{-1}T^{-1}\right\|_{HS} \le 2N \left(\sum_{j=1}^k \sigma_{n,k,j}^{-4}\right)^{1/2} \le CN \left(\sum_{j=1}^k (n-k+1)^2\right)^{1/2} \le CNnk^{1/2},$$

and Lemma 12.2 follows. $\qquad \square$

Thus far, combining the previous two Lemma with (12.4), we've shown that

$$d\left(\Lambda_k, \mathcal{N}\left(\mu_{n,k}, \Sigma_{n,k}\right)\right) \le \frac{Ck^{7/4}}{N^{1/2}} + c_0 \delta k^{1/4}(nN)^{1/2} + 2\mathbb{P}\left(\left\|S_k - \widehat{S}_k\right\| > \delta\right). \tag{12.10}$$

So it remains to estimate

$$\mathbb{P}\left(\left\|S_k - \widehat{S}_k\right\|_2 \ge \delta\right)$$

and optimize over δ. To do this, write $S_{k,j}, \widehat{S}_{k,j}$ for the jth components of $S_k, \widehat{S}_k$. By (10.9), there exists $C > 0$ so that for $1 \le j \le k \le n$,

$$\mathbb{P}\left(|S_{k,j} - \widehat{S}_{k,j}| \ge s\right) \le 2e^{-sN/4}, \qquad s \ge C\frac{j}{N}\log\left(\frac{en}{j}\right).$$

For any collection positive real numbers $\delta_j > C\frac{j}{N}\log\left(\frac{en}{j}\right)$ we therefore have,

$$\mathbb{P}\left(\left\|S_k - \widehat{S}_k\right\|_2 \geq \left(\sum_{j=1}^k \delta_j^2\right)^{1/2}\right) \leq \sum_{j=1}^k \mathbb{P}\left(|S_{k,j} - \widehat{S}_{k,j}| \geq \delta_j\right) \leq 2\sum_{j=1}^k e^{-\delta_j N/4}.$$

Setting

$$\delta_j := \frac{Cj}{N}\log\left(\frac{en}{j}\right)\log\left(\frac{N}{n}\right),$$

for a sufficiently large constant C we find

$$\mathbb{P}\left(\left|S_{k,j} - \widehat{S}_{k,j}\right| \geq \delta_j\right) \leq 2e^{-Cj\log(en/j)\log(N/n)} \leq 2(n/N)^{j/2}.$$

Hence, as soon as $N > n$, we have

$$\mathbb{P}\left(\left\|S_k - \widehat{S}_k\right\|_2 \geq \delta\right) \leq C\left(\frac{n}{N}\right)^{1/2}$$

where

$$\delta := \left(\sum_{j=1}^k \delta_j^2\right)^{1/2} \leq \frac{Ck^{3/2}\log(n)\log(N/n)}{N}.$$

In conjunction with (12.10) yields

$$d\left(\Lambda_k, \mathcal{N}\left(\mu_{n,k}, \Sigma_{n,k}\right)\right) \leq \frac{Ck^{7/4}}{N^{1/2}} + \left(\frac{Ck^{7/2}n\log^2(n)\log^2(N/n)}{N}\right)^{1/2} + C\left(\frac{n}{N}\right)^{1/2}$$

$$\leq \left(\frac{4Ck^{7/2}n\log^2(n)\log^2(N/n)}{N}\right)^{1/2},$$

as claimed.

Acknowledgments

References

[AAGM15]   S. Artstein-Avidan, A. Giannopoulos, and V.D. Milman. *Asymptotic geometric analysis, Part I, Volume 202*. American Mathematical Society, Providence (2015).

[AB12]   G. Akemann and Z. Burda. Universal microscopic correlation functions for products of independent Ginibre matrices. *Journal of Physics A: Mathematical and Theoretical*, (46)45 (2012), 465201

[ABK14]   G. Akemann, Z. Burda, and M. Kieburg. Universal distribution of Lyapunov exponents for products of Ginibre matrices. *Journal of Physics A: Mathematical and Theoretical*, (39)47 (2014), 395202

[ABK19]   G. Akemann, Z. Burda, and M. Kieburg. From integrable to chaotic systems: universal local statistics of Lyapunov exponents. *EPL (Europhysics Letters)*, (4)126 (2019), 40001

[ABKN14]   G. Akemann, Z. Burda, M. Kieburg, and T. Nagao. Universal microscopic correlation functions for products of truncated unitary matrices. *Journal of Physics A: Mathematical and Theoretical*, (25)47 (2014), 255202

[AG97]   G.B. Arous and A. Guionnet. Large deviations for Wigner's law and voiculescu's non-commutative entropy. *Probability Theory and Related Fields*, (4)108 (1997), 517–542

[Ahn19]   A. Ahn. Fluctuations of beta-Jacobi product processes. *arXiv preprint* 1910.00743 (2019).

[AI15]   G. Akemann and J.R. Ipsen. Recent exact and asymptotic results for products of independent random matrices. *arXiv preprint* 1502.01667 (2015).

[ALPTJ10]   R. Adamczak, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *Journal of the American Mathematical Society*, (2)23 (2010), 535–561

[AS64]   M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, Ninth Dover Printing, Tenth GPO Printing Edition (1964).

[Bal93]   K. Ball. The reverse isoperimetric problem for gaussian measure. *Discrete & Computational Geometry*, (4)10 (1993), 411–420

[BBCC11]   T. Banica, S.T. Belinschi, M. Capitaine, and B. Collins. Free Bessel laws. *Canadian Journal of Mathematics*, (1)63 (2011), 3–37

[Ben03]   V. Bentkus. On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference*, (2)113 (2003), 385–402

[Ben05]   V. Bentkus. A Lyapunov-type bound in RD. *Theory of Probability & Its Applications*, (2)49 (2005), 311–323

[BLR85]   P. Bougerol, P. Lacroix, Jean as well as Huber, and Murray (eds) Rosenblatt. *The Concentration of Measure Phenomenon*. Progress in Probability (1985).

[Car82]   R. Carmona. Exponential localization in one dimensional disordered systems. *Duke Mathematical Journal*, (1)49 (1982), 191–213

[CN84]   J.E. Cohen and C.M. Newman. The stability of large random matrices and their products. *The Annals of Probability*, (1984), 283–310

[Dam11]   D. Damanik. A short course on one-dimensional random schrodinger operators. *arXiv preprint* 1107.1094 (2011).

[Dys62]   F.J. Dyson. Statistical theory of the energy levels of complex systems. I. *Journal of Mathematical Physics*, (1)3 (1962), 140–156

[Fil19]   S. Filip. Notes on the multiplicative ergodic theorem. *Ergodic Theory and Dynamical Systems*, (5)39 (2019), 1153–1189

[FK60]   H. Furstenberg and H. Kesten. Products of random matrices. *The Annals of Mathematical Statistics*, (2)31 (1960), 457–469

[FL16]   P.J. Forrester and D.-Z. Liu. Singular values for products of complex ginibre matrices with a source: hard edge limit and phase transition. *Communications in Mathematical Physics*, (1)344 (2016), 333–368

[For13]   P.J. Forrester. Lyapunov exponents for products of complex gaussian random matrices. *Journal of Statistical Physics*, (5)151 (2013), 796–808

[For14]   P.J. Forrester. Eigenvalue statistics for product complex wishart matrices. *Journal of Physics A: Mathematical and Theoretical*, (34)47 (2014), 345202

[GJ18]   F. Götze and J. Jalowy. Rate of convergence to the circular law via smoothing inequalities for log-potentials. *arXiv preprint* 1807.00489 (2018).

[GS18]   V. Gorin and Y. Sun. Gaussian fluctuations for products of random matrices. *arXiv preprint* 1812.06532 (2018)

[GT10]   F. Götze and A. Tikhomirov. On the asymptotic spectrum of products of independent random matrices. *arXiv preprint* 1012.2710 (2010).

[HN19]   B. Hanin and M. Nica. Products of many large random matrices and gradients in deep neural networks. *Communications in Mathematical Physics*, (2019) 1–36

[HNWTW20] D. Huang, J. Niles-Weed, J.A. Tropp, and R. Ward. Matrix concentration for products. *arXiv preprint* 2003.05437 (2020).

[HW20]   A. Henriksen and R. Ward. Concentration inequalities for random matrix products. *Linear Algebra and Its Applications*, 594 (2020), 81–94

[IN92]   M. Isopi and C.M. Newman. The triangle law for Lyapunov exponents of large random matrices. *Communications in Mathematical Physics*, (3)143 (1992), 591–598

[Jal19]   J. Jalowy. Rate of convergence for products of independent non-Hermitian random matrices. *arXiv preprint* 1912.09300 (2019).

[Joh04]   K. Johansson. Determinantal processes with number variance saturation. *Communications in Mathematical Physics*, (1-3)252 (2004), 111–148

[Kal02]   O. Kallenberg. *Foundations of Modern Probability.* Springer, Berlin (2002).

[Kar08]   V. Kargin. Lyapunov exponents of free operators. *Journal of Functional Analysis*, (8)255 (2008), 1874–1888

[KK16]   M. Kieburg and H. Kösters. Exact relation between singular value and eigenvalue statistics. *Random Matrices: Theory and Applications*, (04)5 (2016), 1650015

[KMS20]   T. Kathuria, S. Mukherjee, and N. Srivastava. On concentration inequalities for random matrix products. *arXiv preprint* 2003.06319 (2020).

[Lat97]   R. Latała. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, (3)25 (1997), 1502–1513

[LP82]   É. Le Page. Théoremes limites pour les produits de matrices aléatoires. In: *Probability Measures on Groups.* Springer, Berlin (1982), pp. 258–303.

[LW19]   D.-Z. Liu and Y. Wang. Phase transitions for infinite products of large non-hermitian random matrices. *arXiv preprint* 1912.11910 (2019).

[LWW18]   D.-Z. LIU, D. WANG, and Y. WANG. Lyapunov exponent, universality and phase transition for products of random matrices. *arXiv preprint* 1810.00433 (2018).

[LWZ16]   D.-Z. LIU, D. WANG, and L. ZHANG. Bulk and soft-edge universality for singular values of products of ginibre random matrices. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, Vol. 52 (2016). Institut Henri Poincaré, pp. 1734–1762.

[Naz03]   F. NAZAROV. On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a gaussian measure. In: *Geometric Aspects of Functional Analysis*. Springer, Berlin (2003), pp. 169–187.

[Nem17]   Y. NEMISH. Local law for the product of independent non-Hermitian random matrices with independent entries. *Electronic Journal of Probability*, 22 (2017).

[New86]   C.M. NEWMAN. The distribution of Lyapunov exponents: exact results for random matrices. *Communications in Mathematical Physics*, (1)103 (1986), 121–126

[OS11]   S. O'ROURKE and A. SOSHNIKOV. Products of independent non-Hermitian random matrices. *Electronic Journal of Probability*, 16 (2011), 2219–2245

[Ose68]   V.I. OSELEDETS. A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 19 (1968) 179–210

[Pao06]   G. PAOURIS. Concentration of mass on convex bodies. *Geometric & Functional Analysis GAFA*, (5)16 (2006), 1021–1049

[PP13]   G. PAOURIS and P. PIVOVAROV. Small-ball probabilities for the volume of random convex sets. *Discrete & Computational Geometry*, (3)49 (2013), 601–646

[Red16]   N.K. REDDY. Lyapunov exponents and eigenvalues of products of random matrices. *arXiv preprint* 1606.07704 (2016).

[Red19]   N.K. REDDY. Equality of Lyapunov and stability exponents for products of isotropic random matrices. *International Mathematics Research Notices*, (2)2019 (2019), 606–624

[Rud14]   M. RUDELSON. Lecture notes on non-asymptotic random matrix theory. In: *Modern Aspects of Random Matrix Theory—AMS Proceedings of Symposia in Applied Mathematics* (2014), pp. 83–121.

[Rud17]   M. RUDELSON. Delocalization of eigenvectors of random matrices. lecture notes. *arXiv preprint* 1707.08461 (2017).

[RV08]   M. RUDELSON and R. VERSHYNIN. The Littlewood–Offord problem and invertibility of random matrices. *Advances in Mathematics*, (2)218 (2008), 600–633

[RV09]   M. RUDELSON and R. VERSHYNIN. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, (12)62 (2009) 1707–1739

[Spi70]   M.D. SPIVAK. *A Comprehensive Introduction to Differential Geometry*. Publish or Perish (1970).

[Tao10]   T. TAO. *An Epsilon of Room, II: Pages from Year Three of a Mathematical Blog*. American Mathematical Society, Providence (2010).

[Tik20]   K. Tikhomirov. Singularity of random bernoulli matrices. *Annals of Mathematics*, (2)191 (2020), 593–634.

[Tro15]   J. Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, (1-)8 (2015), 1–230

[Tuc10]   G.H. Tucci. Limits laws for geometric means of free random variables. *Indiana University Mathematics Journal*, (2010) 1–13

[Ver12]   R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. pages 210–268. Cambridge University Press, Cambridge (2012).

B. Hanin and G. Paouris
Department of Mathematics, Texas A&M University, College Station, TX 77843, USA.
bhanin@princeton.edu

G. Paouris
grigoris@math.tamu.edu

B. Hanin
Google, Mountain View, CA 94043, USA.