RANDOM MATRIX PRODUCTS: UNIVERSALITY AND LEAST SINGULAR VALUES

BY PHIL KOPEL[1,*], SEAN O'ROURKE[1,**] AND VAN VU[2]

[1]*Department of Mathematics, University of Colorado Boulder,* [*]*philip.kopel@colorado.edu*; [**]*sean.d.orourke@colorado.edu*
[2]*Department of Mathematics, Yale University, van.vu@yale.edu*

We establish, under a moment matching hypothesis, the local universality of the correlation functions associated with products of M independent i.i.d. random matrices, as M is fixed, and the sizes of the matrices tend to infinity. This generalizes an earlier result of Tao and the third author for the case $M = 1$.

We also prove Gaussian limits for the centered linear spectral statistics of products of M independent i.i.d. random matrices. This is done in two steps. First, we establish the result for product random matrices with Gaussian entries, and then extend to the general case of non-Gaussian entries by another moment matching argument. Prior to our result, Gaussian limits were known only for the case $M = 1$. In a similar fashion, we establish Gaussian limits for the centered linear spectral statistics of products of independent truncated random unitary matrices. In both cases, we are able to obtain explicit expressions for the limiting variances.

The main difficulty in our study is that the entries of the product matrix are no longer independent. Our key technical lemma is a lower bound on the least singular value of the translated linearization matrix associated with the product of M normalized independent random matrices with independent and identically distributed sub-Gaussian entries. This lemma is of independent interest.

1. Introduction and statement of results. Random matrices with independent entries have been among the most central objects of study in random matrix theory since the pioneering work of Ginibre in 1965 [22], and many important advances and applications have been established since that time (for a partial overview see [12], and the excellent collection of references contained therein).

DEFINITION 1. An *i.i.d. random matrix (with sub-Gaussian decay)* is a $n \times n$ random matrix $M_n = (\xi_{i,j})_{1 \le i, j \le n}$ where each entry $\xi_{i,j}$ is given by an independent copy of a sub-Gaussian random variable ξ with mean zero and unit variance, and whose real and imaginary parts are independent.

The distribution of a single entry of an i.i.d. random matrix is commonly called the *atom distribution*. The requirement that atom distributions be *sub-Gaussian* means that there exists some constants $C, v > 0$ (independent of n) such that for each $t > 0$:

$$(1) \qquad \mathbf{P}(|\xi_{1,1}| > t) \le Ce^{-vt^2}.$$

If each entry $\xi_{i,j}$ is given by a complex standard Gaussian random variable, the i.i.d. matrix is said to belong to the complex Ginibre ensemble, or more briefly GinUE. If the entries are

real Gaussians, then the matrix is said to belong to the real Ginibre ensemble, or GinOE. The ensemble of matrices with independent centered Bernoulli entries provides another natural example of an i.i.d. random matrix.

One of the most salient and important features of i.i.d. matrices is that they are non-Hermitian, and the methods used to study them are therefore frequently different from those employed in Hermitian random matrix theory. Notice that unlike in the Hermitian case we no longer have strictly real eigenvalues, which introduces a number of difficulties (for instance, limiting the usefulness of polynomials in studying general functions of the eigenvalues, which in turn limits the usefulness of combinatorial methods based on moment methods [8, 12, 27]). On the other hand, the lack of symmetry is occasionally useful and the mutual independence of all rows (or all columns) can simplify matters considerably in certain situations [54].

In this paper, we study the product of M independent i.i.d. random matrices. The spectrum of such a product has been an object of keen interest in random matrix theory, and many important advances have been made in recent years. We refer the reader to [1–6, 11, 17, 20, 23–26, 32, 33, 35, 36] and references therein; however, since the subject has been so intensively studied, the collection of references above as well as our discussion below if far from complete. This subfield of random matrix theory is not only motivated by its own intrinsic interest, as the study of products of matrices is a very natural generalization of the study of a single random matrix, but is also motivated by connections to several of the applied sciences (the interested reader is directed to the impressive survey of applications assembled in [25]).

The first important result concerning random matrix products is the M-fold circular law, which occupies the same position as either the semicircular law for Wigner matrices or circular law for independent entry matrices, and governs the limiting distribution of eigenvalues. This remarkable result says that as $n \to \infty$ the limiting distribution of the spectrum of the product of M normalized n by n independent i.i.d. matrices (obeying certain moment assumptions), $X = n^{-M/2}X^{(1)} \cdots X^{(M)}$, is supported on the unit disc $|z| < 1$, and has the following density [24, 36]:

$$(2) \qquad \frac{1}{M\pi}|z|^{2/M-2}.$$

Notice that, although for even somewhat sizable M the distribution of a single entry of the product matrix X may be quite complicated, the limiting distribution is miraculously simple.

Once the global distribution has been established, the next natural questions are the limiting laws of linear statistics and universality of local statistics. In this paper, we attack both problems and solve them under certain moment conditions. Throughout the paper, we always assume all i.i.d. random matrices have atom distribution with sub-Gaussian decay as described in Definition 1.

Our first result concerns the *linear statistics* of random product matrices, which describe *fluctuations* of the spectrum of a random matrix around its limiting distribution—in this case, the M-fold circular law. The centered linear statistic associated with the random matrix X (with eigenvalues denoted by $\lambda_1, \ldots, \lambda_n$) and some test function f (whose regularity depends on the ensemble under consideration) is defined by the following formula:

$$(3) \qquad N_n[f] = \sum_{j=1}^{n} f(\lambda_j) - \mathbf{E}\left[\sum_{j=1}^{n} f(\lambda_j)\right].$$

The behavior of linear statistics is well understood for Wigner matrices (the unnormalized statistic converges to a normal distribution, which implies fluctuations comparable to a constant, in striking opposition to the fluctuations seen in the case of the sum of i.i.d. random

variables, which are comparable to $\sqrt{n}$) up to some delicate questions about the optimal regularity constraints on the test function f [31, 43–45]. Much progress has been made in recent years on the linear statistics of independent entry ensembles [27, 34, 37], even though due to difficulties associated with the failure of the Hermitian condition there is still much work to be done in that area (the failure of Hermiticity and the complex eigenvalues it implies means that analytic test functions are no longer dense in the space of smooth test functions, for instance, which limits the efficacy of trace methods), and substantial new ideas are likely to be required. To the best of the authors' knowledge, the linear statistics of products of i.i.d. matrices have not yet been investigated.

Before we can state our result, we will need to introduce the concept of *moment matching*. Two i.i.d. random matrices X and X' with entries $\xi_{i,j}$ and $\xi'_{i,j}$, respectively, are said to match moments to order k if, for all $1 \leq i, j \leq n$, $a, b \geq 0$, $a + b \leq k$:

$$\mathbf{E}\big[\mathrm{Re}(\xi_{i,j})^a \, \mathrm{Im}(\xi_{i,j})^b\big] = \mathbf{E}\big[\mathrm{Re}(\xi'_{i,j})^a \, \mathrm{Im}(\xi'_{i,j})^b\big].$$

We will say that two product matrices, $X^{(1,1)} \cdots X^{(1,M)}$ and $X^{(2,1)} \cdots X^{(2,M)}$, match moments to order k if each individual pair of factor matrices (meaning $X^{(1,j)}$ and $X^{(2,j)}$ for $1 \leq j \leq M$) do. We will also need to define a relevant Sobolev norm.

DEFINITION 2. Let f be a real valued function defined on the complex plane, and let $\hat{f}(k)$ denote the kth Fourier coefficient of the restriction of a function f to the circle $|z| = R$:[1]

$$(4) \qquad \hat{f}(k) = \frac{1}{2\pi} \int_0^{2\pi} f\big(\mathrm{Re}^{\sqrt{-1}\theta}\big) e^{-\sqrt{-1}k\theta} \, d\theta.$$

The inner product $\langle f, g \rangle_{H^{1/2}(|z|=R)}$ is then defined to be

$$(5) \qquad \langle f, g \rangle_{H^{1/2}(|z|=R)} = \sum_{k \in \mathbb{Z}} |k| \hat{f}(k) \overline{\hat{g}(k)}.$$

Finally, we set the Sobolev norm $\|\cdot\|_{H^{1/2}(|z|=R)}$ to be the norm induced by this inner product.

This Sobolev norm makes an appearance in the expression of the limiting variance of linear statistics of products of matrices matching the GinUE ensemble to four moments.

THEOREM 3. *Let $f : \mathbb{C} \to \mathbb{R}$ be a test function with at least two continuous derivatives, supported in the spectral bulk $\{z \in \mathbb{C} : \tau_0 < |z| < 1 - \tau_0\}$ for some fixed $\tau_0 > 0$. Fix an integer $M \geq 1$, and let $n^{-M/2} X^{(1)} \cdots X^{(M)}$ be a matrix product such that each factor $X^{(i)}$ is a n by n i.i.d. random matrix* (which are all jointly independent) *with an atom distribution matching the standard complex Gaussian distribution to four moments. Then the centered linear statistic associated with the test function f and the product matrix $n^{-M/2} X^{(1)} \cdots X^{(M)}$, denoted $N_n[f]$, converges in distribution as $n \to \infty$ to the mean zero normal distribution with the following limiting variance*:

$$(6) \qquad \frac{1}{4\pi} \int_{|z|<1} |\nabla f(z)|^2 \, d^2z + \frac{1}{2} \|f\|_{H^{1/2}(|z|=1)}^2.$$

Notice that the limiting variance does not depend on M.

We will prove this result by way of a more general result establishing four moment universality for linear statistics of products of i.i.d. matrices, along with an entirely separate result establishing the appropriate central limit theorem for products of Ginibre matrices (the proof

---

[1]Here, we use $\sqrt{-1}$ to denote the imaginary unit and reserve i as an index.

of which is based on the rotary flow combinatorial approach pioneered by Rider and Virág for a single Ginibre matrix [38]). If a central limit theorem for linear statistics was established for products of matrices with some known sub-Gaussian distribution besides the complex Gaussian distribution (for instance, the real Gaussian), then our machinery (specifically, Theorem 30 appearing below) would establish the same central limit theorem for all matrices which match that base case to four moments.

There are four remarks which must be made about the technical assumptions appearing in the statement of Theorem 3. The first is that the condition requiring the entries of each matrix to be identically distributed is needed only to provide an appropriate least singular value estimate (specifically, we will need this assumption to be able to apply Theorem 7 below, otherwise the condition is not needed—the substitution of an alternate condition that preserves the statement of Theorem 7 will not affect the proof of this result). The second remark concerns the sub-Gaussian decay condition on the atom distributions from Definition 1. Naturally, we expect our main results to hold under weaker assumptions (such as subexponential tails or the existence of a high enough number of finite moments). The sub-Gaussian decay condition is mostly used to control certain events with high enough probability, and it seems likely that more technical methods could be utilized to relax this assumption. Due to the already technical nature of our proofs, we have not pursued this direction.

The third remark is that the assumption of two derivatives on f, as well as the condition requiring support in the spectral bulk, is necessary only for the moment matching argument. If one is only interested in the linear statistics in the specific case of complex Gaussian atom distribution, then as will be clear from the proof, both of these conditions can be relaxed (see Theorem 26 for details).

The fourth remark concerns the assumption on the support of f, specifically that f is supported away from both the origin and the spectral edge $|z| = 1$. Under this assumption, the term $\|f\|^2_{H^{1/2}(|z|=1)}$ appearing in (6) is zero; we include this term in the limiting variance since it matches the analogous result obtained for products of complex Ginibre matrices (Theorem 26 below), which does not feature this restriction on f. For Theorem 3, the condition on the support of f appears to be an artifact of the proof, and is necessary mainly to ensure good behavior of the entries of the resolvent of the Hermitian linearization of the product matrix. Lifting this requirement is a direction for future work (a similar condition encountered in the course of the proof of the local M-fold circular law [32] has recently been lifted [23], at least at the origin, but the methods employed there do not appear to directly extend to our case), and we suspect (6) to still be the limiting variance once this condition is lifted.

The same technique that establishes Theorem 3 in the base case of complex Ginibre matrices can also be applied to other product matrix models, as long as the eigenvalues constitute a rotationally invariant determinantal point process and the ensuing combinatorics work out reasonably. As an illustration, we obtain a limit theorem for linear statistics in the case of *truncated unitary random matrices*. A truncated unitary random matrix is a random matrix produced by taking the top-left n by n submatrix of a random unitary K by K matrix (distributed according to the relevant Haar measure), where $\tau = (K - n)/n$ is a fixed parameter. The study of these matrices is motivated in part by connections to topics such as time evolution in quantum mechanics or chaotic scattering on mesoscopic devices; the interested reader is directed to [3, 4, 57] and the references within.

THEOREM 4. *Let $f : \mathbb{C} \to \mathbb{R}$ be a real valued polynomial test function, and let $U^{(1)} \cdots U^{(M)}$ be the product of M n by n jointly independent truncated unitary random matrices, with fixed truncation ratio $\tau \in (1/2, 1)$. Then, for fixed M and τ, in the large dimensional limit $n \to \infty$ the associated centered linear statistic, $N_n[f]$, converges in distribution to the*

normal distribution with mean zero and the following limiting variance:

$$(7) \qquad \frac{1}{4\pi} \int_{|z| \le (1/(1+\tau))^{M/2}} |\nabla f(z)|^2 \, d^2 z + \frac{1}{2} \|f\|^2_{H^{1/2}(|z|=|1+\tau|^{-M/2})}.$$

Theorem 4 deals only with polynomial test functions to simplify technical aspects of the argument; we believe a similar result should hold for more general test functions but this is outside the scope of the paper. The restriction to the disc $|z| \le (1/(1+\tau))^{M/2}$ is explained by the limiting distribution of eigenvalues of products of truncated unitary matrices, which is defined by the following formula:

$$\phi^{M,\tau}(z) = \frac{1}{\pi M} \frac{\tau}{|z|^{2(1-1/M)}(1-|z|^{2/M})^2} \quad \text{on } |z| \le \left(\frac{1}{\tau+1}\right)^{M/2},$$

where $\phi^{M,\tau}(z)$ vanishes on $|z| > (\frac{1}{\tau+1})^{M/2}$. This formula is already in the literature [3], but we encounter it during the course of the proof of Theorem 4.

Our next result will be the local universality of the k-point correlation functions throughout the spectral bulk for products of i.i.d. factor matrices matching to four moments. Define the k-point correlation function $p^{(k)}$, for $1 \le k \le n$, of the random points $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ by way of the following formula (with $\phi : \mathbb{C}^k \to \mathbb{C}$ denoting any arbitrary continuous, compactly supported test function):

$$\mathbf{E}\left[\sum_{i_1,\ldots,i_k \text{ distinct}} \phi(\lambda_{i_1}, \ldots, \lambda_{i_k}) \right] = \int_{\mathbb{C}^k} \phi(z_1, \ldots, z_k) p^{(k)}(z_1, \ldots, z_k) \, d^2 z_1 \cdots d^2 z_k.$$

We are abusing notation here by writing $p^{(k)}(z_1, \ldots, z_k) \, d^2 z_1 \cdots d^2 z_k$, as $p^{(k)}$ is in general a measure and not a function (see [51] for a complete discussion of this issue).

The following universality result holds in the case of products $Z_n = X_n^{(1)} \cdots X_n^{(M)}$ of $n \times n$ i.i.d. matrices with complex atom distributions. Before discussing the correlation functions for the eigenvalues of Z_n, we note that the largest eigenvalue of Z_n is on the order of $n^{M/2}$, and we will need to first rescale the eigenvalues so that the typical spacing in the bulk is on the order of a constant. There are several possible ways to rescale the eigenvalues. We adopt the convention of taking the Mth root of each eigenvalue of Z_n to create a new point process with Mn points. Formally, given the product $Z_n = X_n^{(1)} \cdots X_n^{(M)}$, we define the Mth *root eigenvalue process associated with Z_n* as the point process on $\mathbb{C}$ consisting of the Mn roots (counted with algebraic multiplicity) of the polynomial $z \mapsto \det(z^M I - Z_n)$, where I denotes the identity matrix. Clearly, the eigenvalues of Z_n can be recovered from this process by raising each root to the Mth power. Our next result proves universality of the correlation functions corresponding to the Mth root eigenvalue process.

We define the open ball of radius $r > 0$ centered at $z_0 \in \mathbb{C}$ as

$$B(z_0, r) := \{z \in \mathbb{C} : |z - z_0| < r\}.$$

THEOREM 5.   *For $\beta = 1, 2$, let $Z_n^\beta = X_n^{(\beta,1)} \cdots X_n^{(\beta,M)}$ be the product of M independent i.i.d. n by n matrices with complex-valued entries and sub-Gaussian decay. Assume that the factor matrices $X_n^{(1,i)}$ and $X_n^{(2,i)}$ (for $1 \le i \le M$) match to four moments. Let $z_1, \ldots, z_k$ be complex numbers (which may depend on n) located in the spectral bulk $\tau_0 \le n^{-1/2}|z_i| \le 1 - \tau_0$ for some fixed positive constant $\tau_0 > 0$.*

Let $p_\beta^{(k)}$ be the k-point correlation function of the Mth root eigenvalue process associated with Z_n^β. If $G : \mathbb{C}^k \to \mathbb{C}$ is a smooth function supported on $B(0, r_0)^k$ (for a small $r_0 > 0$ which

depends on τ_0) then for any sufficiently small positive constant $c_0 > 0$ the following estimate holds:

$$\left| \int_{\mathbb{C}^k} G(w_1, \ldots, w_k) p_1^{(k)}(z_1 + w_1, \ldots, z_k + w_k) \, d^2 w_1 \cdots d^2 w_k \right.$$

$$\left. - \int_{\mathbb{C}^k} G(w_1, \ldots, w_k) p_2^{(k)}(z_1 + w_1, \ldots, z_k + w_k) \, d^2 w_1 \cdots d^2 w_k \right|$$

$$\leq C_{[\tau_0, k, G]} n^{-c_0}.$$

The constant $C_{[\tau_0, k, G]}$ depends on τ_0, k, M, and (linearly) on the maximal L^∞-norm bound on the first $2k + 6$ derivatives of G.

While this result focuses on the Mth root eigenvalue process, our methods can be generalized to other scaling conventions for the eigenvalues as well; see Remark 36 for further details.

The proof of Theorem 5 is based on an application of Theorem 2.1 in [51], and may be particularly of interest in the case of complex Ginibre factor matrices. Indeed, the derivation of the correlation functions for the product of independent complex Ginibre matrices can be found in [2], which finds that the limiting correlation functions match those of a single Ginibre matrix away from the origin. Specifically, the eigenvalues of the product of M independent $n \times n$ complex Ginibre matrices form a determinantal point process with kernel $K_n^{(M)}$ defined by the formula

$$(8) \qquad K_n^{(M)}(z, w) = \sqrt{w_M(z) w_M(w)} \sum_{k=0}^{n-1} \frac{1}{(\pi k!)^M} (z\bar{w})^k.$$

The weight function $w_M(z)$ appearing here depends only on the modulus of z and is defined through the Meijer G-function; see equation (2.3) in [2]. As shown in [2], the kernel K_n^M obeys the following scaling limit in the spectral bulk, away from the origin:

$$\lim_{n \to \infty} M^{2-M} |\xi_i \xi_j|^{M-1} K_n^{(M)}\left((\xi_i/\sqrt{M})^M, (\xi_j/\sqrt{M})^M\right)$$

$$= \frac{1}{\pi} \left(\frac{\xi_i \overline{\xi_j}}{|\xi_i \xi_j|} \right)^{(1-M)/2} \exp\left(-\frac{1}{2} (|\xi_i|^2 + |\xi_j|^2 + \xi_i \overline{\xi_j}) \right).$$

Up to the phase factor in front (which is irrelevant after taking determinants), this limiting kernel is the same as the limiting bulk behavior of a single Ginibre matrix (when $M = 1$). Theorem 5 implies that this behavior should be universal for other matrix products matching the Ginibre ensemble to four moments. Both the spectral edge and the origin are dealt with in [2] as well, but our universality result does not extend to these cases. It would be interesting to see if it can be generalized to include these cases as well.

In the case of real entries, the spectrum and, therefore, the correlation functions split into real and imaginary components. For a thorough exposition on the spectrum in the real case see, for instance, [51]. For such a point process, we define $p^{k,l}$ to be the correlation function for k real (which we will denote $\zeta_{i,\mathbb{R}}$) and l complex points in the upper half place (which we will denote $\zeta_{j,\mathbb{C}_+}$). To be precise, for continuous test functions $\phi : \mathbb{R}^k \times \mathbb{C}^l \to \mathbb{C}$ with compact support we require

$$\mathbf{E}\left[\sum_{i_1, \ldots, i_k \text{ distinct}} \sum_{j_1, \ldots, j_l \text{ distinct}} \phi(\zeta_{i_1,\mathbb{R}}, \ldots, \zeta_{i_k,\mathbb{R}}, \zeta_{j_1,\mathbb{C}_+}, \ldots, \zeta_{j_l,\mathbb{C}_+}) \right]$$

$$= \int_{\mathbb{R}^k} \int_{\mathbb{C}_+^l} \phi(x_1, \ldots, x_k, z_1, \ldots, z_l)$$

$$\times p^{k,l}(x_1, \ldots, x_k, z_1, \ldots, z_l) \, d^2 z_1 \cdots d^2 z_l \, dx_1 \cdots dx_k.$$

This defines the mixed correlation functions $p^{k,l}$ over regions of the form $\mathbb{R}^k \times \mathbb{C}_+^l$. These function are then extended to include $z_i \in \mathbb{C}_-$ by requiring that they be symmetric under conjugation of any complex argument.

The following universality result for the M-th root eigenvalue process holds in the case of products of matrices whose entries have real distributions.

THEOREM 6.   *For $\beta = 1, 2$, let $Z_n^\beta = X_n^{(\beta,1)} \cdots X_n^{(\beta,M)}$ be the product of M independent i.i.d. n by n matrices with identically distributed, real-valued entries and sub-Gaussian decay, and let $x_1, \ldots, x_k \in \mathbb{R}$ and $z_1, \ldots, z_l \in \mathbb{C}$ be numbers (which are allowed to depend on n) located in the spectral bulk (meaning $\tau_0 \leq n^{-1/2}|z_i| \leq 1 - \tau_0$ and $\tau_0 \leq n^{-1/2}|x_i| \leq 1 - \tau_0$ for some fixed $\tau_0 > 0$).*

Let $p_\beta^{k,l}$ be the mixed (k, l)-correlation function for the Mth root eigenvalue process associated with Z_n^β, and assume that the factor matrices $X_n^{(1,i)}$ and $X_n^{(2,i)}$ (for $1 \leq i \leq M$) match to four moments, and also that the correlation functions for one of the two processes satisfy the following estimate in the spectral bulk (which is the annulus defined by the equations $\tau_0 \leq n^{-1/2}|z_i| \leq 1 - \tau_0$ and $\tau_0 \leq n^{-1/2}|x_i| \leq 1 - \tau_0$):

$$(9) \qquad\qquad\qquad p_\beta^{2,0}(x_1, x_2) < C,$$

$$(10) \qquad\qquad\qquad p_\beta^{0,1}(z_1) < C.$$

If $G : \mathbb{R}^k \times \mathbb{C}^l \to \mathbb{C}$ is a smooth function supported on $[-r_0, r_0]^k \times B(0, r_0)^l$ (for some small $r_0 > 0$ depending on τ_0), then for any sufficiently small $c_0 > 0$ the following estimate holds (where we write $d\mu = dy_1 \cdots dy_k \, d^2 w_1 \cdots d^2 w_l$):

$$\left| \int_{\mathbb{R}^k} \int_{\mathbb{C}^l} G(y_1, \ldots, y_k, w_1, \ldots, w_l) \right.$$
$$\times\, p_1^{k,l}(x_1 + y_1, \ldots, x_k + y_k, z_1 + w_1, \ldots, z_l + w_l) \, d\mu$$
$$- \int_{\mathbb{R}^k} \int_{\mathbb{C}^l} G(y_1, \ldots, y_k, w_1, \ldots, w_l)$$
$$\left. \times\, p_2^{k,l}(x_1 + y_1, \ldots, x_k + y_k, z_1 + w_1, \ldots, z_l + w_l) \, d\mu \right|$$
$$\leq C_{[\tau_0, k, l, G]} n^{-c_0}.$$

The constant $C_{[\tau_0, k, l, G]}$ depends on C, k, l, τ_0, M and (linearly) on the maximal L^∞-norm bound on the first $2(k+l) + 6$ derivatives of G.

Analogously to the complex case, this result may be of particular interest when comparing to the product of M real Ginibre factor matrices; concrete expressions for the correlation kernels (which are expressed in terms of Pfaffians) as well as the correlations between real eigenvalues and the correlations between complex eigenvalues are derived and can be found in [21, 26].

As was the case with Theorem 3, the spectral bulk requirement (namely, that z_i is chosen such that the estimate $\tau_0 \leq n^{-1/2}|z_i| \leq 1 - \tau_0$ holds) which appears in both Theorem 5 and Theorem 6 is necessary to invoke aspects of Nemish's proof of the local M-fold circular law for the spectral bulk [32]. In particular, this restriction appears to be an artifact of the method of proof and the authors believe that it should be possible to lift this requirement; an inspection of the proofs of Theorems 5 and 6 should make clear exactly where the difficulties lie.

The proofs of the preceding results (with the exception of Theorem 4) rely in large part on a least singular value estimate, Theorem 7 below, which is the main technical advance in the present volume and may be useful in a variety of other contexts. To present this result properly, we first must discuss *linearization*, one of the most useful techniques in the study of products of random matrices, which allows one to bypass many of the difficulties associated with the product structure. In a linearization argument, one studies the product $X = X^{(1)} X^{(2)} \cdots X^{(M)}$ by considering instead an associated Mn by Mn linearization block matrix Y, defined below:

$$Y = \begin{pmatrix} 0 & X^{(1)} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & X^{(2)} & \cdots & \cdots & 0 \\ 0 & 0 & 0 & X^{(3)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & X^{(M-1)} \\ X^{(M)} & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

As observed by Burda, Janik and Waclaw [17], if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of X, then each λ_k is an eigenvalue of Y^M with multiplicity M. Indeed, to see this it suffices to raise Y to the Mth power and notice that the diagonal block entries are cyclic permutations of $X^{(1)} \cdots X^{(M)}$. It is this link that, in some situations, allows one to study the linearization matrix in lieu of the actual product matrix (which may be substantially more cumbersome).

Frequently, it happens the matrix of interest is not Y but is actually $Y - zI$, where I is the identity matrix and z is some parameter in the complex plane. Our next result deals primarily with the smallest singular value of such matrices.

THEOREM 7. *Let $X^{(i)}$, for $1 \le i \le M$, denote a family of jointly independent n by n i.i.d. random matrices with sub-Gaussian decay (with potentially different atom distributions) and, for any $z \in \mathbb{C}$, define the Mn by Mn matrix $Y(z) = Y - zI$ as below*:

$$Y(z) = \begin{pmatrix} -zI & X^{(1)} & 0 & \cdots & 0 \\ 0 & -zI & X^{(2)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -zI & X^{(M-1)} \\ X^{(M)} & 0 & \cdots & 0 & -zI \end{pmatrix}.$$

Let $\sigma_1(Y(z))$ denote the least singular value of $Y(z)$. For any sufficiently small constant $A > 0$, there exists $t_0 > 0$ (depending on A, M and the distributions of the entries) such that if $n^{1/2 - t_0} \le |z| \le n^{1/2 + t_0}$, then

$$(11) \qquad \mathbf{P}\{\sigma_1(Y(z)) \le n^{-1/2 - A}\} \le Cn^{-KA}.$$

Here, the constant C depends only on M, A and the distributions of the entries; K depends only on M and the distributions of the entries. The requisite smallness of A depends on M and the distributions of the entries.

Previous efforts to control the smallest singular value of this matrix only established a lower bound of the form n^{-B}, where B was (potentially a quite large) positive constant which was not explicitly determined [24, 35, 36].

The random matrix theory literature contains many other results concerning the least singular value and related singularity probability bounds for various models of random matrices including [9, 12, 16, 18, 19, 29, 30, 39, 40, 42, 46–49, 53–55] and references therein. The approach to proving Theorem 7 is roughly modeled on the approach taken to prove a similar

result for a single independent entry random matrix in [49], however the block structure introduces substantial new difficulties that must be overcome. Namely, to control the smallest singular value we would like to control the distance between a given row of $Y(z)$ (say, the last row) and the span of the remaining rows, exploiting the independence between the last row and others through a suitable anti-concentration estimate. The difficulty is that the last row of $Y(z)$ contains quite a lot of zeros placed there by the block structure itself, and if the normal to the hyperplane spanned by the remaining rows places too much weight away from coordinates on which the last row is supported then no anticoncentration estimate could possibly be applied and the entire argument collapses. On the one hand, such a scenario does not sound particularly probable (as it would seem to require significant coordination between the independent factor matrices), and on the other hand there does not appear to be a simple way to easily rule out such occurrences. The bulk of the argument is occupied with this problem, which is not encountered in the case of a single i.i.d. random matrix (where all entries of all rows are random) and whose resolution depends on careful consideration of the interplay between the block structure and the linear spaces spanned by the small singular vectors of the individual factor matrices.

As a corollary, Theorem 7 implies a least singular value bound on the original product matrix.

COROLLARY 8. *Let* $X_1, \ldots, X_M$ *be independent n by n i.i.d. random matrices with sub-Gaussian decay* (*featuring potentially different atom distributions*). *For any* $z \in \mathbb{C}$, *let* $\sigma_1(X(z))$ *denote the smallest singular value of the random matrix* $X(z) = X_1 \cdots X_M - z^M I$, *where I denotes the identity matrix. For any sufficiently small constant* $A > 0$, *there exists* $t_0 > 0$ (*depending on A, M and the distributions of the entries*) *such that if* $n^{1/2-t_0} \le |z| \le n^{1/2+t_0}$, *then*

$$(12) \qquad \mathbf{P}\{\sigma_1(X(z)) \le |z|^{M-1} n^{-1/2-A}\} \le Cn^{-KA}.$$

Here, the constant C depends only on M, A and the distributions of the entries; K depends only on M and the distributions of the entries. The requisite smallness of A depends on M and the distributions of the entries.

The remainder of this paper is organized as follows. In Section 3, we present the proof of Theorem 7 and the proof of its corollary for matrix products. In Section 4, we present a version of Theorem 3 in the base case of Gaussian entries; the proof of this result along with the proof of Theorem 4, which uses similar techniques, are given in the Supplementary Material [28]. In Section 5, we prove Theorem 3 in full generality, and in Section 6 we prove Theorems 5 and 6. The proofs of Theorems 5 and 6 reference the proof of Theorem 3 to some degree; otherwise, these sections may be read independently of one another.

2. Notation. In this section, we will collect some elementary notation which we will use throughout the paper.

We will write $f_n = O(g_n)$, $f_n \ll g_n$, $g_n \gg f_n$ or $g_n = \Omega(f_n)$ if $|f_n| \le C|g_n|$ holds for all $n > C$ and for some fixed $C > 0$. If the value of the constant C depends on some parameters, we will denote this with subscripts, for example, if the constant depends on the parameters $a_1, \ldots, a_k$, we would write $f_n = O_{a_1, \ldots, a_k}(g_n)$. We will write $f_n = o(g_n)$ to denote $f_n/g_n \to 0$ as $n \to \infty$. We will also make heavy use of the following two basic definitions from probability theory.

DEFINITION 9. A sequence of events E_n holds with *overwhelming probability* if for every fixed $A > 0$ there exists a constant $C_A > 0$ such that the estimate $\mathbf{P}(E_n) \ge 1 - C_A n^{-A}$ holds for all n.

DEFINITION 10. A sequence of events E_n holds with *exponential probability* (or *exponentially high probability*) if there exist constants $C, c > 0$ such that the estimate $\mathbf{P}(E_n) \geq 1 - C \exp(-cn)$ holds for all n.

We will also make extensive use of the 1-, 2- and ∞-norms for vectors. Letting $v = (v_1, \ldots, v_n) \in \mathbb{C}^n$:

$$
(13) \qquad \|v\|_1 = |v_1| + |v_2| + \cdots + |v_n|,
$$

$$
(14) \qquad |v| = \|v\|_2 = \sqrt{|v_1|^2 + \cdots + |v_n|^2},
$$

$$
(15) \qquad \|v\|_\infty = \max_j |v_j|.
$$

For a matrix M, we define the operator norm $\|M\|_{\mathrm{op}}$ as follows:

$$
(16) \qquad \|M\|_{\mathrm{op}} = \sup_{\|v\|_2 = 1} \|Mv\|_2.
$$

For complex functions f defined on the measure space (X, Σ, μ), we define the usual 1-, 2- and ∞-norms:

$$
(17) \qquad \|f\|_1 = \int_X |f(x)| \, d\mu,
$$

$$
(18) \qquad \|f\|_2 = \sqrt{\int_X |f(x)|^2 \, d\mu},
$$

$$
(19) \qquad \|f\|_\infty = \operatorname{ess\,sup} |f(x)|.
$$

In the last definition, we use the essential supremum of $|f(x)|$ to denote the smallest constant which bounds $|f(x)|$ almost everywhere with respect to the appropriate measure. On the linear algebra side of things, we will let $V^\perp$ denote the vector space of all vectors orthogonal to V, and will frequently use $X - z$ to denote the matrix $X - zI$, where X is some matrix, I is the identity matrix of matching dimension and $z \in \mathbb{C}$. For a $p \times n$ matrix A with $p \leq n$, we let $\sigma_1(A) \leq \cdots \leq \sigma_p(A)$ denote the ordered singular values of A. In particular,

$$
\sigma_1(A) = \inf_{\|v\|_2 = 1} \|Av\|_2.
$$

We let $[m]$ denote the discrete interval $\{1, \ldots, m\}$. For a finite set S, $|S|$ is the cardinality of S. We often use $\sqrt{-1}$ for the imaginary unit and reserve i as an index.

3. Smallest singular value.

3.1. *Preliminary definitions and lemmas.* In this section, we provide the proof of Theorem 7 and Corollary 8. We begin by collecting some elementary definitions and lemmas which we will require, and also make some preliminary reductions. Let $X^{(i)}$ for $1 \leq i \leq M$ and $Y(z)$ be as in the statement of Theorem 7, and let the small positive parameter $A > 0$ be arbitrary. In several parts of the proof, we assume $M > 1$ as Theorem 7 and Corollary 8 both follow from [49], Theorem 3.2, in the case $M = 1$.

Our argument will rely on the following estimate on the singular values of the factor matrices $X^{(i)}$, which follows from the proof of Lemma 4.11 in [12].

LEMMA 11. *There exists a constant $\gamma \in (0, 1)$ and a constant $c_0 > 0$ such that with overwhelming probability, the following estimate hold simultaneously for the singular values*

$\sigma_1(X^{(i)}) \leq \cdots \leq \sigma_n(X^{(i)})$ *of each factor matrix* $X^{(i)}$ *and for all choices of* $0 < \tau < \gamma$ *(such that* $n^{1-\tau}$ *is an integer):*

$$\sigma_{n^{1-\tau}}(X^{(i)}) \geq c_0 n^{1/2-\tau}. \tag{20}$$

The constants γ *and* c_0 *depend only on the atom distributions of the matrices in question.*

We will also need the small ball probability bound, which appears as Corollary 6.3 in [49].

LEMMA 12. *Let* $\xi_1, \ldots, \xi_n$ *be i.i.d. random variables with mean zero and variance one. Then there exists a constant* $c > 0$ *such that*

$$\mathbf{P}\{|\xi_1 v_1 + \cdots + \xi_n v_n - z| < c\} \leq 1 - c$$

for all $z \in \mathbb{C}$ *and all unit vectors* $v = (v_1, \ldots, v_n) \in \mathbb{C}^n$.

It is technically possible to operate in greater generality here: the assumption of identical distribution can be weakened to the more general technical condition that the random variables in question be κ-controlled, but we will not seek such refinements here. For more information on κ-controlled distributions, the reader is referred to [49].

We will also need the classical Chernoff bound for indicator variables; see, for example, [13].

LEMMA 13. *Let* $\mathcal{X}_1, \ldots, \mathcal{X}_n$ *be independent indicator variables, and let* $\mu = \mathbf{E}[\sum_{i=1}^n \mathcal{X}_i]$ *denote the expectation of the sum. Then for any* $0 < \delta \leq 1$,

$$\mathbf{P}\left\{\sum_{i=1}^n \mathcal{X}_i \geq (1+\delta)\mu\right\} \leq e^{-\mu\delta^2/3} \tag{21}$$

and

$$\mathbf{P}\left\{\sum_{i=1}^n \mathcal{X}_i \leq (1-\delta)\mu\right\} \leq e^{-\mu\delta^2/2}. \tag{22}$$

We will frequently use the Chernoff bound in conjunction with a well-known cardinality bound on epsilon nets. A set of vectors $\mathcal{N}_\epsilon \subseteq \Omega \subseteq \mathbb{C}^n$ is said to constitute an ϵ-net for Ω if for any vector $v \in \Omega$ there exists $u \in \mathcal{N}_\epsilon$ such that $\|u - v\|_2 \leq \epsilon$.

LEMMA 14. *For any* $\epsilon > 0$, *the unit sphere of* $\mathbb{C}^n$ *contains an* ϵ-net $\mathcal{N}_\epsilon$ *whose cardinality satisfies*

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

The reader is referred to the excellent resource [54] for a proof. We will also require the following estimate on the sizes of nets

LEMMA 15. *For* $a, b \in (0, 1)$, *the set of unit vectors in* $\mathbb{C}^n$ *with at most* an *nonzero coordinates admits a* b-net *with cardinality at most*

$$\binom{n}{an}\left(\frac{C}{b}\right)^{2an} \leq \left(\frac{eC}{ab}\right)^{2an}.$$

Here, $C > 0$ *is an absolute constant.*

This follows from Lemma 14 along with the well-known bound on binomial coefficients $\binom{n}{k} \leq (ne/k)^k$ (valid for all k less than n).

During the course of the argument, we will find it necessary to work with so-called *compressible* and *incompressible* vectors separately. Intuitively, this distinction captures whether or not a vector is well approximated by a sparse vector. Precisely, we define a compressible vector, with parameters $a, b \in (0, 1)$, to be a vector v' such that there exists a vector v supported on at most an coordinates with $\|v - v'\|_2 \leq b$. Let $\text{Comp}(a, b)$ denote the set of compressible unit vectors with parameters a and b, and let $\text{Incomp}(a, b)$, the set of incompressible unit vectors, denote its complement on the unit sphere.

Without loss of generality, we now place some assumptions on the matrices we will be working with, ruling out some very unlikely pathologies. By the bounds on the operator norm of i.i.d. random matrices with sub-Gaussian entries [54] we will often work on the event

$$(23) \qquad \max\{\|X^{(1)}\|_{\text{op}}, \ldots, \|X^{(M)}\|_{\text{op}}\} \leq \sqrt{n}(1 + r_0),$$

which holds with exponential probability for r_0 sufficiently large.

Let $\tilde{X}^{(i)}$ be the $n - 1$ by n rectangular matrix formed by deleting the last row of $X^{(i)}$, and let $\tilde{Y}(z)$ be the $Mn - 1$ by Mn rectangular matrix formed by deleting the last row of $Y(z)$. By the upper bound on the operator norm of $X^{(i)}$, for sufficiently large and fixed $r_0 > 0$, the following holds:

$$(24) \qquad \max_i\{\|X^{(i)}\|_{\text{op}}, \|\tilde{X}^{(i)}\|_{\text{op}}\} \leq \sqrt{n}(1 + r_0)$$

with exponential probability.

3.2. *The compressible case.* Our starting observation is that it is quite unlikely for any fixed selection of unit vector to be orthogonal to the row space of $Y(z)$.

LEMMA 16. *The following estimate holds for all $z \in \mathbb{C}$ and any fixed unit vector u:*

$$\mathbf{P}\{\|Y(z)u\|_2 \leq c_1\sqrt{n}\} \leq O(\exp(-c_0 n)).$$

The implied constants featured in this bound, as well as $c_0 > 0$ and $c_1 > 0$, can be taken to depend only on M and the atom distributions of the factor matrices.

PROOF. Begin by letting u be any fixed unit vector, and write $u = (u^{(1)}, u^{(2)}, \ldots, u^{(M)})$, where each $u^{(l)} \in \mathbb{C}^n$. The estimate $\|u^{(l)}\|_2 \geq M^{-1/2}$ must hold for at least one l: without loss of generality, we will take this estimate to hold for $l = 2$. Let us then expand (for $1 \leq k \leq n$):

$$|(Y(z)u)_k| = \left|\left(\sum_{j=1}^n X_{k,j}^{(1)} u_j^{(2)}\right) - z u_k^{(1)}\right|.$$

Define T_k to be the indicator variable associated with the event that the following estimate holds:

$$(25) \qquad \left|\left(\sum_{j=1}^n X_{k,j}^{(1)} u_j^{(2)}\right) - z u_k^{(1)}\right| \leq \frac{c}{M^{1/2}}.$$

Since by assumption u is a vector whose first n coordinates have 2-norm at least $M^{-1/2}$, we are able to choose the constant $c > 0$ appropriately (in a manner that depends only on the distribution of $X_{j,k}^{(1)}$) such that Lemma 12, the small ball probability bound, implies that

$$(26) \qquad \mathbf{P}\{T_k = 1\} \leq 1 - c.$$

We would like to bound the probability of the following rare event:

$$\sum_{i=1}^{n} T_i \geq n\left(1 - \frac{c}{2}\right).$$

Clearly, this probability is maximized when the inequality in (26) is an equality, so we may as well assume that $T_k = 1$ with probability $1 - c$. By the Chernoff bound for indicator variables, we have that for any $\delta > 0$:

$$\mathbf{P}\left\{\sum_{i=1}^{n} T_i \geq n(1 - c)(1 + \delta)\right\} \leq \exp(-n(1 - c)\delta^2/3).$$

Taking the parameter δ to be $c/(2(1 - c))$ (we can safely assume $c < 1/2$, as we can always choose smaller c) and $c_0 = c^2/(12(1 - c))$:

$$\mathbf{P}\left\{\sum_{i=1}^{n} T_i \geq n\left(1 - \frac{c}{2}\right)\right\} \leq \exp(-c_0 n).$$

This means that, with exponential probability, for at least $(cn/4)$ coordinates, the following estimate holds:

$$|(Y(z)u)_k| \geq \frac{c}{M^{1/2}}.$$

Since we can bound the norm of $Y(z)u$ from below by the norm of its first $cn/4$ coordinates:

$$\mathbf{P}\left\{\|Y(z)u^{(1)}\|_2^2 \leq \frac{c^3 n}{4M}\right\} \leq O(\exp(-c_0 n)).$$

Setting $c_1 = \sqrt{\frac{c^3}{4M}}$ concludes the argument. □

As a quick corollary, let us use this lemma to handle the case of compressible vectors. The thrust of the argument is that, since the set of compressible vectors can be well approximated by a relatively small ϵ-net, the union bound may be applied and yield nontrivial bounds.

LEMMA 17. *For every $\epsilon > 0$, there exists $t_0, \theta > 0$ (depending on ϵ, M, the constant r_0 from (23) and the atom distribution of the factor matrices $X^{(1)}, \ldots, X^{(M)}$) such that the following estimate holds for all $|z| \leq n^{1/2 + t_0}$, for $a = 1/\log(n)$ and for $b = \theta n^{-\epsilon}$:*

$$(27) \qquad \mathbf{P}\left\{\min_{u \in \mathrm{Comp}(a,b)} \|Y(z)u\|_2 \leq c_1 \sqrt{n}\right\} \leq O(\exp(-c_0 n)).$$

The implied constants featured in this bound, and also the constants $c_0 > 0$ and $c_1 > 0$, can be taken to depend only on ϵ, M and the atom distributions of the matrices $X^{(i)}$.

PROOF. It suffices to prove the bound for all sufficiently small values of $\epsilon > 0$, as the bound then trivially holds for all larger values of ϵ (since $\mathrm{Comp}(a, b') \subset \mathrm{Comp}(a, b)$ for all $b' \leq b$). Let $\Omega_{a,b}$ denote a b-net of the set of all vectors with unit length and at most an nonzero coordinates. Using Lemma 15, we can bound the cardinality of this net:

$$|\Omega_{a,b}| \leq \binom{n}{an}\left(\frac{C}{b}\right)^{2an} \leq \left(\frac{e}{a}\frac{C}{b}\right)^{2an} = \exp\left(2an\log\left(\frac{Ce}{ba}\right)\right).$$

Set $a = 1/\log(n)$ and $b = \theta n^{-\epsilon}$ for ϵ sufficiently small (to be determined momentarily) and $\theta > 0$ to be chosen shortly. Our net can then be chosen to satisfy the cardinality bound:

$$|\Omega_{a,b}| \leq \exp\left(\frac{2n}{\log(n)}\log(Ce\theta^{-1}n^{\epsilon}\log(n))\right) \leq O(\exp(c_3 n)),$$

where c_3 can be chosen arbitrarily small by taking ϵ sufficiently small (and because $\log(\log(n))/\log(n)$ tends to zero). By the union bound and Lemma 16, if we take c_3 to be small enough (and hence ϵ small enough), we have the following probability estimate:

$$(28) \qquad \mathbf{P}\left\{ \min_{u \in \Omega_{a,b}} \|Y(z)u\|_2 \le c_1 \sqrt{n} \right\} \le O\big(\exp(-cn)\big).$$

This proves the result over the net, it remains to extend this bound to the rest of the set of compressible vectors. Recall that a vector u in $\mathrm{Comp}(a,b)$ is distance at most b away from a vector with at most an nonzero coordinates and, therefore, is at most distance $2b$ away from a vector u' in $\Omega_{a,b}$. Suppose there exists a vector $u \in \mathrm{Comp}(a,b)$ such that $\|Y(z)u\| \le \frac{c_1}{10}\sqrt{n}$ and that $\|Y(z)\| \le (2+r_0)n^{1/2+t_0}$. Then there exists a vector $u' \in \Omega_{a,b}$ such that $\|u - u'\| \le 2b$. Hence, by the triangle inequality, taking $\theta = \frac{c_1}{30(2+r_0)}$ and $t_0 = \epsilon$ gives

$$\|Y(z)u'\| \le 2b(2+r_0)n^{1/2+t_0} + \frac{c_1}{10}\sqrt{n} \le c_1\sqrt{n}.$$

In view of (28), we conclude that

$$\mathbf{P}\left\{ \min_{u \in \mathrm{Comp}(a,b)} \|Y(z)u\|_2 \le \frac{c_1}{10}\sqrt{n} \text{ and } \|Y(z)\| \le (2+r_0)n^{1/2+t_0} \right\}$$

$$\le \mathbf{P}\left\{ \min_{u \in \Omega_{a,b}} \|Y(z)u\|_2 \le c_1\sqrt{n} \right\}$$

$$\le O\big(\exp(-cn)\big).$$

Since the norm bound $\|Y(z)\| \le (2+r_0)n^{1/2+t_0}$ holds with exponential probability (due to (23) and the assumption that $|z| \le n^{1/2+t_0}$), the claim follows. $\square$

We will utilize Lemma 17 at the end of the section. Notice that the proof to Lemma 17 goes over if we delete a row from either a factor matrix $X^{(i)}$ or $Y(z)$ (simply by replacing n with $n-1$ as appropriate gives the equivalent results in this rectangular case).

LEMMA 18. *For every $\epsilon > 0$, there exists $t_0, \theta > 0$ (depending on ϵ, M, the constant r_0 from (23) and the atom distribution of the factor matrices $X^{(1)}, \ldots, X^{(M)}$) such that the following estimate holds for all $|z| \le n^{1/2+t_0}$, for $a = 1/\log(n)$ and for $b = \theta n^{-\epsilon}$:*

$$\mathbf{P}\left\{ \min_{u \in \mathrm{Comp}(a,b)} \|\tilde{Y}(z)u\|_2 \le c_0\sqrt{n} \right\} \le O\big(\exp(-cn)\big).$$

The constants in this bound depend only on ϵ, M and the atom distributions of the matrices $X^{(i)}$.

This extension will be useful for some of the arguments we will make later on as a quick way of ruling out certain problematic scenarios without repeating all the details of this section.

3.3. *The incompressible case, Part I.* We need to collect some preliminary lemmas in preparation for dealing with incompressible vectors.

First, an overview. Recall that $\tilde{Y}(z)$ denotes the rectangular matrix produced by deleting the last row of $Y(z)$. In this section, we will endeavor to show that, with high probability, any unit vector v which is orthogonal to the rows of $\tilde{Y}(z)$ has the property that its first n coordinates are not too close, in a certain sense, to the singular vectors corresponding to the small singular values of the product of the first $M-1$ factor matrices. The end game is to control the unit normal to the row space of $\tilde{Y}(z)$, and to show that it does not place the vast majority of its weight away from its leading n coordinates.

Observe that if v is orthogonal to the first $Mn - 1$ rows of $Y(z)$, the block structure which results from linearization implies that

$$v^{(1)} = \frac{1}{z^{M-1}} X^{(1)} \cdots X^{(M-1)} v^{(M)}$$

and

$$\tilde{I} v^{(M)} = \frac{1}{z} \tilde{X}^{(M)} v^{(1)},$$

where $\tilde{I}$ is the $(n - 1) \times n$ matrix formed from the $n \times n$ identity matrix by removing the last row. In view of these identities, we would like to show that a specific class of unit vectors v satisfy an estimate which looks like

$$\tag{29} \left\| \prod_{i=1}^{M-1} \left(\frac{1}{z} X^{(i)} \right) v \right\|_2 > n^{-\epsilon_0}$$

for an appropriate set of values of $z \in \mathbb{C}$ (this set of complex numbers will depend on M). One of the main technical challenges will be showing that ϵ_0 can be chosen independent of M.

We will need to get a handle on the small singular values of the matrices $X^{(i)}$ to accomplish our goal, so let us start there. We will let $\sigma_1^{(i)} \leq \sigma_2^{(i)} \leq \cdots \leq \sigma_n^{(i)}$ denote the ordered singular values of $X^{(i)}$, and let $u_j^{(i)}$ denote some choice of associated unit singular vectors:

$$\tag{30} (X^{(i)})^* (X^{(i)}) u_j^{(i)} = (\sigma_j^{(i)})^2 u_j^{(i)}.$$

Since we have a decent estimate on the number of small singular values of our factor matrices, we will try to isolate all vectors which interact with these singular values into a space of modest dimension. The advantage here is that we may now apply epsilon-net arguments which are not admissible when dealing with incompressible vectors more generally. The construction of this space is straightforward and is the occupation of the next lemma. The parameter τ determines exactly which singular values we will categorize as small, and an explicit selection of τ will be made later in the argument.

LEMMA 19. *Let $c, \tau > 0$ be sufficiently small constants. With overwhelming probability, there exists a linear subspace V_τ with dimension at most $O(n^{1-\tau})$ such that for all unit vectors v orthogonal to V_τ:*

$$\tag{31} \| X^{(1)} \cdots X^{(M-1)} v \|_2 \geq \frac{c n^{(1/2)(M-1)}}{n^{(M-1)\tau}}.$$

In addition, V_τ can be taken to be the linear subspace spanned by the singular vectors corresponding to the $O(n^{1-\tau})$ smallest singular values of $X^{(1)} \cdots X^{(M-1)}$.

PROOF. We first construct a subspace V_τ with the desired properties; in the second half of the proof, we will show that V_τ can be taken to be the linear subspace spanned by the singular vectors corresponding to the $O(n^{1-\tau})$ smallest singular values of $X^{(1)} \cdots X^{(M-1)}$. Indeed, for each $i \in \{1, \ldots, M - 1\}$, let $W_\tau^{(i)}$ denote the vector space which is given by the span of $u_j^{(i)}$ for $1 \leq j \leq n^{1-\tau}$. With overwhelming probability, we may assume that each of the factor matrices $X^{(1)}, \ldots, X^{(M-1)}$ is both invertible (by Proposition 27 in [52]) and satisfies (20) (with the same τ as appears in the statement of the lemma, as we can take τ to be sufficiently small.)

Define the space $\tilde{V}_\tau^{(M-1)}$ to simply be equal to $W_\tau^{(M-1)}$, and define $Z_\tau^{(M-1)}$ to be the image of the vector space $(W_\tau^{(M-1)})^\perp$ under the matrix $(X^{(M-1)})^{-1}$, which we may

write $(X^{(M-1)})^{-1}(W_\tau^{(M-1)})^\perp$. Define $\tilde{V}_\tau^{(M-2)}$ to be the span of the union of $\tilde{V}_\tau^{(M-1)}$ with $(Z_\tau^{(M-1)})^\perp$. Then the dimension of $\tilde{V}_\tau^{(M-2)}$ is at most $2n^{1-\tau}$ (because it is the span of the union of two spaces each with dimension at most $n^{1-\tau}$).

Because $\tilde{V}_\tau^{(M-2)}$ contains $\tilde{V}_\tau^{(M-1)}$, Lemma 11 implies that every unit vector v orthogonal to $\tilde{V}_\tau^{(M-2)}$ meets the following condition:

$$\|X^{(M-1)}v\|_2 \geq cn^{1/2-\tau}.$$

Further, the orthogonality between v and $\tilde{V}_\tau^{(M-2)}$ ensures that $v \in Z_\tau^{(M-1)}$ by construction and, therefore, that the vector $X^{(M-1)}v$ is orthogonal to the space $W_\tau^{(M-2)}$. Lemma 11 then guarantees that

$$\|X^{(M-2)}X^{(M-1)}v\|_2 \geq cn^{1/2-\tau}\|X^{(M-1)}v\|_2.$$

Combining the preceding two estimates,

$$\|X^{(M-2)}X^{(M-1)}v\|_2 \geq c^2 n^{1-2\tau}.$$

Continuing the construction iteratively, we construct a space $\tilde{V}_\tau^{(1)}$ with dimension at most $O(n^{1-\tau})$ such that for all unit vectors v orthogonal to $\tilde{V}_\tau^{(1)}$:

$$(32) \qquad \|X^{(1)}\cdots X^{(M-2)}X^{(M-1)}v\|_2 \geq c^{M-1}n^{(1/2)(M-1)-(M-1)\tau}.$$

This shows that the subspace V_τ has exactly the properties we require. To complete the proof, we need to show that V_τ can be taken to be spanned by the singular vectors of $X^{(1)}\cdots X^{(M-1)}$. Indeed, we now use V_τ to construct another linear subspace Σ_τ which satisfies the same properties, but which is spanned by the singular vectors of the product. Let d be the dimension of V_τ. Fix a realization in which (31) holds and in which $d = O(n^{1-\tau})$. Let Σ_τ denote the linear subspace spanned by the singular vectors of $X^{(1)}\cdots X^{(M-1)}$ corresponding to singular values which are strictly smaller than $\frac{cn^{(1/2)(M-1)}}{n^{(M-1)\tau}}$ (if there are no such singular values, take Σ_τ to be the trivial subspace). By the orthogonality of the singular vectors, it follows that if v is a unit vector orthogonal to Σ_τ, then

$$\|X^{(1)}\cdots X^{(M-1)}v\|_2 \geq \frac{cn^{(1/2)(M-1)}}{n^{(M-1)\tau}}.$$

It remains to show that the dimension of Σ_τ is $O(n^{1-\tau})$. This follows from (31). Indeed, by the minimax principle for singular vectors (see, for instance, [10], Problem III.6.1), it follows that

$$\sigma_{d+1}\big(X^{(1)}\cdots X^{(M-1)}\big) \geq \min_{v \in V_\tau^\perp, \|v\|=1} \|X^{(1)}\cdots X^{(M-1)}v\|$$

$$\geq \frac{cn^{(1/2)(M-1)}}{n^{(M-1)\tau}}.$$

This implies that $\dim(\Sigma_\tau) \leq d = O(n^{1-\tau})$, completing the proof of the lemma. $\square$

Next, we leverage the relatively small size of the space V_τ to control its behavior much in the same way we controlled compressible vectors. Specifically, we will see that it is unlikely for a unit vector in V_τ to be the first n coordinates of a vector approximately normal to both the first $(M-1)n$ rows of $\tilde{Y}(z)$ (which we may condition on) and the remaining rows (which are still random).

LEMMA 20. *Let $\tilde{I}_n$ be the $n - 1$ by n matrix formed by deleting the last row of the n-dimensional identity matrix, and assume that $z \in \mathbb{C}$ is such that $|z| \in [n^{1/2-\delta}, n^{1/2+\delta}]$ for some choice of $\delta \in (0, \frac{1}{4(M+1)})$. Then with overwhelming probability the matrices $X^{(1)}, \ldots, X^{(M-1)}$ are such that the following probability estimate holds with respect to the random matrix $\tilde{X}^{(M)}$:*

$$\mathbf{P}\left\{ \min_{\|v\|=1, v \in V_\tau} \left\| \left(z\tilde{I}_n - \tilde{X}^{(M)} \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)}\right) \right) v \right\|_2 \leq C_\delta n^{1/2-(M+1)\delta} \right\}$$

$$\leq C_\delta^{-1} \exp(-c_\tau n).$$

The constant $C_\delta > 0$ depends on δ, on M and on the distribution of the matrix entries, while c_τ depends on the distribution of the matrix entries and also on the choice of τ.

PROOF. By Lemma 19, the dimension of V_τ is at most $O(n^{1-\tau})$ with overwhelming probability. Assume this is so, and choose a $(1/\sqrt{n})$-net of the unit ball of V_τ, which we will denote by Ω_τ. By Lemma 14, we can take this net to have cardinality at most

$$(33) \qquad |\Omega_\tau| \leq (1 + 2n^{1/2})^{2n^{1-\tau}} \leq \exp([(3/2)n^{-\tau}\log(n)]n)$$

for n sufficiently large. Crucially, as long as $\tau > 0$ this cardinality will grow slower than $\exp(cn)$ for any fixed choice of $c > 0$. Now, fix an arbitrary choice of $v \in \Omega_\tau$ and, for notational simplicity, define the vector y_v as follows:

$$y_v = \left(\frac{1}{z} X^{(1)}\right)\left(\frac{1}{z} X^{(2)}\right)\left(\frac{1}{z} X^{(3)}\right) \cdots \left(\frac{1}{z} X^{(M-1)}\right) v.$$

For $y_v \neq 0$, Lemma 12, the small ball probability bound, implies the following anticoncentration estimate for $1 \leq j \leq n - 1$:

$$\mathbf{P}\left\{ \left| \sum_{i=1}^{n} \tilde{X}_{j,i}^{(M)}(y_v)_i - zv_j \right| \leq c\|y_v\|_2 \right\} \leq 1 - c.$$

Let T_j, for $1 \leq j \leq n-1$, be the indicator variable associated with the event that the following estimate holds:

$$\left| \sum_{i=1}^{n} \tilde{X}_{j,i}^{(M)}(y_v)_i - zv_j \right|^2 \geq c^2\|y\|_2^2.$$

By the small ball probability bound, each T_j equals 1 with probability at least c, and by the Chernoff inequality (Lemma 13) for sums of indicator variables,

$$\mathbf{P}\left\{ \sum_{j=1}^{n-1} T_j \leq \frac{c}{2}(n-1) \right\} \leq \exp(-c(n-1)/8).$$

On the complement of this event at least $(c/2)(n-1)$ many of the T_j must then be equal to 1. Consequently,

$$\mathbf{P}\left\{ \|\tilde{X}^{(M)} y_v - z\tilde{I}_n v\|_2^2 \leq \frac{c^3}{2}\|y_v\|_2^2(n-1) \right\}$$

$$\leq \exp(-cn/8)$$

for a slightly different constant $c > 0$ and all n sufficiently large. To proceed further, we need to have an estimate of the quantity $\|y_v\|_2$, which we will handle by inspecting various cases. If we assume that $\|y_v\|_2 \geq (n^{1/2}/|z|)^{M-1}$, then we can conclude that there exist constants

$C_0, c_0 > 0$ depending only on the entry distributions such that for this choice of unit vector v:

(34) $$\mathbf{P}\{\|\tilde{X}^{(M)}y_v - z\tilde{I}_n v\|_2 \le C_0|z|(n^{1/2}/|z|)^M\} \le C_0^{-1}\exp(-c_0 n).$$

If we instead assume that $\|y_v\|_2 \ge n^{-2\delta}$, then we can similarly conclude that there exist constants $C_0, c_0 > 0$ depending only on the entry distributions such that for this choice of unit vector v:

(35) $$\mathbf{P}\{\|\tilde{X}^{(M)}y_v - z\tilde{I}_n v\|_2 \le C_0 n^{1/2-2\delta}\} \le C_0^{-1}\exp(-c_0 n).$$

These two cases handle the situations where y_v is relatively large, we now consider the case where y_v is relatively small. Specifically, assume that

(36) $$\|y_v\|_2 \le \min\{(n^{1/2}/|z|)^{M-1}, n^{-2\delta}\}.$$

The Chernoff bound method employed above is not helpful for such y_v, as the bounds we would obtain would be too weak to establish the desired result. We therefore employ a different approach in this case, which takes advantage of the largeness of $|z|$ instead of the randomness of the rows of $\tilde{X}^{(M)}$. For this method to work, we need to make sure that v does not put too much mass on its last coordinate, so for now also assume that the following estimate holds:

(37) $$\|\tilde{I}_n v\|_2 \ge n^{-\delta/2}.$$

We will justify this assumption in due course, but first notice that the bound on the operator norm of $\tilde{X}^{(M)}$, along with the smallness of $\|y_v\|_2$, imply

$$\|\tilde{X}^{(M)}y_v\|_2 \le (1 + r_0)n^{1/2-2\delta}$$

with exponential probability. Comparing magnitudes, we have the estimate

$$\|z\tilde{I}_n v - \tilde{X}^{(M)}y_v\|_2 \ge (|z|n^{-\delta/2} - (1 + r_0)n^{1/2-2\delta}).$$

Using our assumptions on the magnitude of the parameter $|z|$ and also bounding $3\delta/2$ by 2δ, we obtain the following estimate:

(38) $$\|z\tilde{I}_n v - \tilde{X}^{(M)}y_v\|_2 \ge cn^{1/2-2\delta}$$

for a constant $c > 0$.

It remains to justify (37), which as in the case of compressible vectors, boils down to the fact that the product of our random matrices with any fixed unit vector is not small with very high probability. To that end, let $e_n \in \mathbb{C}^n$ denote the unit Cartesian coordinate vector supported only on the nth coordinate. By the same argument, as we have used in the compressible case (specifically, by the Chernoff bound calculation in the proof of Lemma 16)

$$\mathbf{P}\{\|X^{(M-1)}e_n\|_2 \le C\sqrt{n}\} \le O(\exp(-cn)).$$

Iterating this argument and applying our assumptions on the magnitude of $|z|$, it is exponentially likely that

$$\left\|\left(\frac{1}{z}X^{(1)}\right)\left(\frac{1}{z}X^{(2)}\right)\left(\frac{1}{z}X^{(3)}\right)\cdots\left(\frac{1}{z}X^{(M-1)}\right)e_n\right\|_2 \ge cn^{(M-1)/2}|z|^{-(M-1)}.$$

On the other hand, if v is a vector such that $\|\omega e_n - v\|_2 \le Cn^{-\delta/2}$ for some $\omega \in \mathbb{C}$ with $|\omega| = 1$, then (by the operator norm bound on the matrices $X^{(i)}$)

$$\left\|\left(\frac{1}{z}X^{(1)}\right)\left(\frac{1}{z}X^{(2)}\right)\left(\frac{1}{z}X^{(3)}\right)\cdots\left(\frac{1}{z}X^{(M-1)}\right)(v - \omega e_n)\right\|_2$$
$$\le C(1 + r_0)^{M-1}|z|^{-(M-1)}n^{(M-1)/2}n^{-\delta/2}.$$

Using the fact that $C(1 + r_0)^{M-1}n^{-\delta/2} \to 0$ as $n \to \infty$, as well as the triangle inequality,

$$\left\|\left(\frac{1}{z}X^{(1)}\right)\left(\frac{1}{z}X^{(2)}\right)\left(\frac{1}{z}X^{(3)}\right)\cdots\left(\frac{1}{z}X^{(M-1)}\right)v\right\|_2 \geq cn^{(M-1)/2}|z|^{-(M-1)}.$$

This in turn implies (37), as we can assume that if v is such that (37) fails then $\|y_v\|_2 \geq c(n^{1/2}/|z|)^{M-1}$ necessarily, and we can take $c < 1$.

Combining (38) with (34) and (35), we see that for any choice of unit vector v in the net Ω_τ,

$$\mathbf{P}\{\|\tilde{X}^{(M)}y_v - z\tilde{I}_n v\|_2 \leq C_0 \min\{|z|(n^{1/2}/|z|)^M, n^{1/2-2\delta}\}\} \leq \exp(-c_0 n).$$

Taking the union bound over all v in our net and using (33), we obtain

(39)
$$\mathbf{P}\Big\{\min_{v\in\Omega_\tau}\|\tilde{X}^{(M)}y_v - z\tilde{I}_n v\|_2 \leq C_0 \min\{|z|(n^{1/2}/|z|)^M, n^{1/2-2\delta}\}\Big\}$$
$$\leq O(\exp(-c_0 n)).$$

To complete the proof, we approximate an arbitrary vector in V_τ with a vector in our net. For any unit $v' \in V_\tau$ and $v \in \Omega_\tau$, the triangle inequality and the definition of y_v imply

$$\|\tilde{X}^{(M)}y_v - z\tilde{I}_n v\|_2 - \left\|\left(z\tilde{I}_n - \tilde{X}^{(M)}\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)\right)\right\|_{\text{op}}\|v - v'\|_2$$

$$\leq \left\|\left(z\tilde{I}_n - \tilde{X}^{(M)}\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)\right)v'\right\|_2.$$

By construction of our net, we may take $\|v - v'\|_2 \leq n^{-1/2}$, and by the triangle inequality again we also have the operator norm estimate

$$\left\|\left(z\tilde{I}_n - \tilde{X}^{(M)}\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)\right)\right\|_{\text{op}} \leq |z|(1 + (1+r_0)^M(n^{1/2}/|z|)^M),$$

which holds with exponential probability by (23). In view of (39), it remains to bound the following quantity from below:

$$\min\{|z|(n^{1/2}/|z|)^M, n^{1/2-2\delta}\} - \frac{|z|}{\sqrt{n}}(1 + (1+r_0)^M(n^{1/2}/|z|)^M).$$

To accomplish this, we will inspect both instances of the minimum. On one hand, we have

$$|z|(n^{1/2}/|z|)^M - \frac{|z|}{\sqrt{n}}(1 + (1+r_0)^M(n^{1/2}/|z|)^M)$$

$$= |z|\left(\left(1 - \frac{(1+r_0)^M}{\sqrt{n}}\right)(n^{1/2}/|z|)^M - \frac{1}{\sqrt{n}}\right)$$

$$\geq cn^{1/2-\delta}n^{-M\delta} = cn^{1/2-(M+1)\delta}$$

for n sufficiently large. On the other hand,

$$n^{1/2-2\delta} - \frac{|z|}{\sqrt{n}}(1 + (1+r_0)^M(n^{1/2}/|z|)^M)$$

$$\geq n^{1/2-2\delta} - cn^{(M-1)\delta}$$

$$\geq cn^{1/2-2\delta}.$$

To obtain the last inequality, we have used the assumption $\delta < \frac{1}{4(M+1)}$. Since $M + 1 \geq 2$, we can therefore conclude

$$cn^{1/2-(M+1)\delta} \leq \left\| \left(z\tilde{I}_n - \tilde{X}^{(M)} \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) \right) v' \right\|_2.$$

Since v' was an arbitrary unit vector in V_τ, this concludes the proof. $\square$

We now have a level of control which we can live with over the space V_τ, and it remains to take similar control over vectors which lie largely but not entirely in V_τ. The epsilon-net methods which we have been employing will not work in this case, as the portion of such a vector which does not lie in V_τ may lie in one of any number of directions. Instead, our next lemma follows from an approximation argument.

LEMMA 21. *Let $\epsilon_0 > 0$ be a sufficiently small constant, and suppose that the estimate $n^{1/2-\epsilon_0/16M} \leq |z| \leq n^{1/2+\epsilon_0/16M}$ holds. Let $V[\epsilon_0]$ denote the set of all unit vectors v which satisfy*

$$\left\| \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) v \right\| \leq n^{-\epsilon_0}.$$

Then there exists a positive constant C_{ϵ_0} such that with overwhelming probability $X^{(1)}, \ldots, X^{(M-1)}$ are such that the following probability bound holds with respect to the random matrix $\tilde{X}^{(M)}$:

$$\mathbf{P}\left\{ \min_{v \in V[\epsilon_0]} \left\| \left(z\tilde{I}_n - \tilde{X}^{(M)} \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) \right) v \right\|_2 \leq C_{\epsilon_0} \right\}$$

$$\leq O\left(\exp(-c_{\epsilon_0} n) \right).$$

The choice of ϵ_0 depends only on the atom distributions of the factor matrices. Here, C_{ϵ_0} depends on ϵ_0 and M, and c_{ϵ_0} depends on ϵ_0, the atom distributions of the factor matrices and M.

PROOF. Choose $0 < \tau \leq \epsilon_0/4M$. Applying Lemma 20 with $\delta = \epsilon_0/8(M+1)$, it is overwhelmingly probable that $X^{(1)}, \ldots, X^{(M-1)}$ are such that with probability at least $1 - C_0 \exp(-c_0 n)$ (with respect to $\tilde{X}^{(M)}$) the following event occurs:

$$(40) \qquad \min_{\|v\|=1, v \in V_\tau} \left\| \left(z\tilde{I}_n - \tilde{X}^{(M)} \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) \right) v \right\|_2 \geq Cn^{1/2-\epsilon_0/8}.$$

We will obtain our result by essentially just observing that (40) implies

$$(41) \qquad \min_{v \in V[\epsilon_0]} \left\| \left(z\tilde{I}_n - \tilde{X}^{(M)} \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) \right) v \right\|_2 \geq C_{\epsilon_0}.$$

Indeed, let v be any unit vector such that

$$(42) \qquad \left\| \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) v \right\|_2 \leq n^{-\epsilon_0}.$$

We can decompose $v = v^{(1)} + v^{(2)}$, where $v^{(1)} \in V_\tau$ and $v^{(2)}$ is orthogonal to V_τ (and therefore to $v^{(1)}$ as well). We would like to show that $v^{(2)}$ must be small, and that therefore our

vector v is well approximated by its projection $v^{(1)}$. Using the fact that $v^{(2)}$ is orthogonal to V_τ (and, therefore, lies entirely in the span of the singular vectors associated with the "large" singular values), as well as our assumptions about the magnitude of the complex parameter z, we obtain

$$n^{-(M-1)\epsilon_0/(2M)}\|v^{(2)}\|_2 \leq \left\|\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)v^{(2)}\right\|_2.$$

Following Lemma 19, we see that V_τ is spanned by the singular vectors corresponding to the smallest singular values of the product $X^{(1)}\cdots X^{(M-1)}$. Thus, $v^{(1)}$ can be expressed as a linear combination of singular vectors in V_τ, and $v^{(2)}$ can be expressed as linear combinations of singular vectors from $V_\tau^\perp$. Hence, the orthogonality of singular vectors implies

$$\left\|\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)v\right\|_2 \geq n^{-(M-1)\epsilon_0/(2M)}\|v^{(2)}\|_2.$$

Substituting in (42),

$$n^{-(M-1)\epsilon_0/(2M)}\|v^{(2)}\|_2 \leq \left\|\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)v\right\|_2 \leq n^{-\epsilon_0}$$

and, therefore,

(43)
$$\|v^{(2)}\|_2 \leq n^{-\epsilon_0/2}.$$

Applying the triangle inequality and $\|v - v^{(1)}\|_2 = \|v^{(2)}\|_2$,

$$\left\|\left(z\tilde{I}_n - \tilde{X}^{(M)}\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)\right)v^{(1)}\right\|_2 - \left\|\left(z\tilde{I}_n - \tilde{X}^{(M)}\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)\right)\right\|_{op}\|v^{(2)}\|_2$$

$$\leq \left\|\left(z\tilde{I}_n - \tilde{X}^{(M)}\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)\right)v\right\|_2.$$

By (24), (40) and (43), with exponential probability,

$$\left\|\left(z\tilde{I}_n - \tilde{X}^{(M)}\prod_{h=1}^{M-1}\left(\frac{1}{z}X^{(h)}\right)\right)v\right\|_2 \geq C_{\epsilon_0}n^{1/2-\epsilon_0/8} - Cn^{1/2-\epsilon_0/4}.$$

This is what we wanted to show, as (41) is now established. □

We are now in a position to prove the following.

LEMMA 22. *Suppose that $n^{1/2-\epsilon_0/16M} \leq |z| \leq n^{1/2+\epsilon_0/16M}$, and let u be any unit vector orthogonal to the subspace spanned by the first $Mn - 1$ rows of $Y(z)$. If we write $u = (u^{(1)}, \ldots, u^{(M)})$, where each $u^{(i)}$ has n entries, it is overwhelmingly probable that $X^{(1)}, \ldots, X^{(M-1)}$ are such that the following probability estimate holds (with respect to $\tilde{X}^{(M)}$):*

(44)
$$\mathbf{P}\{\|u^{(1)}\|_2 \leq n^{-2\epsilon_0}\} \leq \exp(-c_{\epsilon_0}n).$$

Here, $\epsilon_0 > 0$ is sufficiently small constant (where the maximal legal choice of ϵ_0 depends on the atom distributions of the factor matrices).

PROOF. By Lemma 21, we may safely work on the event that

$$(45) \qquad \min_{v \in V[\epsilon_0]} \left\| \left(z \tilde{I}_n - \tilde{X}^{(M)} \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) \right) v \right\|_2 \geq C_{\epsilon_0}.$$

This is because the first $M - 1$ factor matrices are, with overwhelming probability, such that this event is exponentially likely with respect to $\tilde{X}^{(M)}$. Since u is normal to the span of the first $nM - 1$ rows of $Y(z)$ by assumption, the block structure of $Y(z)$ implies that for $1 \leq j \leq M - 1$,

$$u^{(j)} = \left(\prod_{h=j}^{M-1} \frac{1}{z} X^{(h)} \right) u^{(M)}$$

and, in particular,

$$(46) \qquad u^{(1)} = \left(\prod_{h=1}^{M-1} \frac{1}{z} X^{(h)} \right) u^{(M)}.$$

This is because, by the block structure and the orthogonality of u, we can express each $u^{(i)}$ in terms of $u^{(i+1)}$. Since $\|z^{-1} X^{(h)}\|_{\mathrm{op}}$ is at most $(1 + r_0) n^{\epsilon_0/16M}$, with exponential probability, the fact that u is a unit vector implies that the 2-norm of $u^{(M)}$ cannot be too small. Specifically, we can assume that

$$(47) \qquad n^{-\epsilon_0} \leq \|u^{(M)}\|_2.$$

By (46) and the orthogonality of u to the last $n - 1$ rows of $\tilde{Y}(z)$,

$$\left\| \left(z \tilde{I}_n - \tilde{X}^{(M)} \prod_{h=1}^{M-1} \left(\frac{1}{z} X^{(h)} \right) \right) u^{(M)} \right\|_2 = 0.$$

Comparing with (45), we have that $u^{(M)}/\|u^{(M)}\|_2$ cannot lie in $V[\epsilon_0]$, and this (along with (47) and (46)) implies

$$\|u^{(1)}\|_2 = \|u^{(M)}\|_2 \left\| \left(\prod_{h=1}^{M-1} \frac{1}{z} X^{(h)} \right) \frac{u^{(M)}}{\|u^{(M)}\|_2} \right\|_2 \geq n^{-2\epsilon_0}.$$

This is what we wanted to show. $\square$

We will also need the following lemma, which follows from a line of reasoning similar to the one used in establishing Lemma 16.

LEMMA 23. *For every sufficiently small constant $\theta > 0$ and for every sufficiently small $t_0 > 0$, the following holds with $a = 1/\log n$, $b = \theta n^{-100(M-1)t_0}$ and $n^{1/2-t_0} \leq |z| \leq n^{1/2+t_0}$. Let $u \in \mathbb{C}^{Mn}$ be a unit vector orthogonal to the first $Mn - 1$ rows of the random matrix $Y(z)$. Write $u = (u^{(1)}, u^{(2)}, \ldots, u^{(M)})$ where $u^{(j)} \in \mathbb{C}^n$ for $1 \leq j \leq M$. Then, with overwhelming probability, any choice of u must be such that $u^{(1)}$ is not identically zero and the normalized vector $u^{(1)}/\|u^{(1)}\|_2$ does not lie in $\mathrm{Comp}(a, b)$. Here, sufficient smallness of θ depends only on M, the constant r_0 from (23) and the distributions of the entries of the factor matrices.*

PROOF. Let $\theta, t_0 \in (0, 1)$ be a sufficiently small constants to be chosen later. The claim that $u^{(1)} \neq 0$ is trivial, as the linear structure of $Y(z)$ and orthogonality assumption on u would then imply that $X^{(1)}$ is noninvertible, which can be ruled out with exponentially high probability [54]; alternatively, one could also use Lemma 22.

It therefore suffices to show that $u^{(1)}/\|u^{(1)}\|_2$ is not compressible. We will assume $M > 1$; in fact, the $M = 1$ case can be deduced from Lemma 18. Let $v = (v^{(1)}, \ldots, v^{(M)})$ be some nonzero vector in $\mathbb{C}^{Mn}$ with each $v^{(i)} \in \mathbb{C}^n$ and $\|v^{(1)}\|_2 = 1$. We will use a net argument to show that, with overwhelming probability, v cannot be orthogonal to the first $Mn - 1$ rows of $Y(z)$ if $v^{(1)} \in \mathrm{Comp}(a, b)$.

If v is orthogonal to the first $Mn - 1$ rows of $Y(z)$, the block structure which results from linearization implies that

$$
(48) \qquad v^{(1)} = \frac{1}{z^{M-1}} X^{(1)} \cdots X^{(M-1)} v^{(M)}
$$

and

$$
(49) \qquad \tilde{I} v^{(M)} = \frac{1}{z} \tilde{X}^{(M)} v^{(1)},
$$

where $\tilde{I}$ is the $(n - 1) \times n$ matrix formed from the $n \times n$ identity matrix by removing the last row; these identities also appeared in the proof of Lemma 22. In particular, since $\|v^{(1)}\|_2 = 1$, with exponentially high probability, we use (23) and (48) to deduce that

$$
1 \le \left(\frac{(1 + r_0) n^{1/2}}{n^{1/2 - t_0}} \right)^{M-1} \|v^{(M)}\|_2,
$$

and hence

$$
(50) \qquad \|v^{(M)}\|_2 \ge C_{r_0} n^{-(M-1)t_0}
$$

for some constant $C_{r_0} > 0$ which depends only on r_0 and M.

Inductively repeating the argument from the proof of Lemma 16 (using Lemma 12 to control the size of each coordinate and then applying the Chernoff bound), we have

$$
\mathbf{P}\{\|z^{M-1} y - X^{(1)} \cdots X^{(M-1)} x\|_2 \le c_1 n^{(M-1)/2 - 3(M-1)t_0}\} = O(\exp(-cn))
$$

for any fixed vectors x, y with $\|x\|_2 \ge \frac{C_{r_0}}{2} n^{-3(M-1)t_0}$.

Let a and b be as in the statement of the lemma, and let $\Omega_{a,b}$ be a $3b$-net of $\mathrm{Comp}(a, b)$, the set of compressible unit vectors in $\mathbb{C}^n$. By Lemma 15, $\Omega_{a,b}$ can be chosen so that

$$
|\Omega_{a,b}| \le \exp\left(2an \log\left(\frac{eC}{ab} \right) \right) = \exp(o(n) + 200n(M - 1)t_0).
$$

Let $\mathcal{N}$ be a b-net of $\{z \in \mathbb{C} : |z| \le n^2\}$. A simple volume argument shows that $\mathcal{N}$ can be chosen so that

$$
l := |\mathcal{N}| \le O(n^{O_M(1)}).
$$

Let $\omega_1, \ldots, \omega_l$ be an enumeration of the elements in $\mathcal{N}$. To each $y \in \Omega_{a,b}$, we associate the vectors $x_{y,1}, \ldots, x_{y,l}$ such that

$$
\tilde{I} x_{y,k} = \frac{1}{z} \tilde{X}^{(M)} y
$$

for each $1 \le k \le l$, and the last coordinate of $x_{y,k}$ is given by ω_k. The vectors $x_{y,1}, \ldots, x_{y,l}$ are random, but only depend on $\tilde{X}^{(M)}$. In particular, these vectors are independent of $X^{(1)}, \ldots, X^{(M-1)}$. For notational simplicity, define

$$
\Omega'_{a,b} = \left\{ (y, x_{y,k}) : y \in \Omega_{a,b}, 1 \le k \le l, \|x_{y,k}\|_2 \ge \frac{C_{r_0}}{2} n^{-3(M-1)t_0} \right\},
$$

so

$$
|\Omega'_{a,b}| \ll \exp(o(n) + 200n(M - 1)t_0).
$$

Taking t_0 sufficiently small and applying the union bound, we conclude that

(51)
$$\mathbf{P}\Big\{ \min_{(y,x)\in\Omega'_{a,b}} \big\| z^{M-1}y - X^{(1)}\cdots X^{(M-1)}x \big\|_2 \le c_1 n^{(M-1)/2-3(M-1)t_0} \Big\}$$
$$= O\big(\exp(-c'n)\big).$$

Here, we have exploited the fact that $\Omega'_{a,b}$ depends only on $\tilde{X}^{(M)}$ and is independent of $X^{(1)}, \ldots, X^{(M-1)}$, while the probability above is only in terms of $X^{(1)}, \ldots, X^{(M-1)}$. In particular, the probability bound above holds uniformly for any realization of $\tilde{X}^{(M)}$.

Now suppose $v = (v^{(1)}, \ldots, v^{(M)}) \in \mathbb{C}^{Mn}$ is normal to the first $Mn - 1$ rows of $Y(z)$ with $\|v^{(1)}\|_2 = 1$, $v^{(1)} \in \mathrm{Comp}(a, b)$, and $C_{r_0} n^{-(M-1)t_0} \le \|v^{(M)}\|_2 \le n$. Then clearly $v^{(1)}$ and $v^{(M)}$ must satisfy (48) and (49). In addition, there exists $y' \in \Omega_{a,b}$ such that $\|v^{(1)} - y'\|_2 \le 3b$. By (49), with exponentially high probability and for all $1 \le k \le l$,

$$\big\| \tilde{I}v^{(M)} - \tilde{I}x_{y',k} \big\|_2 \le 3(1+r_0)n^{t_0}b.$$

By the assumption that $\|v^{(M)}\|_2 \le n$, the last coordinate of $v^{(M)}$ cannot be larger than n. Thus, there exists k such that $\|v^{(M)} - x_{y',k}\|_2 \le 5(1+r_0)n^{t_0}b$. Taking θ sufficiently small (in particular, choosing $\sqrt{\theta}$ small enough), we obtain

$$\big\| v^{(M)} - x_{y',k} \big\|_2 < \frac{C_{r_0}}{2}\sqrt{\theta}n^{-99(M-1)t_0}.$$

In particular, by the lower bound assumption on $\|v^{(M)}\|_2$, this implies that $\|x_{y',k}\|_2 \ge \frac{C_{r_0}}{2}n^{-3(M-1)t_0}$. Hence, $(y', x_{y',k}) \in \Omega'_{a,b}$. Applying (48), (23) and taking θ sufficiently small, we conclude that

$$\big\| z^{M-1}y' - X^{(1)}\cdots X^{(M-1)}x_{y',k} \big\|_2$$
$$\le |z|^{M-1}3b + \big[(1+r_0)\sqrt{n}\big]^{M-1}\frac{C_{r_0}}{2}\sqrt{\theta}n^{-99(M-1)t_0}$$
$$< c_1 n^{(M-1)/2-3(M-1)t_0}$$

with exponentially high probability. Comparing to (51), we obtain

$$\mathbf{P}\{\exists v \text{ orthogonal to rows of } \tilde{Y}(z) \text{ with } v^{(1)} \in \mathrm{Comp}(a, b),$$
$$C_{r_0} n^{-(M-1)t_0} \le \big\| v^{(M)} \big\|_2 \le n\}$$
$$= O\big(\exp(-c''n)\big).$$

To complete the proof, it remains to show, with overwhelming probability, that every vector v orthogonal to the first $Mn - 1$ rows of $Y(z)$ satisfies $C_{r_0} n^{-(M-1)t_0} \le \|v^{(M)}\|_2 \le n$. Indeed, the lower bound follows, with exponentially high probability, due to (50). We now prove the upper bound holds with overwhelming probability. Indeed, in view of (48) and (49), $v^{(M)}$ must satisfy

$$\Big(\tilde{I} - \frac{1}{z^M}\tilde{X}^{(M)}X^{(1)}\cdots X^{(M-1)} \Big)\frac{v^{(M)}}{\|v^{(M)}\|_2} = 0.$$

(Note that $v^{(M)}$ cannot be zero by (48) since $v^{(1)}$ is assumed to be a unit vector.) Taking t_0 sufficiently small and applying Lemma 21, we find that, with overwhelming probability, $v^{(M)}/\|v^{(M)}\|_2 \notin V[16Mt_0]$. Hence, by definition of $V[16Mt_0]$ and (48),

$$\Big\| \frac{1}{z^{M-1}}X^{(1)}\cdots X^{(M-1)}\frac{v^{(M)}}{\|v^{(M)}\|_2} \Big\|_2 = \frac{\|v^{(1)}\|_2}{\|v^{(M)}\|_2} \ge n^{-16Mt_0},$$

which by rearranging (and since $\|v^{(1)}\|_2 = 1$) yields

$$\|v^{(M)}\|_2 \leq n^{16Mt_0} \leq n$$

for t_0 sufficiently small. The proof of the lemma is complete. □

3.4. *The incompressible case, Part II.* Let $\theta > 0$ be a sufficiently small constant satisfying Lemma 23. We now prove the following result for incompressible vectors.

LEMMA 24. *For any sufficiently small constant $A > 0$, there exists a constant $t_0 > 0$ such that the following holds for $a = 1/\log n$ and $b = \theta n^{-100(M-1)t_0}$. Suppose that z lies in the following annulus:*

(52) $$n^{1/2-t_0} \leq |z| \leq n^{1/2+t_0}.$$

Then the following probability bound holds:

$$\mathbf{P}\left\{\min_{v\in\text{Incomp}(a,b)} \|Y(z)v\|_2 \leq cn^{-1/2-A}\right\} \leq O\big((\log n)n^{-KA}\big).$$

Here, c and K are positive constants depending only on the atom distributions of the factor matrices and on M (as does in the implied constant on the right-hand side of the probability estimate).

The argument presented here will in large part follow along the same lines as the argument in [49], and will be organized around the following result of Rudelson and Vershynin [41].

LEMMA 25. *Let dist_k denote the distance between the kth row of $Y(z)$ and the hypersurface spanned by the other $Mn - 1$ rows. Then the following estimate holds for any constants $1 > a, b > 0$:*

(53) $$\mathbf{P}\left\{\min_{v\in\text{Incomp}(a,b)} \|Y(z)v\|_2 \leq \epsilon'b(Mn)^{-1/2}\right\} \leq \frac{1}{aMn}\sum_{k=1}^{Mn} \mathbf{P}\{\text{dist}_k \leq \epsilon'\},$$

where $\epsilon' > 0$ is any arbitrary positive constant.

PROOF OF LEMMA 24. Our task is now to estimate dist_k, which can be formulated as the magnitude of an inner product. It is sufficient to establish the result for the very last row of $Y(z)$, the changes required to establish the result for the other rows being purely notational. Make the following definitions, for constants $c_1, c_2 > 0$ to be determined later:

$$\epsilon = c_1 b n^{-A}, \qquad \rho = c_2 n^{-A}.$$

Define also $S_{\epsilon,\rho}$ to be the set of unit vectors $v \in \mathbb{C}^n$ satisfying the following inequality (where X is a random vector with the same distribution as a row of $X^{(M)}$):

(54) $$\sup_{\xi\in\mathbb{C}}\big[\mathbf{P}(|X \cdot v - \xi| \leq \epsilon)\big] \geq \rho.$$

By Lemma 6.7 in [49], we have that for any $t_0, \theta > 0$, for n large enough, and $A < 1/2$ (and choosing c_1 and c_2 appropriately):

$$S_{\epsilon,\rho} \subset \text{Comp}\left(O\left(\frac{1}{n\rho^2}\right), O\left(\frac{\epsilon}{\rho}\right)\right) \subset \text{Comp}(a, b).$$

Let Φ be an Mn-dimensional vector $\Phi = (\Phi^{(1)}, \ldots, \Phi^{(M)})$ (not necessarily of unit length) orthogonal to first $Mn - 1$ rows of $Y(z)$, with each $\Phi^{(i)} \in \mathbb{C}^n$ and with $\|\Phi^{(1)}\|_2 = 1$. Notice

that Φ can be chosen to depend only on the first $Mn - 1$ rows of $Y(z)$ and is independent of the last row of $X^{(M)}$. By Lemma 23, with overwhelming probability (with respect to $X^{(1)}, \ldots, X^{(M-1)}$), we have that such a normal vector exists and $\Phi^{(1)}$ is not in $S_{\epsilon,\rho}$. This implies that

$$(55) \qquad \mathbf{P}\{\|X_n^{(M)}\Phi^{(1)} - z\Phi^{(M)}\|_2 \le \epsilon\} \le \rho.$$

Here, since Φ is orthogonal to the first $Mn - 1$ rows of $Y(z)$, $\|X_n^{(M)}\Phi^{(1)} - z\Phi^{(M)}\|_2$ is simply the magnitude of the dot product of Φ with the last row of $Y(z)$. For A and t_0 sufficiently small, Lemma 22, with overwhelming probability, guarantees that $\|(\Phi^{(1)}/\|\Phi\|_2)\|_2 \ge n^{-A}$. In other words, (since $\|\Phi^{(1)}\|_2 = 1$), $\|\Phi\|_2 \le n^A$. Rescaling, we obtain

$$(56) \qquad \mathbf{P}\{\text{dist}_{Mn} \le \epsilon n^{-A}\} \le \mathbf{P}\{\|X_n^{(M)}\Phi^{(1)} - z\Phi^{(M)}\|_2 \le \epsilon n^{-A}\|\Phi\|_2\} \le O(\rho).$$

Applying Lemma 25, we conclude that

$$\mathbf{P}\left\{\min_{v \in \text{Incomp}(a,b)} \|Y(z)v\|_2 \le c_1'\theta^2 n^{-2A-1/2-200(M-1)t_0}\right\} \le O((\log n)\rho),$$

where $c_1' = c_1/\sqrt{M}$. Taking t_0 sufficiently small (in terms of A and M), we deduce that

$$\mathbf{P}\left\{\min_{v \in \text{Incomp}(a,b)} \|Y(z)v\|_2 \le c_1'\theta^2 n^{-3A-1/2}\right\} \le O((\log n)\rho).$$

Since this is true for any sufficiently small positive constant A, replacing A by $A/3$ completes the proof of Lemma 24. $\square$

The proof of Theorem 7 can now be completed by combining Lemma 17 with Lemma 24 and bounding the lower order terms (such as bounding $\log n$ above by $n^{KA/2}$ for n sufficiently large). We now turn to the proof of Corollary 8. From the block structure of $Y(z)$, one can compute each $n \times n$ block of the inverse $Y(z)^{-1}$. Indeed, the top-left $n \times n$ minor of $Y(z)^{-1}$ is given by

$$\left(\frac{1}{z^{M-1}}X_1 \cdots X_M - z\right)^{-1}.$$

By Theorem 7,

$$\mathbf{P}\{\|Y(z)^{-1}\|_{\text{op}} \ge n^{1/2+A}\} \le Cn^{-KA},$$

and consequently (using the fact that the operator norm of a matrix bounds the operator norm of any submatrix):

$$\mathbf{P}\left\{\left\|\left(\frac{1}{z^{M-1}}X_1 \cdots X_M - z\right)^{-1}\right\|_{\text{op}} \ge n^{1/2+A}\right\} \le Cn^{-KA}.$$

We conclude that with probability at least $1 - Cn^{-KA}$ the smallest singular value of the translated product matrix $X_1 \cdots X_M - z^M$ is no smaller than $|z|^{M-1}n^{-(1/2+A)}$. This establishes Corollary 8.

4. Linear statistics of product matrices. The main result of this section will be a central limit theorem for linear statistics of products of independent complex Ginibre matrices. Recall that $\mathbb{U}$ represents the unit disk in the complex plane centered at the origin.

THEOREM 26. *Let $f : \mathbb{C} \to \mathbb{R}$ be a function with continuous partial derivatives and at most polynomial growth at infinity. Additionally, let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of the matrix product $n^{-M/2} X^{(1)} \cdots X^{(M)}$, where each $X^{(i)}$ is an independent $n \times n$ complex Ginibre matrix, and let $N_n[f]$ denote the associated centered linear statistic: $N_n[f] = \sum_{j=1}^{n} f(\lambda_j) - \mathbb{E}[\sum_{j=1}^{n} f(\lambda_j)]$. Then $N_n[f]$ converges in distribution to the mean zero normal distribution with the following limiting variance*:

$$(57) \qquad \frac{1}{4\pi} \int_{\mathbb{U}} |\nabla f(z)| \, d^2 z + \frac{1}{2} \|f\|^2_{H^{1/2}(\partial \mathbb{U})}.$$

We will follow the approach developed by Rider and Virág [38] to prove a similar result for the Ginibre ensemble, which will come down to using rotary flow combinatorial machinery to control the cumulants of the statistic in the case of polynomial test functions, and then producing a variance bound to extend the result to the case of more general test functions. The key property which we are exploiting here is the fact that the Ginibre product matrix, like the Ginibre ensemble itself, is a rotationally invariant determinantal point process. The proof of Theorem 26 is given in the Supplementary Material [28]. As the proof of Theorem 4 follows similar arguments, the details are also presented in [28].

Theorem 26 will provide us with a base case to which we will then apply four moment universality.

5. Four moment universality. In this section, we prove Theorem 3 by way of four moment universality (with the bound on the smallest singular value, Theorem 7, playing a crucial role in the argument). The development in this section is based on an approach previously employed in [27] for independent entry matrices, which was itself based on the argument put forward in [52].

We will on several occasions appeal to technical results obtained by Nemish during the course of his proof of the local M-fold circular law [32].

THEOREM 27 (Nemish). *Let $f : \mathbb{C} \to \mathbb{R}$ be a fixed smooth function with compact support. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $n^{-M/2} X^{(1)} \cdots X^{(M)}$, where each jointly independent factor $X^{(i)}$ is an $n \times n$ i.i.d. random matrix. If $|z_0|, (1 - |z_0|) \geq \tau_0$ for some $\tau_0 > 0$, then for any $d \in (0, 1/2)$,*

$$(58) \qquad \left(\frac{1}{n} \sum_{j=1}^{n} f_{z_0}(\lambda_j) - \frac{1}{M\pi} \int_{|z| \leq 1} f_{z_0}(z) |z|^{2/M - 2} \right) \prec n^{-1 + 2d} \|\Delta f\|_1.$$

Here, f_{z_0} is the n^{-d} rescaling of $f(z)$ around z_0:

$$f_{z_0}(z) = n^{2d} f(n^d (z - z_0)).$$

Similar local law results for independent-entry matrices have also been obtained in [7, 14, 15, 52, 56]. The notation $A \prec B$ appearing in the local M-fold circular law denotes stochastic domination.

DEFINITION 28. Let $A_n, B_n \in \mathbb{C}$ be two sequences of random variables for $n \in \mathbb{N}$. The sequence B_n is said to stochastically dominate A_n (written $A_n \prec B_n$) if, for any $\epsilon > 0$ and $D > 0$, the following holds:

$$(59) \qquad \mathbf{P}\{|A_n| \geq n^{\epsilon} |B_n|\} \leq C_D n^{-D}.$$

This explains the presence of the spectral bulk condition $\tau_0 < |z| < 1 - \tau_0$ in Theorem 3; if the arguments in [32] could be extended to the spectral edge then our argument would be extended to the spectral edge as well. Notice that this is asking more than the extension of the *result* in [32]; indeed this has recently been accomplished at the origin in [23], but the arguments used do not appear to translate into the context of the argument presented here.

5.1. *Overview of the argument.* Recall that the linearization matrix of the product $n^{-M/2}X^{(1)} \cdots X^{(M)}$, denoted Y, is defined as follows:

$$
Y = \frac{1}{\sqrt{n}}
\begin{pmatrix}
0 & X^{(1)} & 0 & \cdots & 0 \\
0 & 0 & X^{(2)} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & X^{(M-1)} \\
X^{(M)} & 0 & \cdots & 0 & 0
\end{pmatrix}.
$$

For $\beta = 1, 2$, let $Y^{(\beta)}$ be the linearization matrix associated with the product $n^{-M/2} \times X^{(\beta,1)} \cdots X^{(\beta,M)}$ (with factor matrices obeying the assumptions in the statement of Theorem 29 below). For any choice of complex z in the spectral bulk (defined as in the statement of Theorem 29), define $Y^{(\beta)}(z) = Y^{(\beta)} - zI$, and define also

$$
(60) \qquad W^{(\beta)}(z) = \begin{pmatrix} 0 & Y^{(\beta)}(z) \\ (Y^{(\beta)}(z))^* & 0 \end{pmatrix}.
$$

We let $\lambda_j^{(\beta)}(z)$, for $1 \le j \le 2Mn$, denote the eigenvalues of $W^{(\beta)}(z)$.

Our argument will rest in large part on the classical Girko Hermitization trick, which we will use to get around various complications which stem from the failure of Hermiticity. The trick relies on the following identity for twice continuously differentiable test functions with compact support:

$$
(61) \qquad f(\lambda) = \frac{1}{2\pi} \int_{\mathbb{C}} \Delta f(z) \log|\lambda - z| \, d^2 z.
$$

If $\iota_1, \ldots, \iota_{Mn}$ are the eigenvalues of $Y^{(\beta)}$, then this formula becomes

$$
\sum_{j=1}^{Mn} f(\iota_j) = \frac{1}{2\pi} \int_{\mathbb{C}} \Delta f(z) \log|\det(Y^{(\beta)} - z)| \, d^2 z
$$

$$
= \frac{1}{4\pi} \int_{\mathbb{C}} \Delta f(z) \log|\det W^{(\beta)}(z)| \, d^2 z.
$$

The upshot here is that the matrix $W^{(\beta)}(z)$ is Hermitian, and can therefore be analyzed by the tools of Hermitian random matrix theory. The cost we have incurred is the presence of the integral, which we will need to deal with. Using this formulation, we prove the following four moment universality result for the linear statistics of the linearization matrix.

THEOREM 29. *Suppose $Y^{(1)}$ is the linearization matrix associated with the product matrix $n^{-M/2}X^{(1,1)} \cdots X^{(1,M)}$ and $Y^{(2)}$ is the linearization matrix associated with the product $n^{-M/2}X^{(2,1)} \cdots X^{(2,M)}$ (where all factor matrices are mutually independent n by n i.i.d. matrices), and suppose that the atom distributions of the factors $X^{(1,i)}$ and $X^{(2,i)}$ match to four moments for $1 \le i \le M$. Let $f : \mathbb{C} \to \mathbb{R}$ be a fixed function with two continuous derivatives, supported in the spectral bulk $\tau_0 < |z| < 1 - \tau_0$ for some fixed $\tau_0 > 0$. If the linear statistic generated by the eigenvalues of $Y^{(1)}$ and f, denoted $N_n^{(1)}[f]$, converges in distribution to some limiting distribution χ, then the linear statistic generated by the eigenvalues of $Y^{(2)}$ and f, $N_n^{(2)}[f]$, necessarily converges in distribution to χ as well.*

We can connect this linear statistic with the linear statistic of the product matrix $n^{-M/2} X^{(\beta,1)} \cdots X^{(\beta,M)}$ (with eigenvalues $\mu_1, \ldots, \mu_n$) through the formula (see, for instance, [36] for a complete derivation)

$$\sum_{j=1}^{n} f(\mu_j) = \sum_{j=1}^{Mn} \left[\frac{1}{M} f(\iota_j^M) \right], \tag{62}$$

where $\iota_1, \ldots, \iota_{Mn}$ are the eigenvalues of the linearization matrix of the product $n^{-M/2} \times X^{(\beta,1)} \cdots X^{(\beta,M)}$. Applying universality to the test function $M^{-1} f(z^M)$ immediately extends fourth moment universality for linear statistics of linearization matrices to fourth moment universality for linear statistics of product matrices. Theorem 29 then implies the following corollary.

THEOREM 30. *Suppose that the product matrices $X^{(1)} = n^{-M/2} X^{(1,1)} \cdots X^{(1,M)}$ and $X^{(2)} = n^{-M/2} X^{(2,1)} \cdots X^{(2,M)}$ are both products of mutually independent n by n i.i.d. random matrices, and suppose that the atom distributions of the factors $X^{(1,i)}$ and $X^{(2,i)}$ match to four moments for $1 \leq i \leq M$. Let $f : \mathbb{C} \to \mathbb{R}$ be a fixed function with two continuous derivatives, supported in the spectral bulk $\tau_0 < |z| < 1 - \tau_0$ for some fixed $\tau_0 > 0$. If the linear statistic generated by the eigenvalues of $X^{(1)}$ and f, denoted $N_n^{(1)}[f]$, converges in distribution to some limiting distribution χ, then the linear statistic generated by the eigenvalues of $X^{(2)}$ and f, $N_n^{(2)}[f]$, converges in distribution to χ as well.*

Combining Theorem 30 with the Gaussian result, Theorem 26, one immediately obtains Theorem 3, which is our objective. It remains then to prove Theorem 29, which the rest of this section is dedicated. Throughout the proof, we will assume that all matrices under consideration feature exclusively real entries; the same proof goes forward for matrices with complex entries but with slightly more cumbersome notation.

5.2. *Preliminaries.* Here, we collect some preliminary results and definitions which we will need in order to prove Theorem 29. Define an *elementary matrix* to be a Hermitian matrix featuring one or at most two entries equal to 1, and all the other entries set to zero. Therefore, adding a multiple of an elementary matrix to a Hermitian matrix H changes either a single diagonal entry or two conjugate off-diagonal entries of H, and leaves the other entries undisturbed.

For a $n \times n$ Hermitian matrix H and an elementary matrix V, define

$$H_t = H + \frac{1}{\sqrt{n}} t V, \tag{63}$$

$$R_0(\zeta) = (H - \zeta)^{-1}, \tag{64}$$

$$R_t(\zeta) = (H_t - \zeta)^{-1}, \tag{65}$$

$$s_t(\zeta) = \frac{1}{n} \operatorname{Tr} R_t(\zeta). \tag{66}$$

We will also need to define an appropriate matrix norm:

$$\|A\|_{(\infty,1)} = \max_{1 \leq i,j \leq n} |A_{ij}|.$$

The following Taylor expansion type lemma is due to Tao and Vu (see Proposition 13 in [50]), and is proven by iterating the classical resolvent identity.

LEMMA 31. *Let H be a Hermitian matrix, V an elementary matrix, t and E real numbers and $\eta > 0$. Take $\zeta = E + \sqrt{-1}\eta$. Let $k \geq 0$ be fixed. Suppose we have*

$$|t| \times \|R_0(\zeta)\|_{(\infty,1)} = o(\sqrt{n}).$$

Then we have the following Taylor expansion to order k of the quantity $s_t(\zeta)$:

$$s_0 + \sum_{j=1}^{k} n^{-j/2} c_j t^j$$

$$+ O\left(n^{-(k+1)/2}|t|^{k+1}\|R_0(\zeta)\|_{(\infty,1)}^{k+1} \min\left(\|R_0(\zeta)\|_{(\infty,1)}, \frac{1}{n\eta}\right)\right).$$

The coefficients c_j are independent of t and obey the following estimate:

$$|c_j| \ll \|R_0(\zeta)\|_{(\infty,1)}^{j} \min\left(\|R_0(\zeta)\|_{(\infty,1)}, \frac{1}{n\eta}\right).$$

We will also need the following Monte Carlo sampling lemma (see Lemma 6.1 in [51]), a consequence of Chebyshev's inequality.

LEMMA 32. *Let (X, μ) be a probability space and F a square integrable function from (X, μ) to the real line. For m independent x_i, distributed according to μ, define the empirical average:*

$$S_m = \frac{1}{m} \sum_{i=1}^{m} F(x_i).$$

Then for any $\delta > 0$ the following estimate holds with probability at least $1 - \delta$:

(67) $$\left|S_m - \int_X F \, d\mu\right| \leq \frac{1}{\sqrt{\delta m}}\left(\int_X \left(F - \int_X F \, d\mu\right)^2 d\mu\right)^{1/2}.$$

We will require the following technical lemma, which will take the place of Propositions 29 and 31 in [52].

LEMMA 33. *Let $N_I^{(\beta)} = \text{Card}\{i, \lambda_i^{(\beta)} \in I\}$ be the counting function of the number of eigenvalues in an interval I of $W^{(\beta)}(z)$, let $R_0(\zeta)$ denote the matrix $(W^{(\beta)}(z) - \zeta)^{-1}$ for any $\zeta \in \mathbb{C}$, and suppose that $\tau_0 \leq |z| \leq 1 - \tau_0$ for some fixed $\tau_0 > 0$. Then we have, uniformly and with overwhelming probability, the following bounds:*

(68) $$N_I \leq n^{o(1)}(1 + Mn|I|)$$

for all intervals I (where $|I|$ here denotes the length of the interval I) and

(69) $$\left|R_0(\sqrt{-1}\eta)_{i,j}\right| \leq n^{o(1)}\left(1 + \frac{1}{Mn\eta}\right)$$

for all $\eta > 1/n$ and $1 \leq i, j \leq 2Mn$. In addition, for any sufficiently small constant c_0, there exists an event which holds with probability at least $1 - O(n^{-\Omega(c_0)+o(1)})$ such that conditioned on this event,

(70) $$\sup_{\eta>0} \|R_0(\sqrt{-1}\eta)\|_{(\infty,1)} \leq O(n^{O(c_0)})$$

with overwhelming probability.

PROOF. Both (68) and (69) are immediate from the proof of the local M-fold circular law [32]. First, (68) follows from the fact that the number of classical eigenvalue locations of $W^{(\beta)}(z)$ in an interval is proportional to the length of the interval, and from the eigenvalue rigidity argument which was used to obtain (26) from Theorem 5 in [32] (see also Lemma 5.1 in [14]). Second, (69) follows immediately from (ii) in the proof of Lemma 17 in [32].

Equation (70) follows from the proof of Lemma 46 in [52], where instead of applying Proposition 31 from [52], one applies (69). We omit the details. □

5.3. *Proof of Theorem* 29. In this section, we will prove Theorem 29 by way of the following, somewhat more general result.

THEOREM 34. *Let $G : \mathbb{R} \to \mathbb{C}$ be any smooth function with five bounded derivatives, and let $f : \mathbb{C} \to \mathbb{R}$ be a function with two continuous derivatives supported in the spectral bulk $\tau_0 < |z| < 1 - \tau_0$ for some fixed $\tau_0 > 0$. Let $N_n^{(\beta)}[f]$ (for $\beta = 1, 2$) denote the linear statistics of the linearization matrices corresponding to two products of M independent n by n i.i.d. random matrices: $n^{-M/2} X^{(1,1)} \cdots X^{(1,M)}$ and $n^{-M/2} X^{(2,1)} \cdots X^{(2,M)}$. Assume furthermore that the entry distributions of $X^{(1,i)}$ match the entry distributions of $X^{(2,i)}$ to four moments for $1 \le i \le M$. Then there exist constants $C, A > 0$ such that*

$$(71) \qquad |\mathbf{E}G(N_n^{(1)}[f]) - \mathbf{E}G(N_n^{(2)}[f])| \le Cn^{-A}.$$

Notice that Theorem 29 follows as a simple corollary by the Fourier inversion formula (as we can take G such that $\mathbf{E}[G(N_n^{(\beta)}[f])]$ is the characteristic function of $N_n^{(\beta)}[f]$; see [52] or [27]). The following proof of Theorem 34 is a combination of the proof of a similar result in [52] with a Monte Carlo sampling argument.

PROOF OF THEOREM 34. The proof is divided into three steps: the first is a preprocessing step which reformulates the statement we wish to prove into a statement about finite sums. The second step reduces the problem into a statement about Stieltjes transforms and the third uses resolvent swapping and a Taylor expansion to conclude the argument.

Step 1: Reformulating the problem. We will first need a variance bound which follows from the proof of the local M-fold circular law [32]. By (2.9) in [32], we have that, for any $D > 0$ and any $\epsilon > 0$, with probability at least $1 - O_{D,\epsilon}(n^{-D})$, the following estimate holds:

$$(72) \qquad \int_{\mathbb{C}} |\Delta f(z)|^2 \left| \sum_{j=1}^{2Mn} \log |\lambda_j^{(\beta)}(z)| - \log |\gamma_j(z)| \right|^2 d^2 z \le O_{D,\epsilon}(n^\epsilon).$$

Here, $\gamma_j(z)$ represent the classical locations of the eigenvalues of $W^\beta(z)$, as defined in [32] (the exact definition of $\gamma_j(z)$ and properties thereof will not be essential to our argument, so we do not provide an overview of this material here; the important feature will simply be that these are deterministic quantities).

By the Girko Hermitization trick, to prove the desired result it suffices to establish the following estimate (see the discussion in Section 5.1):

$$\left| \mathbf{E}G\left(\int_{\mathbb{C}} \Delta f(z) \sum_{j=1}^{2Mn} [\log |\lambda_j^{(1)}(z)|] d^2 z \right) - \mathbf{E}G\left(\int_{\mathbb{C}} \Delta f(z) \sum_{j=1}^{2Mn} [\log |\lambda_j^{(2)}(z)|] d^2 z \right) \right|$$

$$\le Cn^{-A}.$$

Fix some choice of positive constant $k_0 > 0$ to be determined later, and for $K = \lceil n^{k_0} \rceil$, let $z_1, \ldots, z_K$ denote independent random elements selected uniformly at random from the support of f, independent of the product matrices. Let $L > 0$ be the Lebesgue measure of the

support of f; since f is supported in the spectral bulk, it follows that $L = O(1)$. Define the two stochastic Riemann sums $S_K^{(1)}$ and $S_K^{(2)}$:

$$S_K^{(\beta)} = \frac{L}{K} \sum_{i=1}^{K} \left(\sum_{j=1}^{2Mn} \Delta f(z_i) [\log|\lambda_j^{(\beta)}(z_i)| - \log|\gamma_j(z_i)|] \right).$$

By (72) (with some choice of $D > 0$ to be determined) and the Monte Carlo sampling lemma (Lemma 32),

$$\mathbf{P}\left\{ \left| \left(\int_{\mathbb{C}} \Delta f(z) \sum_{j=1}^{2Mn} [\log|\lambda_j^{(\beta)}(z)| - \log|\gamma_j(z)|] d^2z \right) - S_K^{(\beta)} \right|^2 \le O\left(\frac{n^\epsilon}{\delta K} \right) \right\}$$

$$\ge 1 - O(n^{-D}) - \delta.$$

Choosing $\epsilon = k_0/4$ and $\delta = n^{-k_0/4}$ and D sufficiently large, this becomes

$$\mathbf{P}\left\{ \left| \left(\int_{\mathbb{C}} \Delta f(z) \sum_{j=1}^{2Mn} [\log|\lambda_j^{(\beta)}(z)| - \log|\gamma_j(z)|] d^2z \right) - S_K^{(\beta)} \right|^2 \le O\left(\frac{1}{n^{k_0/8}} \right) \right\}$$

$$\ge 1 - O(n^{-k_0/8}).$$

Since we are dealing with the expectations of bounded functions, this estimate will suffice for our purposes. Indeed, it follows that we may write

$$\mathbf{E}G\left(\int_{\mathbb{C}} \Delta f(z) \sum_{j=1}^{Mn} [\log|\lambda_j^{(1)}(z)| - \log|\gamma_j(z)|] d^2z \right)$$

$$- \mathbf{E}G\left(\int_{\mathbb{C}} \Delta f(z) \sum_{j=1}^{Mn} [\log|\lambda_j^{(2)}(z)| - \log|\gamma_j(z)|] d^2z \right)$$

$$= \mathbf{E}G(S_K^{(1)}) - \mathbf{E}G(S_K^{(2)}) + O_G\left(\frac{1}{n^{k_0/8}} \right).$$

Step 2: Additional reductions. We have now replaced the integral which resulted from the Girko Hermitization trick with a sum of K terms; this more or less reduces the problem to the one solved in [52] (as we may deal with each summand separately and just add the resulting errors), and the remainder of the proof just follows the argument made in [52]. We may condition on the precise choice of points $z_1, \ldots, z_K$, and do so now.

Expanding the logarithm by way of the fundamental theorem of calculus, one obtains (where we use $s_\beta(z, \sqrt{-1}\eta)$ to denote quantity $\frac{1}{2Mn} \operatorname{Tr}[W^{(\beta)}(z) - \sqrt{-1}\eta]^{-1})$ for any $z \in \mathbb{C}$ with $|z| \le n^2$ (say)

$$\log|\det(W^\beta(z))| = \log|\det(W^\beta(z) - in^{100})| - 2Mn \operatorname{Im} \int_0^{n^{100}} s_\beta(z, \sqrt{-1}\eta) d\eta$$

$$= 200Mn \log(n) + O(n^{-10}) - 2Mn \operatorname{Im} \int_0^{n^{100}} s_\beta(z, \sqrt{-1}\eta) d\eta.$$

Since G has a bounded first derivative and may be translated, in order to prove Theorem 34 it is sufficient to show that

$$\left| \mathbf{E}G\left(\frac{nL}{K} \sum_{i=1}^{K} \operatorname{Im} \int_0^{n^{100}} s_1(z_i, \sqrt{-1}\eta) d\eta \right) \right.$$

$$\left. - \mathbf{E}G\left(\frac{nL}{K} \sum_{i=1}^{K} \operatorname{Im} \int_0^{n^{100}} s_2(z_i, \sqrt{-1}\eta) d\eta \right) \right| \le Cn^{-A}.$$

Step 3: Resolvent swapping. We would like to simplify this last expression via the resolvent swapping lemma, Lemma 31; however, we must first make sure that the matrices $W^{(1)}(z_i)$ and $W^{(2)}(z_i)$ satisfy the assumptions of the lemma for each z_i, which is the one part of the proof where we will require control over the smallest singular value of linearization matrices. The argument here is the exact clone of the same argument in [52], and we do not reproduce it here. Indeed, by the arguments given in Section 8 of [52] (with our Theorem 7 in the place of their Proposition 27 and our Lemma 33 in place of their Propositions 29 and 31) we can conclude that, with probability $1 - O(n^{-\Theta_0 c_0})$ the matrices $W^{(1)}(z_i)$ and $W^{(2)}(z_i)$, for any fixed i, are such that resolvent swapping lemma applies. Here, the constant $c_0 > 0$ is sufficiently small and $\Theta_0 > 0$ is absolute. Since $K = O(n^{k_0})$, by taking k_0 small enough and using the union bound we may safely assume that this condition is satisfied for each z_i, and thus every summand in $S_K^{(\beta)}$.

We may now safely swap entries. We demonstrate this process by first swapping the very first entry in the very first pair of factor matrices, $X_{1,1}^{(1,1)}$ and $X_{1,1}^{(2,1)}$. This will require some additional notation, which we develop presently. Let $s_{(1,1)}(z_i, \sqrt{-1}\eta)$ denote the Stieltjes transform of $W^{(1)}(z_i)$, and let $s_{(2,1)}(z_i, \sqrt{-1}\eta)$ denote the Stieltjes transform of the matrix formed by taking $W^{(1)}(z_i)$ and replacing the distribution of the entry $X_{1,1}^{(1,1)}$ (which we will denote $\xi_{1,1}^{(1,1)}$) with the distribution of the entry $X_{1,1}^{(2,1)}$ (which we will denote $\xi_{1,1}^{(2,1)}$). Also let $s'_{(1,1)}(z_i, \sqrt{-1}\eta)$ denote the Stieltjes transform of the matrix formed by taking $W^{(1)}(z_i)$ and replacing the distribution of the entry $X_{1,1}^{(1,1)}$ with the distribution of the random variable which is identically equal to zero.

An application of Lemma 31 (with $W^{(\beta)}(z_i)$ in the role of the Hermitian matrix H) produces the expansion (for $\beta = 1, 2$):

$$s_{(\beta,1)}(z_i, \sqrt{-1}\eta) = s'_{(1,1)}(z_i, \sqrt{-1}\eta) + \sum_{j=1}^{4} (\xi_{1,1}^{(\beta,1)})^j n^{-j/2} c_j(\eta)$$

$$+ O\left(n^{-5/2+O(c_0)} \min\left(1, \frac{1}{n\eta}\right)\right).$$

Define the constants $\tilde{c}_j$ as follows:

$$\tilde{c}_j = n \operatorname{Im} \int_0^{n^{100}} c_j(\eta) \, d\eta.$$

By Lemma 31 and (70), the coefficients $\tilde{c}_j$ satisfy

$$|\tilde{c}_j| \leq O(n^{O(c_0)}). \tag{73}$$

By a trivial integration, we have

$$n \int_0^{n^{100}} \min\left(1, \frac{1}{n\eta}\right) d\eta \leq O(\log(n)) \tag{74}$$

and, therefore, we may write the expansion

$$\frac{n}{K} \sum_{i=1}^{K} \operatorname{Im} \int_0^{n^{100}} s_{(\beta,1)}(z_i, \sqrt{-1}\eta) \, d\eta$$

$$= \frac{n}{K} \sum_{i=1}^{K} \operatorname{Im} \int_0^{n^{100}} s'_{(1,1)}(z_i, \sqrt{-1}\eta) \, d\eta + \sum_{j=1}^{4} ((\xi_{1,1}^{(\beta,1)})^j n^{-j/2} \tilde{c}_j)$$

$$+ O(n^{-5/2+O(c_0)}).$$

Computing the fourth-order Taylor expansion of the function G, one obtains

$$\mathbf{E}G\left(\frac{nL}{K}\sum_{i=1}^{K}\operatorname{Im}\int_0^{n^{100}} s_{(\beta,1)}(z_i, \sqrt{-1}\eta)\, d\eta\right)$$

$$= \mathbf{E}G\left(\frac{nL}{K}\sum_{i=1}^{K}\operatorname{Im}\int_0^{n^{100}} s'_{(1,1)}(z_i, \sqrt{-1}\eta)\, d\eta\right)$$

$$+ \mathbf{E}\sum_{k=1}^{4}\frac{L^k}{k!}G^{(k)}\left(\frac{nL}{K}\sum_{i=1}^{K}\operatorname{Im}\int_0^{n^{100}} s'_{(1,1)}(z_i, \sqrt{-1}\eta)\, d\eta\right)$$

$$\times\left(\sum_{j=1}^{4}(\xi_{11}^{(\beta,1)})^j n^{-j/2}\tilde{c}_j\right)^k$$

$$+ O_G(n^{-5/2+O(c_0)}).$$

Using the moment matching condition (i.e., the fact that the random variables $\xi_{11}^{(2,1)}$ and $\xi_{11}^{(1,1)}$ match to four moments) and bounding the remaining terms using (73),

$$\left|\mathbf{E}G\left(\frac{nL}{K}\sum_{i=1}^{K}\operatorname{Im}\int_0^{n^{100}} s_{(1,1)}(z_i, \sqrt{-1}\eta)\, d\eta\right)\right.$$

$$\left. - \mathbf{E}G\left(\frac{nL}{K}\sum_{i=1}^{K}\operatorname{Im}\int_0^{n^{100}} s_{(2,1)}(z_i, \sqrt{-1}\eta)\, d\eta\right)\right|$$

$$\leq O(n^{-5/2+O(c_0)}).$$

Repeating this process for all Mn^2 entries, summing and applying the triangle inequality concludes the proof. $\square$

6. Proofs of Theorems 5 and 6. In this section, we will use Theorem 7 to prove Theorem 5, which establishes local universality for the correlation functions. The proof of Theorem 6 is virtually identical, except one swaps the call to Theorem 2.1 in [51] (which deals with polynomials with complex coefficients) with a call to Theorem 3.1 in [51] (which deals with polynomials with real coefficients), and is therefore omitted.

As in the statement of Theorem 5, for $\beta \in \{1, 2\}$ we let Z_n^β denote the product of M independent n by n i.i.d. matrices, with $p_\beta^{(k)}$ denoting the associated k-point correlation function for the Mth root eigenvalue process, and we assume the factor matrices of Z_n^1 and Z_n^2 match to four moments. Let

$$f_n^\beta(z) = \det(z^M I - Z_n^\beta)$$

denote the polynomial whose roots form the Mth root eigenvalue process associated with Z_n^β. Additionally, we let $z_1, \ldots, z_k$ be complex numbers (which are allowed to depend on n) located in the spectral bulk $\tau_0 \leq n^{-1/2}|z_i| \leq 1 - \tau_0$, and let $G : \mathbb{C}^k \to \mathbb{C}$ denote a smooth function supported on the polydisc $B(0, r_0)^k$, where r_0 is a small constant which is allowed to depend on τ_0.

The method of proof consists mainly of an application of Theorem 2.1 in [51], the statement of which we reproduce as Theorem 35 below, a wide ranging result concerning the universality of zeros of random polynomials. We will use $N_{B(z,\rho)}(f)$ to denote the number of zeros of f in $B(z, \rho)$, the disk centered at z with radius ρ.

THEOREM 35 (Tao–Vu). *Let $C_1, r_0 \geq 1$, $1 \geq c_0 \geq 0$ be real constants and let $a_0, k \geq 1$ be integer constants. Set $A = 100ka_0/c_0$. Let f_n^1 and f_n^2 be random polynomials of degree at most n, and let $z_1, \ldots, z_k$ be k points in the complex plane (which are allowed to depend on n). Assume that three conditions holds:*

1. (*Nondegeneracy*) *The probability that either polynomial is identically zero is at most $C_1 n^{-A}$.*

2. (*Nonclustering*) *For $r \geq 1$, one has $N_{B(z_i, r)}(f_n^1) \leq C_1 n^{1/A} r^2$ with probability at least $1 - C_1 n^{-A}$, and similarly for f_n^2.*

3. (*Comparability of log-magnitudes*) *Given any $1 \leq k' \leq n^{c_0}$ and complex number $z_1', \ldots, z_{k'}' \in \bigcup_{i=1}^k B(z_i, 20r_0)$, and any smooth function $F : \mathbb{C}^{k'} \to \mathbb{C}$ with the derivative bound $|\nabla^a F(w)| \leq n^{c_0}$ for all $0 \leq a \leq a_0$ and $w \in \mathbb{C}^{k'}$, one has*

$$|\mathbf{E}F(\log(|f_n^1(z_1')|), \ldots, \log(|f_n^1(z_{k'}')|)) - \mathbf{E}F(\log(|f_n^2(z_1')|), \ldots, \log(|f_n^2(z_{k'}')|))|$$
$$\leq C_1 n^{-c_0},$$

with the convention that F vanishes when one or more of its arguments is undefined.

Let $G : \mathbb{C}^k \to \mathbb{C}$ be a smooth function supported on the polydisc $B(0, r_0)^k$ such that, for some $C_2 > 0$, G obeys the derivative bound $|\nabla^a G(w)| \leq C_2$ for all $0 \leq a \leq a_0 + 2k + 1$ and all $w \in \mathbb{C}^k$. Then (letting $p_{\beta,n}^{(k)}$ be the k-point correlation function of the zeros of the random polynomial f_n^β),

$$\left| \int_{\mathbb{C}^k} G(w_1, \ldots, w_k) p_{1,n}^{(k)}(z_1 + w_1, \ldots, z_k + w_k) \, d^2 w_1 \cdots d^2 w_k \right.$$

$$\left. - \int_{\mathbb{C}^k} G(w_1, \ldots, w_k) p_{2,n}^{(k)}(z_1 + w_1, \ldots, z_k + w_k) \, d^2 w_1 \cdots d^2 w_k \right|$$

$$\leq O(n^{-c_0/4}),$$

where the implied constant depends only on C_1, r_0, c_0, k, a_0 and linearly on C_2.

REMARK 36. Theorem 35 is designed to handle cases where the mean spacing between zeros is on the order of a constant. This is consistent with the Mth root eigenvalue process, where the mean spacing between points in the bulk is comparable to 1. To handle cases where the mean spacing is not on the order of a constant, one would need to generalize Theorem 35; see Remark 2.5 in [51] for further details. If one wishes to prove a version of Theorem 5 using a different scaling convention (rather than the Mth root eigenvalue process), one would need to utilize this more general version of Theorem 35.

The proof of Theorem 5 will boil down to making sure that the three conditions in this preceding theorem hold. Let us show instead that the analogous result holds for the characteristic polynomials of the linearization matrices, $Y^1(0)$ and $Y^2(0)$ defined by

$$Y^\beta(z) = \begin{pmatrix} -zI & X^{(\beta,1)} & 0 & \cdots & 0 \\ 0 & -zI & X^{(\beta,2)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & -zI & X^{(\beta,M-1)} \\ X^{(\beta,M)} & 0 & \cdots & 0 & -zI \end{pmatrix}.$$

Indeed, since we have the identity $|f_n^\beta(z)| = |\det(Y_n^\beta(z))|$ (which follows from induction and the Schur determinant identity), we may assume that f_n^1 and f_n^2 are the characteristic

polynomials of $Y^1(0)$ and $Y^2(0)$, respectively. Choose $r_0 > 0$ so small that $B(z_i, 20r_0)$ is still in the spectral bulk $\tau_0' \le n^{-1/2}|z| \le 1 - \tau_0'$ for some positive τ_0' smaller than τ_0, and let c_0 denote an arbitrarily small positive constant. We now verify the three conditions of Theorem 35.

The first condition, the nondegeneracy condition, is immediate since these are both characteristic polynomials of matrices. Second, we need the nonclustering property, ensuring that $N_{B(z_i, \rho)}(f_n^\beta) \le C_1 n^{1/A} \rho^2$ for $\rho \ge 1$ with probability at least $1 - O(n^{-A})$, but this is just a consequence of Nemish's local circular law for the linearization matrices $Y^\beta(0)$, which follows from the results in [32] if we choose the constant C_1 sufficiently large (for instance, by approximating the indicator function of the relevant disc by smooth functions). We therefore focus on establishing the third condition: comparability of log-magnitudes. We will demonstrate this using some of the same arguments we used in establishing Theorem 3, which were originally introduced in [52]. Because the arguments here are similar to those in [52], we mostly provide a sketch of the details and explain those portions which differ from [52].

We need to show that given any $k' \le n^{c_0}$ and any collection of complex numbers $z_1', \ldots, z_{k'}'$ in the spectral bulk, and for any function F satisfying the estimate $|\Delta^a F| \le n^{c_0}$ for $0 \le a \le 5$, we have

$$\left| \mathbf{E} F\left(\log|f_n^1(z_1')|, \ldots, \log|f_n^1(z_{k'}')|\right) - \mathbf{E} F\left(\log|f_n^2(z_1')|, \ldots, \log|f_n^2(z_{k'}')|\right) \right|$$
$$\le C_1 n^{-c_0}.$$

We will also first assume that $k' = 1$ to keep the presentation and the notations simple, but the argument generalizes easily. We will deal with the linearization matrices using the same strategy employed during the proof of Theorem 34, albeit without the Monte Carlo sampling step. Define the matrices $W^{(\beta)}(z_1)$, for $\beta \in \{1, 2\}$, as in (60). Let $s_\beta(z_1', \sqrt{-1}\eta)$ denote the Stieltjes transform of $W^{(\beta)}(z_1')$, which is defined as $\frac{1}{2Mn} \operatorname{tr}(W^{(\beta)}(z_1') - \sqrt{-1}\eta)^{-1}$. We have (by the argument presented in Step 2 of the proof of Theorem 34)

$$\left| \mathbf{E} F\left(\log|\det Y^1(z_1')|\right) - \mathbf{E} F\left(\log|\det Y_n^2(z_1')|\right) \right|$$
$$= \left| \mathbf{E} F\left(\frac{1}{2}\log|\det W^1(z_1')|\right) - \mathbf{E} F\left(\frac{1}{2}\log|\det W_n^2(z_1')|\right) \right|$$
$$= \left| \mathbf{E}\tilde{F}\left(Mn \operatorname{Im} \int_0^{n^{100}} s_1(z_1', \sqrt{-1}\eta)\, d\eta\right) \right.$$
$$\left. - \mathbf{E}\tilde{F}\left(Mn \operatorname{Im} \int_0^{n^{100}} s_2(z_1', \sqrt{-1}\eta)\, d\eta\right) \right| + O(n^{-10})$$

by taking c_0 sufficiently small. Here, $\tilde{F}$ is a translation of F (exactly as was done in Step 2 of the proof of Theorem 34). Applying the same argument as in Step 3 of the proof of Theorem 34 (and also as in [52]), and in particular using Theorem 7, we see that Lemma 31, the resolvent swapping lemma, applies, and that Mn^2 separate applications of the the resolvent swapping lemma (combined with a Taylor expansion of $\tilde{F}$, again exactly as in Step 3 of the proof of Theorem 34), provides

$$\left| \mathbf{E} F\left(\log|\det Y^1(z_1')|\right) - \mathbf{E} F\left(\log|\det Y_n^2(z_1')|\right) \right| \le C_F Mn^2 n^{-5/2 + O(c_0)}$$
$$\le O(n^{-c_0})$$

by taking c_0 sufficiently small, which establishes all three conditions from Theorem 35. If $k' > 1$, we simply apply this argument k' times (one for each argument of F), and use the upper bound $k' \le O(n^{c_0})$ (by taking c_0 smaller if necessary) to show that the error is still sufficiently small. By Theorem 35, the result is then established.

SUPPLEMENTARY MATERIAL

Supplement to "Random matrix products: Universality and least singular values" (DOI: 10.1214/19-AOP1396SUPP; .pdf). The supplemental material contains the proofs of Theorems 4 and 26. The proofs are based on the approach developed by Rider and Virág [38], which comes down to using rotary flow combinatorial machinery to control the cumulants of the linear statistics.

REFERENCES

[1] ADHIKARI, K., KISHORE REDDY, N., RAM REDDY, T. and SAHA, K. (2016). Determinantal point processes in the plane from products of random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 16–46. MR3449292 https://doi.org/10.1214/14-AIHP632

[2] AKEMANN, G. and BURDA, Z. (2012). Universal microscopic correlation functions for products of independent Ginibre matrices. *J. Phys. A* **45** 465201, 18. MR2993423 https://doi.org/10.1088/1751-8113/45/46/465201

[3] AKEMANN, G., BURDA, Z., KIEBURG, M. and NAGAO, T. (2014). Universal microscopic correlation functions for products of truncated unitary matrices. *J. Phys. A* **47** 255202, 26. MR3224113 https://doi.org/10.1088/1751-8113/47/25/255202

[4] AKEMANN, G. and IPSEN, J. R. (2015). Recent exact and asymptotic results for products of independent random matrices. *Acta Phys. Polon. B* **46** 1747–1784. MR3403839 https://doi.org/10.5506/APhysPolB.46.1747

[5] ALEXEEV, N., GÖTZE, F. and TIKHOMIROV, A. (2010). Asymptotic distribution of singular values of powers of random matrices. *Lith. Math. J.* **50** 121–132. MR2653641 https://doi.org/10.1007/s10986-010-9074-4

[6] ALEXEEV, N. V., GÖTZE, F. and TIKHOMIROV, A. N. (2010). On the singular spectrum of powers and products of random matrices. *Dokl. Math.* **82** 505–507. https://doi.org/10.1134/S1064562410040010 https://doi.org/10.1134/S1064562410040010

[7] ALT, J., ERDŐS, L. and KRÜGER, T. (2018). Local inhomogeneous circular law. *Ann. Appl. Probab.* **28** 148–203. MR3770875

[8] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

[9] BASAK, A. and RUDELSON, M. (2019). The circular law for sparse non-Hermitian matrices. *Ann. Probab.* **47** 2359–2416. MR3980923

[10] BHATIA, R. (1997). *Matrix Analysis*. *Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 https://doi.org/10.1007/978-1-4612-0653-8

[11] BORDENAVE, C. (2011). On the spectrum of sum and product of non-Hermitian random matrices. *Electron. Commun. Probab.* **16** 104–113. MR2772389 https://doi.org/10.1214/ECP.v16-1606

[12] BORDENAVE, C. and CHAFAÏ, D. (2012). Around the circular law. *Probab. Surv.* **9** 1–89. MR2908617 https://doi.org/10.1214/11-PS183

[13] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. With a foreword by M. Ledoux. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[14] BOURGADE, P., YAU, H.-T. and YIN, J. (2014). Local circular law for random matrices. *Probab. Theory Related Fields* **159** 545–595. MR3230002 https://doi.org/10.1007/s00440-013-0514-z

[15] BOURGADE, P., YAU, H.-T. and YIN, J. (2014). The local circular law II: The edge case. *Probab. Theory Related Fields* **159** 619–660. MR3230004 https://doi.org/10.1007/s00440-013-0516-x

[16] BOURGAIN, J., VU, V. H. and WOOD, P. M. (2010). On the singularity probability of discrete random matrices. *J. Funct. Anal.* **258** 559–603. MR2557947 https://doi.org/10.1016/j.jfa.2009.04.016

[17] BURDA, Z., JANIK, R. A. and WACLAW, B. (2010). Spectrum of the product of independent random Gaussian matrices. *Phys. Rev. E* (3) **81** 041132, 12. MR2736204 https://doi.org/10.1103/PhysRevE.81.041132

[18] COOK, N. A. (2018). Lower bounds for the smallest singular value of structured random matrices. *Ann. Probab.* **46** 3442–3500. MR3857860

[19] EDELMAN, A. (1988). Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** 543–560. MR0964668 https://doi.org/10.1137/0609045

[20] FORRESTER, P. J. (2014). Probability of all eigenvalues real for products of standard Gaussian matrices. *J. Phys. A* **47** 065202, 12. MR3159513 https://doi.org/10.1088/1751-8113/47/6/065202

[21] FORRESTER, P. J. and IPSEN, J. R. (2016). Real eigenvalue statistics for products of asymmetric real Gaussian matrices. *Linear Algebra Appl.* **510** 259–290. MR3551633 https://doi.org/10.1016/j.laa.2016.08.015

[22] GINIBRE, J. (1965). Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.* **6** 440–449. MR0173726 https://doi.org/10.1063/1.1704292

[23] GÖTZE, F., NAUMOV, A. and TIKHOMIROV, A. (2017). On local laws for non-Hermitian random matrices and their products. Available at arXiv:1708.06950.

[24] GÖTZE, F. and TIKHOMIROV, A. (2010). On the asymptotic spectrum of products of independent random matrices. Available at arXiv:1012.2710.

[25] IPSEN, J. R. (2015). Products of independent Gaussian random matrices. Available at arXiv:1510.06128.

[26] IPSEN, J. R. and KIEBURG, M. (2014). Weak commutation relations and eigenvalue statistics for products of rectangular random matrices. *Phys. Rev. E* **89** 032106. https://doi.org/10.1103/PhysRevE.89.032106

[27] KOPEL, P. (2015). Linear statistics of non-Hermitian matrices matching the real or complex Ginibre ensemble to four moments. Available at arXiv:1510.02987.

[28] KOPEL, P., O'ROURKE, S. and VU, V. (2020). Supplement to "Random matrix products: Universality and least singular values." https://doi.org/10.1214/19-AOP1396SUPP.

[29] LITVAK, A. E., LYTOVA, A., TIKHOMIROV, K., TOMCZAK-JAEGERMANN, N. and YOUSSEF, P. (2017). Adjacency matrices of random digraphs: Singularity and anti-concentration. *J. Math. Anal. Appl.* **445** 1447–1491. MR3545253 https://doi.org/10.1016/j.jmaa.2016.08.020

[30] LITVAK, A. E., PAJOR, A., RUDELSON, M. and TOMCZAK-JAEGERMANN, N. (2005). Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* **195** 491–523. MR2146352 https://doi.org/10.1016/j.aim.2004.08.004

[31] LYTOVA, A. and PASTUR, L. (2009). Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37** 1778–1840. MR2561434 https://doi.org/10.1214/09-AOP452

[32] NEMISH, Y. (2017). Local law for the product of independent non-Hermitian random matrices with independent entries. *Electron. J. Probab.* **22** Paper No. 22, 35. MR3622892 https://doi.org/10.1214/17-EJP38

[33] NEMISH, Y. (2018). No outliers in the spectrum of the product of independent non-Hermitian random matrices with independent entries. *J. Theoret. Probab.* **31** 402–444. MR3769819 https://doi.org/10.1007/s10959-016-0708-2

[34] O'ROURKE, S. and RENFREW, D. (2016). Central limit theorem for linear eigenvalue statistics of elliptic random matrices. *J. Theoret. Probab.* **29** 1121–1191. MR3540493 https://doi.org/10.1007/s10959-015-0609-9

[35] O'ROURKE, S., RENFREW, D., SOSHNIKOV, A. and VU, V. (2015). Products of independent elliptic random matrices. *J. Stat. Phys.* **160** 89–119. MR3357969 https://doi.org/10.1007/s10955-015-1246-5

[36] O'ROURKE, S. and SOSHNIKOV, A. (2011). Products of independent non-Hermitian random matrices. *Electron. J. Probab.* **16** 2219–2245. MR2861673 https://doi.org/10.1214/EJP.v16-954

[37] RIDER, B. and SILVERSTEIN, J. W. (2006). Gaussian fluctuations for non-Hermitian random matrix ensembles. *Ann. Probab.* **34** 2118–2143. MR2294978 https://doi.org/10.1214/009117906000000403

[38] RIDER, B. and VIRÁG, B. (2007). The noise in the circular law and the Gaussian free field. *Int. Math. Res. Not. IMRN* **2** Art. ID rnm006, 33. MR2361453 https://doi.org/10.1093/imrn/rnm006

[39] RUDELSON, M. (2008). Invertibility of random matrices: Norm of the inverse. *Ann. of Math.* (2) **168** 575–600. MR2434885 https://doi.org/10.4007/annals.2008.168.575

[40] RUDELSON, M. and VERSHYNIN, R. (2008). The least singular value of a random square matrix is $O(n^{-1/2})$. *C. R. Math. Acad. Sci. Paris* **346** 893–896. MR2441928 https://doi.org/10.1016/j.crma.2008.07.009

[41] RUDELSON, M. and VERSHYNIN, R. (2008). The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600–633. MR2407948 https://doi.org/10.1016/j.aim.2008.01.010

[42] SANKAR, A., SPIELMAN, D. A. and TENG, S.-H. (2006). Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.* **28** 446–476. MR2255338 https://doi.org/10.1137/S0895479803436202

[43] SHCHERBINA, M. (2011). Central limit theorem for linear eigenvalue statistics of the Wigner and sample covariance random matrices. *Zh. Mat. Fiz. Anal. Geom.* **7** 176–192, 197, 199. MR2829615

[44] SINAĬ, YA. G. and SOSHNIKOV, A. B. (1998). A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices. *Funktsional. Anal. i Prilozhen.* **32** 56–79, 96. MR1647832 https://doi.org/10.1007/BF02482597

[45] SOSOE, P. and WONG, P. (2013). Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices. *Adv. Math.* **249** 37–87. MR3116567 https://doi.org/10.1016/j.aim.2013.09.004

[46] TAO, T. and VU, V. (2007). On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.* **20** 603–628. MR2291914 https://doi.org/10.1090/S0894-0347-07-00555-3

[47] TAO, T. and VU, V. (2008). Random matrices: The circular law. *Commun. Contemp. Math.* **10** 261–307. MR2409368 https://doi.org/10.1142/S0219199708002788

[48] TAO, T. and VU, V. (2010). Random matrices: The distribution of the smallest singular values. *Geom. Funct. Anal.* **20** 260–297. MR2647142 https://doi.org/10.1007/s00039-010-0057-8

[49] TAO, T. and VU, V. (2010). Smooth analysis of the condition number and the least singular value. *Math. Comp.* **79** 2333–2352. MR2684367 https://doi.org/10.1090/S0025-5718-2010-02396-8

[50] TAO, T. and VU, V. (2012). A central limit theorem for the determinant of a Wigner matrix. *Adv. Math.* **231** 74–101. MR2935384 https://doi.org/10.1016/j.aim.2012.05.006

[51] TAO, T. and VU, V. (2015). Local universality of zeroes of random polynomials. *Int. Math. Res. Not. IMRN* **13** 5053–5139. MR3439098 https://doi.org/10.1093/imrn/rnu084

[52] TAO, T. and VU, V. (2015). Random matrices: Universality of local spectral statistics of non-Hermitian matrices. *Ann. Probab.* **43** 782–874. MR3306005 https://doi.org/10.1214/13-AOP876

[53] TAO, T. and VU, V. H. (2009). Inverse Littlewood–Offord theorems and the condition number of random discrete matrices. *Ann. of Math.* (2) **169** 595–632. MR2480613 https://doi.org/10.4007/annals.2009.169.595

[54] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

[55] VERSHYNIN, R. (2015). Estimation in high dimensions: A geometric perspective. In *Sampling Theory, a Renaissance*. *Appl. Numer. Harmon. Anal.* 3–66. Birkhäuser/Springer, Cham. MR3467418

[56] YIN, J. (2014). The local circular law III: General case. *Probab. Theory Related Fields* **160** 679–732. MR3278919 https://doi.org/10.1007/s00440-013-0539-3

[57] ŻYCZKOWSKI, K. and SOMMERS, H.-J. (2000). Truncations of random unitary matrices. *J. Phys. A* **33** 2045–2057. MR1748745 https://doi.org/10.1088/0305-4470/33/10/307