Submitted to the Annals of Applied Probability

PRODUCTS OF RANDOM MATRICES: DIMENSION AND GROWTH IN NORM

By Vladislav Kargin[1]

Abstract

Suppose that $X_1, \ldots, X_n, \ldots,$ are independent, identically-distributed, rotationally invariant $N \times N$ matrices. Let $\Pi_n = X_n \ldots X_1$. It is known that $n^{-1} \log \|\Pi_n\|$ converges to a non-random limit. We prove that under certain additional assumptions on matrices X_i the speed of convergence to this limit does not decrease when the size of matrices, N, grows.

1. Introduction. Let X_i be a sequence of independent $N \times N$ random matrices and $\Pi_n = X_n \ldots X_1$. In a celebrated paper [2], Furstenberg and Kesten proved that $n^{-1} \log \|\Pi_n\|$ converges provided that $E \log^+ (\|X_i\|) < \infty$. Later, Oseledec in [7] proved convergence for other singular values of Π_n, and Cohen and Newman in [1] studied the behavior of the limit in the situation when N approaches infinity. This paper investigates the question of how the speed of convergence depends on the dimension of matrices N.

Consider a dynamical system (a gas, an economy, an ecosystem, etc.). Its evolution can be described by a mapping $\psi_i \to X_i (\psi_i)$, where ψ_i is a vector that describes the state of the system at time i. We can often model the mapping as a multiplication by a random matrx X_i. Stability and other long-run properties of the system depend on the growth in the norm of the product $\Pi_n = X_n \ldots X$, which we can measure by calculating the quantity $n^{-1} \log (\|\Pi_n\|)$.

The sub-multiplicativity property of the norm ($\|X_2 X_1\| \leq \|X_2\| \|X_1\|$) ensures that $n^{-1} \log (\|\Pi_n\|)$ converges to $E \log \|X_1 u\|$, where u is an arbi-

[1]Stanford University; kargin@stanford.edu

imsart-aap ver. 2008/08/29 file: swp0000.tex date: January 5, 2009

trary vector. Intuitively, this means that it is not important what was the starting vector of the system. After some time, all products grow at the same rate independently of the initial state.

It is of interest to investigate whether this erasure of memory about the initial state occurs slower in more complex systems, that is, in systems, which are described by matrices of larger size.

Of course, when we compare long-run properties of systems, we should only look at the systems that are comparable in the short run, that is, the system that have comparable one-step behavior. Roughly, the difference between one-step growth of a specially-chosen and a random vector can be measured by the ratio of $\|X_1\|^2$ to $N^{-1}tr\left(X_1^* X_1\right)$, where N is the dimension of the matrix X_i. Indeed, $\|X_1\|^2$ is the square of the maximal possible increase in the length of the state vector, and $N^{-1}tr\left(X_1^* X_1\right)$ is the average of the squared singular values of X_1, hence it can be considered as a measure of the increase in the length of a random state vector.

Hence, if we want systems to be comparable in the short run, then we should restrict this ratio by a constant that does not depend on the dimension of the system. We will call this property uniform boundedness of singular values.

We also want to look at sufficiently symmetric systems, that is, systems without preferential directions. We codify this by requiring that matrices X_i are rotationally invariant, that is, the distribution of matrix elements does not depend on the choice of basis.

The main result of this paper is that under these assumptions the speeed with which the memory of the initial state is erased does not decrease as the dimension of the system grows.

Intuitively, the asymptotic behavior of $n^{-1}\log\|\Pi_n\|$ depends on three factors. First of all, for a fixed vector v,

$$n^{-1}\log\|\Pi_n v\| = n^{-1}\sum_{i=1}^{n}\log\|X_i v_i\|$$

for a certain sequence of vectors v_i and this averaging is likely to concentrate the distribution of $n^{-1}\log\|\Pi_n v\|$. This factor does not depend on the dimension N. On the other hand, we are interested in the convergence of the supremum of $n^{-1}\log\|\Pi_n v\|$ over all $v \in S^N$, and to ensure the convergence of this supremum we have to make sure that variables $n^{-1}\log\|\Pi_n v\|$ are all close to the limit $E\log\|X_1 u\|$ for a sufficiently dense set of vectors v. The number of elements in such a set is likely to grow exponentially in N, and this might make the convergence of $n^{-1}\log\|\Pi_n\|$ slower for large N.

The third factor appears because for every fixed vector v, the norm $\|X_i v\|$ becomes concentrated around some particular value as $N \to \infty$. This factor is likely to speed up the convergence of $n^{-1}\log\|\Pi_n v\|$ and therefore of $n^{-1}\log\|\Pi_n\|$.

We will show in this paper that the third factor dominates and the speed of convergence of $n^{-1}\log\|\Pi_n\|$ is not slowed down by the growth in the dimension N.

Previously, the speed of convergence in the Furstenberg-Kesten theorem was investigated in [8], [5], and [4]. They proved a central limit theorem for $n^{-1/2}\log\|\Pi_n\|$ and studied large deviations of $n^{-1}\log\|\Pi_n\|$ for a large class of random matrices. However, the results in these papers do not provide effectively computable bounds on the rate of convergence in limit theorems, and, as a consequence, do not help us to investigate how the speed of convergence changes as the dimension of matrices grows. One of the contributions of this paper is deriving more explicit bounds on the speed of convergence in limit theorems.

Let us describe the problem in a more formal fashion. Consider independent identically-distributed N-by-N matrices $X_i^{(N)}$. We are interested in the behavior of the norm of the product $\Pi_n = X_n^{(N)} \dots X_1^{(N)}$, and we will make the following assumptions about matrices $X_i^{(N)}$. First of all, we assume that random matrices $X_i^{(N)}$ are rotationally invariant; that is, the distribution of their entries does not depend on the choice of coordinates. Formally, we use the following definition:

Definition 1 *A random matrix X is rotationally invariant if for every integer $k \geq 1$, for every collection of vectors $\{v_i, w_i\}$, $i = 1, ..., k$, and for every orthogonal matrix U, the joint distributions of random vectors $\{\langle w_i, Xv_i \rangle\}_{i=1}^k$ and $\{\langle Uw_i, XUv_i \rangle\}_{i=1}^k$ are the same.*

Assumption A ("rotational invariance") Matrices $X_i^{(N)}$ are rotationally invariant.

We also impose an assumption needed for the validity of the Furstenberg-Kesten theorem.

Assumption B ("Furstenberg-Kesten") For all N, $E \log^+ \left\| X_i^{(N)} \right\|$ exists.

Second, we restrict our study to two important cases. The first one is the case of (real) Gaussian matrices $X_i^{(N)}$, that is, independent random N-by-N matrices with independent entries distributed according to the Gaussian distribution with zero mean and variance σ^2/N, i.e., as $\mathcal{N}\left(0, \sigma^2/N\right)$.

The second case is that of independent rotationally invariant N-by-N matrices $X_i^{(N)}$ that satisfy the following assumptions. Let $s_k^{(i,N)}$ be the eigenvalues of $X_i^{(N)*}X_i^{(N)}$ (i.e., squared singular values of $X_i^{(N)}$), and let

$$\overline{s}^{(i,N)} = \frac{1}{N} \sum_{k-1}^{N} s_k^{(i,N)} = \frac{1}{N} \operatorname{Tr}\left(X_i^{(N)*} X_i^{(N)}\right).$$

(We will sometimes omit superscripts to lighten the notation.)

Assumption C ("uniformly bounded singular values") With probability 1, $\max_k s_k^{(i,N)} \leq b\overline{s}^{(i,N)}$, where the constant b does not depend on N.

In other form, Assumption C says that

$$\left\| X_i^{(N)} \right\|^2 \leq b \frac{1}{N} tr\left(X_i^{(N)*} X_i^{(N)} \right)$$

with probability 1.

Assumption D ("comparability across N") $Var\left[\log \overline{s}^{(i,N)}\right]$ exists and bounded by a constant which does not depend on N.

One example of a matrix family that satisfies these assumptions is Hermitian matrices $X_i^{(N)}$ which are generated in the following way. Sample N independent values from a distribution supported on $[\alpha, \beta]$, where $\beta > \alpha > 0$, and construct a diagonal matrix $D^{(N)}$ by putting these values on the main diagonal. Then take a Haar-distributed random orthogonal matrix $U_i^{(N)}$ and define $X_i^{(N)}$ as $D^{(N)} U_i^{(N)}$. A sequence of these matrices (with independent $U_i^{(N)}$) will satisfy all the assumptions.

The main result is as follows:

Theorem 2 *Let $X_i^{(N)}$ be independent, identically distributed $N \times N$ matrices, which satisfy assumptios A and B and which are either Gaussian with independent entries $\mathcal{N}\left(0, \sigma^2/N\right)$, or satisfy Assumptions C and D . Let $\Pi_n = X_n^{(N)} \ldots X_1^{(N)}$ and let v be an arbitrary unit vector. Then $n^{-1} \log \|\Pi_n\|$ converges in probability to $E \log \left\| X_1^{(N)} v \right\|$ and the convergence is uniform in N. That is, for each $\delta > 0$, there exists an $n_0(\delta)$ such that for all $n \geq n_0$ and all $N \geq 1$,*

$$\Pr\left\{ \left| n^{-1} \log \|\Pi_n\| - E \log \left\| X_1^{(N)} v \right\| \right| \geq \delta \right\} \leq \delta. \tag{1}$$

The assumptions of the theorem are sufficient but not necessary. The assumption that $s_k \leq b\overline{s}$ is used in the proof of Proposition 3 below, where it is used to estimate the probability of large deviations of $\log \left\| X_i^{(N)} v \right\|$ and to show that the rate in the corresponding exponential inequality is proportional to N. It is likely that this assumption can be somewhat relaxed by requiring instead that $\Pr\left\{s_k/\overline{s} > b + u\right\} \leq ce^{-c'Nu}$.

One particular implication of the assumption $s_k \leq b\overline{s}$ is that the bound on singular values does not depend on the dimension of the matrix. In order to understand this assumption better, consider the following example. Let

$$X_i^{(N)} = \sqrt{N} \left|y_i\right\rangle \left\langle x_i\right|,$$

where $\left\langle x_i\right|$ is a Haar-distributed row N-dimensional vector, and $\left|y_i\right\rangle$ is a Haar-distributed column N-dimensional vector. (Vectors $\left\langle x_i\right|$ and $\left|y_i\right\rangle$ are assumed to be independent.) Then the squared singular values of X_i are all zero except one, which equals N. Hence $\overline{s}^{(i,N)} = 1$ and $\log \overline{s}^{(i,N)} = 0$. We can conclude that Assumptions A, B, and D are satisfied, and Assumption C is not satisfied.

Next, consider $\left\|X_i^{(N)} v\right\|^2$, where v is an arbitrary vector. It is easy to see that this random variable is distributed as

$$N\left(u_1\right)^2,$$

where u_1 is the first coordinate of a Haar-distributed vector u. In other words ξ_i is distributed as

$$\frac{Y_1^2}{\left(Y_1^2 + \ldots Y_N^2\right)/N},$$

where Y_i are independent standard Gaussian variables. Using this facts, it is possible to check that

$$\lim_{N\to\infty} E\log\left\|X_1 v\right\|^2 = E\log\left(Y_1\right)^2 \in (-\infty, 0).$$

Next, let us compute $n^{-1}\log\left\|\Pi_n\right\|^2$. Note that

$$\Pi_n = N^{n/2}\left|y_n\right\rangle\left\langle x_n\right|\ldots\left|y_2\right\rangle\left\langle x_2|y_1\right\rangle\left\langle x_1\right|,$$

and

$$\Pi_n^*\Pi_n = N^n\left|x_1\right\rangle\left\langle x_n|y_{n-1}\right\rangle^2\ldots\left\langle x_2|y_1\right\rangle^2\left\langle x_1\right|.$$

Hence,

$$n^{-1}\log\|\Pi_n\|^2 = \frac{\log N}{n} + n^{-1}\sum_{i=1}^{n-1}\log\xi_i,$$

where ξ_i are independent and distributed as $N(u_1)^2$ above. Hence, ξ_i converges in distribution to Y_1^2 as $N\to\infty$. It is clear that

$$n^{-1}\sum_{i=1}^{n-1}\log\xi_i \to E\log\|X_1 v\|^2$$

in probability as $n\to\infty$. Therefore, for large N,

$$n^{-1}\log\|\Pi_n\|^2 - E\log\|X_1 v\|^2 \sim \frac{\log N}{n}.$$

This bias term cannot be made small uniformly in N by an increase in n. This means that the claim of Theorem 2 fails in this case.

Later, in Section 3, we will prove a necessary condition for the uniform convergence by using the basic idea of this example.

In order to understand the role of the rotational invariance assumption, consider the following example.

Let X_i be independent, identically distributed, diagonal matrices. The diagonal elements of a matrix X_i are independent Bernoulli variables that take values a and b. That is, a diagonal element takes the value $b>0$ with probability p and the value $a>0$ with probability $q=1-p$. Assume that $b>a$.

It is easy to see that the norm of $\Pi_n = X_1\dots X_n$ is given by the following expression:

$$\|\Pi_n\| = \max\left\{a^{\alpha_1}b^{\beta_1},\dots,a^{\alpha_N}b^{\beta_N}\right\},$$

where $\alpha_i+\beta_i=n$, and β_i are independent random variables with the binomial distribution $B(p,n)$.

Taking the logarithm and dividing by n, we get:

$$\frac{1}{n}\log\|\Pi_n\| = \log a + \log(b/a)\max\left\{\widetilde{\beta}_1,\dots,\widetilde{\beta}_N\right\},$$

where $\widetilde{\beta}_i = \beta_i/n$. Note that as n grows, each $\widetilde{\beta}_i$ approaches the Gaussian distribution $\mathcal{N}(p, pq/n)$.

If N is fixed, then $\lim n^{-1} \log \|\Pi_n\| = \log a + p \log(b/a)$. However, if N grows simultaneously with n, then the limit of $n^{-1} \log \|\Pi_n\|$ may be non-existent, or may depend on the speed of growth in N relative to the speed of growth in n. Hence, the conclusion of Theorem 2 is invalid in this case.

It is an interesting problem whether the assumption of rotational invariance can be relaxed so that the result in Theorem 2 holds for a larger class of matrices, for example, for matrices with i.i.d. non-Gaussian entries (i.e., Wigner matrices). However, this problem appears to be hard since at this moment very little is known about effective bounds on the rate of convergence in the Furstenberg-Kesten theorem.

Let me now explain two results which will be used as tools in the proof of Theorem 2. The proofs of these results will be given in later sections.

Our main tool is the following proposition.

Proposition 3 *(i) Suppose that all X_i are Gaussian with independent entries $\mathcal{N}(0, \sigma^2/N)$. Then for all sufficiently small t, all $N \geq N_1(t)$ and all $n \geq 1$,*

$$\Pr\left\{\left|\frac{1}{n}\log\|\Pi_n v\| - \log\sigma\right| > t\right\} \leq 2\exp\left(-\frac{1}{8}Nnt^2\right).$$

(ii) Suppose that i.i.d. N-by-N matrices X_i are rotationally invariant and satisfy Assumption C with constant b. Let

$$\overline{s}^{(i,N)} = \frac{1}{N}\sum_{k-1}^{N} s_k^{(i,N)} = \frac{1}{N}\operatorname{Tr}(X_i^* X_i).$$

Then for all $t \in (0, 1/4)$, all $N \geq N_1(t)$ and all $n \geq 1$,

$$\Pr\left\{\left|\frac{1}{n}\log\|\Pi_n v\| - \frac{1}{n}\sum_{i=1}^{n}\log\overline{s}^{(i,N)}\right| > t\right\} \leq 2\exp\left(-\frac{1}{32b^2}Nnt^2\right).$$

In its essence, Proposition 3 is a large deviation result which quantifies the speed of convergence of $n^{-1}\log\|\Pi_n v\|^2$ for a fixed vector v. Its main point is that the rate in this large deviation estimate is proportional to the dimension N. The proof of this proposition will be given in Section 2.

The other tool is as follows. Let a set of points on the unit sphere in $\mathbb{R}^N$ be called an ε-net if the sphere is covered by spherical caps with centers at these points and angular radius ε.

Proposition 4 *Let A be an arbitrary N-by-N matrix. Suppose that the end-points of vectors v_i form an ε-net of the unit sphere in $\mathbb{R}^N$. Then, for all sufficiently small ε*

$$\log \|A\| \leq \max_i \log \|Av_i\| + 2\varepsilon.$$

This proposition allows us to control the matrix norm $\|\Pi_n\|$ by the norms of vectors $\|\Pi_n v_i\|$, where v_i runs through a finite set of values.

Proof: Let v_i be a vector in the net which is closest to a unit vector v. Then

$$\begin{aligned} \|Av\| &\leq \|Av_i\| + \|A(v - v_i)\| \\ &\leq \|Av_i\| + \varepsilon \|A\|. \end{aligned}$$

Taking the supremum over v, we obtain that

$$(1 - \varepsilon) \|A\| \leq \max_i \|Av_i\|.$$

Hence,

$$\log \|A\| \leq \max_i \log \|Av_i\| - \log (1 - \varepsilon),$$

and the claim of the proposition follows. QED.

This proposition is useful in conjunction with the following result about the size of sphere coverings. By Lemma 2.6 on page 7 of [6], for ε smaller than a certain constant, there exists an ε-net with cardinality $M \leq \exp(N \log(3/\varepsilon))$.

Now let us prove Theorem 2 by using Propositions 3 and 4.

Proof of Theorem 2: We focus on the case when Assumptions C and D hold. The proof for the case of Gaussian matrices goes along a similar route and it is simpler.

First of all, note that is enough to prove that (1) holds for all sufficiently large N, i.e., for all $N \geq N_0(\delta)$. Indeed, for each $N \leq N_0$ we can apply results in [3] and find that inequality (1) holds if $n \geq n(\delta, N)$. Hence, inequality (1) holds for all $N \leq N_0$, provided that

$$n \geq n_0(\delta) = \max_{N \leq N_0(\delta)} \{n(\delta, N)\}.$$

We will choose an appropriate $N_0(\delta)$ later.

We are going to prove that for all sufficiently large N and n, (i.e., $N \geq N_2(\delta)$ and all $n \geq n_2(\delta)$), it is true that

$$\Pr\left\{\left|\frac{1}{n}\log\|\Pi_n\|^2 - \frac{1}{n}\sum_{i=1}^{n}\log\overline{s}^{(i,N)}\right| > \frac{\delta}{10}\right\} < \frac{\delta}{10}. \tag{2}$$

Let vectors v_j, $j = 1, \ldots, M$, form an $(\delta/100)$-net on the unit sphere. Then, by using Propositions 4 and 3, the union bound and the estimate on the number of elements in the net we obtain:

$$\begin{aligned}
& \Pr\left\{\left|\frac{1}{n}\log\|\Pi_n\|^2 - \frac{1}{n}\sum_{i=1}^{n}\log\overline{s}^{(i,N)}\right| > \frac{\delta}{10}\right\} \\
\leq\ & \Pr\left\{\max_{v_i}\left|\frac{1}{n}\log\|\Pi_n v_i\|^2 - \frac{1}{n}\sum_{i=1}^{n}\log\overline{s}^{(i,N)}\right| > \frac{\delta}{100}\right\} \\
\leq\ & 2\exp\left\{\left(\log\left(\frac{300}{\delta}\right) - cn\left(\frac{\delta}{100}\right)^2\right)N\right\},
\end{aligned}$$

where c is a certain constant. Clearly, we can choose $n_2(\delta)$ in such a way that for all $n \geq n_2(\delta)$, it is true that

$$\log\left(\frac{300}{\delta}\right) - cn\left(\frac{\delta}{100}\right)^2 < \alpha < 0$$

for some α, and then choose $N_2(\delta)$, such that for all $N > N_2(\delta)$ it is true that

$$2\exp\{\alpha N\} < \frac{\delta}{10}.$$

This choice of $n_2(\delta)$ and $N_2(\delta)$ is sufficient to ensure that (2) holds.

Next, let $d_N = E \log \overline{s}^{(i,N)}$. Since variance of $\log \overline{s}_i^{(N)}$ is bounded above by a finite constant which does not depend on N (Assumption D), therefore we can find such $n_3(\delta)$ that for all $n \geq n_3(\delta)$, it is true that

$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\log \overline{s}_i^{(N)} - d_N\right| > \frac{\delta}{100}\right\} < \frac{\delta}{100}. \tag{3}$$

for all N.

It follows that for all $n \geq n_4(\delta)$ and $N \geq N_2(\delta)$, it is true that

$$\Pr\left\{\left|\frac{1}{n}\log\|\Pi_n\|^2 - d_N\right| > \frac{\delta}{5}\right\} < \frac{\delta}{5}. \tag{4}$$

Note that by the Furstenberg-Kesten theorem,

$$\Pr\left\{\left|\frac{1}{n}\log\|\Pi_n\|^2 - E\log\left\|X_i^{(N)}u\right\|^2\right| > \frac{\delta}{5}\right\} < \frac{\delta}{5} \tag{5}$$

for all $n \geq n_5(\delta, N)$. This implies that for all $N \geq N_2(\delta)$, there exists such n, that both inequalities (4) and (5) hold. This implies that for all such N, and for all $\delta < 1$, the following inequality holds.

$$\left|d_N - E\log\left\|X_i^{(N)}u\right\|^2\right| < \frac{2\delta}{5}. \tag{6}$$

Otherwise, the sum of the events in (4) and (5) would cover all probability space and hence the sum of probabilities in (4) and (5), $2\delta/5$, would have to be greater then 1. This contradicts to the assumption that $\delta < 1$.

Inequalities (2), (3)and (6) imply that

$$\Pr\left\{\left|\frac{1}{n}\log\|\Pi_n\|^2 - E\log\|X_i u\|^2\right| > \delta\right\} < \delta$$

for all $n > n_0(\delta)$ and $N > N_0(\delta)$, where n_0 and N_0 are sufficiently large functions of δ. QED.

It remains to complete the proof by proving Proposition 3. We will do this in the next section.

The rest of the paper consists of Section 2, which is devoted to the proof of Proposition 3, Section 3, which gives a necessary condition for uniform convergence in Furstenberg-Kesten theorem, and Section 4, which concludes.

2. A large deviation bound for the dilation of a fixed vector . Everywhere in this section, we assume that random matrices X_i are independent, identically distributed, and rotationally invariant, and that $\Pi_i = X_i X_{i-1} ... X_1$. Let us consider the following random variables:

$$y_i = \log\left(\frac{\|X_i \Pi_{i-1} v\|}{\|\Pi_{i-1} v\|}\right).$$

It is known (e.g., [1]) that the random variables y_i are independent and identically distributed. Their distribution coincides with the distribution of $\log(\|X_1 v\|)$, where v is an arbitrary unit vector.

2.1. *Gaussian matrices.* In this section we consider an important case when each matrix X_i has independent Gaussian entries distributed according to $\mathcal{N}(0, \sigma^2/N)$. In this case, $\log\|X_1 v\|^2$ is distributed in the same way as the random variable

$$y = \log\left(\frac{\sigma^2}{N}\sum_{k=1}^{N} Y_k^2\right),$$

where Y_k are independent standard Gaussian variables. In order to prove Proposition 3 for this case, it is enough to show that the following result holds.

Proposition 5 *Let y_i be independent copies of the variable*

$$y = \log\left[\frac{1}{N}\sum_{k=1}^{N} Y_k^2\right].$$

If $t \leq 1$, then there exists a function $N_0(t)$ such that for all $N \geq N_0(t)$ and all n, the following inequality holds:

$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} y_i\right| \geq t\right\} \leq 2e^{-t^2 nN/8}.$$

Proof: First of all, let us compute

$$Ee^{yz} = E\left(\frac{1}{N}\sum_{k=1}^{N} Y_k^2\right)^z,$$

where z is a real number. By explicit calculation,

$$E\left(\sum_{i=1}^{N} Y_i^2\right)^z = \frac{2^z \Gamma\left(\frac{N}{2}+z\right)}{\Gamma\left(\frac{N}{2}\right)},$$

where $\Gamma(z)$ is the Gamma function. This formula is valid for $z > -N/2$.

Let $z = \alpha N$, where $\alpha > -1/2$. Then using the Stirling formula for large N, we can write:

(7)

$$E\left(e^{yz}\right) = E\left(\sum_{i=1}^{N} Y_i^2\right)^z \sim \frac{1}{\sqrt{1+2\alpha}} N^{\alpha N} \exp\left\{\left[\left(\frac{1}{2}+\alpha\right)\log(1+2\alpha) - \alpha\right] N\right\}.$$

Note that for all $\alpha \geq 0$, $\left(\frac{1}{2}+\alpha\right)\log(1+2\alpha) - \alpha \leq \alpha^2$, and for all $\alpha > -1/2$, $\left(\frac{1}{2}+\alpha\right)\log(1+2\alpha) - \alpha \leq 2\alpha^2$ with equalities only for $\alpha = 0$.

If $t > 0$, we set $\alpha = t/2$ and $z = (t/2)N$, and use the fact that for all sufficiently large N, the asymptotic term in (7) dominates all other terms. Hence, we obtain the estimate:

$$e^{-tz} E e^{yz} \leq \exp\left(-t^2 N/4\right).$$

If $t \in (-1, 0)$ then we can take $\alpha = t/4$ and $z = (t/4)N$, and we obtain:

$$e^{-tz} E e^{yz} \leq \exp\left(-t^2 N/8\right).$$

By standard arguments we can translate these inequalities into statements about probabilities of large deviations. If $0 \leq t < 1$, then

$$\Pr\left\{\left|\frac{1}{n}\sum_{i=1}^{n} y_i\right| \geq t\right\} \leq 2e^{-t^2 Nn/8}.$$

QED.

2.2. *Matrices with uniformly bounded singular values.* In this section we are going to prove the second part of Proposition 3. Since X_i are i.i.d and rotationally invariant, therefore the distribution of $y_i = \log\left(\|\Pi_i v\|^2 / \|\Pi_{i-1} v\|^2\right)$

coincides with the distribution of $\log\left(\|X_1 v\|^2\right)$ and equals the distribution of the random variable $y = \sum_{k=1}^N s_k u_k^2$. Here u_k are components of the random vector u, which is uniformly distributed on the unit sphere and which is independent of s_k.

Let us start with considering large deviations of $x = \sum_{k=1}^N s_k u_k^2$. Let $\overline{s}^{(N)} =: N^{-1} \sum_{k=1}^N s_k$.

Proposition 6 *Suppose that with probability 1, $|s_k| \leq B$ for all k. Then for all $t > 0$,*

(8)

$$\max\left[\Pr\left\{\sum_{k=1}^N s_k u_k^2 \leq \overline{s} - t\right\}, \Pr\left\{\sum_{k=1}^N s_k u_k^2 \geq \overline{s} + t\right\}\right] \leq \exp\left\{-\frac{Nt^2}{4B\left(B+t\right)}\right\}.$$

Proof: Let x denote $\sum_{k=1}^N s_k u_k^2$ and let us estimate $\Pr\left\{x \geq \overline{s}^{(N)} + t\right\}$. We will estimate the conditional probability $\Pr\left\{x \geq \overline{s}^{(N)} + t \mid s_1, ..., s_N\right\}$, which we denote as $\Pr\left\{x \geq \overline{s} + t\right\}$ for simplicity. Note that

$$\begin{aligned} \Pr\left\{x \geq \overline{s} + t\right\} &\leq e^{-z(\overline{s}+t)} E e^{zx} \\ &= e^{-z(\overline{s}+t)}\left(1 + M_1 z + \frac{1}{2!} M_2 z^2 + ...\right), \end{aligned}$$

where $z > 0$ and $M_p = E x^p$.

Let us use von Neumann's formulas from [9] (pages 373-375) for the uncentered moments of the random variable x. Namely, let

$$\alpha_l = \frac{1}{2l} \sum_{i=1}^N \left(s_i\right)^l,$$

and let

$$1 + \beta_1 z + \beta_2 z^2 + \beta_3 z^3 + \ldots = e^{\alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 + \ldots}.$$

Then, von Neumann's result is that

$$E x^k = \frac{2^k k!}{N\left(N+2\right) \ldots \left(N + 2k - 2\right)} \beta_k.$$

Using this result, we write:

$$\begin{aligned}
1+M_1 z+\frac{1}{2!}M_2 z^2+\ldots &= 1+\frac{2}{N}\beta_1 z+\frac{2^2}{N\,(N+2)}\beta_2 z^2+\ldots \\
&\leq 1+\beta_1\left(\frac{2z}{N}\right)+\beta_2\left(\frac{2z}{N}\right)^2+\ldots \\
&= \exp\left\{\alpha_1\left(\frac{2z}{N}\right)+\alpha_2\left(\frac{2z}{N}\right)^2+\ldots\right\}.
\end{aligned}$$

Next, note that $2\alpha_1/N=\overline{s}$, and that $\alpha_k\leq k^{-1}\,(N/2)\,B^k$. This implies that

$$\begin{aligned}
e^{-z(\overline{s}+t)}\left(1+M_1 z+\frac{1}{2!}M_2 z^2+\ldots\right) &\leq e^{-zt}\exp\left\{\frac{N}{4}\left[\left(\frac{2Bz}{N}\right)^2+\left(\frac{2Bz}{N}\right)^3+\ldots\right]\right\} \\
&= e^{-zt}\exp\left\{\frac{N}{4}\frac{\left(\frac{2Bz}{N}\right)^2}{1-\frac{2Bz}{N}}\right\} \\
&= e^{-zt}\exp\left\{\frac{B^2 z^2}{N-2Bz}\right\}.
\end{aligned}$$

Let

$$z_0=\frac{Nt}{2B\,(B+t)}.$$

Then

$$\frac{B^2 z_0^2}{N-2Bz_0}-z_0 t=-\frac{Nt^2}{4B\,(B+t)}.$$

Altogether, we get:

$$\Pr\{x\geq\overline{s}+t\}\leq\exp\left\{-\frac{Nt^2}{4B\,(B+t)}\right\}.$$

The proof of the inequality for $\Pr\{x\leq\overline{s}-t\}$ is similar. QED.

Corollary 7 *Suppose that with probability 1, $s_k\leq b\overline{s}$ for all k. Then for all $t>0$,*

(9)

$$\max\left[\Pr\left\{\sum_{k=1}^{N}s_k u_k^2\leq\overline{s}\,(1-t)\right\},\Pr\left\{\sum_{k=1}^{N}s_k u_k^2\geq\overline{s}\,(1+t)\right\}\right]\leq\exp\left\{-\frac{Nt^2}{4b\,(b+t)}\right\}.$$

Corollary 8 *Let* $0 \le s_k \le b$ *for each* $k,$ *and* $t \in (0, 1/2)$. *Then,*

(i)

$$\Pr\left\{\log \sum_{k=1}^{N} s_k u_k^2 \ge \log \overline{s} + t\right\} \le \exp\left\{-\frac{Nt^2}{4b\,(b+t)}\right\}; \tag{10}$$

(ii)

$$\Pr\left\{\log \sum_{k=1}^{N} s_k u_k^2 \ge \log \overline{s} - (2\log 2)\, t\right\} \le \exp\left\{-\frac{Nt^2}{4b\,(b+t)}\right\}; \tag{11}$$

(iii)

$$\Pr\left\{\left|\log \sum_{k=1}^{N} s_k u_k^2 - \log \overline{s}\right| \ge t\right\} \le 2\exp\left\{-\frac{Nt^2}{4c\,(c+t)}\right\}, \tag{12}$$

where $c = (2\log 2)\, b$.

Proof: Let x denote $\sum_{k=1}^{N} s_k u_k^2$. Then

$$\begin{aligned}\Pr\{x \ge \overline{s} + t\} &= \Pr\left\{\log x \ge \log \overline{s} + \log\left(1 + \frac{t}{\overline{s}}\right)\right\} \\ &\ge \Pr\left\{\log x \ge \log \overline{s} + \frac{t}{\overline{s}}\right\}.\end{aligned}$$

This and (8) proves the first inequality The second inequality is proved similarly, and the third one is a consequence of the first two inequalities. QED.

Lemma 9 *Suppose that* X *is a random variable such that*

$$\Pr\{|X| \ge t\} \le 2\exp\left\{-\frac{Nt^2}{4c\,(c+t)}\right\},$$

where $c > 0$. *Let* $|z| < N/(16c)$. *Then*

$$Ee^{zX} \le \sqrt{32\pi}\sqrt{\frac{c^2 z^2}{N}}\exp\left(2\frac{c^2 z^2}{N}\right) + 3e^{|z|/\sqrt{N}} + 2\exp\left(-\frac{N}{16}\right).$$

Proof: Consider the case when $z \geq 0$. First, let us estimate $\int_{1/\sqrt{N}}^{\infty} e^{zt}\mu(dt)$, where μ is the distribution measure of X. Let $F(t) =: \Pr\{X \geq t\}$. Then, by integrating by parts and using the inequalities

$$F(t) \leq 2\exp\left\{-\frac{Nt^2}{4c(c+t)}\right\}$$

and $N \geq 1$, we get

$$\begin{aligned}
\int_{1/\sqrt{N}}^{\infty} e^{zt}\mu(dt) &= F\left(\frac{1}{\sqrt{N}}\right)e^{z/\sqrt{N}} + z\int_{1/\sqrt{N}}^{\infty} e^{zt}F(t)\,dt \\
&\leq 2e^{-1/[4c(c+1)]}e^{z/\sqrt{N}} + 2z\int_{1/\sqrt{N}}^{\infty} e^{zt}\exp\left\{-\frac{Nt^2}{4c(c+t)}\right\}dt.
\end{aligned}$$

In order to estimate the integral in the last line, we divide it into two pieces, $\int_{1/\sqrt{N}}^{b}$ and $\int_{b}^{\infty}$. Then,

$$\begin{aligned}
\int_{1/\sqrt{N}}^{b} e^{zt}\exp\left\{-\frac{Nt^2}{4c(c+t)}\right\}dt &\leq \int_{-\infty}^{\infty} e^{zt}\exp\left\{-\frac{Nt^2}{8c^2}\right\}dt \\
&= \exp\left(\frac{2c^2}{N}z^2\right)\int_{-\infty}^{\infty}\exp\left\{-\frac{N}{8c^2}\left(t-\frac{4c^2}{N}z\right)^2\right\}dt \\
&= \sqrt{\frac{8\pi c^2}{N}}\exp\left(\frac{2c^2}{N}z^2\right).
\end{aligned}$$

Next, for the second piece, we have:

$$\begin{aligned}
\int_{b}^{\infty} e^{zt}\exp\left\{-\frac{Nt^2}{4c(c+t)}\right\}dt &\leq \int_{b}^{\infty} e^{zt}\exp\left\{-\frac{Nt}{8c}\right\}dt \\
&= \frac{1}{N/(8c)-z}\exp\left(-\left(\frac{N}{8c}-z\right)c\right) \\
&\leq \frac{16c}{N}\exp\left(-\frac{N}{16}\right),
\end{aligned}$$

where we used the assumption that $z \leq N/(16c)$.

Hence, combining the previous inequalities and using the assumption that

$z \leq N/(16c)$ again, we get:

$$\int_{1/\sqrt{N}}^{\infty} e^{zt}\mu(dt) \leq 2e^{-1/[4c(c+1)]}e^{z/\sqrt{N}} + 2z\sqrt{\frac{8\pi c^2}{N}}\exp\left(\frac{2c^2}{N}z^2\right) +2\exp\left(-\frac{N}{16}\right).$$

In addition,

$$\int_{-\infty}^{1/\sqrt{N}} e^{zt}\mu(dt) \leq e^{z/\sqrt{N}}.$$

Combining all the parts, we get:

$$\int_{-\infty}^{\infty} e^{zt}\mu(dt) \leq \sqrt{\frac{32\pi c^2 z^2}{N}}\exp\left(\frac{2c^2}{N}z^2\right) + \left(1+2e^{-1/[4c(c+1)]}\right)e^{z/\sqrt{N}} + 2\exp\left(-\frac{N}{16}\right),$$

from which the claim of the lemma follows for $z \geq 0$. The case when $z \leq 0$ is similar. QED.

Corollary 10 *Let* $X = \log\left(\sum_{k=1}^{N} s_k u_k^2\right) - \log(\overline{s})$ *and let* $|z| \leq N/(16c)$, *where* $c = (2\log 2)\, b$. *Then*

$$Ee^{zX} \leq \sqrt{32\pi}\sqrt{\frac{c^2z^2}{N}}\exp\left(2\frac{c^2z^2}{N}\right) + 3e^{|z|/\sqrt{N}} + 2\exp\left(-\frac{N}{16}\right).$$

Proof: This follows directly from Lemma 9 and inequality (10). QED.

Proof of the second part of Proposition 3: Note that

$$\log\|\Pi_n u\|^2 = \sum_{i=1}^{n}\log\left[\sum_{k=1}^{N} s_k^{(i,N)}\left(u_k^{(i,N)}\right)^2\right]$$

where $u_k^{(i,N)}$ are components of independent Haar-distributed N-vectors $u^{(i,N)}$. Let

$$Y_i = \log\left(\sum_{k=1}^{N} s_k^{(i,N)}\left(u_k^{(i,N)}\right)^2\right) - \log\left(\overline{s}^{(i,N)}\right).$$

We aim to estimate

$$\Pr\left\{\left|\sum_{i=1}^{n} Y_i\right| > nt\right\}.$$

As usual,

$$\Pr\left\{\sum_{i=1}^{n} Y_i > nt\right\} \le e^{-nzt}\left(Ee^{zY_i}\right)^n,$$

where $z > 0$.

Note that by Assumption B,

$$s_k^{(i,N)} \le b\overline{s}^{(i,N)},$$

hence, our previous lemmas are applicable.

We set $z = tN/\left(4c^2\right)$ and assume that $N \ge 4/t^2$. (Note that the assumption that $t \in (0, 1/4]$ implies that $z \le N/\left(16c\right)$.) Then, by the previous Lemma, we have:

$$\begin{aligned} Ee^{zY_i} &\le \sqrt{32\pi}\sqrt{\frac{c^2z^2}{N}}\exp\left(2\frac{c^2z^2}{N}\right) + 3e^{z/\sqrt{N}} + 2\exp\left(-\frac{N}{16}\right) \\ &\le \sqrt{2\pi}\sqrt{\frac{t^2N}{c^2}}\exp\left(\frac{1}{8}\frac{t^2N}{c^2}\right) + 3\exp\left(\frac{t\sqrt{N}}{4c^2}\right) + 2\exp\left(-\frac{N}{16}\right). \end{aligned}$$

Since $N \ge 4/t^2$, then the first term dominates the other two terms, and we can write:

$$Ee^{zY_i} \le \left(\sqrt{2\pi}\sqrt{\frac{t^2N}{c^2}} + 5\right)\exp\left(\frac{t^2N}{8c^2}\right).$$

Hence,

$$\begin{aligned} e^{-nzt}\left(Ee^{zY_i}\right)^n &\le \left(\sqrt{2\pi}\sqrt{\frac{t^2N}{c^2}} + 5\right)^n \exp\left(-\frac{t^2N}{8c^2}n\right) \\ &= \exp\left\{-n\left[-\log\left(\sqrt{2\pi}\left(t/c\right)\sqrt{N} + 5\right) + \left(t^2/8c^2\right)N\right]\right\}. \end{aligned}$$

Clearly we can find an $N_0(t)$ such that for all $N > N_0(t)$

$$e^{-nzt}\left(Ee^{zY_i}\right)^n \le \exp\left\{-n\left(t^2/16c^2\right)N\right\}.$$

Hence, for all $N > N_0(t)$

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n}\left[y_i^{(N)} - \log\left(\overline{s}^{(i,N)}\right)\right] > t\right\} \le \exp\left\{-n\left(t^2/16c^2\right)N\right\}. \tag{13}$$

The case of the inequality

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n}\left[y_i^{(N)}-\log\left(\overline{s}^{(i,N)}\right)\right]<-t\right\}\leq\exp\left\{-n\left(t^2/16c^2\right)N\right\}. \tag{14}$$

is similar. Finally, note that $16c^2 \leq 32b^2$. QED.

3. Necessary condition. Let us introduce the following assumption.

Assumption D' $E\left[\log\left\|X_i^{(N)}u\right\|\right]^2$ exists and bounded by a constant that does not depend on N.

Theorem 11 *Let Assumptions A, B, and D′ hold. Suppose that for every $\delta > 0$ there exists such an $n_0(\delta)$ that*

$$\Pr\left\{\left|n^{-1}\log\|\Pi_n\|-E\log\left\|X_1^{(N)}v\right\|\right|\geq\delta\right\}\leq\delta \tag{15}$$

for all N and all $n \geq n_0(\delta)$. Let $b(N)$ is an arbitrary function of N such that $\lim_{N\to\infty} b(N) = +\infty$. Then

$$\lim_{N\to\infty}\Pr\left\{\left\|X_1^{(N)}\right\|\geq b(N)\right\}=0.$$

Proof: Let v_0 be such a unit vector that $\left\|X_1^{(N)}\right\| = \left\|X_1^{(N)}v_0\right\|$. Note that $X_1^{(N)}v_0$ has the Haar distribution by assumption of rotational invariance. By using the fact that $\|\Pi_n\| \geq \|\Pi_n v_0\|$, we can write the inequality

$$n^{-1}\log\|\Pi_n\|\geq\frac{\log\left\|X_1^{(N)}\right\|}{n}+\frac{1}{n}\sum_{i=2}^{n}\log\left(X_i^{(N)}u_i\right),$$

where u_i are independent Haar-distributed vectors. By using assumption D', we can conclude that $n^{-1}\sum_{i=2}^{n}\log\left(X_i^{(N)}u_i\right)$ converges in probability to $E\log\left\|X_1^{(N)}v\right\|$ and that the convergence is uniform in N. This fact and the supposition of the theorem imply that $n^{-1}\log\left\|X_1^{(N)}\right\|$ must converge in probability to zero as $n\to\infty$, and that the convergence must be uniform in N. If the conclusion of the theorem were invalid, then for some $\delta > 0$ and all n, we could find an $N = N(n,\delta)$ such that $\Pr\left\{\log\left\|X_1^{(N)}\right\|\geq n\delta\right\}\geq\delta$, and this would contradict the uniform convergence of $n^{-1}\log\left\|X_1^{(N)}\right\|$ to zero. QED.

4. Conclusion. In this paper, we found sufficient conditions that ensure that the convergence rate in the Furstenberg-Kesten theorem is uniform with respect to the dimension of the space in which matrices operate. Let us call this phenomenon dimensional uniformity of convergence.

Several interesting questions remain to be answered. First, is it possible to prove the dimensional uniformity of convergence for random matrices which are not rotationally invariant, for example, for Wigner matrices?

Second, assuming rotational invariance, what characterises the laws of singular values $s_k^{(i,N)}$, for which the dimensional uniformity of convergence holds? In other words, what are necessary and sufficient conditions for dimensional uniformity of convergence?

REFERENCES

[1] Joel E. Cohen and Charles M. Newman, *The stability of large random matrices and their products*, Annals of Probability **12** (1984), 283–310.

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

[3] Harry Furstenberg, *Noncommuting random products*, Transactions of the American Mathematical Society **108** (1963), 377–428.

[4] Y. Guivarch and A. Raugi, *Frontiere de furstenberg, proprietes de contraction et theoremes de convergence*, Zeit. Fur Wahrscheinlichkeitstheorie und Verw. Gebiete **67** (1985), 265–278.

[5] E. Le Page, *Theoremes limites pour le produits de matrices aleatoires*, Lecture Notes in Mathematics, vol. 928, Springer-Verlag, New York, 1982, pp. 258–303.

[6] Vitali D. Milman and Gideon Schechtman, *Asymptotic theory of finite dimensional normed spaces*, corrected second printing ed., Lecture Notes in Mathematics, vol. 1200, Springer, 2001.

[7] V. I. Oseledec, *A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems*, Transactions of the Moscow Mathematical Society **19** (1968), 197–231.

[8] V. N. Tutubalin, *On limit theorems for products of random matrices*, Theory of Probability and its Applications **10** (1965), 15–27.

[9] John von Neumann, *Distribution of the ratio of the mean square successive difference to the variance*, Annals of Mathematical Statistics **12** (1941), 367–395.