

ASYMPTOTIC LYAPUNOV EXPONENTS FOR LARGE RANDOM MATRICES

HOI H. NGUYEN

ABSTRACT. Suppose that A_1, \dots, A_N are independent random matrices whose atoms are iid copies of a random variable ξ of mean zero and variance one. It is known from the works of Newman et. al. in the late 80s that when ξ is gaussian then $N^{-1} \log \|A_N \dots A_1\|$ converges to a non-random limit. We extend this result to more general matrices with explicit rate of convergence. Our method relies on a simple connection between structures and dynamics.

1. INTRODUCTION

Let $A_i, i \geq 1$ be a sequence of independent identically distributed square random matrices of a given distribution μ in $M_{\mathbf{R}}(n)$. Let B_N be the matrix product

$$B_N = A_N \dots A_1.$$

Furstenberg and Kesten [11] (see also [5, Theorem 4.1, p.11]) proved in 1960 that

Theorem 1.1. *Assume that $\mathbf{E} \log^+(\|A_i\|) < \infty$ (where $\log^+ x = \max\{0, \log x\}$) then with probability one $\frac{1}{N} \log \|B_N\|$ converges to a deterministic number γ .*

Here and later, if not specified, our norm is always the $\|\cdot\|_2$ norm. The limit γ is called the *top Lyapunov exponent*. If we assume the common distribution μ of the A_i to be strongly irreducible (i.e. there does not exist a finite family of proper linear subspaces V_1, \dots, V_k of \mathbf{R}^n such that $M_\mu(V_1 \cup \dots \cup V_k) = V_1 \cup \dots \cup V_k$, where M_μ is the smallest closed subgroup which contains the support of μ), then Furstenberg showed in [10] (see also [5, Corollary 3.4, p.53]) that

Theorem 1.2 (Furstenberg's theorem). *Assume that $\mathbf{E} \log^+(\|A_i\|) < \infty$ and that μ is strongly irreducible, then*

- $\lim_{N \rightarrow \infty} \frac{1}{N} \log \|B_N \mathbf{x}\| = \gamma$ uniformly on $\mathbf{x} \in S^{n-1}$;
- for any μ -invariant distribution ν on $\mathbf{P}(\mathbf{R}^n)$ (i.e. $\nu(A) = \int \int 1_A(M\bar{\mathbf{x}}) d\mu(M) d\nu(\bar{\mathbf{x}})$) we have

$$\gamma = \int \int \log \frac{\|M\mathbf{x}\|}{\|B_N \mathbf{x}\|} d\mu(M) d\nu(\bar{\mathbf{x}}).$$

2000 *Mathematics Subject Classification.* 15A52, 60B10.

The author is supported by grants DMS-1600782, DMS-1128155, and CCF-1412958. The most part of this note was done while the author was visiting IAS and VIASM. The author thanks the two institutions for their hospitality.

There are also important extensions when M_μ is replaced by T_μ , the smallest closed semi-group which contains the support of μ ; and when strongly irreducibility is reduced to irreducibility, see for instance [5, 10, 13].

We next introduce other Lyapunov exponents by the use of exterior powers \wedge^k .

Definition 1.3. Assume that $\mathbf{E} \log^+(\|A_i\|) < \infty$. The Lyapunov exponents $\gamma_1, \dots, \gamma_n$ associated to A_i are defined inductively by $\gamma_1 = \gamma$ and for $k \geq 2$,

$$\sum_{i=1}^k \gamma_i = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{E} \log \|\wedge^k B_N\|.$$

In [20] (see also [13, Theorem 1.2]), Oseledec showed the following extremely powerful theorem on the convergence of Lyapunov exponents.

Theorem 1.4 (Oseledec's multiple ergodic theorem). *Assume that $\mathbf{E} \log^+(\|A_i\|) < \infty$, then the followings hold.*

- *With probability one,*

$$\gamma_k = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{E} \log \sigma_k(B_N), \quad (1)$$

where $\sigma_1(B_N) \geq \dots \geq \sigma_n(B_N)$ are the singular values of B_N .

- *With probability one, the matrix limit $(B_N B_N^T)^{1/2N}$ converges to a matrix $M \in M_{\mathbf{R}}(n)$ whose eigenvalues coincide with $\exp(\gamma_i)$ counting multiplicities.*
- *Let $\exp(\alpha_1(M)) < \dots < \exp(\alpha_k(M))$ denote the different eigenvalues of M with multiplicities $n_1(M), \dots, n_k(M)$, and let U_1, \dots, U_k be the corresponding eigensubspaces, and set $V_i = U_1 \oplus \dots \oplus U_i$. Then the pair $(\alpha_i(M), n_i(M))$ is μ -invariant, and for any unit vector $\mathbf{x} \in V_i \setminus V_{i-1}$, with probability one*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \|B_N \mathbf{x}\| = \alpha_i.$$

In practice, the issues when the top exponent γ_1 is strictly positive or when all of the Lyapunov exponents are distinct are extremely important. We refer the reader to [13] for further discussion on these accounts.

Following the two celebrated results of Furstenberg and Oseledec above, for some *nice* distribution μ it is also natural to ask

Question 1.5. *Can we give*

- (i) *(invariant measure) fine approximation for the Lyapunov's exponents?*
- (ii) *(large deviation type) quantification of the rate of convergence ?*

These aspects have been widely studied by many researchers, especially for unimodular and/or symplectic matrices of fixed size in connection to the theory of Schrödinger operators. For a

thorough introduction to these topics, we refer the reader to the books by Figotin and Pastur [8] and by Bourgain [3]. For the sake of completeness, allow us to insert here a large deviation type result for the shift model from [3] (see also [4] and [15]).

Theorem 1.6. *Assume that ω is an element of the one dimensional torus \mathbf{T} such that*

$$\text{dist}(k\omega, \mathbf{Z}^2) > c \frac{1}{|k| \log^3(1 + |k|)} \text{ for all } k \in \mathbf{Z}/\{0\}.$$

Let E be a fixed parameter and let f be a real analytic function on \mathbf{T} . Let \mathbf{x} be sampled uniformly at random from \mathbf{T} , and consider the random matrix product

$$B_N = \prod_{j=1}^N \begin{pmatrix} f(\mathbf{x} + j\omega) - E & -1 \\ 1 & 0 \end{pmatrix}.$$

Then for $t > N^{-1/10}$

$$\mathbf{P}_{\mathbf{x}} \left(\left| \frac{1}{N} \log \|B_N\| - \frac{1}{N} \mathbf{E} \log \|B_N\| \right| > t \right) < C e^{-ct^2 N},$$

for some absolute constants C and c .

One can use result of this type to study the decay of the corresponding Green function.

1.7. The iid model with large dimension. Our main focus is on a simple model of random matrices of very *large* dimension which are not necessarily unimodular. Especially, we will consider those A_i random matrices where the entries are iid copies of a common real random variable ξ of mean zero and variance $1/n$. This ensemble had been considered by Cohen, Isopi and Newman in the 80s [6, 12, 19] in connection to May's proposal of a specific quantitative relationship between complexity and stability within certain ecological models. We cite here a result by Newman which might best suit our discussion.

Theorem 1.8. [19, Equation (6)] *Assume that the entries of A_i are iid copies of $\frac{1}{\sqrt{n}}N(0, 1)$. Let $\mu_1 \geq \dots \geq \mu_n$ be the Lyapunov's exponents of the matrix product B_N . Then*

$$\mu_i = \frac{1}{2} (\log 2 + \Psi(\frac{n-i+1}{2}) - \log n),$$

where $\Psi(d) = \Gamma'(d)/\Gamma(d)$ is the digamma function.

This result was also generalized in [12] to ξ having bounded density and $\mathbf{E}((\sqrt{n}\xi)^4) < \infty$. We also refer the reader to a more recent result by Forester [9] and the survey [1] by Akerman and Ispen for more references. These results address the first part of Question 1.5 for various random matrices of *invariance* type.

For the large deviation part of Question 1.5, the only result we found for the iid model is due to Kargin [17] who considered the rate of convergence of the top exponents.

Theorem 1.9. [17, Proposition 3] *Let $\varepsilon > 0$ be given. Assume that the entries of A_i are iid copies of $\frac{1}{\sqrt{n}}N(0, 1)$. Then for all sufficiently small t , and all $n \geq n_0(t)$ and $N \geq 1$*

$$\mathbf{P}\left(\frac{1}{N} \log \|B_N\| > t + \varepsilon/N\right) \leq 2(1 + 2/\varepsilon)^n \exp\left(-\frac{1}{8}Nnt^2\right).$$

Remark 1.10. *To be more precise, Proposition 3 of [17] shows that $\mathbf{P}(\frac{1}{N} \log \|B_N \mathbf{x}\| > t) \leq \exp(-\frac{1}{8}Nnt^2)$ for any fixed $\mathbf{x} \in S^{n-1}$, from which one can deduce Theorem 1.9 by an ε -net argument, see for instance Claim 2.2.*

1.11. **Our results.** As far as we are concerned, all of the results in the literature with respect to this iid model assumed the common distribution ξ to be sufficiently smooth (i.e. at least the density function exists and is bounded), so that $\frac{1}{N} \log \|B_N\|$ with $N \rightarrow \infty$ is well defined almost surely.

The smoothness assumption is natural, as if A_i were singular with positive probability, then our chain B_N would become singular with probability one; in this case it might still be reasonable to study the top Lyapunov exponent, but not other exponents. However, and this is important to our study, *would it still be useful to study the "Lyapunov exponents" when N does not necessarily tend to infinity?* In other words, even when the exponents are not well defined, can we still say useful things about the growth of the chain B_N for some *effective* range of N ? Our motivation is based on the following two facts:

- (i) in many practical problems, it is not known a priori that our random matrix model is smooth;
- (ii) to estimate the Lyapunov's exponents using computer, it actually computes $\frac{1}{N} \log \sigma_i(B_N)$ for some sufficiently large (but not too large) N .

Trying to address these questions, and with a universality approach in mind, we will consider the matrix models A_i where the entries of $\sqrt{n}A_i$ are iid copies of a random variable ξ of mean zero, variance one, and that there exists a parameter K such that for all t

$$\mathbf{P}(|\xi| \geq t) = O(\exp(-t^2/K)). \tag{2}$$

One representative example of our matrices are Bernoulli ensembles, where ξ takes value ± 1 with probability $1/2$. As addressed above, there are two main obstacles for this discrete model: first the matrix law is not rotational invariant; and second, with probability one, the product matrix B_N will be the *zero matrix* as $N \rightarrow \infty$ (for instance when n is even then with positive probability the entries of A_1 are all 1s and each row of A_2 has exactly $n/2$ entries 1s).

The first problem is not strictly impossible, as there have been major developments in recent years showing that the spectral behavior of the iid matrices is universal. The second problem is, on the other hand, more subtle. This forces us to put an upper bound on N . The main question then boils down to finding a fine range of N for which one can still achieve non-trivial estimates.

In this note we show that as long as N grows slower than exponential in n , one can have relatively good control on the exponents.

Theorem 1.12 (Main results). *Let $\varepsilon > 0$ be given. Let A_1, A_2, \dots be independent matrices whose entries are iid copies of $\frac{1}{\sqrt{n}}\xi$ with ξ from (2). Then the followings hold.*

(1) (top exponent) *For any $t \geq 1/n$ we have*

$$\mathbf{P}\left(\left|\frac{1}{N} \log \|B_N\|\right| \geq t + \varepsilon/N\right) \leq (1 + 2/\varepsilon)^n [\exp(-c \min\{t^2, t\}Nn) + Nn^{-cn}].$$

(2) (second exponent) *For any $t \geq 1/n$ we have*

$$\mathbf{P}\left(\left|\frac{1}{N} \log \sup_{(\mathbf{x}_1, \mathbf{x}_2) \in \Gamma_2} \|B_N \mathbf{x}_1 \wedge B_N \mathbf{x}_2\|\right| \geq t + \varepsilon/N\right) \leq (1 + 2/\varepsilon)^n [\exp(-c \min\{t^2, t\}Nn) + Nn^{-cn}].$$

(3) (last exponent) *For any constant $\varepsilon > 0$, there exists $C = C(\varepsilon)$ such that*

$$\mathbf{P}\left(\inf_{\mathbf{x} \in S^{n-1}} \frac{1}{N} \log \|B_N \mathbf{x}\| \leq -\left(\frac{1}{2} + \varepsilon\right) \log n\right) \leq C^m \exp(-N/2) + Nn^{-\omega(1)}.$$

In short, (1) of Theorem 1.12 extends Theorem 1.9 to general matrix ensemble, with the extra assumption $N \ll n^{cn}$ and with $n_0(t) = O(1/t)$. It shows that although the chain dies out eventually, one (and the computer) can still see the concentration of the top exponent as long as N is not exceedingly large. This also fits with the simulation presented in [18] (see Fig. 4 and the discussion afterwards.) By taking $t = 1/n$ and $\varepsilon = 1/2$, we obtain

$$\mathbf{P}\left(\left|\frac{1}{N} \log \|B_N\|\right| \geq O(1/n)\right) \leq C^m [\exp(-N/n) + Nn^{-cn}].$$

We also show that the approach can be modified (in a non-trivial way) to control other top Lyapunov's exponents: it follows from (1) and (2) that the asymptotic second exponent γ_2 is also well concentrated around zero, and the method seems to extend to other asymptotic γ_k for any fixed k . Nevertheless, our concentration result is not local enough to see the difference between γ_1 and γ_2 as in Theorem 1.8.

Finally, we show in (3) that the asymptotic least exponent γ_n is approximately at least $-\frac{1}{2} \log n$, which again fits with the calculation of Theorem 1.8. Our control for γ_n , on the other hand, is not as sharp as for the top ones. It is not clear how γ_n fluctuates around its mean, but we believe that it is not very well concentrated. Furthermore, a similar bound for $\mathbf{P}(\inf_{\mathbf{x} \in S^{n-1}} \frac{1}{N} \log \|B_N \mathbf{x}\| \geq -(\frac{1}{2} - \varepsilon) \log n)$ is expected to hold, but we will not address this matter here (it is usual the case that bounding this quantity from below (i.e. (3)) is essentially harder than from bounding from above.)

In our next section we introduce the methods to prove Theorem 1.12.

2. PROOF METHOD

2.1. The top exponent. Here we discuss the method to prove (1) of Theorem 1.12. To estimate $\|B_N\| = \sup_{\mathbf{x}_0 \in S^{n-1}} \|B_N \mathbf{x}_0\|$, it is worth working with a finite collection of unit vectors \mathbf{x}_0 . Let $\varepsilon > 0$ be a parameter, and let \mathcal{N}_{start} be an ε -net of S^{n-1} . It is well-known that one can assume

$|\mathcal{N}_{start}| \leq (1 + 2/\varepsilon)^n$. The following is often used in the context of bounding the largest singular values of random matrices.

Claim 2.2. *We have*

$$\sup_{\mathbf{x}_0 \in \mathcal{N}_{start}} \|B_N \mathbf{x}_0\| \leq \|B_N\| \leq (1 - \varepsilon)^{-1} \sup_{\mathbf{x}_0 \in \mathcal{N}_{start}} \|B_N \mathbf{x}_0\|.$$

With this claim, one hopes to control $\frac{1}{N} \log \|B_N\|$ (up to an approximated factor $1 + \frac{1}{N} \log(1 + \varepsilon)$ and up to a correcting factor $(1 + 2/\varepsilon)^n$ in probability) by establishing a strong concentration result for $\frac{1}{N} \log \|B_N \mathbf{x}_0\|$ for each $\mathbf{x}_0 \in \mathcal{N}_{start}$. This was also the main starting point of [17].

Let \mathbf{x}_0 be an element of \mathcal{N}_{start} . One might write

$$\begin{aligned} \log \|B_N \mathbf{x}_0\| &= \log \|A_N A_{N-1} \dots A_2 A_1 \mathbf{x}_0\| \\ &= \log \|A_N \frac{A_{N-1} \dots A_2 A_1 \mathbf{x}_0}{\|A_{N-1} \dots A_2 A_1 \mathbf{x}_0\|}\| + \log \|A_{N-1} \frac{A_{N-2} \dots A_2 A_1 \mathbf{x}_0}{\|A_{N-2} \dots A_2 A_1 \mathbf{x}_0\|}\| + \dots + \\ &+ \dots + \log \|A_2 \frac{A_1 \mathbf{x}_0}{\|A_1 \mathbf{x}_0\|}\| + \log \|A_1 \mathbf{x}_0\| \\ &= \sum_{i=0}^{N-1} \log \|A_{i+1} \mathbf{x}_i\|, \end{aligned}$$

where

$$\mathbf{x}_i := \frac{A_i \dots A_2 A_1 \mathbf{x}_0}{\|A_i \dots A_2 A_1 \mathbf{x}_0\|}. \quad (3)$$

When ξ has discrete distribution such as Bernoulli, there would be a minor problem that $B_N \mathbf{x}_0$ might be vanishing, but we could rule out this possibility by choosing the net \mathcal{N}_{start} to consist of vectors of "highly irrational" entries (which remain highly irrational under the actions of the matrices A_i).

Now we want to control $\log \|A_{i+1} \mathbf{x}_i\|$ conditioning on A_1, \dots, A_i (and hence on \mathbf{x}_i). Note that

$$\mathbf{E}_{A_{i+1}} \|A_{i+1} \mathbf{x}_i\|^2 = 1.$$

Roughly speaking, to hope for a good concentration of $\log \|A_{i+1} \mathbf{x}_i\|$ around zero, the very first step we have to guarantee is that with high probability with respect to A_{i+1} , the vector norm $\|A_{i+1} \mathbf{x}_i\|$ is being well away from zero.

This probability certainly depends on the structure of \mathbf{x}_i . For instance if $\mathbf{x}_i = (\pm 1/\sqrt{2}, \pm 1/\sqrt{2}, 0, \dots, 0)$ or $\mathbf{x}_i = (\pm 1/\sqrt{n}, \dots, \pm 1/\sqrt{n})$ then the chance that $\|A_{i+1} \mathbf{x}_i\|$ being small (or even being annihilated) is not quite small if we are working with Bernoulli matrices. With this in mind, our general strategy consists of three main steps.

- (Step 1.) (dynamics and structures) find a set \mathcal{S} of S^{n-1} with the following properties:
 - \mathcal{S} covers an ε -net \mathcal{N}_{start} of S^{n-1} ,
 - \mathcal{S} remains stable under the action of each A_i . In other words, with high very high probability all of the normalized vectors \mathbf{x}_i from (3) belong to \mathcal{S} ;
 - for any $\mathbf{x} \in \mathcal{S}$, with high probability with respect to A_{i+1} the norm $\|A_{i+1}\mathbf{x}\|$ is bounded away from zero.
- (Step 2.) (concentration over good vectors) we show that for each $\mathbf{x}_i \in \mathcal{S}$, $\log \|A_{i+1}\mathbf{x}_i\|$ is very well concentrated around zero.
- (Step 3.) (law of large number) use concentration information from Step 2 to prove Theorem 1.12 .

We will lay out the choice of \mathcal{S} in Section 3. Step 2 will be carried out in Section 4, and Step 3 is concluded in Section 5.

2.3. The second exponent. We will extend the ideas of the previous subsection to deal with (2) of Theorem 1.12. First of all, let \mathcal{P}_{start} be some subset of $S^{n-1} \times S^{n-1}$ that covers an ε -net (that is for any $(\mathbf{x}, \mathbf{y}) \in S^{n-1} \times S^{n-1}$ there exists $(\mathbf{x}', \mathbf{y}') \in \mathcal{P}_{start}$ such that $\|\mathbf{x} - \mathbf{x}'\|, \|\mathbf{y} - \mathbf{y}'\| \leq \varepsilon$). Similarly to Claim 2.2, we have the following.

Claim 2.4.

$$\sup_{(\mathbf{x}, \mathbf{y}) \in S^{n-1} \times S^{n-1}} \text{Vol}_2(B_N \mathbf{x}, B_N \mathbf{y}) \leq (1 - 2\varepsilon - \varepsilon^2)^{-1} \sup_{(\mathbf{x}', \mathbf{y}') \in S^{n-1} \times S^{n-1} \cap \mathcal{P}_{start}} \text{Vol}_2(B_N \mathbf{x}', B_N \mathbf{y}').$$

Proof. (of Claim 2.4) Assume that $\sup_{(\mathbf{x}, \mathbf{y}) \in S^{n-1} \times S^{n-1}} \text{Vol}_2(B_N \mathbf{x}, B_N \mathbf{y})$ is attained at (\mathbf{x}, \mathbf{y}) . Let $(\mathbf{x}', \mathbf{y}')$ be an element in \mathcal{P}_{start} such that $\|\mathbf{x} - \mathbf{x}'\| \leq \varepsilon$ and $\|\mathbf{y} - \mathbf{y}'\| \leq \varepsilon$. By the triangle inequality

$$\begin{aligned} \text{Vol}_2(B_N \mathbf{x}, B_N \mathbf{y}) &\leq \text{Vol}_2(B_N \mathbf{x}', B_N \mathbf{y}') + \text{Vol}_2(B_N(\mathbf{x} - \mathbf{x}'), B_N \mathbf{y}) \\ &\quad + \text{Vol}_2(B_N \mathbf{x}', B_N(\mathbf{y} - \mathbf{y}')) + \text{Vol}_2(B_N(\mathbf{x} - \mathbf{x}'), B_N(\mathbf{y} - \mathbf{y}')) \\ &\leq \text{Vol}_2(B_N \mathbf{x}', B_N \mathbf{y}') + (2\varepsilon + \varepsilon^2) \sup_{(\mathbf{z}_1, \mathbf{z}_2) \in S^{n-1} \times S^{n-1}} \text{Vol}_2(B_N \mathbf{z}_1, B_N \mathbf{z}_2). \end{aligned}$$

□

Beside containing an ε -net we will also choose $\mathcal{P}_{start} \subset S^{n-1} \times S^{n-1}$ to satisfy certain non-structured properties (such as $\mathcal{P}_{start} \subset \mathcal{P}$, a broader set to be introduced below). The detail of construction of \mathcal{P}_{start} will be presented in Section 6.

Now let $(\mathbf{x}_0, \mathbf{y}_0) \in \mathcal{P}_{start}$. As customary, one might write

$$\begin{aligned}
\log \|B_N \mathbf{x}_0 \wedge B_N \mathbf{y}_0\| &= \log \|A_N A_{N-1} \dots A_2 A_1 \mathbf{x}_0 \wedge A_N A_{N-1} \dots A_2 A_1 \mathbf{y}_0\| \\
&= \log \|A_N \frac{A_{N-1} \dots A_2 A_1 \mathbf{x}_0 \wedge A_{N-1} \dots A_2 A_1 \mathbf{y}_0}{\|A_{N-1} \dots A_2 A_1 \mathbf{x}_0 \wedge A_{N-1} \dots A_2 A_1 \mathbf{y}_0\|}\| \\
&+ \log \|A_{N-1} \frac{A_{N-2} \dots A_2 A_1 \mathbf{x}_0 \wedge A_{N-2} \dots A_2 A_1 \mathbf{y}_0}{\|A_{N-2} \dots A_2 A_1 \mathbf{x}_0 \wedge A_{N-2} \dots A_2 A_1 \mathbf{y}_0\|}\| + \dots \\
&+ \log \|A_2 \frac{A_1 \mathbf{x}_0 \wedge A_1 \mathbf{y}_0}{\|A_1 \mathbf{x}_0 \wedge A_1 \mathbf{y}_0\|}\| + \log \|A_1 \mathbf{x}_0 \wedge A_1 \mathbf{y}_0\| \\
&= \sum_{i=0}^{N-1} \log \frac{\|A_{i+1} \mathbf{x}_i \wedge A_{i+1} \mathbf{y}_i\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|},
\end{aligned}$$

where

$$\mathbf{x}_i := \frac{A_i \dots A_2 A_1 \mathbf{x}_0}{\|A_i \dots A_2 A_1 \mathbf{x}_0\|} \text{ and } \mathbf{y}_i := \frac{A_i \dots A_2 A_1 \mathbf{y}_0}{\|A_i \dots A_2 A_1 \mathbf{y}_0\|}.$$

To control $\frac{\|A_{i+1} \mathbf{x}_i \wedge A_{i+1} \mathbf{y}_i\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|}$, we first pull out $\|A_{i+1} \mathbf{x}_i\|$ and $\|A_{i+1} \mathbf{y}_i\|$

$$\begin{aligned}
\log \frac{\|A_{i+1} \mathbf{x}_i \wedge A_{i+1} \mathbf{y}_i\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|} &= \log \|A_{i+1} \mathbf{x}_i\| + \log \|A_{i+1} \mathbf{y}_i\| \\
&+ \log \frac{\|A_{i+1} \mathbf{x}_i / \|A_{i+1} \mathbf{x}_i\| \wedge A_{i+1} \mathbf{y}_i / \|A_{i+1} \mathbf{y}_i\|\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|} \\
&= \log \|A_{i+1} \mathbf{x}_i\| + \log \|A_{i+1} \mathbf{y}_i\| + \log \frac{\|\mathbf{x}_{i+1} \wedge \mathbf{y}_{i+1}\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|}
\end{aligned}$$

By our treatment of the top exponent, one has very good control on $\log \|A_{i+1} \mathbf{x}_i\| + \log \|A_{i+1} \mathbf{y}_i\|$, thus the main task is to study the remaining term. To hope for a good concentration of $\log \frac{\|\mathbf{x}_{i+1} \wedge \mathbf{y}_{i+1}\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|}$ around zero, among other things we have to guarantee that $\mathbf{x}_i \wedge \mathbf{y}_i \neq 0$ with high probability, and within this event that $\frac{\|\mathbf{x}_{i+1} \wedge \mathbf{y}_{i+1}\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|}$ is close to one. Thus compared to the previous section, beside the task of bounding $\|\mathbf{x}_{i+1}\|, \|\mathbf{y}_{i+1}\|, \|\mathbf{x}_i\|, \|\mathbf{y}_i\|$ away from zero, we will have to show that the angles between these vectors are highly stable under the process, and this task is more complicated. Nevertheless, our plan remains the same in principle.

- (Step 1.) (dynamics and structures) find a set \mathcal{P} of pair vectors in \mathbf{R}^n such that $\mathcal{P}_{start} \subset \mathcal{P}$ and which remains stable under the action of the A_i 's with given $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{P}$: with very high probability with respect to A_{i+1}

$$(\mathbf{x}_{i+1}, \mathbf{y}_{i+1}) = \left(\frac{A_{i+1} \mathbf{x}_i}{\|A_{i+1} \mathbf{x}_i\|}, \frac{A_{i+1} \mathbf{y}_i}{\|A_{i+1} \mathbf{y}_i\|} \right) \in \mathcal{P}.$$

- (Step 2.) (concentration over good vectors) show that for given $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{P}$, with very high probability with respect to A_{i+1} the norm $\frac{\|\mathbf{x}_{i+1} \wedge \mathbf{y}_{i+1}\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|}$ is very well concentrated around one.

- (Step 3.) (law of large number) use concentration information from Step 2 to prove (2) of Theorem 1.12.

We will present a full proof of (2) of Theorem 1.12 with a more detailed description of the three steps above in Section 6.

2.5. The last exponent. Now we discuss the method to prove (3) of Theorem 1.12. Here the net argument does not work at all. We will have to relate the smallest Lyapunov exponent to the distances among the rows of the matrices A_i .

Let $\varepsilon > 0$ be a given small constant, and consider the event

$$\mathcal{E}_\varepsilon = \left\{ \inf_{\mathbf{x} \in S^{n-1}} \|B_N \mathbf{x}\| \leq T_\varepsilon := ((1 - \varepsilon)/\sqrt{n})^N \right\}.$$

As $\mathbf{x} = (x_1, \dots, x_n) \in S^{n-1}$, there exists $i_0 \in [n]$ such that $|x_{i_0}| \geq 1/\sqrt{n}$. With $\mathbf{c}_i = B_N \mathbf{e}_i$ being the i -th column vector of B_N , it follows from $\|\sum_i x_i \mathbf{c}_i\| \leq T_\varepsilon$ that

$$\text{dist}(\mathbf{c}_{i_0}, \text{span}(\mathbf{c}_i, i \neq i_0)) \leq \sqrt{n} T_\varepsilon.$$

Let $\mathcal{E}_{\varepsilon,1}$ be the event that

$$\mathcal{E}_{\varepsilon,1} = \left\{ \log \text{dist}(\mathbf{c}_n, \text{span}(\mathbf{c}_i, i \neq n)) \leq \log \sqrt{n} + \log T_\varepsilon \right\}.$$

We then have

$$\mathbf{P}(\mathcal{E}_\varepsilon) \leq n \mathbf{P}(\mathcal{E}_{\varepsilon,1}).$$

Thus for the upper bound, the main focus is to estimate $\mathbf{P}(\mathcal{E}_{\varepsilon,1})$. We will show that this probability is so small that the extra factor n will not affect at all.

In general, for any general non-degenerate tuple $(\mathbf{v}_1, \dots, \mathbf{v}_n)$,

$$\left\| \frac{A\mathbf{v}_1 \wedge \dots \wedge A\mathbf{v}_n}{\mathbf{v}_1 \wedge \dots \wedge \mathbf{v}_n} \right\| = \frac{\sqrt{\det(V^* A^* A V)}}{\sqrt{V^* V}} = |\det(A)|.$$

Also, for any non-degenerate tuple $(\mathbf{v}_1, \dots, \mathbf{v}_{n-1})$, with $\mathbf{v}_n \in S^{n-1}$ being orthogonal to all other $\mathbf{v}_i, 1 \leq i \leq n-1$, we write

$$\begin{aligned}
\left\| \frac{A\mathbf{v}_1 \wedge \cdots \wedge A\mathbf{v}_{n-1}}{\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{n-1}} \right\| &= \frac{|\det(A\mathbf{v}_1, \dots, A\mathbf{v}_{n-1})|}{|\det(\mathbf{v}_1, \dots, \mathbf{v}_{n-1})|} \\
&= \frac{|\det(A\mathbf{v}_1, \dots, A\mathbf{v}_{n-1}, A\mathbf{v}_n)| / \text{dist}(A\mathbf{v}_n, H_{A\mathbf{v}_1, \dots, A\mathbf{v}_{n-1}})}{|\det(\mathbf{v}_1, \dots, \mathbf{v}_{n-1}, \mathbf{v}_n)|} \\
&= \frac{|\det(A)|}{\text{dist}(A\mathbf{v}_n, H_{A\mathbf{v}_1, \dots, A\mathbf{v}_{n-1}})},
\end{aligned}$$

where $H_{A\mathbf{v}_1, \dots, A\mathbf{v}_{n-1}}$ is the subspace spanned by $A\mathbf{v}_1, \dots, A\mathbf{v}_{n-1}$. Taking \mathbf{v}_i to be the standard normal basis \mathbf{e}_i we thus obtain,

$$\begin{aligned}
\log \text{dist}(\mathbf{c}_n, \text{span}(\mathbf{c}_i, i \neq n)) &= \log \text{dist}(B_N \mathbf{e}_n, H_{B_N \mathbf{e}_1, \dots, B_N \mathbf{e}_{n-1}}) \\
&= \log \left\| \frac{B_N \mathbf{e}_1 \wedge \cdots \wedge B_N \mathbf{e}_n}{\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n} \right\| - \log \left\| \frac{B_N \mathbf{e}_1 \wedge \cdots \wedge B_N \mathbf{e}_{n-1}}{\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_{n-1}} \right\|.
\end{aligned}$$

Now as $B_N = A_N \dots A_1$, we can rewrite the second term of the above formula as

$$\begin{aligned}
\log \left\| \frac{B_N \mathbf{e}_1 \wedge \cdots \wedge B_N \mathbf{e}_{n-1}}{\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_{n-1}} \right\| &= \log \left\| \frac{A_N \dots A_1 \mathbf{e}_1 \wedge \cdots \wedge A_N \dots A_1 \mathbf{e}_{n-1}}{\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_{n-1}} \right\| \\
\log \left\| \frac{A_N (A_{N-1} \dots A_1 \mathbf{e}_1) \wedge \cdots \wedge A_N (A_{N-1} \dots A_1 \mathbf{e}_{n-1})}{A_{N-1} \dots A_1 \mathbf{e}_1 \wedge \cdots \wedge A_{N-1} \dots A_1 \mathbf{e}_{n-1}} \right\| &+ \log \left\| \frac{A_{N-1} \dots A_1 \mathbf{e}_1 \wedge \cdots \wedge A_{N-1} \dots A_1 \mathbf{e}_{n-1}}{A_{N-2} \dots A_1 \mathbf{e}_1 \wedge \cdots \wedge A_{N-2} \dots A_1 \mathbf{e}_{n-1}} \right\| \\
+ \cdots + \log \left\| \frac{A_1 \mathbf{e}_1 \wedge \cdots \wedge A_1 \mathbf{e}_{n-1}}{\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_{n-1}} \right\|. &
\end{aligned}$$

Decomposing similarly for $\log \|B_N \mathbf{e}_1 \wedge \cdots \wedge B_N \mathbf{e}_n / \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n\|$, we obtain

$$\log \text{dist}(B_N \mathbf{e}_n, H_{B_N \mathbf{e}_1, \dots, B_N \mathbf{e}_{n-1}}) = \sum_i \log \text{dist}(A_i \mathbf{v}_i, H_{A_i \dots A_1 \mathbf{e}_1, \dots, A_i \dots A_1 \mathbf{e}_{n-1}}), \quad (4)$$

where \mathbf{v}_i is a unit vector that is orthogonal to the vectors $\mathbf{u}_{i1}, \dots, \mathbf{u}_{i(n-1)}$ satisfying

$$\mathbf{u}_{i1} = A_{i-1} \dots A_1 \mathbf{e}_1, \dots, \mathbf{u}_{i(n-1)} = A_{i-1} \dots A_1 \mathbf{e}_{n-1}.$$

Now as $(A_i^{-1} \mathbf{v}_i)^T A_i \mathbf{u}_{ij} = \mathbf{v}_i^T \mathbf{u}_{ij} = 0$, the vector $A_i^{-1} \mathbf{v}_i / \|A_i^{-1} \mathbf{v}_i\|$ is the unit normal vector of the subspace $H_{A_i \mathbf{u}_{i1}, \dots, A_i \mathbf{u}_{i(n-1)}}$, and so

$$\text{dist}(A_i \mathbf{v}_i, H_{A_i \dots A_1 \mathbf{e}_1, \dots, A_i \dots A_1 \mathbf{e}_{n-1}}) = \frac{1}{\|A_i^{-1} \mathbf{v}_i\|}. \quad (5)$$

Note that the vectors $\mathbf{u}_{i1}, \dots, \mathbf{u}_{i(n-1)}$ and \mathbf{v}_i are independent of A_i , and hence it boils down to study the upper bound of $\|A_i^{-1}\mathbf{v}_i\|$ with the randomness with respect to A_i . We will study this in more detail in Section 7.

The rest of the paper is organized as follows. We will provide a detailed treatment for (1) of Theorem 1.12 throughout sections 3, 4 and 5. We then sketch the proof of (2) of Theorem 1.12 in Section 6. The proof of (3) of Theorem 1.12 will be presented in Section 7. We conclude the note by Section 8 with some remarks.

Notation. Throughout this paper, we regard N as an asymptotic parameter going to infinity. We write $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote the claim that $|X| \leq CY$ for some fixed C ; this fixed quantity C is allowed to depend on other fixed quantities such as the sub-gaussian parameter K of ξ unless explicitly declared otherwise. We also use $o(Y)$ to denote any quantity bounded in magnitude by $c(N)Y$ for some $c(N)$ that goes to zero as $N \rightarrow \infty$.

3. STEP 1 FOR (1) OF THEOREM 1.12: STRUCTURES UNDER MATRIX ACTION

Our choice of the set \mathcal{S} is motivated by recent ideas from Tao-Vu [25, 26] and from Rudelson-Vershynin [21, 22] in the context of controlling the small ball probability of random walk. Although this looks surprising at first, the reader will see later that these structures are indeed the right object to work with.

We first introduce the notion of *least common denominator* by Rudelson and Vershynin (see [21]). Fix parameters κ and γ , where $\gamma \in (0, 1)$. For any nonzero vector \mathbf{x} define

$$\mathbf{LCD}_{\kappa, \gamma}(\mathbf{x}) := \inf \left\{ \theta > 0 : \text{dist}(\theta \mathbf{x}, \mathbf{Z}^n) < \min(\gamma \|\theta \mathbf{x}\|, \kappa) \right\}.$$

We record a few easy consequences of \mathbf{LCD} .

Fact 3.1. *We have*

- If $\mathbf{y} = \lambda \mathbf{x}$ with $\lambda \neq 0$, then

$$\mathbf{LCD}_{\kappa, \gamma}(\mathbf{y}) = \frac{1}{|\lambda|} \mathbf{LCD}_{\kappa, \gamma}(\mathbf{x}).$$

- Assume that $\|\mathbf{x}\|, \|\mathbf{y}\| \geq \varepsilon$ with $D = \mathbf{LCD}_{\gamma, \kappa}(\mathbf{x}) \geq 1$ and $\|\mathbf{x} - \mathbf{y}\| \leq D^{-2}$, then

$$\mathbf{LCD}_{\kappa+1, \gamma+\frac{1}{D}}(\mathbf{y}) \leq \mathbf{LCD}_{\kappa, \gamma}(\mathbf{x}).$$

Proof. (of Fact 3.1) Assume that $\text{dist}(D\mathbf{x}, \mathbf{Z}^n) \leq \min\{\gamma \|D\mathbf{x}\|, \kappa\}$ for some $D > 0$. Then as $\|\mathbf{x} - \mathbf{y}\| \leq 1/D^2$, by the triangle inequality we have

$$\text{dist}(D\mathbf{y}, \mathbf{Z}^n) \leq \min\{\gamma(\|D\mathbf{y}\| + D\|\mathbf{x} - \mathbf{y}\|), \kappa + D\|\mathbf{x} - \mathbf{y}\|\} \leq \min\left\{\left(\gamma + \frac{1}{D}\right)\|D\mathbf{y}\|, \kappa + 1\right\}.$$

□

There are two main advantages of working with unit vectors $\mathbf{x} = (x_1, \dots, x_n)$ of large LCD. First, as it turns out, if $\mathbf{LCD}(\mathbf{x})$ is large then the random sum $\xi_1 x_1 + \dots + \xi_n x_n$, where ξ_i are iid copies of ξ from (2), behaves like a continuous random variable of bounded density (even when the ξ_i are discrete.) This statement is the content of the following result.

Theorem 3.2. [21] *For every*

$$\varepsilon \geq \frac{1}{\mathbf{LCD}_{\kappa, \gamma}(\mathbf{x})}$$

we have

$$\sup_x \mathbf{P}(|\xi_1 x_1 + \dots + \xi_n x_n - x| \leq \varepsilon) = O\left(\frac{\varepsilon}{\gamma} + e^{-\Theta(\kappa^2)}\right),$$

where the implied constants depend on ξ .

Second, vectors with small LCD can be well approximated by rational vectors $\mathbf{p}/\|\mathbf{p}\|$ with $\mathbf{p} \in \mathbf{Z}^n$ and $\|\mathbf{p}\|$ small.

Theorem 3.3. *Let $D \geq c\sqrt{n}$. Then the set $\{\mathbf{x} \in S^{n-1} : c\sqrt{n} \leq \mathbf{LCD}_{\kappa, \gamma}(\mathbf{x}) \leq D\}$ has a $(2\kappa/D)$ -net \mathcal{N}_D of cardinality at most*

$$(C_0 D / \sqrt{n})^n \log_2 D,$$

for some absolute constant C_0 .

To show this result, it suffices to establish nets for the level sets $S_{D_0} := \{x \in S^{m-1} : D_0 \leq \mathbf{LCD}_{\kappa, \gamma}(x) \leq 2D_0\}$.

Lemma 3.4. [22, Lemma 4.7] *There exists a $(2\kappa/D_0)$ -net of S_{D_0} of cardinality at most $(C_0 D_0 / \sqrt{n})^n$, where C_0 is an absolute constant.*

Subdividing these nets into $(2\kappa/D)$ -nets and taking the union as D_0 ranges over powers of two, we thus obtain Theorem 3.3. As the proof of Lemma 3.4 is short and uses the important notion of \mathbf{LCD} , we include it here for the reader's convenience.

Proof. (of Lemma 3.4) For $x \in S_{D_0}$, denote

$$D(x) := \mathbf{LCD}_{\kappa, \gamma}(x).$$

By definition, $D_0 \leq D(x) \leq 2D_0$ and there exists $p \in \mathbf{Z}^m$ with

$$\left\| x - \frac{p}{D(x)} \right\| \leq \frac{\kappa}{D(x)} = O\left(\frac{n^{2c}}{n^{1-c}}\right) = o(1).$$

As $\|x\| = 1$, this implies that $\|p\| \approx D(x)$, more precisely

$$1 - \frac{\kappa}{D(x)} \leq \left\| \frac{p}{D(x)} \right\| \leq 1 + \frac{\kappa}{D(x)}. \quad (6)$$

This implies that

$$\|p\| \leq (1 + o(1))D(x) < 3D_0. \quad (7)$$

It also follows from (6) that

$$\left\| x - \frac{p}{\|p\|} \right\| \leq \left\| x - \frac{p}{D(x)} \right\| + \left\| \frac{p}{\|p\|} \left(\frac{\|p\|}{D(x)} - 1 \right) \right\| \leq 2 \frac{\kappa}{D(x)} \leq \frac{2\kappa}{D_0}. \quad (8)$$

Now set

$$\mathcal{N}_0 := \left\{ \frac{p}{\|p\|}, p \in \mathbf{Z}^m \cap B(0, 3D_0) \right\}.$$

By (7) and (8), \mathcal{N}_0 is a $\frac{2\kappa}{D_0}$ -net for S_{D_0} . On the other hand, it is known that the size of \mathcal{N}_0 is bounded by $(C_0 \frac{D_0}{\sqrt{m}})^m$ for some absolute constant C_0 . \square

As Theorem 3.2 and Theorem 3.3 suggest, we will choose \mathcal{S} to be the collection of unit vectors with large LCD: for $\gamma = 1/2$, $\kappa = n^c$ and $D = \exp(n^c)$ with a sufficiently small constant c to be chosen we set

$$\mathcal{S} := \{\mathbf{x} \in S^{n-1}, \mathbf{LCD}_{\gamma, \kappa}(\mathbf{x}) \geq D\}. \quad (9)$$

We next show that this set contains "most" of the vectors of S^{n-1} .

Lemma 3.5. *With $\kappa = n^c$ with some $c < 1/6$ we have*

$$\text{Vol}(S^{n-1} \setminus \mathcal{S}) \leq \text{Vol}_{n-1}(\mathbf{B}(0, n^{-1/2+5c}))$$

Proof. (of Lemma 3.5) Let \mathcal{N}_D be one of the sets obtained from Theorem 3.3. Then by definition

$$\begin{aligned}
\text{Vol}(S^{n-1} \setminus \mathcal{S}) &\leq \text{Vol}_{n-1}(\mathcal{N}_D + \mathbf{B}(0, \kappa/D) \cap S^{n-1}) \leq \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} (\kappa/D)^{n-1} |\mathcal{N}_D| \\
&\leq \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} (C\kappa/D)^{n-1} \times (C_0 D/\sqrt{n})^n \log_2 D \\
&\leq \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} (CC_0\kappa/\sqrt{n})^n D^2 \\
&\leq \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} (n^{-1/2+4c})^n \\
&\leq \text{Vol}_{n-1}(\mathbf{B}(0, n^{-1/2+5c})).
\end{aligned}$$

□

Lemma 3.5 implies that for any $n^{-1/2+5c} \leq \varepsilon \leq 1$, there exists a ε -net \mathcal{N}_{start} of S^{n-1} with size $(C/\varepsilon)^n$ that belongs to \mathcal{S} . This set \mathcal{N}_{start} will be the collection of our starting vectors \mathbf{x}_0 discussed in Section 2.

Now we will proceed to our main result of Step 1. For this, we will find the following lemma useful.

Lemma 3.6. [21, Lemma 2.2] *Let ζ_1, \dots, ζ_n be independent non-negative random variables, and let $K, t_0 > 0$. If one has*

$$\mathbf{P}(\zeta_k < t) \leq Kt$$

for all $k = 1, \dots, n$ and all $t \geq t_0$, then one also has

$$\mathbf{P}\left(\sum_{k=1}^n \zeta_k^2 < t^2 n\right) \leq O((Kt)^n)$$

for all $t \geq t_0$.

For our analysis below we recall the definition of \mathcal{S} from (9).

Theorem 3.7 (key estimate, stability of non-structures). *Assume that $A = (a_{ij})_{1 \leq i, j \leq n}$ is a random matrix of size n whose entries are iid copies of ξ satisfying (2). Let $\mathbf{x} = (x_1, \dots, x_n)$ be any deterministic vector from \mathcal{S} . Then*

$$\mathbf{P}_A\left(\frac{A\mathbf{x}}{\|A\mathbf{x}\|} \notin \mathcal{S}\right) \leq n^{-cn/8}.$$

Proof. (of Theorem 3.7) We first consider the event \mathcal{E}_1 that $\|A\mathbf{x}\|^2 \leq n^{1-c/2}$. As $\mathbf{LCD}_{\gamma, \kappa}(\mathbf{x}) \geq D = \exp(n^c) \gg \sqrt{n}$, by Theorem 3.2

$$\mathbf{P}(|a_{i1}x_1 + \dots + a_{in}x_n| \leq n^{-c/4}) = O(n^{-c/4}).$$

Thus by Lemma 3.6

$$\mathbf{P}(\mathcal{E}_1) = \mathbf{P}(\|A\mathbf{x}\|^2 \leq n^{1-c/2}) \leq C^n n^{-cn/4}. \quad (10)$$

Now for the event \mathcal{E}_2 that $\|\mathbf{Ax}\| \geq n^{1-2c}$, by the standard Chernoff deviation result (as $\|\mathbf{Ax}\|^2 = \sum_i (\sum_j a_{ij}x_j)^2$) we have

$$\mathbf{P}(\|\mathbf{Ax}\| \geq n^{1-2c}) \leq \exp(-n^{1-4c}). \quad (11)$$

On the complement of \mathcal{E}_1 and \mathcal{E}_2 , for each $n^{1/2-c/4} \leq r \leq n^{1-2c}$ let us look at the event \mathcal{E}_r that $\|\mathbf{Ax} - \mathbf{y}\| = O(r\kappa/D)$ for some vector $\mathbf{y} = (y_1, \dots, y_n)$ of norm $\|\mathbf{y}\| = r$ in the net $r \cdot \mathcal{N}_D$ (with \mathcal{N}_D obtained from Theorem 3.3). Clearly this covers the event that $\mathbf{LCD}(\mathbf{Ax}/\|\mathbf{Ax}\|) \leq D$ and $r - n^{-c} \leq \|\mathbf{Ax}\| \leq r + n^{-c}$.

Again, as $\mathbf{LCD}(\mathbf{x}) \geq D$ and that $r\kappa \geq n^{1/2+3c/4} > n^{1/2}$, by Theorem 3.2

$$\mathbf{P}(|a_{i1}x_1 + \dots + a_{in}x_n - y_i| \leq \frac{r\kappa}{D\sqrt{n}}) = O(\frac{r\kappa}{D\sqrt{n}}).$$

Thus by Lemma 3.6,

$$\mathbf{P}(\|\mathbf{Ax} - \mathbf{y}\|^2 \leq \frac{r^2\kappa^2}{D^2}) \leq (\frac{Cr\kappa}{D\sqrt{n}})^n.$$

We have thus obtained (taking into account of the size of \mathcal{N}_D from Theorem 3.3)

$$\begin{aligned} \mathbf{P}\left(\exists \mathbf{y} \in r \cdot \mathcal{N}_D, \|\mathbf{Ax} - \mathbf{y}\| \leq O(r\kappa/D), \mathbf{LCD}(\mathbf{Ax}/\|\mathbf{Ax}\|) \leq D\right) &\leq |\mathcal{N}_D| \left(\frac{Cr\kappa}{D\sqrt{n}}\right)^n \leq D^2 \left(\frac{Cr\kappa}{n}\right)^n \\ &\leq n^{-cn}, \end{aligned}$$

where we used the assumption $D = \exp(n^c)$ and $r \leq n^{1-2c}$.

Let \mathcal{E}_3 be the event that $\mathbf{LCD}(\mathbf{Ax}/\|\mathbf{Ax}\|) \leq D$ and $n^{1/2-c/4} \leq \|\mathbf{Ax}\| \leq n^{1-2c}$. Then by taking any $\kappa/D^{O(1)}$ -net $\{r_1, \dots, r_m\}$ of the segment $n^{1/2-c/4} \leq r \leq n^{1-2c}$ we have

$$\begin{aligned} \mathbf{P}(\mathcal{E}_3) &\leq \mathbf{P}\left(\exists i, \exists \mathbf{y} \in r_i \cdot \mathcal{N}_D, \|\mathbf{Ax} - \mathbf{y}\| \leq O(r_i\kappa/D), \mathbf{LCD}(\mathbf{Ax}/\|\mathbf{Ax}\|) \leq D\right) \\ &\leq ((n^{1-2c} - n^{1/2-c/4})D/\kappa)n^{-cn} \\ &\leq n^{-cn/2}. \end{aligned} \quad (12)$$

The proof is complete by (10), (11) and (12). \square

4. STEP 2 FOR (1) OF THEOREM 1.12: CONCENTRATION OF MAGNITUDE OVER
NON-STRUCTURED VECTORS

Recall that

$$\frac{1}{N} \log \|B_N \mathbf{x}_0\| = \frac{1}{N} \sum_{i=0}^{N-1} \log \|A_{i+1} \mathbf{x}_i\|,$$

where

$$\mathbf{x}_i = \frac{A_i \dots A_2 A_1 \mathbf{x}_0}{\|A_i \dots A_2 A_1 \mathbf{x}_0\|}.$$

By Theorem 3.7, we can assume that $\mathbf{x}_i \in \mathcal{S}$ (i.e. $\mathbf{LCD}_{\gamma, \kappa}(\mathbf{x}_i) \geq D$) for all $1 \leq i \leq N$ with a loss of $N \exp(-cn/8)$ in probability.

In this section we study the concentration of $\log \|A_{i+1} \mathbf{x}_i\|$ around its zero mean (here again the randomness is with respect to A_{i+1} , conditioning on all A_1, \dots, A_i .)

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a vector in \mathcal{S} . Let $A = (a_{ij})_{1 \leq i \leq n}$ be a random square matrix whose entries are iid copies of ξ from (2). For short, set $\xi_i := a_{i1}x_1 + \dots + a_{in}x_n$ and

$$y := \log\left(\frac{1}{n} \|A\mathbf{x}\|_2^2\right) = \log\left(\frac{1}{n} (\xi_1^2 + \dots + \xi_n^2)\right).$$

Before stating our estimates, we note that as a_{ij} are subgaussian random variables of parameter K , so are the normalized random variables ξ_1, \dots, ξ_n . This implies that ξ_i^2 are exponential random variables (since $\mathbf{P}(\xi_i^2 \geq t) = \mathbf{P}(|\xi_i| \geq \sqrt{t}) \leq O(\exp(-t/K))$). As a consequence, for any $x \geq 0$ (see for instance [29, Proposition 5.16])

$$\mathbf{P}(|\xi_1^2 + \dots + \xi_n^2 - n| \geq nx) \leq 2e^{-c \min\{nx^2/K^2, nx/K\}}. \quad (13)$$

Theorem 4.1 (concentration over non-structured vectors). *We have*

(i) for any $t > 0$,

$$\mathbf{P}(y \geq t) \leq e^{-c't^2n}$$

for some absolute constant $c' > 0$;

(ii) for any $0 \leq t \leq 2 \log D$

$$\mathbf{P}(y \leq -t) \leq \min\{(Ke^{-t/2})^n, 1\}$$

where K is the parameter from (2);

(iii) for any $t \leq O(1)$

$$\mathbf{P}(|y| \geq t) \leq e^{-c''t^2n}$$

for some absolute constant $c'' > 0$.

Proof. (of Theorem 4.1) For (i), with the parameter $\lambda = ctn/2K^2 + 1$

$$\mathbf{P}(y \geq t) = \mathbf{P}(e^{\lambda y} \geq e^{\lambda t}) \leq e^{-\lambda t} \mathbf{E}((\frac{1}{n}(\xi_1^2 + \dots + \xi_n^2))^\lambda) = e^{-\lambda t} \lambda \int_0^\infty x^{\lambda-1} \mathbf{P}(z > x) dx, \quad (14)$$

where $z := \frac{1}{n}(\xi_1^2 + \dots + \xi_n^2)$. Note that we can trivially bound

$$\int_0^1 x^{\lambda-1} \mathbf{P}(z > x) dx \leq 1/\lambda. \quad (15)$$

For the integral corresponding to $x \geq 1$ we use (13)

$$\int_1^\infty x^{\lambda-1} \mathbf{P}(z > x) dx = \int_0^\infty (1+x)^{\lambda-1} \mathbf{P}(z > x+1) dx \leq 2 \int_0^\infty (1+x)^{\lambda-1} e^{-c \min\{nx^2/K^2, nx/K\}} dx.$$

To this end,

$$\int_0^1 (1+x)^{\lambda-1} e^{-\frac{c}{K^2} nx^2} dx \leq \int_0^1 e^{-\frac{c}{K^2} nx^2 + (\lambda-1)x} dx \leq 1 + e^{K^2(\lambda-1)^2/cn} \leq e^{\lambda t/2}. \quad (16)$$

Furthermore, for $t < 2$ we have

$$\int_1^\infty (1+x)^{\lambda-1} e^{-\frac{c}{K^2} nx} dx \leq \int_1^\infty e^{x \frac{c}{2K^2} n(t-2)} dx = O(1). \quad (17)$$

For $t \geq 2$, let x' be such that $(\lambda-1)/n = x'/\log(1+x')$ (thus $x' \asymp t \log t$) then

$$\int_1^\infty (1+x)^{\lambda-1} e^{-\frac{c}{K^2} nx} dx \leq \int_1^{x'} (1+x)^{\lambda-1} dx + 1 \leq e^{\lambda t/2}. \quad (18)$$

Combining (14),(15),(16), (17) and (18) we obtain

$$\mathbf{P}(y \geq t) \leq e^{-\lambda t/2} \leq e^{-c't^2n}.$$

For the lower tail (ii), we recall that $\mathbf{LCD}_{\gamma,\kappa}(\mathbf{x}) \geq D$. By Lemma 3.6, for any t such that $e^{-t/2} \geq 1/D$

$$\mathbf{P}(y \leq -t) = \mathbf{P}(e^y \leq e^{-t}) = \mathbf{P}(\xi_1^2 + \dots + \xi_n^2 \leq ne^{-t}) \leq (Ke^{-t/2})^n.$$

For (iii), we need to estimate $\mathbf{P}(y \leq -t)$ with $0 \leq t = O(1)$. We have

$$\begin{aligned} \mathbf{P}(y \leq -t) &= \mathbf{P}(z \leq e^{-t}) \leq \mathbf{P}(z \leq 1 - \min\{t, 1\}/2) \leq \mathbf{P}(|z - 1| \geq \min\{t, 1\}/2) \\ &\leq 2e^{-\frac{c''}{K^2}nt^2} \end{aligned}$$

where we used (13) in the last estimate. □

5. STEP 3 FOR (1) OF THEOREM 1.12: CONCLUDING THE PROOF

Let \mathbf{x}_0 be any vector from \mathcal{N}_{start} . We will show

Lemma 5.1. *For any $t \geq 1/n$ we have*

$$\mathbf{P}\left(\left|\frac{1}{N} \log \|B_N \mathbf{x}_0\|\right| \geq t\right) \leq \exp(-c \min\{t^2, t\} N n^{1-2c}) + N n^{-cn}.$$

It is clear that Theorem 1.12 follows from Lemma 5.1 after taking union bound over \mathcal{N}_{start} .

Proof. (of Lemma 5.1) First, by Theorem 3.7, the event \mathcal{F}_1 that $\mathbf{x}_i \in \mathcal{S}$ for all $1 \leq i \leq N$ holds with probability

$$\mathbf{P}(\mathcal{F}_1) \geq 1 - N n^{-cn}.$$

Consider the random sum

$$S = \frac{1}{N}(y_1 + \cdots + y_N),$$

where $y_i = \log \|A_{i+1} \mathbf{x}_i\|$.

Basing on Theorem 4.1, the event \mathcal{F}_2 such that $|y_i| \leq 2 \log D$ for all $y_i, 1 \leq i \leq N$ satisfies

$$\mathbf{P}(\mathcal{F}_2) \geq 1 - N D^{-n}.$$

Introduce the new random variables $y'_i := y_i 1_{|y_i| \leq 2 \log D}$ and $y''_i := y'_i - \mathbf{E}_{A_i} y'_i$. As customary, in what follows our probability is with respect to A_i , conditioning on A_1, \dots, A_{i-1} . By (iii) of Theorem 4.1, for $|t| = O(1)$

$$\mathbf{P}(|y'_i| \geq t) \leq \mathbf{P}(|y_i| \geq t) \leq e^{-ct^2n}. \tag{19}$$

Also, by (i) and (ii) of Theorem 4.1, for $O(1) \leq t \leq 2 \log D$

$$\mathbf{P}(|y'_i| \geq t) \leq \mathbf{P}(|y_i| \geq t) \leq K^n e^{-tn/2} + e^{-ct^2n}. \quad (20)$$

Consequently,

$$|\mathbf{E}_{A_i} y'_i| \leq \int_0^{2 \log D} t \mathbf{P}(|y'_i| \geq t) \leq O\left(\int_0^{1/\sqrt{n}} t dt\right) = O(1/n).$$

Next consider the martingale sum

$$S'' := \frac{1}{N}(y''_1 + \cdots + y''_N).$$

By definition, $|y''_i| \leq 2 \log D + O(1/n) < 3 \log D$. Also by (19) and (20), for $t \geq 1/n$

$$\mathbf{P}(|y''_i| \geq t) \leq \mathbf{P}(|y'_i| \geq t) \leq \exp(-c \min\{t^2, t\}n).$$

This implies the following conditional estimate for $\lambda = ctn$

$$e^{-2\lambda t} \mathbf{E}(e^{\lambda y''_i} | A_1, \dots, A_{i-1}), e^{-2\lambda t} \mathbf{E}(e^{-\lambda y''_i} | A_1, \dots, A_{i-1}) \leq \exp(-c \min\{t, t^2\}n).$$

Following the proof of Azuma's martingale concentration, for $t \geq 1/n$

$$\begin{aligned} \mathbf{P}(S'' \geq 2t) &= \mathbf{P}(y''_1 + \cdots + y''_N \geq 2Nt) = \mathbf{P}(\exp((y''_1 + \cdots + y''_N)\lambda) \geq \exp(2\lambda Nt)) \\ &\leq \exp(-2\lambda Nt) \mathbf{E}(\exp((y''_1 + \cdots + y''_N)\lambda)) \\ &\leq \exp(-c' \min\{t^2, t\}Nn). \end{aligned}$$

We also obtain the same bound for $\mathbf{P}(S'' \leq -2t)$. Thus, as $\mathbf{P}(|S| \geq 2t) \leq \mathbf{P}(|S| \geq 2t | \mathcal{F}_1 \cap \mathcal{F}_2) \mathbf{P}(\mathcal{F}_1 \cap \mathcal{F}_2) + \mathbf{P}(\bar{\mathcal{F}}_1 \cup \bar{\mathcal{F}}_2)$, we have

$$\begin{aligned} \mathbf{P}(|S| \geq 2t + O(1/n)) &\leq \mathbf{P}(|S''| \geq 2t) + \mathbf{P}(\bar{\mathcal{F}}_1 \cup \bar{\mathcal{F}}_2) \leq \exp(-c \min\{t^2, t\}Nn) + \mathbf{P}(\bar{\mathcal{F}}_1) + \mathbf{P}(\bar{\mathcal{F}}_2) \\ &\leq \exp(-c \min\{t^2, t\}Nn) + Nn^{-cn}. \end{aligned}$$

□

6. PROOF OF (2) OF THEOREM 1.12 WITH THE MODIFIED THREE-STEP PLAN

Although our treatment here is analogous to that of the top exponent, the argument is far more complicated as we have to take care of the angles of the pair vectors.

We will first introduce some additional structures. The definition of LCD can be naturally extended to joint structure of two vectors. Let γ, κ be given parameters and let \mathbf{x}, \mathbf{y} be two vectors. Define

$$\mathbf{LCD}_{\gamma, \kappa}(\mathbf{x}, \mathbf{y}) := \inf_{\theta_1^2 + \theta_2^2 = 1} \mathbf{LCD}_{\gamma, \kappa}(\theta_1 \mathbf{x} + \theta_2 \mathbf{y}).$$

For the rest of this section we will choose γ to be a sufficiently small and $\kappa = n^c$ for some constant $c < 1/16$.

6.1. Step 1. Set $D = \exp(n^c)$. First of all we will have to choose $\mathcal{P}_{start} \subset S^{n-1} \times S^{n-1}$ to satisfy the following conditions

- for all $(\mathbf{x}', \mathbf{y}') \in \mathcal{P}_{start}$ we have $\|\mathbf{x}' \wedge \mathbf{y}'\| \geq \varepsilon$,
- for all $(\mathbf{x}', \mathbf{y}') \in \mathcal{P}_{start}$ we have

$$\mathbf{LCD}_{\kappa, \gamma}(\mathbf{x}' / \|\mathbf{x}' \wedge \mathbf{y}'\|, \mathbf{y}' / \|\mathbf{x}' \wedge \mathbf{y}'\|) \geq D, \quad (21)$$

- for any $(\mathbf{x}, \mathbf{y}) \in S^{n-1} \times S^{n-1}$ there exists $(\mathbf{x}', \mathbf{y}') \in \mathcal{P}_{start}$ such that $\|\mathbf{x} - \mathbf{x}'\|, \|\mathbf{y} - \mathbf{y}'\| \leq \varepsilon$.

Remark that a direct choice of $\mathcal{N}_{start} \times \mathcal{N}_{start}$ (with \mathcal{N}_{start} from Section 3) would not work because there were no information on the joint structure.

Lemma 6.2. *There exists a set \mathcal{P}_{start} with the above properties.*

Proof. (of Lemma 6.2) In what follows we will be focusing on the set $\mathcal{S}_{separate}$ of pairs of unit vectors $\mathbf{x}, \mathbf{y} \in S^{n-1}$ with $\|\mathbf{x} \wedge \mathbf{y}\| \geq \varepsilon$.

Assume that $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}_{separate}$ which violates (21). In other words there exist $\theta_1^2 + \theta_2^2 = 1$ with

$$\mathbf{LCD}\left(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \theta_1 \mathbf{x} + \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \theta_2 \mathbf{y}\right) \leq D. \quad (22)$$

In the next step we $1/D^{O(1)}$ -approximate the parameters $\theta_1, \theta_2, \|\mathbf{x} \wedge \mathbf{y}\|$ by numbers of form $k \frac{1}{D}, k \in \mathbf{Z}$. Thus by losing a factor of $D^{O(1)}$ in probability at most, by using Fact 3.1 and that $\|\mathbf{x} \wedge \mathbf{y}\| \geq \varepsilon$, we can treat $\theta_1, \theta_2, \|\mathbf{x} \wedge \mathbf{y}\|$ as constants. Furthermore, by changing the vector direction if needed, without loss of generality we can assume $\theta_2 \geq \theta_1 > 0$.

By Theorem 3.3, we thus have three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ where $\mathbf{z} = t_1 \mathbf{x} + t_2 \mathbf{y}$ with $\mathbf{z} \in k \cdot \mathcal{N}_D$ and $t_1^2 + t_2^2 \gg 1$ as well as $t_2 \geq t_1 \geq 0$.

Solving for \mathbf{y} ,

$$\mathbf{y} = \frac{1}{t_2} \mathbf{z} - \frac{t_1}{t_2} \mathbf{x}.$$

We conclude that there exists an absolute constant C such that for any given $\mathbf{x} \in S^{n-1}$, the vectors $\mathbf{y} \in S^{n-1}$ for which (22) holds belong to a set $\mathcal{S}_{\mathbf{x}}$ of volume at most

$$\text{Vol}(\mathcal{S}_{\mathbf{x}}) \leq \text{Vol}_{n-1}(B(0, C\kappa/D)) \times |\mathcal{N}_D| \times D^{O(1)}.$$

where the first two factors come from the approximation of \mathbf{z} and the magnifying factor t_1/t_2 , while the third factor comes from approximations of the parameters $\theta_1, \theta_2, \|\mathbf{x} \wedge \mathbf{y}\|$ by numbers of the form $k\frac{1}{D}, k \in \mathbf{Z}$ as above.

Varying $\mathbf{x} \in S^{n-1}$, the total volume Vol_T of such a pair $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}_{\text{separate}}$ satisfying (22) is at most

$$\begin{aligned} \text{Vol}_T &\leq \text{Vol}_{n-1}(S^{n-1}) \text{Vol}_{n-1}(\mathbf{B}(0, C\kappa/D)) \times |\mathcal{N}_D| \times D^{O(1)} \\ &\leq \text{Vol}_{n-1}(S^{n-1}) \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} (C\kappa/D)^{n-1} \times (C_0 D/\sqrt{n})^n \log_2 D \times D^{O(1)} \\ &\leq \text{Vol}_{n-1}(S^{n-1}) \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} (CC_0\kappa/\sqrt{n})^n D^{O(1)} \\ &\leq \text{Vol}_{n-1}(S^{n-1}) \times \text{Vol}_{n-1}(S^{n-1}) \times (CC_0\kappa/\sqrt{n})^n D^{O(1)}. \end{aligned} \quad (23)$$

Next, notice that the total volume of an $\varepsilon/2$ -neighborhood of any point on $S^{n-1} \times S^{n-1}$ is at least

$$\begin{aligned} V_{\varepsilon/2} &\geq \text{Vol}_{n-1}(\mathbf{B}(0, \varepsilon/2) \cap S^{n-1}) \times \text{Vol}_{n-1}(\mathbf{B}(0, \varepsilon/2) \cap S^{n-1}) \\ &\geq \text{Vol}_{n-1}(S^{n-1}) \times \text{Vol}_{n-1}(S^{n-1}) \times (\varepsilon/2C)^{2n}. \end{aligned} \quad (24)$$

Thus $V_{\varepsilon/2} > \text{Vol}_T$ if we choose $\kappa = n^c$ for some $c < 1/16$.

It follows that for any $\varepsilon/2$ -neighborhood of any point on $S^{n-1} \times S^{n-1}$, there exists a point $(\mathbf{x}, \mathbf{y}) \in \mathcal{S}_{\text{separate}}$ such that (21) holds. The proof of Lemma 6.2 is then complete by considering a maximal $\varepsilon/4$ -packing of $\mathcal{S}_{\text{separate}}$. \square

Let $\mathcal{P} := \mathcal{P}_D$ be the collection of vector pairs (\mathbf{x}, \mathbf{y}) in $\mathbf{R}^n \times \mathbf{R}^n$ such that

$$\text{LCD}_{\kappa, \gamma} \left(\frac{\mathbf{x}}{\|\mathbf{x} \wedge \mathbf{y}\|}, \frac{\mathbf{y}}{\|\mathbf{x} \wedge \mathbf{y}\|} \right) \geq D.$$

Note that here the vectors of \mathcal{P} do not need to be unit, and by definition $\mathcal{P}_{\text{start}} \subset \mathcal{P} \cap S^{n-1} \times S^{n-1}$. Our next key result is an analog of Theorem 3.7 for joint-structures.

Theorem 6.3 (stability of non-structures, jointly). *Assume that $A = (a_{ij})_{1 \leq i, j \leq n}$ is a random matrix of size n whose entries are iid copies of ξ satisfying (2). Let (\mathbf{x}, \mathbf{y}) be any deterministic vector pair from \mathcal{P} . Then*

$$\mathbf{P}_A \left(\left(\frac{A\mathbf{x}}{\|A\mathbf{x}\|}, \frac{A\mathbf{y}}{\|A\mathbf{y}\|} \right) \notin \mathcal{P} \right) \leq n^{-cn/8}.$$

By Theorem 6.8 (to be proved separately later), with a probability at least $1 - n^{-cn/16}$ we can assume that

$$\|A\mathbf{x}/\|A\mathbf{x}\| \wedge A\mathbf{y}/\|A\mathbf{y}\|\| \leq n^{c/16} \|\mathbf{x} \wedge \mathbf{y}\|.$$

Thus by definition of LCD (see Fact 3.1), for Theorem 6.3 it suffices to show

$$\mathbf{P}_A \left(\mathbf{LCD}_{\gamma, \kappa} \left(\frac{A\mathbf{x}/\|A\mathbf{x}\|}{\|\mathbf{x} \wedge \mathbf{y}\|}, \frac{A\mathbf{y}/\|A\mathbf{y}\|}{\|\mathbf{x} \wedge \mathbf{y}\|} \right) \leq D' \right) \leq n^{-cn/8}, \quad (25)$$

where $D' = Dn^{c/16}$.

Proof. (of Theorem 6.3) By the proof of Theorem 3.7, it suffices to focus on the event \mathcal{E}_1 that $\|A\mathbf{x}\|^2, \|A\mathbf{y}\|^2 \geq n^{1-c/2}$.

For each $n^{1/2-c/4} \leq r, s \leq n^{1-2c}$ (which can be approximated by integral points of the form $r_i = in^{-c}, i \in \mathbf{Z}$) let us look at the event \mathcal{E}_{r_i, s_j} that $r_i \leq \|A\mathbf{x}\| \leq r_{i+1}$ and $s_j \leq \|A\mathbf{y}\| \leq s_{j+1}$ and such that

$$\mathbf{LCD}_{\gamma, \kappa} \left(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} A\mathbf{x}/r_i, \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} A\mathbf{y}/s_j \right) \leq (1 + o(1))D'. \quad (26)$$

(Note that by Fact 3.1 $\mathbf{LCD}_{\gamma, \kappa}(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} A\mathbf{x}/r, \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} A\mathbf{y}/s)$ are comparable in the range $r_i \leq r \leq r_{i+1}, s_j \leq s \leq s_{j+1}$.)

In other words, by definition of joint LCD there exist θ_1, θ_2 with $\theta_1^2 + \theta_2^2 = 1$ such that

$$\mathbf{LCD}_{\gamma, \kappa} \left(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \theta_1 A\mathbf{x}/r_i + \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \theta_2 A\mathbf{y}/s_j \right) \leq (1 + o(1))D'. \quad (27)$$

We can write

$$\theta_1 A\mathbf{x}/r_i + \theta_2 A\mathbf{y}/s_j = \theta A(\theta'_1 \mathbf{x} + \theta'_2 \mathbf{y}).$$

Here

$$\theta = \sqrt{(\theta_1/r_i)^2 + (\theta_2/s_j)^2} \text{ and } \theta'_1 = \frac{\theta_1/r_i}{\theta}, \theta'_2 = \frac{\theta_2/s_j}{\theta}.$$

Because θ is not too small ($\theta \approx n^{1/2}$), we can again assume it to have the form $iD^{-O(1)}$ and relax the constrain $\theta_1'^2 + \theta_2'^2 = 1$ to $1 - D^{-O(1)} \leq \theta_1'^2 + \theta_2'^2 \leq 1 + D^{-O(1)}$.

Now we look at the event $\mathbf{LCD}_{\gamma, \kappa}(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \theta A(\theta'_1 \mathbf{x} + \theta'_2 \mathbf{y})) \leq (1 + o(1))D'$ from (27). By Theorem 3.3, there exists $\mathbf{u} \in \theta^{-1} \cdot \mathcal{N}_{D'}$ such that

$$\left\| \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} A(\theta'_1 \mathbf{x} + \theta'_2 \mathbf{y}) - \mathbf{u} \right\| \leq O(\theta^{-1} \kappa / D'). \quad (28)$$

By passing to numbers of the form $i(\sqrt{n}D')^{-1}, i \in \mathbf{Z}$, up to a multiplicative factor in the RHS of (28), we can assume θ'_1 and θ'_2 to be fixed so that we can take a union bound over the set of these integral points, which obviously has cardinality $D^{O(1)}$. In other words, by passing to those approximated points, with a loss of a factor of $n^{O(1)}D^{O(1)}$ in probability we will be arriving at (28) with fixed θ'_1, θ'_2 and $\mathbf{x}, \mathbf{y}, \mathbf{u}$.

Now we analyze the probability of the event from (28) by invoking the argument from the proof of Theorem 3.7. As $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$, we have $\mathbf{LCD}(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \mathbf{x}, \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \mathbf{y}) \geq D$, henceforth

$$\mathbf{LCD}\left(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \theta'_1 \mathbf{x} + \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \theta'_2 \mathbf{y}\right) \geq D.$$

Note that $\theta^{-1} \kappa \geq n^{1/2+3c/4} > n^{1/2}$. As $\|\theta'_1 \mathbf{x} + \theta'_2 \mathbf{y}\| = \Omega(\|\mathbf{x} \wedge \mathbf{y}\|)$ for any θ'_1, θ'_2 with $\theta_1'^2 + \theta_2'^2 = 1 + o(1)$, by Theorem 3.2

$$\mathbf{P}\left(|a_{i1}z_1 + \dots + a_{in}z_n - u_i| \leq \frac{r\kappa}{D\sqrt{n}}\right) = O\left(\frac{r\kappa}{D\sqrt{n}}\right),$$

where for short we set $\mathbf{z} = (z_1, \dots, z_n) := \theta'_1 \mathbf{x} + \theta'_2 \mathbf{y}$.

Thus by Lemma 3.6, for any fixed \mathbf{u}

$$\mathbf{P}\left(\left\| \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} A\mathbf{z} - \mathbf{u} \right\|^2 \leq \frac{r^2 \kappa^2}{D^2}\right) \leq \left(\frac{Cr\kappa}{D\sqrt{n}}\right)^n.$$

We have thus obtained (taking into account of the size of $\mathcal{N}_{D'}$ from Theorem 3.3)

$$\begin{aligned} & \mathbf{P}\left(\exists \mathbf{u} \in \theta^{-1} \cdot \mathcal{N}_D, \left\| \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \mathbf{A}\mathbf{z} - \mathbf{u} \right\| \leq O(r\kappa/D), \mathbf{LCD}\left(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|} \mathbf{A}\mathbf{z}\right) \leq (1 + o(1))D\right) \\ & \leq |\mathcal{N}_{D'}| \left(\frac{Cr\kappa}{D\sqrt{n}}\right)^n \leq n^{-cn}. \end{aligned}$$

To complete our proof, we take the union bound over all the choices of $r_i, s_j, \theta, \theta'_1, \theta'_2$ to obtain a bound $n^{-cn} D^{O(1)} \leq n^{-cn/2}$, completing the proof of (25). \square

6.4. Step 2. Now we turn to the second step of the plan to control $\frac{\|\mathbf{A}\mathbf{x}/\|\mathbf{A}\mathbf{x}\| \wedge \mathbf{A}\mathbf{y}/\|\mathbf{A}\mathbf{y}\|\|}{\|\mathbf{x} \wedge \mathbf{y}\|}$ for given $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$. Note that if $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$, then

$$\begin{aligned} \|\mathbf{u} \wedge \mathbf{v}\|^2 &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \langle \mathbf{u}, \mathbf{v} \rangle^2 = 1 - \langle \mathbf{u}, \mathbf{v} \rangle^2 = 1 - (1 - \|\mathbf{u} - \mathbf{v}\|^2/2)^2 \\ &= \|\mathbf{u} - \mathbf{v}\|^2 - \left(\frac{\|\mathbf{u} - \mathbf{v}\|^2}{2}\right)^2. \end{aligned} \quad (29)$$

We will be finding the following fact useful.

Fact 6.5. *Let $f(x) = x - x^2/4, 0 \leq x \leq 1$, and let $0 < x_1, x_2 < 1$. Let $t > 0$ be a parameter.*

- *If $f(x_2)/f(x_1) \geq 1 + t$ then $x_2/x_1 \geq 1 + t$.*
- *If $f(x_2)/f(x_1) \leq t < 1$ then $x_2/x_1 \leq t$.*

To bound $\frac{\|\mathbf{A}\mathbf{x}/\|\mathbf{A}\mathbf{x}\| \wedge \mathbf{A}\mathbf{y}/\|\mathbf{A}\mathbf{y}\|\|}{\|\mathbf{x} \wedge \mathbf{y}\|}$ from below, we invoke the following result.

Theorem 6.6. *For any $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$ and any $\delta \geq 1/D$,*

$$\mathbf{P}_A\left(\frac{\|\mathbf{A}\mathbf{x}/\|\mathbf{A}\mathbf{x}\| \wedge \mathbf{A}\mathbf{y}/\|\mathbf{A}\mathbf{y}\|\|}{\|\mathbf{x} \wedge \mathbf{y}\|} \leq \delta^2\right) \leq (C\delta)^n.$$

As by Theorem 4.1 $\|\mathbf{A}\mathbf{x}\|, \|\mathbf{A}\mathbf{y}\| \geq \delta\sqrt{n}$ with probability at least $1 - (C\delta)^n$, we obtain

Corollary 6.7.

$$\mathbf{P}_A\left(\frac{\|\mathbf{A}\mathbf{x} \wedge \mathbf{A}\mathbf{y}\|}{\|\mathbf{x} \wedge \mathbf{y}\|} \leq n\delta^4\right) \leq 2(C\delta)^n.$$

Proof. (of Theorem 6.6) By (29), the assumption $\|\frac{\mathbf{A}\mathbf{x}}{\|\mathbf{A}\mathbf{x}\|} \wedge \frac{\mathbf{A}\mathbf{y}}{\|\mathbf{A}\mathbf{y}\|}\| \leq \delta^2 \|\mathbf{x} \wedge \mathbf{y}\|$ implies that

$$\left\| \mathbf{A}\left(\frac{\mathbf{x}}{\|\mathbf{A}\mathbf{x}\|} - \frac{\mathbf{y}}{\|\mathbf{A}\mathbf{y}\|}\right) \right\| \ll \delta^2 \|\mathbf{x} \wedge \mathbf{y}\|.$$

Again by Theorem 4.1, with probability at least $1 - (C\delta)^n$ we can assume that

$$C\delta\sqrt{n} \leq \|\mathbf{A}\mathbf{x}\|, \|\mathbf{A}\mathbf{y}\| \leq \delta^{-1}\sqrt{n}.$$

Thus the event $\|A(\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|})\| \leq \delta^2 \|\mathbf{x} \wedge \mathbf{y}\|$ implies that $\|\frac{1}{\sqrt{n}}A(\alpha_1\mathbf{x} + \alpha_2\mathbf{y})\| \ll \delta \|\mathbf{x} \wedge \mathbf{y}\|$ with some coefficients α_1, α_2 satisfying $\alpha_1^2 + \alpha_2^2 = 1$.

In what follows we will consider this event. Notice that

$$\|\alpha_1\mathbf{x} + \alpha_2\mathbf{y}\|^2 \geq 1 - |\langle \mathbf{x}, \mathbf{y} \rangle| \geq \frac{1}{2}(1 - |\langle \mathbf{x}, \mathbf{y} \rangle|^2) \geq \frac{1}{2}\|\mathbf{x} \wedge \mathbf{y}\|^2. \quad (30)$$

Now we pass to consider a $1/D^{O(1)}$ -net \mathcal{M} with respect to (θ_1, θ_2) over the ellipsoid $\theta_1\mathbf{x}/\|\mathbf{x} \wedge \mathbf{y}\| + \theta_2\mathbf{y}/\|\mathbf{x} \wedge \mathbf{y}\|, \theta_1^2 + \theta_2^2 = 1$. As this set is one-dimensional, one can take $|\mathcal{M}| = D^{O(1)}$.

Because we can assume that $\|A\| = O(\sqrt{n})$,

$$\inf_{\alpha_1^2 + \alpha_2^2 = 1} \left\| \frac{1}{\sqrt{n}}A(\alpha_1\mathbf{x}/\|\mathbf{x} \wedge \mathbf{y}\| + \alpha_2\mathbf{y}/\|\mathbf{x} \wedge \mathbf{y}\|) \right\| \geq \inf_{\mathbf{u} \in \mathcal{M}} \left\| \frac{1}{\sqrt{n}}A\mathbf{u} \right\| - O(D^{-O(1)}).$$

But $\delta \geq 1/D$, we conclude that the event $\|A(\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|})\| \leq \delta^2 \|\mathbf{x} \wedge \mathbf{y}\|$ implies the event \mathcal{E} where

$$\mathcal{E} := \left\{ \inf_{\mathbf{u} \in \mathcal{M}} \left\| \frac{1}{\sqrt{n}}A\mathbf{u} \right\| \leq 2\delta \right\}.$$

To estimate this event, choose any point \mathbf{u} from \mathcal{M} . As $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$, we have $\mathbf{LCD}(\mathbf{x}/\|\mathbf{x} \wedge \mathbf{y}\|, \mathbf{y}/\|\mathbf{x} \wedge \mathbf{y}\|) \geq D$, and so we also have

$$\mathbf{LCD}(\mathbf{u}) \geq D.$$

By Theorem 4.1 and by (30), as $\|\mathbf{u}\| \geq 1/\sqrt{2}$, as long as $\delta \gg 1/D$ we have

$$\mathbf{P}(\left\| \frac{1}{\sqrt{n}}A\mathbf{u} \right\| \leq (C\delta)^n).$$

Thus

$$\mathbf{P}(\mathcal{E}) \leq D^{O(1)}(C\delta)^n \leq (C'\delta)^n. \quad \square$$

In our next theorem we give an analog of Theorem 4.1.

Theorem 6.8. *There exist constants C, c such that for any $t > 0$ and any $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$ we have*

(i)

$$\mathbf{P}_A\left(\frac{\|\mathbf{Ax}/\|\mathbf{Ax}\| \wedge \mathbf{Ay}/\|\mathbf{Ay}\|\|}{\|\mathbf{x} \wedge \mathbf{y}\|} \geq C\right) \leq \exp(-cn).$$

(ii)

$$\mathbf{P}_A\left(\frac{\|A\mathbf{x} \wedge A\mathbf{y}\|}{\|\mathbf{x} \wedge \mathbf{y}\|} \geq ne^t\right) \leq e^{-ct^2n}.$$

(iii) Furthermore, if $t = o(\log n)$ then one also has

$$\mathbf{P}_A\left(\frac{\|A\mathbf{x} \wedge A\mathbf{y}\|}{\|\mathbf{x} \wedge \mathbf{y}\|} \leq ne^{-t}\right) = O(K^n e^{-tn/2} + e^{-c''t^2n}).$$

Proof. (of Theorem 6.8) First we prove (i). By Fact 6.5 the assumption $\|\frac{A\mathbf{x}}{\|A\mathbf{x}\|} \wedge \frac{A\mathbf{y}}{\|A\mathbf{y}\|}\| \geq C\|\mathbf{x} \wedge \mathbf{y}\|$ implies that

$$\|A\left(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\right)\| \geq C\|\mathbf{x} - \mathbf{y}\|.$$

Note that

$$\frac{\|A\left(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\right)\|}{\|\mathbf{x} - \mathbf{y}\|} = \frac{\|A\left(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\right)\|}{\left\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\right\|} \frac{\left\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\right\|}{\|\mathbf{x} - \mathbf{y}\|}.$$

Here, by (i) and (ii) of Theorem 4.1 the following holds with probability at least $1 - \exp(-cn)$

$$\begin{aligned} \frac{\left\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\right\|}{\|\mathbf{x} - \mathbf{y}\|} &\leq \frac{\left\|\frac{\mathbf{x}-\mathbf{y}}{\|A\mathbf{x}\|} + \mathbf{y}(1/\|A\mathbf{x}\| - 1/\|A\mathbf{y}\|)\right\|}{\|\mathbf{x} - \mathbf{y}\|} \\ &\leq \frac{1}{\|A\mathbf{x}\|} + \frac{\|A\mathbf{x}\| - \|A\mathbf{y}\|}{\|A\mathbf{x}\|\|A\mathbf{y}\|\|\mathbf{x} - \mathbf{y}\|} \\ &\leq \frac{1}{\|A\mathbf{x}\|} + \frac{\|A(\mathbf{x} - \mathbf{y})\|}{\|A\mathbf{x}\|\|A\mathbf{y}\|\|\mathbf{x} - \mathbf{y}\|} \\ &\leq \frac{1}{\|A\mathbf{x}\|} + \frac{\|A\|}{\|A\mathbf{x}\|\|A\mathbf{y}\|} \leq \frac{C'}{\sqrt{n}}. \end{aligned}$$

Thus the event $\|A\left(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\right)\| \geq C\|\mathbf{x} - \mathbf{y}\|$ implies that there exists \mathbf{z} such that $\|A\mathbf{z}\|/\|\mathbf{z}\| \geq C/C'$, and this holds with probability $\exp(-cn)$ if C/C' is sufficiently large.

Now we prove (ii). By changing the size of \mathbf{x} or \mathbf{y} when necessary, without loss of generality we assume

$$-1 < \langle \mathbf{x}, \mathbf{y} \rangle \leq 0.$$

We write

$$\frac{\|A\mathbf{x} \wedge A\mathbf{y}\|}{\|\mathbf{x} \wedge \mathbf{y}\|} = \|A\mathbf{x}\| \|A\mathbf{y}\| \frac{\|A\mathbf{x}/\|A\mathbf{x}\| \wedge A\mathbf{y}/\|A\mathbf{y}\|\|}{\|\mathbf{x} \wedge \mathbf{y}\|}.$$

Thus our assumption implies that

$$\|A\mathbf{x}/\|A\mathbf{x}\| \wedge A\mathbf{y}/\|A\mathbf{y}\|\| \geq \frac{ne^t}{\|A\mathbf{x}\| \|A\mathbf{y}\|} \|\mathbf{x} \wedge \mathbf{y}\|.$$

By Fact 6.5, we then have

$$\|A(\mathbf{x}/\|A\mathbf{x}\| - \mathbf{y}/\|A\mathbf{y}\|)\| \geq \frac{ne^t}{\|A\mathbf{x}\| \|A\mathbf{y}\|} \|\mathbf{x} - \mathbf{y}\|. \quad (31)$$

Now we argue as in the proof of (i),

$$\begin{aligned} \frac{\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\|}{\|\mathbf{x} - \mathbf{y}\|} &= \frac{\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\|}{\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|} \frac{\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|}{\|\mathbf{x} - \mathbf{y}\|} \\ &= \frac{\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\|}{\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|} \sqrt{1/\|A\mathbf{x}\|^2 + 1/\|A\mathbf{y}\|^2} \frac{1/\sqrt{1/\|A\mathbf{x}\|^2 + 1/\|A\mathbf{y}\|^2} \|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|}{\|\mathbf{x} - \mathbf{y}\|}. \end{aligned} \quad (32)$$

Note that as $\langle \mathbf{x}, \mathbf{y} \rangle \leq 0$, for any $\alpha^2 + \beta^2 = 1$,

$$\| \alpha \mathbf{x} - \beta \mathbf{y} \|^2 = 1 - 2\alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle \leq 1 - \langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

Thus

$$\frac{1/\sqrt{1/\|A\mathbf{x}\|^2 + 1/\|A\mathbf{y}\|^2} \|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|}{\|\mathbf{x} - \mathbf{y}\|} \leq \frac{1}{\sqrt{2}}. \quad (33)$$

It follows from (31), (32) and (33) that

$$\frac{\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\|}{\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|} \sqrt{\|A\mathbf{x}\|^2 + \|A\mathbf{y}\|^2} \geq \sqrt{2}ne^t. \quad (34)$$

Again by (i) and (ii) of Theorem 4.1 we can assume that $\|A\mathbf{x}\|, \|A\mathbf{y}\| \leq \sqrt{ne}^{t/4}$ with probability at least $1 - e^{-ct^2n}$. Within this event

$$\frac{\|A(\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|})\|}{\|\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|}\|} \geq e^{t/2}\sqrt{n}.$$

Claim 6.9. *Let $\mathbf{x}, \mathbf{y} \in S^{n-1}$ be given such that $\mathbf{x} \wedge \mathbf{y} \neq 0$. Then*

$$\mathbf{P}\left(\exists \alpha, \beta, \alpha^2 + \beta^2 = 1, \frac{\|A(\alpha\mathbf{x} + \beta\mathbf{y})\|}{\|\alpha\mathbf{x} + \beta\mathbf{y}\|} \geq \sqrt{ne^{t/2}}\right) \leq e^{-ct^2n}.$$

It remains to verify Claim 6.9. To this end, we first find a n^{-C} -net \mathcal{M} of the unit circle $S_{\mathbf{x}, \mathbf{y}}^1$ of the plane spanned by \mathbf{x} and \mathbf{y} . As this set is one-dimensional, one can choose $|\mathcal{M}| = n^C$. With C chosen sufficiently large, one pass from the event $\{\exists \mathbf{z} \in S_{\mathbf{x}, \mathbf{y}}^1, \frac{\|\mathbf{Az}\|}{\|\mathbf{z}\|} \geq \sqrt{ne^t}\}$ to the event $\{\exists \mathbf{z} \in \mathcal{M}, \frac{\|\mathbf{Az}\|}{\|\mathbf{z}\|} \geq \sqrt{ne^t}\}$ without any essential loss. However, for each fixed \mathbf{z} , by (i) of Theorem 4.1 we have $\mathbf{P}(\frac{\|\mathbf{Az}\|}{\|\mathbf{z}\|} \geq \sqrt{ne^t}) \leq e^{-ct^2n}$. The claim then just follows after taking union bound over $n^{O(1)}$ elements of \mathcal{M} .

We complete the proof by proving (iii). This time, without loss of generality we assume

$$0 < \langle \mathbf{x}, \mathbf{y} \rangle \leq 1.$$

We write

$$\frac{\|\mathbf{Ax} \wedge \mathbf{Ay}\|}{\|\mathbf{x} \wedge \mathbf{y}\|} = \|\mathbf{Ax}\| \|\mathbf{Ay}\| \frac{\|\mathbf{Ax}/\|\mathbf{Ax}\| \wedge \mathbf{Ay}/\|\mathbf{Ay}\|\|}{\|\mathbf{x} \wedge \mathbf{y}\|}.$$

Thus our assumption implies that

$$\|\mathbf{Ax}/\|\mathbf{Ax}\| \wedge \mathbf{Ay}/\|\mathbf{Ay}\|\| \leq \frac{ne^{-t}}{\|\mathbf{Ax}\| \|\mathbf{Ay}\|} \|\mathbf{x} \wedge \mathbf{y}\|.$$

By Fact 6.5, we then have

$$\|A(\mathbf{x}/\|\mathbf{Ax}\| - \mathbf{y}/\|\mathbf{Ay}\|)\| \leq \frac{ne^{-t}}{\|\mathbf{Ax}\| \|\mathbf{Ay}\|} \|\mathbf{x} - \mathbf{y}\|. \quad (35)$$

Now use (32)

$$\frac{\|A(\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|})\|}{\|\mathbf{x} - \mathbf{y}\|} = \frac{\|A(\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|})\|}{\|\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|}\|} \frac{1}{\sqrt{1/\|\mathbf{Ax}\|^2 + 1/\|\mathbf{Ay}\|^2}} \frac{\sqrt{1/\|\mathbf{Ax}\|^2 + 1/\|\mathbf{Ay}\|^2}}{\|\mathbf{x} - \mathbf{y}\|} \|\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|}\|. \quad (36)$$

As $\langle \mathbf{x}, \mathbf{y} \rangle \geq 0$, for any $\alpha^2 + \beta^2 = 1$

$$\| |\alpha| \mathbf{x} - |\beta| \mathbf{y} \|^2 = 1 - 2|\alpha\beta| \langle \mathbf{x}, \mathbf{y} \rangle \geq 1 - \langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2.$$

Thus

$$\frac{\frac{1}{\sqrt{1/\|\mathbf{Ax}\|^2 + 1/\|\mathbf{Ay}\|^2}} \left\| \frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|} \right\|}{\|\mathbf{x} - \mathbf{y}\|} \geq \frac{1}{\sqrt{2}}. \quad (37)$$

It follows from (35), (36) and (37) that

$$\frac{\|A(\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|})\|}{\|\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|}\|} \sqrt{\|\mathbf{Ax}\|^2 + \|\mathbf{Ay}\|^2} \leq \sqrt{2}ne^{-t}. \quad (38)$$

Again by (ii) and (iii) of Theorem 4.1 we can assume $\|\mathbf{Ax}\|, \|\mathbf{Ay}\| \geq \sqrt{n}e^{-t/4}$ with probability at least $1 - K^n e^{-tn/2} - e^{-c''t^2n}$. Within this event

$$\frac{\|A(\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|})\|}{\|\frac{\mathbf{x}}{\|\mathbf{Ax}\|} - \frac{\mathbf{y}}{\|\mathbf{Ay}\|}\|} \leq e^{-t/2} \sqrt{n}.$$

Claim 6.10. *Let $\mathbf{x}, \mathbf{y} \in S^{n-1}$ be given such that $\mathbf{x} \wedge \mathbf{y} \neq 0$. Then*

$$\mathbf{P}\left(\exists \alpha, \beta, \alpha^2 + \beta^2 = 1, \frac{\|A(\alpha \mathbf{x} + \beta \mathbf{y})\|}{\|\alpha \mathbf{x} + \beta \mathbf{y}\|} \leq \sqrt{n}e^{-t/2}\right) \leq e^{-ct^2n}.$$

The proof of Claim 6.10 is similar to that of Claim 6.9. In fact, consider the n^{-C} -net \mathcal{M} of the unit circle $S_{\mathbf{x}, \mathbf{y}}^1$ of the plane spanned by \mathbf{x} and \mathbf{y} with size $|\mathcal{M}| = n^C$ and with sufficiently large C . As $t = o(\log n)$, one can pass the event $\{\exists \|\mathbf{z}\| = 1, \frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \leq \sqrt{n}e^{-t/2}\}$ to the event $\{\exists \mathbf{z} \in \mathcal{M}, \frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \leq \sqrt{n}e^{-t/2}\}$ without any essential loss. However, for each fixed \mathbf{z} , by (ii) and (iii) of Theorem 4.1 we have $\mathbf{P}\left(\frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \leq \sqrt{n}e^{-t/2}\right) \leq \max\{K^n e^{-nt/2}, e^{-c''t^2n}\}$. The claim then just follows after taking union bound over $n^{O(1)}$ elements of \mathcal{M} . □

Remark 6.11. *Although the behavior of $\frac{\|\mathbf{Ax} \wedge \mathbf{Ay}\|}{\|\mathbf{x} \wedge \mathbf{y}\|}$ is more relevant to our study, we had to pass to $\frac{\|\mathbf{Ax}\| \|\mathbf{Ax} \wedge \mathbf{Ay}\| \|\mathbf{Ay}\|}{\|\mathbf{x} \wedge \mathbf{y}\|}$ in both Theorem 6.6 and Theorem 6.8 to make use of the convenient identity (29) (which is valid only for unit vectors.)*

6.12. **Step 3.** Let $(\mathbf{x}_0, \mathbf{y}_0)$ be any vector pair from \mathcal{P}_{start} . We will show

Lemma 6.13. *For any $t \geq 1/n$ we have*

$$\mathbf{P}\left(\left|\frac{1}{N} \log \|B_N \mathbf{x}_0 \wedge B_N \mathbf{y}_0\|\right| \geq t\right) \leq \exp(-c \min\{t^2, t\} N n) + N n^{-cn}.$$

It is clear that Theorem 1.12 follows from Lemma 6.13 after taking union bound over \mathcal{P}_{start} .

To prove this result, we first give an analog of Theorem 4.1. Recall the notion of $\mathbf{x}_i, \mathbf{y}_i$ from Subsection 2.3. For short, denote

$$y_i := \log \frac{\|A \mathbf{x}_i \wedge A \mathbf{y}_i\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|} - \log n.$$

Proof. (of Lemma 6.13) We will follow the proof of Lemma 5.1. First, by Theorem 6.3, the event \mathcal{G}_1 that $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{P}$ for all $1 \leq i \leq N$ holds with probability

$$\mathbf{P}(\mathcal{G}_1) \geq 1 - N n^{-cn}.$$

Consider the random sum

$$S = \frac{1}{N} (y_1 + \cdots + y_N).$$

Basing on Corollary 6.7 and Theorem 6.8, the event \mathcal{G}_2 such that $|y_i| \leq 2 \log D$ for all $y_i, 1 \leq i \leq N$ satisfies

$$\mathbf{P}(\mathcal{G}_2) \geq 1 - N D^{-n}.$$

Introduce the new random variables $y'_i := y_i 1_{|y_i| \leq 2 \log D}$ and $y''_i := y'_i - \mathbf{E}_{A_i} y'_i$. As usual, in the sequel we will be conditioning on A_1, \dots, A_{i-1} . By Theorem 6.8, for any positive $t = O(1)$

$$\mathbf{P}_{A_i}(|y'_i| \geq t) \leq \mathbf{P}_{A_i}(|y_i| \geq t) \leq e^{-ct^2 n}. \quad (39)$$

Also, by Theorem 6.6 and Theorem 6.8, for $O(1) \leq t \leq 2 \log D$

$$\mathbf{P}_{A_i}(|y'_i| \geq t) \leq \mathbf{P}_{A_i}(|y_i| \geq t) \leq C^n e^{-tn/2} + e^{-ct^2 n}. \quad (40)$$

Consequently,

$$\mathbf{E}_{A_i} |y'_i| \leq \int_0^{2 \log D} t \mathbf{P}(|y'_i| \geq t) \leq O\left(\int_0^{1/\sqrt{n}} t dt\right) = O(1/n).$$

Consider the martingale sum $S'' := \frac{1}{N}(y_1'' + \cdots + y_N'')$. By definition, $|y_i''| \leq 2 \log D$. Also by (39) and (40), for $t \geq 1/n$

$$\mathbf{P}_{A_i}(|y_i''| \geq t) \leq \mathbf{P}_{A_i}(|y_i'| \geq t) \leq \exp(-c \min\{t^2, t\}n).$$

This implies that for $\lambda = ctn$,

$$e^{-2\lambda t} \mathbf{E}(e^{\lambda y_i''} | A_1, \dots, A_{i-1}), e^{-2\lambda t} \mathbf{E}(e^{-\lambda y_i''} | A_1, \dots, A_{i-1}) \leq \exp(-c \min\{t, t^2\}n).$$

From here, argue similarly as in Section 5, for $t \geq 1/n$

$$\mathbf{P}(|S''| \geq 2t) = \mathbf{P}(|y_1'' + \cdots + y_N''| \geq 2Nt) \leq \exp(-c' \min\{t^2, t\}Nn).$$

Thus,

$$\begin{aligned} \mathbf{P}(|S| \geq 2t + O(1/n)) &\leq \mathbf{P}(|S''| \geq 2t) + \mathbf{P}(\bar{\mathcal{G}}_1 \cup \bar{\mathcal{G}}_2) \leq \exp(-c \min\{t^2, t\}Nn) + \mathbf{P}(\bar{\mathcal{G}}_1) + \mathbf{P}(\bar{\mathcal{G}}_2) \\ &\leq \exp(-c \min\{t^2, t\}Nn) + Nn^{-cn}. \end{aligned}$$

□

7. THE LEAST LYAPUNOV'S EXPONENT: PROOF OF (3) OF THEOREM 1.12

Recall from Subsection 2.5 that

$$\begin{aligned} \log \text{dist}(\mathbf{c}_n, \text{span}(\mathbf{c}_i, i \neq n)) &= \log \text{dist}(B_N \mathbf{e}_n, H_{B_N \mathbf{e}_1, \dots, B_N \mathbf{e}_{n-1}}) \\ &= \sum_{i=1}^N \log \text{dist}(A_i \mathbf{v}_i, H_{A_i \dots A_1 \mathbf{e}_1, \dots, A_i \dots A_1 \mathbf{e}_{n-1}}) \\ &= \sum_{i=1}^N \log d_i, \end{aligned}$$

with

$$d_i^2 := \text{dist}^2(A_i \mathbf{v}_i, H_{A_i \dots A_1 \mathbf{e}_1, \dots, A_i \dots A_1 \mathbf{e}_{n-1}}) = \frac{1}{\|A_i^{-1} \mathbf{v}_i\|_2^2} = \frac{1}{\sum_j \sigma_{ij}^{-2} |\mathbf{v}_i^T \mathbf{u}_{ij}|^2}, \quad (41)$$

where σ_{ij} and \mathbf{u}_{ij} are the singular values and (unit) singular vectors of the matrix A_i , and thus independent of \mathbf{v}_i .

Our main goal is the following estimate on $\mathbf{P}(\mathcal{E}_{\varepsilon,1})$.

Theorem 7.1. *For given $\varepsilon > 0$, there exists an absolute constant C such that the following holds for sufficiently large n and N*

$$\mathbf{P}\left(\frac{1}{N} \sum_{i=1}^N \log d_i \leq -(1/2 + \varepsilon) \log n\right) = \exp(-N/2)C^n + Nn^{-\omega(1)}.$$

We will prove Theorem 7.1 by invoking a series of known results in RMT.

Firstly, we will use the following isotropic delocalization result from [2, Theorem 2.16].

Theorem 7.2. *Let ε and $A > 0$ be given numbers. Then the following holds for $n \geq n(\varepsilon, A)$: for any fixed unit vector \mathbf{v} , with probability at least $1 - n^{-A}$*

$$\sup_{1 \leq j \leq n} |\mathbf{v}^T \mathbf{u}_{ij}| \leq n^{-1/2+\varepsilon}.$$

From now on we will condition on the event $\mathcal{E}_{\varepsilon,A}$ of Theorem 7.2, in which case we have

$$\sum_j \sigma_{ij}^{-2} |\mathbf{v}_i^T \mathbf{u}_{ij}|^2 \leq n^{-1+2\varepsilon} \sum_j \sigma_j^{-2}.$$

Secondly, by [16, Claim 5.1] the following holds with probability $1 - n^{-\omega(1)}$,

$$\sum_{j=1}^{n-O(\log n)} \sigma_j^{-2} \ll \frac{n}{\log n}.$$

Thus

$$\|A_i^{-1} \mathbf{v}_i\|^2 = \sum_j \sigma_{ij}^{-2} |\mathbf{v}_i^T \mathbf{u}_{ij}|^2 \leq n^{-1+2\varepsilon} \left(\frac{n}{\log n} + \sigma_n^{-2} \log n \right) \leq (1 + n^{-1/2} \sigma_n^{-1})^2 n^{3\varepsilon}.$$

Thirdly, we use the following bound from [21] and [16].

Lemma 7.3. *As long as $\delta \geq \exp(-cn)$*

$$\mathbf{P}(\sigma_n \leq \delta/n) \leq C_0 \delta.$$

Thus altogether we have

$$\mathbf{P}\left(\|A_i^{-1} \mathbf{v}_i\|^2 \geq \frac{n^{1+3\varepsilon}}{\delta^2}\right) \ll \delta. \quad (42)$$

Passing to distances, we obtain the following.

Theorem 7.4. *For any $\delta > \exp(-cn)$ we have*

$$\mathbf{P}(d_i n^{1/2+2\varepsilon} \leq \delta) \leq C_0 \delta.$$

Now let

$$X_i := \log(d_i n^{1/2+2\varepsilon}) \mathbf{1}_{d_i \geq \exp(-cn)}.$$

We have shown that for a given $t_0 > 0$ and for any $\delta > 0$

$$\mathbf{P}\left(\mathbf{E}(\exp(-t_0 X_i) | A_1, \dots, A_{i-1}) \geq (1/\delta)^{t_0}\right) \ll \delta.$$

Hence there exists an absolute constant C such that for any $0 < t_0 \leq 1/2$

$$0 \leq \mathbf{E}(\exp(-t_0 X_i) | A_1, \dots, A_{i-1}) \leq C.$$

Next write

$$\begin{aligned} \mathbf{P}(X_1 + \dots + X_N \leq -Nt_0) &= \mathbf{P}(-X_1 - \dots - X_N \geq Nt_0) \leq \exp(-Nt_0) \mathbf{E} \exp(-t_0(X_1 + \dots + X_N)) \\ &\leq \exp(-Nt_0) \mathbf{E}([\exp(-t_0(X_1 + \dots + X_{N-1}))] \mathbf{E}(\exp(-t_0 X_N) | A_1, \dots, A_{N-1})) \\ &\leq C \exp(-Nt_0) \mathbf{E}([\exp(-t_0(X_1 + \dots + X_{N-1}))]) \end{aligned}$$

Repeat the machinery for X_{N-1}, \dots, X_1 , we thus obtain

$$\mathbf{E} \exp(-t_0(X_1 + \dots + X_N)) \leq C^N.$$

In summary,

$$\mathbf{P}\left(\sum_{i=1}^N \log(d_i n^{1/2+2\varepsilon}) \leq -Nt_0\right) \leq \exp(-Nt_0) C^N + N \exp(-cn).$$

Choosing $t_0 = 1/2$, we obtain Theorem 7.1 after a proper scaling of ε (assuming n, N sufficiently large).

8. REMARKS

We have considered product of N iid random matrices where the A_i can be singular with positive probability. Because of this, one has to assume N not to be too large. The main bulk of the paper

develops several ways to balance between the singularity and the generality of the asymptotic Lyapunov' exponents.

There are various models (especially in connection to the study of Schrödinger operators in various lattices or to the study of random band matrices) where it is natural to study the large deviation type problem for general unimodular ensembles with either discrete or continuous atom distribution. One extremely convenient property of this model is that one does not have to worry about N as the product matrices never vanish. On the other hand, the *mean-field* techniques used in our note do not seem to work. One simple candidate for future study is the symplectic model

$$A_i = \begin{pmatrix} \lambda W_n - E & -I_n \\ I_n & 0 \end{pmatrix}$$

with given parameter E, λ , where $W_n = (w_{ij})_{1 \leq i, j \leq n}$ are random Wigner matrices of upper diagonal entries of variance $1/n$.

It has been shown in [13] that the Lyapunov exponents of this model (and for far more general models) are distinct. Furthermore, these exponents were estimated rather precisely by Sadel and Schulz-Baldes (see also [7]) as follows.

Theorem 8.1. [23, Proposition 8] *As long as $E = 2 \cos \kappa \neq 0$ and $|E| < 2$, then for $1 \leq d \leq n$*

$$\gamma_d = \lambda^2 \frac{1 + 2(n-d)}{8 \sin^2 \kappa} + O(\lambda^3).$$

It remains an interesting and challenging problem to obtain large deviation type estimates for this model.

Acknowledgement. The author is grateful to prof. Thomas Spencer for invaluable comments and suggestions.

REFERENCES

- [1] G. Akemann and J. R. Ipsen, Recent exact and asymptotic results for products of independent random matrices, arxiv.org/abs/1502.01667.
- [2] A. Bloemendal, L. Erdos, A. Knowles, H.-T. Yau, and J. Yin, Isotropic local laws for sample covariance and generalized Wigner matrices, *Electron. J. Probab.*, 19(33):1-53, 2014.
- [3] J. Bourgain, Green's function estimates for lattice Schrödinger operators and applications, *Annals of mathematics studies*, 158.
- [4] J. Bourgain and W. Schlag, Anderson localization for Schrödinger operators on \mathbf{Z} with strongly mixing potentials, *Communication in Mathematical Physics*, 2000 (215), 143-175.
- [5] P. Bourgerol and J. Lacroix, *Products of random matrices with applications to Schrödinger operators*, Boston Birkhauser 1985.
- [6] J. E. Cohen and C.M. Newman, The stability of large random matrices and their products. *Ann. Probab.* **12** (1984) 283-310.
- [7] O. Dorokhov, Solvable model of multichannel localization, *Phys. Rev. B* 37 (1988) 10526-10541.
- [8] A. Figotin and L. Pastur, *Spectra of Random and Almost-Periodic Operators*, Springer, Berlin, 1992.

- [9] P. Forster, Asymptotic of finite system Lyapunov exponents for some random matrix ensembles, *J. Phys. A: Math. Theor.* **48** (2015) 215205.
- [10] H. Furstenberg, Non-commuting random products, *Trans. Amer. Math. Soc.* **108** (1963) 377- 428.
- [11] H. Furstenberg and H. Kesten, Products of random matrices. *Ann. Math. Statist.* **31** (1960) 457-469.
- [12] M. Isopi and C. M. Newman, The triangle law for Lyapunov exponents of large random matrices, *Commun. Math. Phys.* **143**, 591-598 (1992).
- [13] I. Goldsheid and G. Margulis, Lyapunov indices of random matrix products, *Uspekhi Mat. Nauk* **44:5** (1989), 13-60.
- [14] Y. Guivarch and A. Raugi, Propriétés de contraction d'un semi-groupe de matrices inversibles, Coefficients de Liapunoff d'un produit de matrices alatoires independantes, *Israel J. Math.* **65** (1989), 165-196.
- [15] M. Goldstein and W. Schlag, Holder continuity of the integrated density of states for quasi-periodic Schrodinger equations and averages of shifts of sub-harmonic functions, *Annals of Mathematics*, 2001 (154), 155-203.
- [16] H. Nguyen and V. Vu, Normal vector of a random hyperplane, submitted.
- [17] V. Kargin, Products of Random Matrices: dimension and growth in norm, *Annals of Applied Probability*, 2010, Vol. 20, No. 3, 890-906.
- [18] V. Kargin, On the Largest Lyapunov exponent for products of Gaussian matrices, *Journal of Statistical Physics*, 2014 (157), 70-83.
- [19] C. M. Newman, The Distribution of Lyapunov Exponents: Exact Results for Random Matrices, *Commun. Math. Phys.* **103**, 121-126 (1986).
- [20] V.I. Oseledec, A multiplicative ergodic theorem. Ljapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.* **19** (1968) 197-231.
- [21] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices, *Advances in Mathematics*, **218** (2008), no. 2, 600-633.
- [22] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix, *Communications on Pure and Applied Mathematics*, **62** (2009), 1707-1739.
- [23] C. Sadel and H. Schulz-Baldes, Random Lie group actions on compact manifolds: a perturbative analysis, *Ann. Probab.* **38** (2010), no. 6, 2224-2257.
- [24] H. Schulz-Baldes, Perturbation theory for Lyapunov exponents of an Anderson model on a strip, *Geom. Funct. Anal.* **14** (2004), no. 5, 1089-1117.
- [25] T. Tao and V. Vu, From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices, *Bulletin of the American Mathematical Society*, **46** (2009), 377-396.
- [26] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random matrices, *Annals of Mathematics* (2), **169** (2009), no 2, 595-632.
- [27] T. Tao and V. Vu, Random matrices: universality of ESDs and the circular law, *Annals of Probability*, **38** (2010), no. 5 2023-2065, with an appendix by M. Krishnapur.
- [28] D. Thouless, Maximum metallic resistance in thin wires, *Phys. Rev. Lett.* **39** (1977), 1167-1169.
- [29] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, www-personal.umich.edu/~romanv/papers/non-asymptotic-rmt-plain.pdf.
- [30] R. Vershynin, Invertibility of symmetric random matrices, *Random Structures & Algorithms*, **44** (2014), no. 2, 135-182.

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, COLUMBUS OH 43210

E-mail address: nguyen.1261@math.osu.edu