The Annals of Applied Probability
2017, Vol. 27, No. 6, 3672–3705
https://doi.org/10.1214/17-AAP1293

ASYMPTOTIC LYAPUNOV EXPONENTS FOR LARGE RANDOM MATRICES

BY HOI H. NGUYEN[1]

The Ohio State University

Suppose that $A_1, \dots, A_N$ are independent random matrices of size n whose entries are i.i.d. copies of a random variable ξ of mean zero and variance one. It is known from the late 1980s that when ξ is Gaussian then $N^{-1} \log \|A_N \dots A_1\|$ converges to $\log \sqrt{n}$ as $N \to \infty$. We will establish similar results for more general matrices with explicit rate of convergence. Our method relies on a simple interplay between additive structures and growth of matrices.

1. Introduction. Let A_i, $i \geq 1$ be a sequence of independent identically distributed random matrices of a given distribution μ in the space of square matrices of size n of real-valued entries. Let B_N be the matrix product

$$B_N = A_N \dots A_1.$$

Furstenberg and Kesten [10] (see also [3], Theorem 4.1, page 11) proved in 1960 the following.

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 [9] (see also [3], Corollary 3.4, page 53) the following.

THEOREM 1.2. *Assume that* $\mathbf{E} \log^+(\|A_i\|) < \infty$ *and that* μ *is strongly irreducible*, *then*:

- $\lim_{N\to\infty} \frac{1}{N} \log \|B_N \mathbf{x}\| = \gamma$ *uniformly on* $\mathbf{x} \in S^{n-1}$;

Received July 2016; revised January 2017.
[1]Supported by NSF Grants DMS-16-00782, DMS-11-28155 and CCF-14-12958.
MSC2010 subject classifications. 15A52, 60B10.
Key words and phrases. Lyapunov exponents, large random matrices.

- *for any μ-invariant distribution ν on $\mathbf{P}(\mathbf{R}^n)$ [i.e., $\nu(A) = \iint 1_A(M\bar{\mathbf{x}})\, d\mu(M)\, d\nu(\bar{\mathbf{x}})$], we have*

$$\gamma = \iint \log \frac{\|M\bar{\mathbf{x}}\|}{\|\bar{\mathbf{x}}\|}\, d\mu(M)\, d\nu(\bar{\mathbf{x}}),$$

where $\bar{\mathbf{x}}$ is the class of $\mathbf{x}$ in the projective space $\mathbf{P}(\mathbf{R}^n)$.

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, [3, 9, 11].

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

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 $k \geq 2$,

$$\sum_{i=1}^{k} \gamma_i = \lim_{N\to\infty} \frac{1}{N} \mathbf{E} \log \left\| \bigwedge^k B_N \right\|.$$

In [18] (see also [11], Theorem 1.2), Oseledec showed the following extremely powerful theorem on the convergence of Lyapunov exponents.

THEOREM 1.4. *Assume that $\mathbf{E}\log^+(\|A_i\|) < \infty$, then the following hold*:

- *With probability one*,

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

  where $\sigma_1(B_N) \geq \cdots \geq \sigma_n(B_N)$ 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 $\exp(\alpha_1(M)) < \cdots < \exp(\alpha_k(M))$ 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 eigen-subspaces, and set $V_i = U_1 \oplus \cdots \oplus U_i$. 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,*

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

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 [11] for further discussion on these topics.

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

QUESTION 1.5. Can we give:

(i) fine approximation for the Lyapunov's exponents?
(ii) 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 [19] and by Bourgain [4]. For the sake of completeness, allow us to insert here a large deviation-type result for the shift model from [4] (see also [5] and [12]).

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

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

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* $B_N = \prod_{j=1}^N \left(\begin{smallmatrix} f(\mathbf{x}+j\omega)-E & -1 \\ 1 & 0 \end{smallmatrix}\right)$. *Then for* $t > N^{-1/10}$

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

for some absolute constants C and c.

1.1. *The i.i.d. model with large dimension.* Our main focus is on a model of random matrices of *large* dimension which are not necessarily unimodular. Especially, we will consider those A_i random matrices where the entries are i.i.d. 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 1980s [6, 13, 16] 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 [16], equation (6), which is directly related to our discussion.

THEOREM 1.7. *Assume that the entries of A_i are i.i.d. copies of* $\frac{1}{\sqrt{n}}N(0,1)$. *Let $\mu_1 \ge \cdots \ge \mu_n$ be the Lyapunov's exponents of the matrix product B_N. Then*

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

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

This result was also generalized in [13] to ξ having bounded density and $\mathbf{E}((\sqrt{n}\xi)^4) < \infty$. We also refer the reader to a recent paper by Forester [8] and to 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 i.i.d. model is due to Kargin [14], Proposition 3, who considered the rate of convergence of the top exponents.

THEOREM 1.8. *Let $\varepsilon > 0$ be given. Assume that the entries of A_i are i.i.d. copies of $\frac{1}{\sqrt{n}}N(0,1)$. Then for all sufficiently small t, and all $n \geq n_0(t)$ and $N \geq 1$,*

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

REMARK 1.9. To be more precise, Proposition 3 of [14] shows that $\mathbf{P}(|\frac{1}{N} \times \log\|B_N\mathbf{x}\|| > t) \leq \exp(-\frac{1}{8}Nnt^2)$ for any fixed $\mathbf{x} \in S^{n-1}$, from which one can deduce Theorem 1.8 by an ε-net argument; see, for instance, Claim 2.1.

1.2. *Our results*. To the best of our knowledge, all of the results in the literature with respect to the i.i.d. 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 $N \to \infty$ 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 is still reasonable to study the top Lyapunov exponent but not other exponents. However, 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? This question is natural because in many practical problems, it is not known a priori that our random matrix model is smooth. In addition, to estimate the Lyapunov's exponents using computer, one actually computes $\frac{1}{N}\log\sigma_i(B_N)$ for some sufficiently large (but not too large) N.

Trying to address these issues, with a universality approach in mind, we will consider the matrix models A_i where the entries of $\sqrt{n}A_i$ are i.i.d. copies of a random variable ξ of mean zero, variance one, and that there exist parameters K, K' such that for all t

$$\mathbf{P}(|\xi| \geq t) \leq K' \exp(-t^2/K). \tag{2}$$

We remark that 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 parameters K, K' of ξ, unless explicitly declared otherwise.

One representative example of our matrices is the Bernoulli ensemble, 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 $N \to \infty$ (for instance, when n is even then the event when the entries of A_1 are all 1's and when all rows of A_2 have exactly $n/2$ entries 1's has positive probability).

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

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

THEOREM 1.10 (Main results). *Let $\varepsilon > 0$ be given. Let $A_1, A_2, \dots$ be independent matrices whose entries are i.i.d. copies of $\frac{1}{\sqrt{n}}\xi$ with ξ satisfying* (2) *for some K, K'. Then there exist constants c, C depending on ε and K, K' such that the following hold*:

(1) (*Top exponent.*) *For any $t \geq 1/n$, we have*

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

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

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

(3) (*Last exponent.*) *We also have*

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

In short, (1) of Theorem 1.10 extends Theorem 1.8 to general matrix ensembles with the extra assumptions that $N \ll n^{cn}$ and $n_0(t) = O(1/t)$. It shows that although the chain dies out eventually, one 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 [15] (see Figure 4 in that paper). By taking $t = 1/n$ and $\varepsilon = 1/2$, we obtain the following from (1):

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

We also show that the approach can be modified in a nontrivial 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 1.7.

In addition, 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 1.7. 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 it is unlikely to be well concentrated. This is supported by the fact that the least singular value of random matrices is not well concentrated (see, for instance, [25]). 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 the upper bound [i.e., (3)] is essentially harder than the lower bound.

In conclusion, our main results consider product of N i.i.d. 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, which will be described in more details below, 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 Shrödinger operators, where it is natural to study the large deviation-type problem for unimodular ensembles with either discrete or continuous entry distributions. One extremely convenient property of these models 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 [11] that the Lyapunov exponents of this model are distinct. Furthermore, these exponents were estimated rather precisely by Sadel and Schulz-Baldes (see also [7]) as follows.

THEOREM 1.11 ([22], 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.

The rest of the paper is organized as follows. We will introduce the methods to prove Theorem 1.10 in the next section. A detailed treatment for (1) will be carried out throughout Sections 3, 4 and 5. We then extend these treatments to complete the proof of (2) in Section 6. The proof of (3) will be presented in Section 7.

2. Proof method.

2.1. *The top exponent.* Here, we discuss the method to prove (1) of Theorem 1.10. To estimate $\|B_N\| = \sup_{\mathbf{x}_0 \in S^{n-1}} \|B_N \mathbf{x}_0\|$, it is sufficient to work with a finite collection of unit vectors $\mathbf{x}_0$. Let $\varepsilon > 0$ be a parameter, and let $\mathcal{N}_{\text{start}}$ be an ε-net of S^{n-1}. It is well known that one can assume $|\mathcal{N}_{\text{start}}| \le (1 + 2/\varepsilon)^n$. The following is often used in the context of bounding the largest singular values of random matrices (see, for instance, [23], Remark 2.3.3, or the proof of Claim 2.2 below).

CLAIM 2.1. *We have*

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

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 $\mathbf{x}_0 \in \mathcal{N}_{\text{start}}$. This was also the main starting point of [14].

Let $\mathbf{x}_0$ be an element of $\mathcal{N}_{\text{start}}$. One writes

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

where

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

When ξ has discrete distribution such as Bernoulli, there is a minor problem that $B_N \mathbf{x}_0$ can be vanishing, but we can rule out this possibility by choosing the net $\mathcal{N}_{\text{start}}$ 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 $\mathcal{N}_{\text{start}}$ of S^{n-1};
  - $\mathcal{S}$ remains stable under the action of each A_i. In other words, with 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 (1) of Theorem 1.10.

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.

2.2. *The second exponent*. We will extend the ideas of the previous subsection to deal with (2) of Theorem 1.10. First of all, let $\mathcal{P}_{\text{start}}$ be some subset of $S^{n-1} \times S^{n-1}$ that covers an ε-net [i.e., for any $(\mathbf{x}, \mathbf{y}) \in S^{n-1} \times S^{n-1}$ there exists $(\mathbf{x}', \mathbf{y}') \in \mathcal{P}_{\text{start}}$ such that $\|\mathbf{x} - \mathbf{x}'\|, \|\mathbf{y} - \mathbf{y}'\| \leq \varepsilon$].

CLAIM 2.2. *We have*

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

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

$$\begin{aligned}\mathrm{Vol}_2(B_N\mathbf{x}, B_N\mathbf{y}) &\leq \mathrm{Vol}_2(B_N\mathbf{x}', B_N\mathbf{y}') + \mathrm{Vol}_2(B_N(\mathbf{x} - \mathbf{x}'), B_N\mathbf{y}) \\ &\quad + \mathrm{Vol}_2(B_N\mathbf{x}', B_N(\mathbf{y} - \mathbf{y}')) + \mathrm{Vol}_2(B_N(\mathbf{x} - \mathbf{x}'), B_N(\mathbf{y} - \mathbf{y}'))\end{aligned}$$

$$\leq \mathrm{Vol}_2(B_N\mathbf{x}', B_N\mathbf{y}') + (2\varepsilon+\varepsilon^2) \sup_{(\mathbf{z}_1,\mathbf{z}_2)\in S^{n-1}\times S^{n-1}} \mathrm{Vol}_2(B_N\mathbf{z}_1, B_N\mathbf{z}_2). \qquad \square$$

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

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

$$\begin{aligned}
&\log \|B_N\mathbf{x}_0 \wedge B_N\mathbf{y}_0\| \\
&\quad = \log \|A_NA_{N-1}\dots A_2A_1\mathbf{x}_0 \wedge A_NA_{N-1}\dots A_2A_1\mathbf{y}_0\| \\
&\quad = \log \left\| A_N \frac{A_{N-1}\dots A_2A_1\mathbf{x}_0 \wedge A_{N-1}\dots A_2A_1\mathbf{y}_0}{\|A_{N-1}\dots A_2A_1\mathbf{x}_0 \wedge A_{N-1}\dots A_2A_1\mathbf{y}_0\|} \right\| \\
&\qquad + \log \left\| A_{N-1} \frac{A_{N-2}\dots A_2A_1\mathbf{x}_0 \wedge A_{N-2}\dots A_2A_1\mathbf{y}_0}{\|A_{N-2}\dots A_2A_1\mathbf{x}_0 \wedge A_{N-2}\dots A_2A_1\mathbf{y}_0\|} \right\| \\
&\qquad + \cdots + \log \left\| A_2 \frac{A_1\mathbf{x}_0 \wedge A_1\mathbf{y}_0}{\|A_1\mathbf{x}_0 \wedge A_1\mathbf{y}_0\|} \right\| + \log \|A_1\mathbf{x}_0 \wedge A_1\mathbf{y}_0\| \\
&\quad = \sum_{i=0}^{N-1} \log \frac{\|A_{i+1}\mathbf{x}_i \wedge A_{i+1}\mathbf{y}_i\|}{\|\mathbf{x}_i \wedge \mathbf{y}_i\|},
\end{aligned}$$

where

$$\mathbf{x}_i := \frac{A_i \dots A_2A_1\mathbf{x}_0}{\|A_i \dots A_2A_1\mathbf{x}_0\|} \quad \text{and} \quad \mathbf{y}_i := \frac{A_i \dots A_2A_1\mathbf{y}_0}{\|A_i \dots A_2A_1\mathbf{y}_0\|}.$$

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 $\|A_{i+1}\mathbf{y}_i\|$:

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

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 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 much more complicated. Nevertheless, our overall plan will remain the same.

- Step 1. (Dynamics and structures.) Find a set $\mathcal{P}$ of pair vectors in $\mathbf{R}^n$ such that $\mathcal{P}_{\text{start}} \subset \mathcal{P}$ 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 1.10.

We will present a full proof of (2) of Theorem 1.10 in Section 6.

2.3. *The last exponent.* Now we discuss the method to prove (3) of Theorem 1.10. 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

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

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

$$\operatorname{dist}\big(\mathbf{c}_{i_0}, \operatorname{span}(\mathbf{c}_i, i \ne i_0)\big) \le \sqrt{n}T_\varepsilon.$$

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

$$\mathcal{E}_{\varepsilon,1} := \big\{\log\operatorname{dist}\big(\mathbf{c}_n, \operatorname{span}(\mathbf{c}_i, i \ne n)\big) \le \log\sqrt{n} + \log T_\varepsilon\big\}.$$

We then have

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

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 nondegenerate tuple $(\mathbf{v}_1, \dots, \mathbf{v}_n)$,

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

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

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

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

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

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

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

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

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

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

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

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

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

Note that the vectors $\mathbf{u}_{i1},\dots,\mathbf{u}_{i(n-1)}$ and $\mathbf{v}_i$ are independent of A_i. 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 carry out this plan in Section 7.

3. Step 1 for (1) of Theorem 1.10: Structures under matrix action. Our choice of the set $\mathcal{S}$ is motivated by recent ideas from Tao–Vu [24, 26] and from Rudelson–Vershynin [20, 21] in the context of controlling the small ball probability of random walk. Although this looks surprising at first, the reader will see that these structures are indeed the right object to work with.

We first introduce the notion of *least common denominator* by Rudelson and Versynin (see [20]). Fix parameters κ and γ, where $\gamma \in (0,1)$. For any nonzero vector $\mathbf{x}$ define

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

We record a few easy consequences of **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. Assume that $\operatorname{dist}(D\mathbf{x}, \mathbf{Z}^n) \leq \min\{\gamma\|D\mathbf{x}\|, \kappa\}$ for some $D > 0$. Then as $\|\mathbf{x}-\mathbf{y}\| \leq 1/D^2$, by the triangle inequality we have

$$\begin{aligned}\operatorname{dist}(D\mathbf{y}, \mathbf{Z}^n) &\leq \min\{\gamma(\|D\mathbf{y}\| + D\|\mathbf{x}-\mathbf{y}\|), \kappa + D\|\mathbf{x}-\mathbf{y}\|\} \\ &\leq \min\left\{\left(\gamma + \frac{1}{D}\right)\|D\mathbf{y}\|, \kappa+1\right\}.\end{aligned}$$ □

There are two main advantages to 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 $\xi_1x_1 + \cdots + \xi_nx_n$, where ξ_i are i.i.d. 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 ([20]). *For every*

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

we have

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

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, if suffices to establish appropriate 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 ([21], 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 **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

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

This implies that

$$\|p\| \leq \big(1 + o(1)\big) D(x) < 3D_0. \tag{7}$$

It also follows from (6) that

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

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

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

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*

$$\mathrm{Vol}(S^{n-1} \backslash \mathcal{S}) \le \mathrm{Vol}_{n-1}(\mathbf{B}(0, n^{-1/2+5c})).$$

PROOF. Let $\mathcal{N}_D$ be one of the sets obtained from Theorem 3.3. Then by definition

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

Lemma 3.5 implies that for any $n^{-1/2+5c} \le \varepsilon \le 1$, there exists a ε-net $\mathcal{N}_{\text{start}}$ of S^{n-1} with size $(C/\varepsilon)^n$ that belongs to $\mathcal{S}$. This set $\mathcal{N}_{\text{start}}$ 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 ([20], Lemma 2.2). *Let $\zeta_1, \dots, \zeta_n$ be independent nonnegative random variables, and let $K, t_0 > 0$. If one has*

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

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

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

for all $t \ge t_0$.

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

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

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

PROOF OF THEOREM 3.7. We first consider the event $\mathcal{E}_1$ that $\|Ax\|^2 \le n^{1-c/2}$. As $\mathbf{LCD}_{\gamma,\kappa}(\mathbf{x}) \ge D = \exp(n^c) \gg \sqrt{n}$, by Theorem 3.2

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

Thus by Lemma 3.6

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

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

$$\mathbf{P}\big(\|A\mathbf{x}\| \ge n^{1-2c}\big) \le \exp\big(-n^{1-4c}\big). \tag{11}$$

On the complement of $\mathcal{E}_1$ and $\mathcal{E}_2$, for each $n^{1/2-c/4} \le r \le n^{1-2c}$ let us look at the event $\mathcal{E}_r$ that $\|A\mathbf{x} - \mathbf{y}\| = O(r\kappa/D)$ 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 $\mathbf{LCD}(Ax/\|Ax\|) \le D$ and $r - n^{-c} \le \|Ax\| \le r + n^{-c}$.

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

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

Hence, by Lemma 3.6,

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

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

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

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

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

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

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

4. Step 2 for (1) of Theorem 1.10: 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

$$\mathbf{x}_i = \frac{A_i\dots A_2A_1\mathbf{x}_0}{\|A_i\dots A_2A_1\mathbf{x}_0\|}.$$

By Theorem 3.7, we can assume that $\mathbf{x}_i \in \mathcal{S}$ [i.e., $\mathbf{LCD}_{\gamma,\kappa}(\mathbf{x}_i) \ge D$] for all $1 \le i \le N$ 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 $A = (a_{ij})_{1\le i\le n}$ be a random square matrix whose entries are i.i.d. copies of ξ from (2). For short, set $\xi_i := a_{i1}x_1 + \dots + a_{in}x_n$ and

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

Before stating our estimates, we note that as a_{ij} are sub-Gaussian random variables of parameter K, so are the normalized random variables $\xi_1,\dots,\xi_n$. This implies that ξ_i^2 are exponential random variables [since $\mathbf{P}(\xi_i^2 \ge t) = \mathbf{P}(|\xi_i| \ge \sqrt{t}) \le$

$O(\exp(-t/K))$]. As a consequence, for any $x \geq 0$ (see, for instance, [27], Proposition 5.16)

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

THEOREM 4.1 (Concentration over nonstructured vectors). *We have*:

(i) *for any* $t > 0$,

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

for some absolute constant $c' > 0$;

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

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

where K *is the parameter from* (2);

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

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

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

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

$$\begin{aligned}\mathbf{P}(y \geq t) &= \mathbf{P}\big(e^{\lambda y} \geq e^{\lambda t}\big) \\ &\leq e^{-\lambda t}\mathbf{E}\bigg(\bigg(\frac{1}{n}(\xi_1^2 + \cdots + \xi_n^2)\bigg)^{\lambda}\bigg) \qquad (14) \\ &= e^{-\lambda t}\lambda \int_0^\infty x^{\lambda-1}\mathbf{P}(z > x)\,dx,\end{aligned}$$

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

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

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

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

To this end,

$$\begin{aligned}\int_0^1 (1+x)^{\lambda-1}e^{-\frac{c}{K^2}nx^2}\,dx &\leq \int_0^1 e^{-\frac{c}{K^2}nx^2 + (\lambda-1)x}\,dx \qquad (16) \\ &\leq 1 + e^{K^2(\lambda-1)^2/cn} \leq e^{\lambda t/2}.\end{aligned}$$

Furthermore, for $t < 2$ we have

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

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

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

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

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

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

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

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

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

where we used (13) in the last estimate. $\square$

5. Step 3 for (1) of Theorem 1.10: Concluding the proof. Let $\mathbf{x}_0$ be any vector from $\mathcal{N}_{\text{start}}$. We will show the following.

LEMMA 5.1. *For any $t \ge 1/n$, we have*

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

It is clear that Theorem 1.10 follows from Lemma 5.1 after taking union bound over $\mathcal{N}_{\text{start}}$.

PROOF. First, by Theorem 3.7, the event $\mathcal{F}_1$ that $\mathbf{x}_i \in \mathcal{S}$ for all $1 \le i \le N$ holds with probability:

$$\mathbf{P}(\mathcal{F}_1) \ge 1 - Nn^{-cn}.$$

Consider the random sum

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

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

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

$$\mathbf{P}(\mathcal{F}_2) \ge 1 - ND^{-n}.$$

Introduce the new random variables $y_i' := y_i 1_{|y_i| \le 2\log D}$ 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 $|t| = O(1)$

$$\mathbf{P}(|y_i'| \ge t) \le \mathbf{P}(|y_i| \ge t) \le e^{-ct^2 n}. \tag{19}$$

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

$$\mathbf{P}(|y_i'| \ge t) \le \mathbf{P}(|y_i| \ge t) \le K^n e^{-tn/2} + e^{-ct^2 n}. \tag{20}$$

Consequently,

$$|\mathbf{E}_{A_i} y_i'| \le \int_0^{2\log D} t\mathbf{P}(|y_i'| \ge t) \le O\left(\int_0^{1/\sqrt{n}} t\, dt\right) = O(1/n).$$

Next, consider the martingale sum

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

By definition, $|y_i''| \le 2\log D + O(1/n) < 3\log D$. Also by (19) and (20), for $t \ge 1/n$

$$\mathbf{P}(|y_i''| \ge t) \le \mathbf{P}(|y_i'| \ge t) \le \exp(-c\min\{t^2, t\}n).$$

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

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

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

$$\begin{aligned}\mathbf{P}(S'' \ge 2t) &= \mathbf{P}(y_1'' + \cdots + y_N'' \ge 2Nt)\\ &= \mathbf{P}(\exp((y_1'' + \cdots + y_N'')\lambda) \ge \exp(2\lambda Nt))\\ &\le \exp(-2\lambda Nt)\mathbf{E}(\exp((y_1'' + \cdots + y_N'')\lambda))\\ &\le \exp(-c'\min\{t^2, t\}Nn).\end{aligned}$$

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

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

□

6. Proof of (2) of Theorem 1.10 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 $\mathcal{P}_{\text{start}} \subset S^{n-1} \times S^{n-1}$ to satisfy the following conditions:

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

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

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

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

LEMMA 6.1. *There exists a set $\mathcal{P}_{\text{start}}$ with the above properties.*

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

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

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

In the next step, we $1/D^{O(1)}$-approximate the parameters θ_1, θ_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 θ_1, θ_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

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

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

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

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

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

Thus $V_{\varepsilon/2} > \mathrm{Vol}_T$ 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 6.1 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

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

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 6.2 (Stability of nonstructures, jointly). *Assume that $A = (a_{ij})_{1 \le i,j \le n}$ is a random matrix of size n whose entries are i.i.d. copies of ξ satisfying (2). Let $(\mathbf{x}, \mathbf{y})$ be any deterministic vector pair from $\mathcal{P}$. Then*

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

By Theorem 6.6 (to be proved separately later), with a probability at least $1 - n^{-cn/16}$ we can assume that

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

Thus by definition of LCD (see Fact 3.1), for Theorem 6.2 it suffices to show

$$\mathbf{P}_A\biggl(\mathbf{LCD}_{\gamma,\kappa}\biggl(\frac{A\mathbf{x}/\|A\mathbf{x}\|}{\|\mathbf{x} \wedge \mathbf{y}\|}, \frac{A\mathbf{y}/\|A\mathbf{y}\|}{\|\mathbf{x} \wedge \mathbf{y}\|}\biggr) \le D'\biggr) \le n^{-cn/8}, \tag{25}$$

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

PROOF OF THEOREM 6.2. By the proof of Theorem 3.7, it suffices to focus on the event $\mathcal{E}_1$ that $\|A\mathbf{x}\|^2, \|A\mathbf{y}\|^2 \ge n^{1-c/2}$.

For each $n^{1/2-c/4} \le r, s \le n^{1-2c}$ (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 $r_i \le \|A\mathbf{x}\| \le r_{i+1}$ and $s_j \le \|A\mathbf{y}\| \le s_{j+1}$ and such that

$$\mathbf{LCD}_{\gamma,\kappa}\biggl(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|}A\mathbf{x}/r_i, \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|}A\mathbf{y}/s_j\biggr) \le \bigl(1 + o(1)\bigr)D'. \tag{26}$$

[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 $r_i \le r \le r_{i+1}$, $s_j \le s \le s_{j+1}$.]

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

$$\mathbf{LCD}_{\gamma,\kappa}\biggl(\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|}\theta_1 A\mathbf{x}/r_i + \frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|}\theta_2 A\mathbf{y}/s_j\biggr) \le \bigl(1 + o(1)\bigr)D'. \tag{27}$$

We can write

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

Here,

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

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 $1 - D^{-O(1)} \le \theta_1'^2 + \theta_2'^2 \le 1 + D^{-O(1)}$.

Now we look at the event $\mathbf{LCD}_{\gamma,\kappa}(\frac{1}{\|\mathbf{x}\wedge\mathbf{y}\|}\theta A(\theta_1'\mathbf{x} + \theta_2'\mathbf{y})) \le (1 + o(1))D'$ from (27). By Theorem 3.3, there exists $\mathbf{u} \in \theta^{-1} \cdot \mathcal{N}_{D'}$ such that

$$\biggl\|\frac{1}{\|\mathbf{x} \wedge \mathbf{y}\|}A\bigl(\theta_1'\mathbf{x} + \theta_2'\mathbf{y}\bigr) - \mathbf{u}\biggr\| \le O\bigl(\theta^{-1}\kappa/D'\bigr). \tag{28}$$

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,

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

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)

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

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). $\square$

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

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

We will be finding the following fact useful.

FACT 6.3. *Let* $f(x) = x - x^2/4, 0 \le x \le 1$, *and let* $0 < x_1, x_2 < 1$. *Let* $t > 0$ *be a parameter*:

- *If* $f(x_2)/f(x_1) \ge 1+t$, *then* $x_2/x_1 \ge 1+t$.
- *If* $f(x_2)/f(x_1) \le t < 1$, *then* $x_2/x_1 \le t$.

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

THEOREM 6.4. *For any* $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$ *and any* $\delta \ge 1/D$,

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

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

COROLLARY 6.5.

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

PROOF OF THEOREM 6.4. By (29), the assumption $\|\frac{A\mathbf{x}}{\|A\mathbf{x}\|} \wedge \frac{A\mathbf{y}}{\|A\mathbf{y}\|}\| \le \delta^2\|\mathbf{x} \wedge \mathbf{y}\|$ implies that

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

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

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

Thus the event $\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\| \le \delta^2\|\mathbf{x}\wedge\mathbf{y}\|$ 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

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

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 $\|A\| = O(\sqrt{n})$,

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

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

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

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

THEOREM 6.6. *There exist constants C, c such that for any $t > 0$ and any $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$ we have*:

(i)

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

(ii)

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

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

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

PROOF. First we prove (i). By Fact 6.3, 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

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

Note that

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

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

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

Thus the event $\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|}-\frac{\mathbf{y}}{\|A\mathbf{y}\|})\|\ge C\|\mathbf{x}-\mathbf{y}\|$ implies that there exists $\mathbf{z}$ such that $\|A\mathbf{z}\|/\|\mathbf{z}\|\ge C/C'$, 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

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

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

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

By Fact 6.3, we then have

$$\big\|A\big(\mathbf{x}/\|A\mathbf{x}\|-\mathbf{y}/\|A\mathbf{y}\|\big)\big\|\ge\frac{ne^t}{\|A\mathbf{x}\|\|A\mathbf{y}\|}\|\mathbf{x}-\mathbf{y}\|. \tag{31}$$

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

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

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

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

Thus

$$\frac{\frac{1}{\sqrt{1/\|A\mathbf{x}\|^2+1/\|A\mathbf{y}\|^2}}\|\frac{\mathbf{x}}{\|A\mathbf{x}\|}-\frac{\mathbf{y}}{\|A\mathbf{y}\|}\|}{\|\mathbf{x}-\mathbf{y}\|} \le \frac{1}{\sqrt{2}}. \tag{33}$$

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

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

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

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

CLAIM 6.7. *Let $\mathbf{x}, \mathbf{y} \in S^{n-1}$ be given such that $\mathbf{x}\wedge\mathbf{y} \neq 0$. Then*

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

It remains to verify Claim 6.7. To this end, we first find a n^{-C}-net $\mathcal{M}$ of the unit circle $S^1_{\mathbf{x},\mathbf{y}}$ 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 $\{\exists \mathbf{z} \in S^1_{\mathbf{x},\mathbf{y}}, \frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \ge \sqrt{n}e^t\}$ to the event $\{\exists \mathbf{z} \in \mathcal{M}, \frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \ge \sqrt{n}e^t\}$ without any essential loss. However, for each fixed $\mathbf{z}$, by (i) of Theorem 4.1 we have $\mathbf{P}(\frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \ge \sqrt{n}e^{t/2}) \le e^{-ct^2n}$. 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

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

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

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

By Fact 6.3, we then have

$$\big\|A\big(\mathbf{x}/\|A\mathbf{x}\| - \mathbf{y}/\|A\mathbf{y}\|\big)\big\| \le \frac{ne^{-t}}{\|A\mathbf{x}\|\|A\mathbf{y}\|}\|\mathbf{x}-\mathbf{y}\|. \tag{35}$$

Now use (32):

$$\begin{aligned} &\frac{\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\|}{\|\mathbf{x}-\mathbf{y}\|} \\ &\quad = \frac{\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\|}{\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|} \\ &\qquad \times \sqrt{1/\|A\mathbf{x}\|^2 + 1/\|A\mathbf{y}\|^2}\,\frac{\frac{1}{\sqrt{1/\|A\mathbf{x}\|^2+1/\|A\mathbf{y}\|^2}}\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|}{\|\mathbf{x}-\mathbf{y}\|}. \end{aligned} \tag{36}$$

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

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

Thus

$$\frac{\frac{1}{\sqrt{1/\|A\mathbf{x}\|^2+1/\|A\mathbf{y}\|^2}}\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|}{\|\mathbf{x}-\mathbf{y}\|} \ge \frac{1}{\sqrt{2}}. \tag{37}$$

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

$$\frac{\|A(\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|})\|}{\|\frac{\mathbf{x}}{\|A\mathbf{x}\|} - \frac{\mathbf{y}}{\|A\mathbf{y}\|}\|}\sqrt{\|A\mathbf{x}\|^2 + \|A\mathbf{y}\|^2} \le \sqrt{2}ne^{-t}. \tag{38}$$

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

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

CLAIM 6.8. *Let $\mathbf{x},\mathbf{y} \in S^{n-1}$ be given such that $\mathbf{x}\wedge\mathbf{y} \ne 0$. Then*

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

The proof of Claim 6.8 is similar to that of Claim 6.7. In fact, consider the n^{-C}-net $\mathcal{M}$ of the unit circle $S^1_{\mathbf{x},\mathbf{y}}$ 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 $\{\exists\|\mathbf{z}\|=1, \frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \le \sqrt{n}e^{-t/2}\}$ to the event $\{\exists \mathbf{z}\in\mathcal{M}, \frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \le \sqrt{n}e^{-t/2}\}$ without any essential loss. However, for each fixed $\mathbf{z}$, by (ii) and (iii) of Theorem 4.1 we have $\mathbf{P}(\frac{\|A\mathbf{z}\|}{\|\mathbf{z}\|} \le \sqrt{n}e^{-t/2}) \le \max\{K^n e^{-nt/2}, e^{-c''t^2 n}\}$. The claim then just follows after taking union bound over $n^{O(1)}$ elements of $\mathcal{M}$. $\square$

REMARK 6.9. Although the behavior of $\frac{\|A\mathbf{x}\wedge A\mathbf{y}\|}{\|\mathbf{x}\wedge\mathbf{y}\|}$ is more relevant to our study, we had to pass to $\frac{\|A\mathbf{x}/\|A\mathbf{x}\|\wedge A\mathbf{y}/\|A\mathbf{y}\|\|}{\|\mathbf{x}\wedge\mathbf{y}\|}$ in both Theorem 6.4 and Theorem 6.6 to make use of the convenient identity (29) (which is valid only for unit vectors).

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

LEMMA 6.10. *For any $t \ge 1/n$, we have*

$$\mathbf{P}\left(\left|\frac{1}{N}\log\|B_N\mathbf{x}_0 \wedge B_N\mathbf{y}_0\|\right| \ge t\right) \le \exp\bigl(-c\min\{t^2, t\}Nn\bigr) + Nn^{-cn}.$$

It is clear that Theorem 1.10 follows from Lemma 6.10 after taking union bound over $\mathcal{P}_{\text{start}}$.

To prove this result, we first give an analog of Theorem 4.1. Recall the notion of $\mathbf{x}_i, \mathbf{y}_i$ from Section 2.2. 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 6.10. We will follow the proof of Lemma 5.1. First, by Theorem 6.2, the event $\mathcal{G}_1$ that $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{P}$ for all $1 \le i \le N$ holds with probability:

$$\mathbf{P}(\mathcal{G}_1) \ge 1 - Nn^{-cn}.$$

Consider the random sum

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

Basing on Corollary 6.5 and Theorem 6.6, the event $\mathcal{G}_2$ such that $|y_i| \le 2\log D$ for all y_i, $1 \le i \le N$ satisfies

$$\mathbf{P}(\mathcal{G}_2) \ge 1 - ND^{-n}.$$

Introduce the new random variables $y_i' := y_i 1_{|y_i| \le 2\log D}$ 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 6.6, for any positive $t = O(1)$,

$$\mathbf{P}_{A_i}\bigl(|y_i'| \ge t\bigr) \le \mathbf{P}_{A_i}\bigl(|y_i| \ge t\bigr) \le e^{-ct^2 n}. \tag{39}$$

Also, by Theorem 6.4 and Theorem 6.6, for $O(1) \le t \le 2\log D$,

$$\mathbf{P}_{A_i}\big(|y_i'| \ge t\big) \le \mathbf{P}_{A_i}\big(|y_i| \ge t\big) \le C^n e^{-tn/2} + e^{-ct^2 n}. \tag{40}$$

Consequently,

$$\mathbf{E}_{A_i}|y_i'| \le \int_0^{2\log D} t\mathbf{P}\big(|y_i'| \ge t\big) \le O\bigg(\int_0^{1/\sqrt{n}} t\,dt\bigg) = O(1/n).$$

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

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

This implies that, for $\lambda = ctn$,

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

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

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

Thus

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

7. The least Lyapunov exponent: Proof of (3) of Theorem 1.10. Recall from Section 2.3 that

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

with

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

where $\sigma_{i1} \geq \cdots \geq \sigma_{in}$ and $\mathbf{u}_{i1}, \dots, \mathbf{u}_{in}$ 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 $\mathbf{P}(\mathcal{E}_{\varepsilon,1})$.

LEMMA 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 Lemma 7.1 by invoking a series of known results in RMT.

First, we will use the following isotropic delocalization result from [2], Theorem 2.16. Let ε and $A > 0$ be given numbers, let $\mathbf{v}$ be a deterministic vector. Let $\mathcal{E}_{\varepsilon,A,\mathbf{v}}^{(i)}$ be the event that

$$\mathcal{E}_{\varepsilon,A,\mathbf{v}}^{(i)} := \Big\{ \sup_{1 \leq j \leq n} |\mathbf{v}^T \mathbf{u}_{ij}| \leq n^{-1/2+\varepsilon} \Big\}.$$

LEMMA 7.2. *The following holds for sufficiently large n:*

$$\mathbf{P}(\mathcal{E}_{\varepsilon,A,\mathbf{v}}^{(i)}) \geq 1 - n^{-A}.$$

Assuming $\mathcal{E}_{\varepsilon,A,\mathbf{v}_i}^{(i)}$, then we have

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

Second, we will use [17], Claim 5.1, to bound the sum involving large singular values.

LEMMA 7.3. *Let $\mathcal{F}^{(i)}$ be the event that $\sum_{j=1}^{n-O(\log n)} \sigma_{ij}^{-2} \leq \frac{n}{\log n}$, then*

$$\mathbf{P}(\mathcal{F}^{(i)}) \geq 1 - n^{-\omega(1)}.$$

Thus

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

Third, we use the following bound from [20], Theorem 1.2.

LEMMA 7.4. *As long as* $\delta \geq \exp(-cn)$,

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

Thus altogether we have

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

Passing to distances, we obtain the following.

COROLLARY 7.5. *On* $\mathcal{E}^{(i)}_{\varepsilon,\mathbf{v}_i}$ *and* $\mathcal{F}^{(i)}$, *for any* $\delta > \exp(-cn)$, *we have*

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

With a cost of $Nn^{-\omega(1)}$ in probability, we assume $\mathcal{E}^{(i)}_{\varepsilon,\mathbf{v}_i}$, $\mathcal{F}^{(i)}$, $1 \leq i \leq N$. Now let

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

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

$$\mathbf{P}\big(\mathbf{E}(\exp(-t_0X_i) \mid A_1, \dots, A_{i-1}) \geq (1/\delta)^{t_0}\big) \ll \delta.$$

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

$$0 \leq \mathbf{E}\big(\exp(-t_0X_i) \mid A_1, \dots, A_{i-1}\big) \leq C.$$

Next, write

$$\begin{aligned}
&\mathbf{P}(X_1 + \cdots + X_N \leq -Nt_0)\\
&\quad = \mathbf{P}(-X_1 - \cdots - X_N \geq Nt_0)\\
&\quad \leq \exp(-Nt_0)\mathbf{E}\exp\big(-t_0(X_1 + \cdots + X_N)\big)\\
&\quad \leq \exp(-Nt_0)\mathbf{E}\big[\exp\big(-t_0(X_1 + \cdots + X_{N-1})\big)\big]\\
&\qquad \times \mathbf{E}\big(\exp(-t_0X_N) \mid A_1, \dots, A_{N-1}\big)\\
&\quad \leq C\exp(-Nt_0)\mathbf{E}\big[\exp\big(-t_0(X_1 + \cdots + X_{N-1})\big)\big].
\end{aligned}$$

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

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

In summary,

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

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

Acknowledgments. The majority of this paper was done while the author was visiting IAS and VIASM. The author thanks the two institutions for their hospitality.

The author would also like to thank Professor T. Spencer for invaluable suggestions. He is also grateful to the anonymous referees for helpful comments that substantially improved the presentation of the paper.

REFERENCES

[1] AKEMANN, G. and IPSEN, J. R. (2015). Recent exact and asymptotic results for products of independent random matrices. *Acta Phys. Polon. B* **46** 1747–1784. MR3403839

[2] BLOEMENDAL, A., ERDÖS, L., KNOWLES, A., YAU, H.-T. and YIN, J. (2014). Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.* **19**(33) 1–53. MR3183577

[3] BOUGEROL, P. and LACROIX, J. (1985). *Products of Random Matrices with Applications to Schrödinger Operators. Progress in Probability and Statistics* **8**. Birkhäuser, Boston, MA. MR0886674

[4] BOURGAIN, J. (2005). *Green's Function Estimates for Lattice Schrödinger Operators and Applications. Annals of Mathematics Studies* **158**. Princeton Univ. Press, Princeton, NJ. MR2100420

[5] BOURGAIN, J. and SCHLAG, W. (2000). Anderson localization for Schrödinger operators on **Z** with strongly mixing potentials. *Comm. Math. Phys.* **215** 143–175. MR1800921

[6] COHEN, J. E. and NEWMAN, C. M. (1984). The stability of large random matrices and their products. *Ann. Probab.* **12** 283–310. MR0735839

[7] DOROKHOV, O. (1988). Solvable model of multichannel localization. *Phys. Rev. B* **37** 10526–10541.

[8] FORRESTER, P. J. (2015). Asymptotics of finite system Lyapunov exponents for some random matrix ensembles. *J. Phys. A* **48** 215205. MR3353003

[9] FURSTENBERG, H. (1963). Noncommuting random products. *Trans. Amer. Math. Soc.* **108** 377–428. MR0163345

[10] FURSTENBERG, H. and KESTEN, H. (1960). Products of random matrices. *Ann. Math. Stat.* **31** 457–469. MR0121828

[11] GOLDSHEID, I. and MARGULIS, G. (1989). Lyapunov indices of random matrix products. *Uspekhi Mat. Nauk* **44** 13–60.

[12] GOLDSTEIN, M. and SCHLAG, W. (2001). Hölder continuity of the integrated density of states for quasi-periodic Schrödinger equations and averages of shifts of subharmonic functions. *Ann. of Math.* (2) **154** 155–203. MR1847592

[13] ISOPI, M. and NEWMAN, C. M. (1992). The triangle law for Lyapunov exponents of large random matrices. *Comm. Math. Phys.* **143** 591–598. MR1145601

[14] KARGIN, V. (2010). Products of random matrices: Dimension and growth in norm. *Ann. Appl. Probab.* **20** 890–906. MR2680552

[15] KARGIN, V. (2014). On the largest Lyapunov exponent for products of Gaussian matrices. *J. Stat. Phys.* **157** 70–83. MR3249905

[16] NEWMAN, C. M. (1986). The distribution of Lyapunov exponents: Exact results for random matrices. *Comm. Math. Phys.* **103** 121–126. MR0826860

[17] NGUYEN, H. and VU, V. (2017). Normal vector of a random hyperplane. *Int. Math. Res. Not. IMRN*. To appear. DOI:10.1093/imrn/rnw273.

[18] OSELEDEC, V. I. (1968). A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. *Tr. Mosk. Mat. Obs.* **19** 179–210. MR0240280

[19] PASTUR, L. and FIGOTIN, A. (1992). *Spectra of Random and Almost-Periodic Operators*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **297**. Springer, Berlin. MR1223779

[20] RUDELSON, M. and VERSHYNIN, R. (2008). The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600–633. MR2407948

[21] RUDELSON, M. and VERSHYNIN, R. (2009). Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** 1707–1739. MR2569075

[22] SADEL, C. and SCHULZ-BALDES, H. (2010). Random Lie group actions on compact manifolds: A perturbative analysis. *Ann. Probab.* **38** 2224–2257. MR2683629

[23] TAO, T. (2012). *Topics in Random Matrix Theory*. Amer. Math. Soc., Providence, RI. MR2906465

[24] TAO, T. and VU, V. (2009). From the Littlewood–Offord problem to the circular law: Universality of the spectral distribution of random matrices. *Bull. Amer. Math. Soc.* (*N.S.*) **46** 377–396. MR2507275

[25] TAO, T. and VU, V. (2010). Random matrices: The distribution of the smallest singular values. *Geom. Funct. Anal.* **20** 260–297. MR2647142

[26] TAO, T. and VU, V. H. (2009). Inverse Littlewood–Offord theorems and the condition number of random discrete matrices. *Ann. of Math.* (2) **169** 595–632. MR2480613

[27] VERSHYNIN, R. Introduction to the non-asymptotic analysis of random matrices. Available at www-personal.umich.edu/~romanv/papers/non-asymptotic-rmt-plain.pdf.

DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210
USA
E-MAIL: nguyen.1261@math.osu.edu