

ON A CONDITION NUMBER OF GENERAL RANDOM POLYNOMIAL SYSTEMS

HOI H. NGUYEN

ABSTRACT. Condition numbers of random polynomial systems have been widely studied in the literature under certain coefficient ensembles of invariant type. In this note we introduce a method that allows us to study these numbers for a broad family of probability distributions. Our work also extends to certain perturbed systems.

1. INTRODUCTION

1.1. **Condition number of random matrices.** Let \mathbf{f} be a system of n linear forms f_1, \dots, f_n in n complex variables $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{C}^n$,

$$f_l(\mathbf{x}) = a_1^{(l)}x_1 + \dots + a_n^{(l)}x_n, 1 \leq l \leq n.$$

The condition number $\mu(\mathbf{f})$ of \mathbf{f} is defined as

$$\mu(\mathbf{f}) := \frac{\sigma_1(\mathbf{f})}{\sigma_n(\mathbf{f})},$$

where $\sigma_1(\mathbf{f})$ and $\sigma_n(\mathbf{f})$ are the largest and smallest singular values of \mathbf{f} .

An important problem with many practical applications is to bound the condition number of a random matrix. As the largest singular value σ_1 is well understood, the main problem is to study the lower bound of the least singular value σ_n . This problem was first raised by Goldstine and von Neumann [7] well back in the 1940s, with connection to their investigation of the complexity of inverting a matrix.

To answer Goldstine and von Neumann's question, Edelman [6] computed the distribution of the least singular value of the random matrix \mathbf{f}^{Gau} where $a_i^{(l)}, 1 \leq i, l \leq n$, are iid standard Gaussian. He showed that for all fixed $\varepsilon > 0$

$$\mathbf{P}(\sigma_n(\mathbf{f}^{Gau}) \leq \varepsilon n^{-1/2}) = \int_0^{\varepsilon^2} \frac{1 + \sqrt{x}}{2\sqrt{x}} e^{-(x/2 + \sqrt{x})} dx + o(1) = \varepsilon - \frac{1}{3}\varepsilon^3 + O(\varepsilon^4) + o(1).$$

2000 *Mathematics Subject Classification.* 12D10, 65H10.

Key words and phrases. Singularity, Condition numbers, System of equations.

The author is supported by research grant DMS-1358648.

Edelman conjectured that this distribution is universal (i.e., it must hold for other distributions of $a_i^{(l)}$, such as Bernoulli.) Note that the same asymptotic continues to hold for any $\varepsilon > 0$ which can go to 0 with n (see also [21, 22])

$$\mathbf{P}(\sigma_n(\mathbf{f}^{Gau}) \leq \varepsilon n^{-1/2}) \leq \varepsilon. \quad (1)$$

Spielman and Teng, in their recent study of smoothed analysis of the simplex method, conjectured that a slightly adjusted bound also holds in the Bernoulli case [21]

$$\mathbf{P}(\sigma_n(\mathbf{f}^{Ber}) \leq \varepsilon) \leq \varepsilon n^{1/2} + c^n, \quad (2)$$

where $0 < c < 1$ is a constant. The term c^n is needed as \mathbf{f}^{Ber} can be singular with exponentially small probability.

Edelman's conjecture has been proved by Tao and Vu in [25]. This work also confirms Spielman and Teng's conjecture for the case ε is fairly large ($\varepsilon \geq n^{-\delta}$ for some small constant $\delta > 0$). For $\varepsilon \geq n^{-3/2}$, Rudelson [13] obtained a strong bound with an extra (multiplicative) constant factor. In a consequent paper, Rudelson and Vershynin [14] show

Theorem 1.2. *There is a constant $C > 0$ and $0 < c < 1$ such that for any $\varepsilon > 0$,*

$$\mathbf{P}(\sigma_n(\mathbf{f}^{Ber}) \leq \varepsilon n^{-1/2}) \leq C\varepsilon n^{1/2} + c^n.$$

This bound is sharp, up to the constant C . It also gives a new proof of Kahn-Komlós-Szemerédi bound [8] on the singularity probability of a random Bernoulli matrix. All these results hold in more general setting, namely that it is enough to assume that the common distribution of the $a_i^{(l)}$ is subgaussian (see (3)) of zero mean and unit variance.

In practice, one often works with random matrices of the type $\mathbf{c} + \mathbf{f}$ where $\mathbf{c} = (c_i^{(l)})$ is deterministic and \mathbf{f} has iid entries. For instance, in their works on smoothed analysis, Spielman and Teng used this to model a large data matrix perturbed by random noise. They proved in [21] (see also Wschebor [26])

Theorem 1.3. *Let $\mathbf{c} = (c_i^{(l)})$ be an arbitrary n by n matrix. Then for any $\varepsilon > 0$,*

$$\mathbf{P}(\sigma_n(\mathbf{c} + \mathbf{f}^{Gau}) \leq \varepsilon n^{-1/2}) = O(\varepsilon).$$

One may ask whether there is an analogue of Theorem 1.2 for this model. The answer is, somewhat surprisingly, negative. However, Tao and Vu managed to prove

Theorem 1.4. *Assume that $\|\mathbf{c}\|_2 \leq n^\gamma$ for some $\gamma > 0$. Then for any $A > 0$, there exists $B = B(A, \gamma)$ such that*

$$\mathbf{P}(\sigma_n(\mathbf{c} + \mathbf{f}^{Ber}) \leq n^{-B}) \leq n^{-A}.$$

For more discussion on this model, we refer to [24]. For applications of Theorem 1.4 in Random Matrix Theory (such as the establishment of the Circular Law) and many related results, we refer to [11] and the references therein.

1.5. Condition numbers for the study of Newton's method. Let $\mathbf{d} = (d_1, \dots, d_{n-1})$ be a degree sequence, and $\mathbf{f} = \{f_1, \dots, f_{n-1}\}$ be a collection of $n - 1$ homogeneous polynomials in n variables of degree d_1, \dots, d_{n-1} respectively,

$$f_l(x_1, \dots, x_n) = \sum_{\substack{\alpha = (\alpha_1, \dots, \alpha_n) \\ \alpha_1 + \dots + \alpha_n = d_l}} \binom{d_l}{\alpha}^{1/2} a_\alpha^{(l)} \mathbf{x}^\alpha,$$

where $\mathbf{x}^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}$.

In their seminal works [16, 17, 18, 19, 20], Shub and Smale initiated a systematic study of Newton's method for finding common roots of the f_i over the unit vectors in \mathbf{C}^n .

Define the Weyl-norm of the system \mathbf{f} by $\|\mathbf{f}\|_W := \sqrt{\|f_1\|_W^2 + \dots + \|f_{n-1}\|_W^2}$, where $\|f_i\|_W^2 := \sum_\alpha |a_\alpha^{(i)}|^2$. For each complex unit vector $\mathbf{x} = (x_1, \dots, x_n)$ in S^{n-1} , we measure the singularity of the system at \mathbf{x} by

$$\mu_{\text{complex}}^{(1)}(\mathbf{f}, \mathbf{x}) = \|\mathbf{f}\|_W \times \|(D_{\mathbf{x}}|_{T_{\mathbf{x}}})^{-1} \Delta\|_2,$$

where $D_{\mathbf{x}}|_{T_{\mathbf{x}}}$ is the Jacobian of the system \mathbf{f} restricted to the tangent space at \mathbf{x} , and Δ is the diagonal matrix of entries $(\sqrt{d_l}, 1 \leq l \leq n - 1)$.

We denote the condition number of the system by

$$\mu_{\text{complex}}^{(1)}(\mathbf{f}) = \sup_{\mathbf{x} \in S^{n-1}, f_1(\mathbf{x}) = \dots = f_{n-1}(\mathbf{x}) = 0} \mu_{\text{complex}}^{(1)}(\mathbf{f}, \mathbf{x}).$$

To analyze the effectiveness of Newton's method for finding common roots of the f_i , Shub and Smale show that, under an invariant probability measure, the condition number of \mathbf{f} is small with high probability.

Theorem 1.6. [17, 9] *Assume that the coefficients $a_\alpha^{(l)}$ are iid standard complex-Gaussian random variables, then*

$$\mathbf{P}(\mu_{\text{complex}}^{(1)}(\mathbf{f}^{\text{Gau}}) > 1/\varepsilon) = O(n^4 N^2 \mathcal{D} \varepsilon^4).$$

Here $\mathcal{D} := \prod d_i$ is the Bezout number and $N := \sum_{i=1}^{n-1} \binom{n-1+d_i}{d_i}$.

Beside finding common complex roots, another important problem is to find common real roots. In a recent series [3, 4, 5], Cucker, Krick, Malajovich and Wschebor have studied this problem in detail. For convenience, Cucker et. al. introduced the following condition number.

For any $\mathbf{x} \in \mathbf{R}^n$, we measure the singularity of the system at \mathbf{x} by

$$\mu_{real}^{(2)}(\mathbf{f}, \mathbf{x}) = \min \left\{ \sqrt{n} \max_i \|f_i\|_W \times \|(D_{\mathbf{x}}|T_{\mathbf{x}})^{-1}\Delta\|_2, \frac{\max_i \|f_i\|_W}{\max_i |f_i(\mathbf{x})|} \right\}.$$

The condition number of the system is then defined as

$$\mu_{real}^{(2)}(\mathbf{f}) := \sup_{\mathbf{x} \in \mathbf{R}^n, \|\mathbf{x}\|_2=1} \mu_{real}^{(2)}(\mathbf{f}, \mathbf{x}).$$

Notice that the definition of $\mu^{(2)}$ is taken over all $\|\mathbf{x}\|_2 = 1$, and thus (with restricted to \mathbf{R}^n) is more general than $\mu^{(1)}$. We recite here a key estimate by Cucker, Krick, Malajovich and Wschebor with respect to $\mu^{(2)}$.

Theorem 1.7. [5] *Assume that $a_{\alpha}^{(i)}$ are iid standard real Gaussian random variables, then*

$$\mathbf{P}(\mu_{real}^{(2)}(\mathbf{f}^{Gau}) > 1/\varepsilon) = O\left(\max_i d_i^2 \sqrt{D} \sqrt{N} n^{5/2} \sqrt{n} \varepsilon \sqrt{\log \frac{1}{\varepsilon \sqrt{n}}}\right),$$

provided that $\varepsilon^{-1} = \Omega(\max_i d_i^2 n^{7/2} N^{1/2})$.

Roughly speaking (see for instance [3] or [2, Section 19]), Cucker, Krick, Malajovich and Wschebor showed that there exists an iterative algorithm that returns the number of real zeros of \mathbf{f} and their approximations and performs $O(\log(nD\mu_{real}^{(2)}(\mathbf{f})))$ iterations with a total cost of

$$O\left([C(n+1)D^2(\mu_{real}^{(2)})^2]^{2(n+1)} N \log(nD\mu_{real}^{(2)}(\mathbf{f}))\right).$$

Henceforth, the probabilistic analysis of $\mu_{real}^{(2)}$, Theorem 1.7, plays a key role in their study.

The proofs of Theorem 1.6 and Theorem 1.7, on the other hand, heavily rely on the invariance property of (real and complex) Gaussian distributions, and are extremely involved.

Motivated by the results discussed in Subsection 1.1, it is natural and important to study the condition numbers μ_1 and μ_2 for polynomial systems under more general distributions such as Bernoulli. This problem is also closely related to a question raised by P. Burgisser and F. Cucker in [2, Problem 7].

Roughly speaking, there are two main technical obstacles of our task: first is the absence of invariance property of distributions and second is the lacking of linear algebra tools (compared to the condition number problem of matrices discussed in Subsection 1.1). As

a result, to our best knowledge, even the following simple and natural question is not even known.

Question 1.8. *Assume that $a_\alpha^{(l)}$ are iid Bernoulli random variables (taking value ± 1 with probability $1/2$). Is it true that with probability tending to 1 (as $n \rightarrow \infty$), there does not exist non-zero vector $\mathbf{x} \in \mathbf{R}^n$ (or $\mathbf{x} \in \mathbf{C}^n$) with $\mathbf{f}(\mathbf{x}) = 0$ and $\text{rank}(D_{\mathbf{x}}|_{T_{\mathbf{x}}}) < n - 1$?*

1.9. Our result. To simplify our work, we will be focusing only on the Kostlan-Shub-Smale model where n is sufficiently large and $d_i = d \geq 2$ for all i . (Note that the case $d_i = 1$ corresponds to rectangular matrices, the reader is invited to consult for instance [15] for related results.) For this uniform system, Theorem 1.6 and Theorem 1.7 read as follows.

Theorem 1.10 (Non-degeneration of uniform homogenous polynomial systems). *Assume that c_α are iid standard complex Gaussian, then*

$$\mathbf{P}(\mu^{(1)}(\mathbf{f}^{Gau}) > 1/\varepsilon) = O\left((n+d)^{O(d)}(d^{n/4}\varepsilon)^4\right).$$

Moreover, if c_α are iid standard real Gaussian random variables, then

$$\mathbf{P}(\mu_{real}^{(2)}(\mathbf{f}^{Gau}) > 1/\varepsilon) = O\left(n^{O(d)}d^{n/2}\varepsilon\sqrt{\log\frac{1}{\varepsilon}}\right).$$

Notice that these bounds are effective only when ε is exponentially small, namely $\varepsilon \ll d^{-n/4}$ in the complex case and $\varepsilon \ll d^{-n/2}$ in the real case (these are the right scaling as the variance of a typical coefficient is d). A closer look at Theorem 1.10 reveals the following.

Heuristic 1.11. *With high probability, for any $\mathbf{x} \in S^{n-1}$, $(\|D_{\mathbf{x}}|_{T_{\mathbf{x}}}\|_2)^{-1}$ and $\|\mathbf{f}^{Gau}(\mathbf{x})\|_2$ cannot be too small at the same time. In other words, such a random system is not “close” to having “double roots” with high probability.*

Although our method can be extended to the complex case, we will be mainly focusing on the real roots to simplify the presentation. Furthermore, as $\mu^{(2)}$ is more general than $\mu^{(1)}$, we will be limited ourself to a quantity similar to $\mu^{(2)}$ only.

Let $d \geq 2$ be an integer. Let $\mathcal{C} = \{c_{i_1 \dots i_d}^{(l)}, 0 \leq i_1, \dots, i_d \leq n, 1 \leq l \leq n-1\}$ be a deterministic system. We consider a random array $\mathcal{A} = \{a_{i_1 \dots i_d}^{(l)}, 0 \leq i_1, \dots, i_d \leq n, 1 \leq l \leq n-1\}$, where $a_{i_1 \dots i_d}^{(l)}$ are iid copies of real random variable ξ with mean zero, variance one, and there exists $T_0 > 0$ such that

$$\forall t > 0 \quad \mathbf{P}(|\xi| \geq t) = O(\exp(-t^2/T_0)). \quad (3)$$

Such subgaussian distributions clearly cover Gaussian and Bernoulli random variables as special cases.

For $\mathbf{x} = (x_1, \dots, x_n) \in S^{n-1}$ of \mathbf{R}^n , we consider a system $\mathbf{f} = (f_1, \dots, f_{n-1})$ of $n-1$ d -linear forms

$$\begin{aligned} f_l(\mathbf{x}) &:= \sum_{1 \leq i_1, \dots, i_d \leq n} c_{i_1 \dots i_d}^{(l)} x_{i_1} \dots x_{i_d} + \sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} x_{i_1} \dots x_{i_d} \\ &:= f_{l,det}(\mathbf{x}) + f_{l,rand}(\mathbf{x}). \end{aligned}$$

In particular, if ξ is the standard Gaussian and the deterministic system vanishes, then for any ordered d -tuples $\alpha = \{i_1 \leq \dots \leq i_d\}$, the coefficient of $\mathbf{x}_\alpha = x_{i_1} \dots x_{i_d}$ is a sum of $\binom{d}{\alpha}$ iid copies of ξ , which in turn can be written as $\sqrt{\binom{d}{\alpha}} \xi_\alpha$ with a standard Gaussian variable ξ_α . This is exactly the model considered by Cucker et. al. as above. Recall that for $\mathbf{x} \in \mathbf{R}^n$, the Jacobian matrix $D_{\mathbf{x}}$ of \mathbf{f} at \mathbf{x} is given by

$$D_{\mathbf{x}} = \left(\frac{\partial f_l(\mathbf{x})}{\partial x_j} \right)_{1 \leq l \leq n-1, 1 \leq j \leq n}.$$

For $1 \leq l \leq n-1$, the gradient of f_l at \mathbf{x} is

$$D_{l,\mathbf{x}}^{(1)} = \left(\frac{\partial f_l}{\partial x_1}, \dots, \frac{\partial f_l}{\partial x_n} \right)$$

while the Hessian is

$$D_{l,\mathbf{x}}^{(2)} = \left(\frac{\partial^2 f_l}{\partial x_i \partial x_j} \right)_{1 \leq i, j \leq n}.$$

In general for $0 \leq k \leq d$, $D_{l,\mathbf{x}}^{(k)}$, the k -th order derivative, is the k -multilinear form

$$D_{l,\mathbf{x}}^{(k)} = \left(\frac{\partial^k f_l}{\partial x_{i_1} \dots \partial x_{i_k}} \right)_{1 \leq i_1, \dots, i_k \leq n}.$$

Define similarly $D_{l,\mathbf{x},det}^{(k)}$, $D_{l,\mathbf{x},rand}^{(k)}$ for the deterministic and random systems respectively.

To control the smallness of $(\|D_{\mathbf{x}}|_{T_{\mathbf{x}}}\|_2)^{-1}$ and $\|\mathbf{f}(\mathbf{x})\|_2$ simultaneously, motivated by [5, p.220], we introduce a function $L(\mathbf{x}, \mathbf{y})$ for $\mathbf{x} \perp \mathbf{y}$ as follows

$$L(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{\|\mathbf{f}(\mathbf{x})\|_2}{(d^5 n)^{1/2}} + \frac{\|D_{\mathbf{x}}(\mathbf{y})\|_2^2}{d^5 n}}.$$

Let L be the minimum value that $L(\mathbf{x}, \mathbf{y})$ can take,

$$L := \min_{\mathbf{x}, \mathbf{y} \in S^{n-1}, \mathbf{x} \perp \mathbf{y}} L(\mathbf{x}, \mathbf{y}).$$

Our first main goal is to show that L cannot be too small with high probability.

Theorem 1.12 (Main theorem, homogeneous system). *Assume that all the coefficients $a_{i_1 \dots i_d}^{(l)}$ are iid copies of a random variable ξ satisfying (3). Then there exist positive constants $K_0 \geq 1$ and c_0 depending only on ξ with $0 < c_0 < 1$ such that*

$$\mathbf{P}(L \leq \varepsilon) \leq K_0^n d^{5n/2} \varepsilon + c_0^n$$

for all $\varepsilon > 0$ and all $2 \leq d \leq n^{\varepsilon_0}$, with ε_0 a sufficiently small constant again depending only on ξ .

We remark that the “error term” c_0^n in Theorem 1.12 is not avoidable in general.

Example 1.13. *With $d = 2$ and $\mathbf{P}(\xi = \pm 1) = 1/2$, it is easy to check that $\mathbf{P}(\mathbf{f}(\mathbf{x}_0) = 0 \wedge D_{\mathbf{x}_0}|T_{\mathbf{x}_0}$ is singular) = $\Omega((3/8)^{-2n})$, where $\mathbf{x}_0 = (1, 1, 0, \dots, 0)$.*

As a consequence of Theorem 1.12, one confirms Question 1.8 and Heuristic 1.11 for a wide range of coefficient distributions.

Corollary 1.14. *With the same assumption as in Theorem 1.16, we have*

- (Non-existence of “double roots” for random discrete systems)

$$\mathbf{P}(\exists \mathbf{x}, \mathbf{y} \in S^{n-1}, \mathbf{x} \perp \mathbf{y} \wedge f(\mathbf{x}) = 0 \wedge D_{\mathbf{x}}(\mathbf{y}) = 0) \leq c_0^n, \quad (4)$$

- (Regularity at roots and non-vanishing at critical points)

$$\begin{aligned} & \max \left\{ \mathbf{P} \left(\exists \mathbf{x}, \mathbf{y} \in S^{n-1}, \mathbf{x} \perp \mathbf{y}, f(\mathbf{x}) = 0 \wedge \|D_{\mathbf{x}}(\mathbf{y})\|_2 \leq d^{5/4} \sqrt{n} \varepsilon \right), \right. \\ & \quad \left. \mathbf{P} \left(\exists \mathbf{x}, \mathbf{y} \in S^{n-1}, \mathbf{x} \perp \mathbf{y}, D_{\mathbf{x}}(\mathbf{y}) = 0 \wedge \|\mathbf{f}(\mathbf{x})\|_2 \leq d^{5/4} n^{1/4} \varepsilon^2 \right) \right\} \\ & \leq K_0^n d^{5n/2} \varepsilon + c_0^n, \end{aligned} \quad (5)$$

- (Simultaneous vanishing)

$$\begin{aligned} & \mathbf{P} \left(\exists \mathbf{x}, \mathbf{y} \in S^{n-1}, \mathbf{x} \perp \mathbf{y}, \|\mathbf{f}(\mathbf{x})\|_2 \leq (d^5 n)^{1/4} \varepsilon \wedge \|\mathbf{f}(\mathbf{y})\|_2 \leq (d^5 n)^{1/4} \varepsilon \right) \\ & \leq K_0^n d^{5n/2} \varepsilon^{1/2} + c_0^n, \end{aligned} \quad (6)$$

where in the last estimate we replaced ε^2 by ε (together with some very generous estimates on $\|D_{\mathbf{x}}(\mathbf{y})\|_2$).

As noted by Example 1.13, (4) is optimal (with respect to exponential decay). Moreover, the RHS of (6) is comparable to the result of Cucker et. al. from Theorem 1.10 in the regime that d is sufficiently large and $d \leq n^{\varepsilon_0}$. Our proof shows that the error term c_0^n from Theorem 1.16 is felt at “sparse” vectors (such as \mathbf{x}_0 from Example 1.13).

More importantly, our method extends to perturbed systems under appropriate assumptions upon the deterministic system \mathcal{C} .

Definition 1.15. We say that the deterministic system \mathcal{C} is γ -controlled if

$$\max \left(\sup_{\mathbf{x} \in S^{n-1}} \|\mathbf{f}_{det}(\mathbf{x})\|_2^2, \sup_{\mathbf{x}, \mathbf{y}_1 \in S^{n-1}} \|D_{\mathbf{x}, det}^{(1)}(\mathbf{y}_1)\|_2^2, \dots, \sup_{\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_d \in S^{n-1}} \|D_{\mathbf{x}, det}^{(d)}(\mathbf{y}_1, \dots, \mathbf{y}_d)\|_2^2 \right) \leq n^\gamma, \quad (7)$$

where $\|D_{\mathbf{x}, det}^{(1)}(\mathbf{y}_1)\|_2 = \sqrt{\sum_{1 \leq l \leq n-1} (D_{l, \mathbf{x}, det}^{(1)} \mathbf{y}_1^T)^2}$ and so on.

Theorem 1.16 (Main theorem, perturbed systems). *Assume that \mathcal{C} is a deterministic system satisfying (7) with $\gamma \leq 19/18$ and that all the coefficients $a_{i_1 \dots i_d}^{(l)}$ are iid copies of a random variable ξ satisfying (3). Then there exist positive constants K_0 and c_0 depending only on ξ and γ with $0 < c_0 < 1$ such that*

$$\mathbf{P}(L \leq \varepsilon) \leq K_0^n (d^{5/2} + n^{\gamma/2-1/2})^n \varepsilon + c_0^n$$

for all $2 \leq d \leq n^{\varepsilon_0}$ with ε_0 a sufficiently small absolute constant depending on ξ and γ .

We have not tried to optimize the constant $19/18$ on γ , but our method does not seem to extend to the whole $\gamma = O(1)$ regime. On the other hand, the result remains valid if we assume d sufficiently large depending on γ (see Remark 3.6).

We believe that our result will be useful for the study of universality problems for roots and critical points of general random polynomial systems. The reader is invited to consult for instance [12, Lemma 6] for a recent application of this type for univariate random polynomials.

The rest of the note is organized as follows. The main ideas to prove Theorem 1.16 is introduced in Section 2. Sections 3, 4 and 5 will be devoted to prove the main ingredients subsequently.

2. PROOF OF THEOREM 1.16: THE IDEAS

Our treatment will be for general $\gamma = O(1)$. The upper bound of γ will be required at the end of Section 3. As there is nothing to prove if $\varepsilon > d^{-5n/2}$, we will assume $\varepsilon \leq d^{-5n/2}$. We will verify Theorem 1.16 for

$$n^{-(\gamma/2-17/36)n} \leq \varepsilon \leq d^{-5n/2}. \quad (8)$$

The result for $\varepsilon \leq n^{-(\gamma/2-17/36)n}$ easily follows as $K_0^n (d^{5/2} + n^{\gamma/2-1/2})^n n^{-(\gamma/2-17/36)n} = o(c_0^n)$, provided that ε_0 is sufficiently small and n is sufficiently large.

2.1. Growth of function. First of all, we will invoke the following bound.

Theorem 2.2. *Assume that ξ has zero mean, unit variance, and satisfies (3). Then there exists an absolute positive constant C_0 independent of d such that the following holds with probability at least $1 - \exp(-dn/2)$*

$$\max \left(\sup_{\mathbf{x} \in S^{n-1}} \|\mathbf{f}_{rand}(\mathbf{x})\|_2^2, \sup_{\mathbf{x}, \mathbf{y}_1 \in S^{n-1}} \|D_{\mathbf{x}, rand}^{(1)}(\mathbf{y}_1)\|_2^2, \sup_{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2 \in S^{n-1}} \|D_{\mathbf{x}, rand}^{(2)}(\mathbf{y}_1, \mathbf{y}_2)\|_2^2 \right) \leq C_0 d^5 n, \quad (9)$$

and

$$\max \left(\sup_{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 \in S^{n-1}} \|D_{\mathbf{x}, rand}^{(3)}(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)\|_2^2, \dots, \sup_{\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_d \in S^{n-1}} \|D_{\mathbf{x}, rand}^{(d)}(\mathbf{y}_1, \dots, \mathbf{y}_d)\|_2^2 \right) \leq n^{\omega(d)}. \quad (10)$$

Notice that (10) is rather straightforward because $d \leq n^{\varepsilon_0}$, and without affecting the probability much, one can assume that all of the coefficients $a_{i_1 \dots i_d}^{(l)}$ are bounded by $n^{O(1)}$. The proof of the less trivial estimate, (9), will be presented in Section 4.

Together with condition (7) of $c_{i_1 \dots i_d}$ and by the triangle inequality, we obtain a similar bound for the perturbed system $\mathbf{f} = \mathbf{f}_{det} + \mathbf{f}_{rand}$.

Theorem 2.3. *With probability at least $1 - \exp(-dn/2)$, the following holds*

$$\max \left(\sup_{\mathbf{x} \in S^{n-1}} \|\mathbf{f}(\mathbf{x})\|_2^2, \sup_{\mathbf{x}, \mathbf{y}_1 \in S^{n-1}} \|D_{\mathbf{x}}^{(1)}(\mathbf{y}_1)\|_2^2, \sup_{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2 \in S^{n-1}} \|D_{\mathbf{x}}^{(2)}(\mathbf{y}_1, \mathbf{y}_2)\|_2^2 \right) \leq C_0 d^5 n + n^\gamma \quad (11)$$

and

$$\max \left(\sup_{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 \in S^{n-1}} \|D_{\mathbf{x}}^{(3)}(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)\|_2^2, \dots, \sup_{\mathbf{x}, \mathbf{y}_1, \dots, \mathbf{y}_d \in S^{n-1}} \|D_{\mathbf{x}}^{(d)}(\mathbf{y}_1, \dots, \mathbf{y}_d)\|_2^2 \right) \leq n^{\omega(d)}. \quad (12)$$

Next, we translate the assumption of $L \leq \varepsilon$ into slow growth of \mathbf{f} .

Claim 2.4 (Growth of function). *With probability at least $1 - \exp(-dn)$, the following holds. Assume that $L(\mathbf{x}, \mathbf{y}) \leq \varepsilon$ for some $\mathbf{x}, \mathbf{y} \in S^{n-1}$ with $\mathbf{x} \perp \mathbf{y}$, then for any $t \in \mathbf{R}$ with $|t| \leq 1$ and any $\mathbf{z} \in \mathbf{R}^n$ with $\|\mathbf{z}\|_2 \leq 1$,*

$$\|\mathbf{f}(\mathbf{x} + \varepsilon t \mathbf{y} + \varepsilon^2 \mathbf{z})\|_2 \leq C'_0 (d^{5/2} + n^{\gamma/2-1/2}) \sqrt{n} \varepsilon^2, \quad (13)$$

where C'_0 is an absolute constant.

Proof. (of Claim 2.4) We condition on the events considered in Theorem 2.2 and Theorem 2.3. First of all, for each $1 \leq l \leq n-1$, by Taylor expansion

$$f_l(\mathbf{x} + \varepsilon t \mathbf{y} + \varepsilon^2 \mathbf{z}) = f_l(\mathbf{x}) + D_{l, \mathbf{x}}^{(1)}(\varepsilon t \mathbf{y} + \varepsilon^2 \mathbf{z})^T + \frac{1}{2}(\varepsilon t \mathbf{y} + \varepsilon^2 \mathbf{z}) D_{l, \mathbf{x}}^{(2)}(\varepsilon t \mathbf{y} + \varepsilon^2 \mathbf{z})^T + o(\varepsilon^2),$$

where we used (12) for the remainder, noting that $\varepsilon \leq d^{-5n/2}$.

By the triangle inequality,

$$\begin{aligned} |f_l(\mathbf{x} + \varepsilon t\mathbf{y} + \varepsilon^2\mathbf{z})| &\leq |f_l(\mathbf{x})| + \varepsilon |D_{l,\mathbf{x}}^{(1)}\mathbf{y}^T| + \varepsilon^2 |D_{l,\mathbf{x}}^{(1)}\mathbf{z}^T| + \frac{1}{2}\varepsilon^2 |(t\mathbf{y} + \varepsilon\mathbf{z})D_{l,\mathbf{x}}^{(2)}(t\mathbf{y} + \varepsilon\mathbf{z})^T| + o(\varepsilon^2) \\ &\leq |f_l(\mathbf{x})| + \varepsilon |D_{l,\mathbf{x}}^{(1)}\mathbf{y}^T| + \varepsilon^2 |D_{l,\mathbf{x}}^{(1)}\mathbf{z}^T| + \varepsilon^2(1 + \varepsilon^2) |\mathbf{u}D_{l,\mathbf{x}}^{(2)}\mathbf{u}^T| + o(\varepsilon^2), \end{aligned}$$

where $\mathbf{u} := (t\mathbf{y} + \varepsilon\mathbf{z})/\sqrt{2(t^2 + \varepsilon^2)}$ (and hence $\|\mathbf{u}\|_2 \leq 1$).

By Theorem 2.3, $\sum_l |D_{l,\mathbf{x}}^{(1)}\mathbf{z}^T|^2$ and $\sum_l |\mathbf{u}D_{l,\mathbf{x}}^{(2)}\mathbf{u}^T|^2$ are smaller than $2(C_0d^{5/2}n + n^\gamma)$. As such, by Cauchy-Schwarz inequality

$$\begin{aligned} \sum_l f_l^2(\mathbf{x} + \varepsilon t\mathbf{y} + \varepsilon^2\mathbf{z}) &\leq 4 \sum_l f_l^2(\mathbf{x}) + 4\varepsilon^2 \sum_l (D_{l,\mathbf{x}}^{(1)}\mathbf{y}^T)^2 \\ &\quad + 4\varepsilon^4 \sum_l (D_{l,\mathbf{x}}^{(1)}\mathbf{z}^T)^2 + 4\varepsilon^4 \sum_l (\mathbf{u}D_{l,\mathbf{x}}^{(2)}\mathbf{u}^T)^2 + o(\varepsilon^4), \\ &\leq 4d^5n\varepsilon^4 + 4d^5n\varepsilon^4 + 8(C_0d^5n + n^\gamma)\varepsilon^4 + 8(C_0d^5n + n^\gamma)\varepsilon^4 + o(\varepsilon^4), \end{aligned}$$

where we used the assumption that

$$\sum_l |f_l(\mathbf{x})|^2 = \|\mathbf{f}(\mathbf{x})\|_2^2 \leq d^5nL^4 \leq d^5n\varepsilon^4$$

and

$$\sum_l |D_{l,\mathbf{x}}^{(1)}\mathbf{y}^T|^2 = \|D_{\mathbf{x}}(\mathbf{y})\|_2^2 \leq d^5nL^2 \leq d^5n\varepsilon^2.$$

Thus

$$\|f(\mathbf{x} + \varepsilon t\mathbf{y} + \varepsilon^2\mathbf{z})\|_2 \leq C'_0(d^{5/2} + n^{\gamma/2-1/2})\sqrt{n}\varepsilon^2. \quad \square$$

Notice that as $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, the distance from $\mathbf{x} + \varepsilon t\mathbf{y} + \varepsilon^2\mathbf{z}$ to S^{n-1} is at most $2\varepsilon^2$, and so

$$\mathbf{x} + \varepsilon\mathbf{y} + \varepsilon^2\mathbf{z} \in S_{\varepsilon^2} := S^{n-1} + B(0, 2\varepsilon^2).$$

With this notation, because the set $\{\mathbf{x} + \varepsilon t\mathbf{y} + \varepsilon^2\mathbf{z}, \|\mathbf{z}\|_2 \leq 1, |t| \leq 1\}$ has volume at least $\frac{\pi^{n/2}}{\Gamma(n/2+1)}\varepsilon^{2(n-1)+1}$, (13) implies that there exists $A \subset S_{\varepsilon^2}$ with volume at least $\frac{\pi^{n/2}}{\Gamma(n/2+1)}\varepsilon^{2(n-1)+1}$ such that $\|f(\mathbf{a})\|_2 \leq C'_0(d^{5/2} + n^{\gamma/2-1/2})\sqrt{n}\varepsilon^2$ for all $\mathbf{a} \in A$. Thus, in order to prove Theorem 1.16 it suffices to show the following.

Theorem 2.5. *There exist K_0, c_0 such that the following holds*

$$\begin{aligned} \mathbf{P}\left(\exists A \subset S_{\varepsilon^2} : \mu(A) \geq \frac{\pi^{n/2}}{\Gamma(n/2+1)} \varepsilon^{2(n-1)+1} \wedge \|f(\mathbf{a})\|_2 \leq C'_0(d^{5/2} + n^{\gamma/2-1/2})\sqrt{n}\varepsilon^2 \quad \forall \mathbf{a} \in A\right) \\ \leq K_0^n (d^{5/2} + n^{\gamma/2-1/2})^n \varepsilon + c_0^n. \end{aligned}$$

2.6. Hypothetical assumption. For $\mathbf{x} \in S_{\varepsilon^2}$, let $E_{\mathbf{x}}$ be the event that $\|f(\mathbf{x})\|_2 \leq C'_0(d^{5/2} + n^{\gamma/2-1/2})\sqrt{n}\varepsilon^2$. Assume that the following holds for all $\mathbf{x} \in S_{\varepsilon^2}$

$$\mathbf{P}(E_{\mathbf{x}}) = \mathbf{P}\left(\|f(\mathbf{x})\|_2 \leq C'_0(d^{5/2} + n^{\gamma/2-1/2})\sqrt{n}\varepsilon^2\right) \leq C_0''^n (d^{5/2} + n^{\gamma/2-1/2})^n \varepsilon^{2(n-1)}, \quad (14)$$

for some absolute constant C_0'' . Then as

$$\text{Vol}(S_{\varepsilon^2}) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} ((1+2\varepsilon^2)^n - (1-2\varepsilon^2)^n) = O\left(\frac{n\pi^{n/2}}{\Gamma(n/2+1)} \varepsilon^2\right),$$

one would have

$$\int_{\mathbf{x} \in S_{\varepsilon^2}} \mathbf{P}(E_{\mathbf{x}}) d\mu(\mathbf{x}) = O\left(\frac{n\pi^{n/2}}{\Gamma(n/2+1)} C_0''^n (d^{5/2} + n^{\gamma/2-1/2})^n \varepsilon^{2n}\right).$$

By using Markov's bound and Fubini, one thus infers that

$$\begin{aligned} \mathbf{P}\left(\mu\{\mathbf{x} \in S_{\varepsilon^2} : E_{\mathbf{x}}\} \geq \frac{\pi^n}{\Gamma(n+1)} \varepsilon^{2(n-1)+1}\right) &\leq n C_0''^n (d^{5/2} + n^{\gamma/2-1/2})^n \varepsilon^{2n} / \varepsilon^{2(n-1)+1} \\ &= n C_0''^n (d^{5/2} + n^{\gamma/2-1/2})^n \varepsilon. \end{aligned}$$

One would then be done with proving Theorem 2.5 by setting $K_0 = 2C_0''$.

However, the assumption (14) is not always true. Our next goal is to characterize those \mathbf{x} with $\mathbf{P}(E_{\mathbf{x}}) > C_0'''^n (d^{5/2} + n^{\gamma/2-1/2})^n \varepsilon^{2(n-1)}$. For short, set

$$M_d := C'_0(d^{5/2} + n^{\gamma/2-1/2}). \quad (15)$$

Recall that $E_{\mathbf{x}}$ is the event $\|f(\mathbf{x})\|_2 \leq C'_0(d^{5/2} + n^{\gamma/2-1/2})\sqrt{n}\varepsilon^2 = M_d\sqrt{n}\varepsilon^2$. This is exactly a concentration event in a small ball. Fortunately, the latter has been studied extensively in the context of random matrix. In what follows we will introduce some key lemmas, our approach follows [14].

2.7. Diophantine Structure. Let y_1, \dots, y_m be real numbers. Rudelson and Vershynin [14] defined the essential *least common denominator* (**LCD**) of $\mathbf{y} = (y_1, \dots, y_m)$ as follows. Fix parameters α and γ_0 , where $\gamma_0 \in (0, 1)$, and define

$$\mathbf{LCD}_{\alpha, \gamma_0}(\mathbf{y}) := \inf \left\{ D > 0 : \text{dist}(D\mathbf{y}, \mathbf{Z}^m) < \min(\gamma_0 \|D\mathbf{y}\|_2, \alpha) \right\}.$$

Here $\text{dist}(\mathbf{A}, \mathbf{Z}^m) := \inf_{\mathbf{a} \in \mathbf{A}, \mathbf{z} \in \mathbf{Z}^m} \|\mathbf{a} - \mathbf{z}\|_2$. One typically assumes γ_0 to be a small constant. The inequality $\text{dist}(D\mathbf{y}, \mathbf{Z}^m) < \alpha$ then yields that most coordinates of $\theta\mathbf{a}$ are within a small distance from non-zero integers.

Theorem 2.8. [14][15, Theorem 3.3] *Consider a sequence $\mathbf{y} = (y_1, \dots, y_m)$ of real numbers which satisfies $\sum_{i=1}^m y_i^2 \geq 1$. Assume that a_i are iid copies of ξ satisfying (3). Then, for every $\alpha > 0$ and $\gamma_0 \in (0, 1)$, and for*

$$\varepsilon \geq \frac{1}{\mathbf{LCD}_{\alpha, \gamma_0}(\mathbf{y})},$$

we have

$$\sup_{y \in \mathbf{C}} \mathbf{P}_{a_1, \dots, a_m} \left(\left| \sum_{1 \leq i \leq m} a_i y_i - y \right| \leq \varepsilon \right) \leq C_1 \left(\frac{\varepsilon}{\gamma_0} + e^{-2\alpha^2} \right),$$

where C_1 is an absolute constant.

In application we will set $m = n^d$, while $\mathbf{y}_{\mathbf{x}} = (x_{i_1} \dots x_{i_d})_{1 \leq i_1, \dots, i_d \leq n}$ and $a_{i_1 \dots i_d}^{(l)}$ will play the role of \mathbf{y} and of the a_i 's respectively. As $\mathbf{x} \in S_{\varepsilon^2}$, one has

$$\|\mathbf{y}_{\mathbf{x}}\|_2^2 = \|\mathbf{x}\|_2^2 \geq (1 - 2\varepsilon^2)^2 = 1 - O(\varepsilon^2).$$

We will choose $\gamma_0 = 1/2$ and

$$\alpha := \begin{cases} n^{7d/16-1/4} & \text{if } 2 \leq d = o(\log n / \log \log n), \\ n^{d/4} & \text{otherwise.} \end{cases} \quad (16)$$

Observe from Theorem 2.8 that if $(\mathbf{LCD}_{\alpha, \gamma_0}(\mathbf{y}_{\mathbf{x}}))^{-1} \leq M_d \varepsilon^2$ (with M_d from (15)) then

$$\mathbf{P}_{a_1, \dots, a_m} \left(\left| \sum_{1 \leq i \leq m} a_i y_i \right| \leq M_d \varepsilon^2 \right) \leq C_1 (2M_d \varepsilon^2 + e^{-2\alpha^2}) \leq 4C_1 M_d \varepsilon^2,$$

as one can check from (8) that $\varepsilon \geq n^{-(\gamma-1/3)n} \geq \exp(-n^{5/4}/2) \geq \exp(-\alpha^2/2)$.

Thus

$$\mathbf{P}(|f_i(\mathbf{x})| \leq M_d \varepsilon^2) \leq 4C_1 M_d \varepsilon^2.$$

In fact, Theorem 2.8 also implies that $\mathbf{P}(|f_i(\mathbf{x})| \leq M_d \delta^2) \leq 4C_1 M_d \delta^2$ for any $\delta \geq \varepsilon$. Before proceeding further, we will need the following tensorization trick.

Lemma 2.9. [14, Lemma 2.2] *Let K, δ_0 be given. Assume that $\mathbf{P}(|f_i(\mathbf{x})| < \delta) \leq K\delta^2$ for all $\delta \geq \delta_0$. Then*

$$\mathbf{P}(\|f(\mathbf{x})\|_2 < \delta\sqrt{n-1}) \leq (C_0K\delta)^{n-1}$$

where C_0 is an absolute constant.

For the sake of completeness, we will present a short proof of Lemma 2.9 in Appendix A.

By independence and by Lemma 2.9, we have

$$\begin{aligned} \mathbf{P}(\|f(\mathbf{x})\|_2 \leq M_d n^{1/2} \varepsilon^2) &= \mathbf{P}(\sqrt{f_1^2(\mathbf{x}) + \cdots + f_{n-1}^2(\mathbf{x})} \leq M_d n^{1/2} \varepsilon^2) \\ &\leq (4C_0C_1)^{n-1} (M_d \varepsilon^2)^{n-1} \\ &\leq K_0^{n-1} (M_d \varepsilon^2)^{n-1}, \end{aligned}$$

with $K_0 := 4C_0C_1$. Thus we have shown the following.

Theorem 2.10. *If $\mathbf{x} \in S_\varepsilon^2$ and $(\mathbf{LCD}_{\alpha,1/2}(\mathbf{y}_\mathbf{x}))^{-1} \geq M_d \varepsilon^2$, then*

$$\mathbf{P}(E_\mathbf{x}) = \mathbf{P}(\|f(\mathbf{x})\|_2 \leq M_d \varepsilon^2 n^{1/2}) \leq K_0^{n-1} (M_d^2 \varepsilon^2)^{n-1}.$$

It remains to focus on \mathbf{x} with relatively small $\mathbf{LCD}(\mathbf{y}_\mathbf{x})$,

$$\mathbf{LCD}_{\alpha,1/2}(\mathbf{y}_\mathbf{x}) < (M_d \varepsilon^2)^{-1} =: \varepsilon'^{-1}. \quad (17)$$

Thus the upper bound $M_d \sqrt{n} \varepsilon^2$ in Claim 2.4 becomes $\sqrt{n} \varepsilon'$. The proof of Theorem 2.5 is complete if one can show the following.

Theorem 2.11. *There exists an absolute constant $c_0 \in (0, 1)$ such that*

$$\mathbf{P}(\exists \mathbf{x} \in S_{\varepsilon^2} : \mathbf{LCD}_{\alpha,1/2}(\mathbf{y}_\mathbf{x}) \leq \varepsilon'^{-1} \wedge \|f(\mathbf{x})\|_2 \leq \sqrt{n} \varepsilon') \leq c_0^n.$$

Indeed, by Theorem 2.11, with probability at least $1 - c_0^n$, for all $\mathbf{a} \in S_{\varepsilon^2}$ with $\mathbf{LCD}_{\alpha,1/2}(\mathbf{y}_\mathbf{a}) \leq \varepsilon'^{-1}$ one has $\|\mathbf{f}(\mathbf{a})\|_2 > \sqrt{n} \varepsilon'$. Conditioning on this event, all of the elements \mathbf{a} of the set A in Theorem 2.5 must have $\mathbf{LCD}_{\alpha,1/2}(\mathbf{y}_\mathbf{a}) \geq \varepsilon'^{-1}$. But then the conclusion of Theorem 2.5 follows from Theorem 2.10 via an application of Fubini and Makov's bound.

Before proving Theorem 2.11, it is important to remark that if there exists $\mathbf{x}_0 \in S_{\varepsilon^2}$ satisfying $\mathbf{LCD}_{\alpha,1/2}(\mathbf{y}_{\mathbf{x}_0}) \leq \varepsilon'^{-1}$ such that $\|f(\mathbf{x}_0)\| \leq \sqrt{n} \varepsilon'$, then the normalized vector $\mathbf{x}_1 = \mathbf{x}_0 / \|\mathbf{x}_0\| \in S^{n-1}$ satisfies

$$\mathbf{LCD}_{\alpha,1/2}(\mathbf{y}_{\mathbf{x}_1}) \leq (1 + 2\varepsilon^2) \varepsilon'^{-1} = (1 + o(1)) \varepsilon'^{-1}$$

and

$$\|f(\mathbf{x}_1)\|_2 \leq (1 + 2\varepsilon^2)^d \sqrt{n}\varepsilon' = (1 + o(1))\sqrt{n}\varepsilon', \quad (18)$$

where we used the assumption that ε is sufficiently small (recall from (8) that $\varepsilon \leq d^{-n/4}$).

Hence it is enough to prove Theorem 2.11 for $\mathbf{x} \in S^{n-1}$ only. We next introduce two different types of vectors depending on their *sparsity*.

Definition 2.12. Let $\delta, \rho \in (0, 1)$ be sufficiently small (depending on d). A vector $\mathbf{x} \in \mathbf{R}^n$ is called *sparse* if $|\text{supp}(\mathbf{x})| \leq \delta n$. A vector $\mathbf{x} \in S^{n-1}$ is called *compressible* if \mathbf{x} is within Euclidean distance ρ from the set of all sparse vectors. A vector $\mathbf{x} \in S^{n-1}$ is called *incompressible* if it is not compressible. The sets of compressible and incompressible vectors will be denoted by $\text{Comp}(\delta, \rho)$ and $\text{Incomp}(\delta, \rho)$ respectively.

In what follows we will choose

$$\delta = \rho = \kappa_0/d^2, \quad (19)$$

where κ_0 is a sufficiently small absolute constant .

Theorem 2.13. *There exists a positive constant $0 < c_0 < 1$ such that the probability that there exists a compressible vector $\mathbf{x} \in \text{Comp}(\delta, \rho)$ with $\|f(\mathbf{x})\|_2 \leq (1 + o(1))\varepsilon'\sqrt{n}$ is bounded by c_0^n .*

The proof of Theorem 2.13 will be presented in Section 5. Notice that this is where the error term c_0^n arises in Theorem 1.16, which is unavoidable owing to Example 1.13. We remark further that Theorem 2.13 holds as long as $\varepsilon' = o(1)$.

Our main analysis lies in the treatment for incompressible structural vectors.

Theorem 2.14. *Conditioning on the event considered in Theorem 2.2, the probability that there exists an incompressible \mathbf{x} in S_{ε^2} with $\mathbf{LCD}_{\alpha, 1/2}(\mathbf{y}_{\mathbf{x}}) \leq (1 + o(1))\varepsilon'^{-1}$ such that $\|f(\mathbf{x})\|_2 \leq (1 + o(1))\varepsilon'\sqrt{n}$ is bounded by $O(n^{-(1/16 - o(1))n})$.*

3. PROOF OF THEOREM 2.14

First of all, incompressible vectors spread out thanks to the following observation.

Fact 3.1. [14, Lemma 3.4] *Let $\mathbf{x} \in \text{Incomp}(\delta, \rho)$. Then there exists a set $\sigma \subset \{1, \dots, n\}$ of cardinality $|\sigma| \geq \rho^2 \delta n / 2$ such that*

$$\frac{\rho}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\delta n}}, \forall k \in \sigma.$$

With the choice of δ and ρ from (19),

$$\frac{\kappa_0}{2d^2\sqrt{n}} \leq |x_k| \leq \frac{d}{\kappa_0\sqrt{n}}. \quad (20)$$

As such, there are at least σ^d product terms $x_{i_1} \dots x_{i_d}$ with $|x_{i_1} \dots x_{i_d}| \geq (\frac{\kappa_0}{2d^2})^d n^{-d/2}$. By definition, the **LCD** of $\mathbf{y}_\mathbf{x}$ is then at least $n^{d/2}/(O(d))^{O(d)}$, where the implied constants depend on κ_0 .

We divide $[n^{d/2}/(O(d))^{O(d)}, \varepsilon'^{-1}]$ into dyadic intervals. For $n^{d/2}/(O(d))^{O(d)} \leq D \leq \varepsilon'^{-1}$, define

$$S_D := \{\mathbf{x} \in S^{n-1}, D \leq \mathbf{LCD}_{\alpha, 1/2}(\mathbf{y}_\mathbf{x}) \leq 2D\}.$$

It follows from the definition of α from (16) that $\alpha \ll n^{d/2}/(O(d))^{O(d)} \leq D$. Our next lemma is an upper bound for any fixed incompressible vector.

Lemma 3.2 (Treatment for a single vector). *Assume that $\mathbf{x} \in S_D$. Then for any $t > 1/D$*

$$\mathbf{P}(\|f(\mathbf{x})\|_2 < t\sqrt{n}) \leq (4C_0C_1)^{n-1} t^{2(n-1)}.$$

Proof. (of Lemma 3.2) The claim follows from the definition of $\mathbf{LCD}_{\alpha, 1/2}(\mathbf{y}_\mathbf{x})$, Theorem 2.8 and Lemma 2.9. \square

3.3. Approximation by structure. Recall that $\mathbf{x} \in S_D$ if $D \leq \mathbf{LCD}_{\alpha, 1/2}(\mathbf{y}_\mathbf{x}) \leq 2D$. Observe that $\mathbf{y}_\mathbf{x}$ is a vector in \mathbf{R}^{n^d} with rich multiplicative structure. The main goal of this section is to translate this piece of diophantine information on $\mathbf{y}_\mathbf{x}$ to \mathbf{x} itself.

Lemma 3.4 (Nets of the level sets). *There exists a $d^{O(d)}\alpha/D$ -net \mathcal{M}_D of S_D in the Euclidean metric of cardinality*

$$|\mathcal{M}_D| \leq \binom{n}{d-1} \left(1 + d^{O(d)}D/\alpha\right)^d \times \left(1 + d^{O(d)}D/n^{d/2}\right)^{n-d+1}.$$

Proof. (of Lemma 3.4) By definition of **LCD**,

$$\sum_{1 \leq i_1, \dots, i_d \leq n} \|D(\mathbf{x})x_{i_1} \dots x_{i_d} - p_{ij}\|_{\mathbf{R}/\mathbf{Z}}^2 \leq \alpha^2$$

for some $D \leq D(\mathbf{x}) \leq 2D$.

As there are $|\sigma| \geq \rho^2 \delta n/2$ indices i satisfying (20), by the pigeon-hole principle, there exist $d-1$ indices i_1, \dots, i_{d-1} where x_{i_j} satisfies (20) and such that

$$\begin{aligned} \sum_{1 \leq j \leq n} \|D(\mathbf{x})x_{i_1} \dots x_{i_{d-1}}x_j\|_{\mathbf{R}/\mathbf{Z}}^2 &\leq \alpha^2/(\rho^2\delta/2)^d n^{d-1} = \alpha^2 d^{6d} \kappa_0^{-3d} / 2^d n^{d-1} \\ &:= f_0^2 \alpha^2 / n^{d-1}. \end{aligned} \quad (21)$$

Without loss of generality, one assumes that $i_1 = 1, \dots, i_{d-1} = d-1$. Fix x_1, \dots, x_{d-1} for the moment. Set

$$D' := D(\mathbf{x})x_1 \dots x_{d-1}.$$

Then as $D \leq D(\mathbf{x}) \leq 2D$ and the x_i 's satisfy (20),

$$\left(\frac{\kappa_0}{2d^2}\right)^{d-1} \frac{D}{n^{(d-1)/2}} := f_1 \frac{D}{n^{(d-1)/2}} \leq |D'| \leq 2\left(\frac{d}{\kappa_0}\right)^{d-1} \frac{D}{n^{(d-1)/2}} := f_2 \frac{D}{n^{(d-1)/2}}. \quad (22)$$

By definition and from (21), with $\mathbf{x}' := (x_d, \dots, x_n)$, there exists $\mathbf{p} = (p_d, \dots, p_n) \in \mathbf{Z}^{n-d+1}$ such that

$$\|D'\mathbf{x}' - \mathbf{p}\|_2 \leq f_0 \alpha / n^{(d-1)/2}.$$

So

$$\|\mathbf{x}' - \frac{1}{D'}\mathbf{p}\|_2 \leq f_0 \alpha / |D'| n^{(d-1)/2} \leq (f_0/f_1) \alpha / D, \quad (23)$$

where we used the lower bound for $|D'|$ from (22).

Notice furthermore that

$$\begin{aligned} \|\mathbf{p}\|_2 &\leq \|D'\mathbf{x}'\|_2 + f_0 \alpha / n^{(d-1)/2} \leq |D'| + f_0 \alpha / n^{(d-1)/2} \\ &\leq f_2 D / n^{(d-1)/2} + f_0 \alpha / n^{(d-1)/2} \\ &\leq d^{O(d)} D / n^{(d-1)/2}. \end{aligned}$$

The collection \mathcal{P} of such integral vectors \mathbf{p} has size at most

$$|\mathcal{P}| \leq \left(1 + (d^{O(1)} D / n^{(d-1)/2}) / \sqrt{n}\right)^{n-d+1} \leq \left(1 + d^{O(d)} D / n^{d/2}\right)^{n-d+1}.$$

Next, for the set $|z| \leq n^{(d-1)/2} / (f_1 D)$ in \mathbf{R} we choose an ε_d -net \mathcal{N}_{local} with $\varepsilon_d = (f_0/f_1) n^{(d-1)/2} \alpha / D (f_1 D + f_0 \alpha)$. Clearly we can choose \mathcal{N}_{local} so that

$$|\mathcal{N}_{local}| \leq 1 + 2n^{(d-1)/2}/(\varepsilon_d f_1 D) = 2(f_1/f_0)D/\alpha + 1 \leq 1 + d^{O(d)}D/\alpha.$$

Define the following set in \mathbf{R}^{n-d+1}

$$\mathcal{N}_{1\dots(d-1)} := \{b\mathbf{p}, b \in \mathcal{N}_{local}, \mathbf{p} \in \mathcal{P}\}.$$

By definition,

$$|\mathcal{N}_{1\dots(d-1)}| \leq \left(1 + d^{O(d)}D/\alpha\right) \times \left(1 + d^{O(d)}(D + \alpha)/n^{d/2}\right)^{n-d+1}. \quad (24)$$

Moreover, as $|1/D'| \leq n^{(d-1)/2}/(f_1 D)$, there exists $b \in \mathcal{N}_{local}$ such that $|1/D' - b| \leq \varepsilon_d$. As such, by (23)

$$\begin{aligned} \|\mathbf{x}' - b\mathbf{p}\|_2 &\leq \|\mathbf{x}' - \frac{1}{D'}\mathbf{p}\|_2 + \left\|\left(\frac{1}{D'} - b\right)\mathbf{p}\right\|_2 \\ &\leq (f_0/f_1)\alpha/D + \varepsilon_d((f_1 D + f_0 \alpha)/n^{(d-1)/2}) \\ &\leq 2(f_0/f_1)\alpha/D. \end{aligned}$$

Thus $\mathcal{N}_{1\dots(d-1)}$ is an $2(f_0/f_1)\alpha/D$ -net for $\mathbf{x}' = (x_d, \dots, x_n)$.

To continue, one approximates (x_1, \dots, x_{d-1}) by an arbitrary $(f_0/f_1)\alpha/D$ -net in $|z| \leq 1$ of \mathbf{R}^{d-1} . We therefore obtain a net $\mathcal{N}'_{1\dots(d-1)}$ that $3(f_0/f_1)\alpha/D$ -approximates the vector (x_1, \dots, x_n) , which has size

$$\begin{aligned} |\mathcal{N}'_{1\dots(d-1)}| &\leq (1 + 2(f_1/f_0)D/\alpha)^{d-1} \times |\mathcal{N}_{1\dots(d-1)}| \\ &\leq \left(1 + d^{O(d)}D/\alpha\right)^d \times \left(1 + d^{O(d)}D/n^{d/2}\right)^{n-d+1}. \end{aligned}$$

In summary, for each $d-1$ tuple i_1, \dots, i_{d-1} , one obtains a net $\mathcal{N}'_{i_1, \dots, i_d}$ (by fixing $x_{i_1}, \dots, x_{i_{d-1}}$ instead of x_1, \dots, x_{d-1}). The union set \mathcal{M}_D of all $\mathcal{N}'_{i_1, \dots, i_{d-1}}$ will satisfy the conclusion of our theorem. \square

3.5. Passing from \mathcal{M}_D to S_D . Assume that there exists \mathbf{x} with $D < \mathbf{LCD}(\mathbf{y}_{\mathbf{x}}) \leq 2D$ such that $\|f(\mathbf{x})\|_2 \leq \alpha\sqrt{n}/D$. Choose $\mathbf{x}_0 \in \mathcal{M}_D$ which is $3(f_0/f_1)\alpha/D$ -approximates \mathbf{x} . By conditioning on the event of Theorem 2.2,

$$\begin{aligned}
\|f(\mathbf{x})\|_2 &\leq \|f(\mathbf{x}_0)\|_2 + (\sqrt{C_0}d^{1/4}\sqrt{n})3(f_0/f_1)\alpha/D \leq \alpha\sqrt{n}/D + (\sqrt{C_0}d^{1/4}\sqrt{n})3(f_0/f_1)\alpha/D \\
&= (1 + 3\sqrt{C_0}d^{1/4}f_0/f_1)\alpha\sqrt{n}/D \\
&:= (f_3\alpha/D)\sqrt{n}.
\end{aligned}$$

On the other hand, it follows from Lemma 3.2 and Lemma 3.4 that

$$\begin{aligned}
\mathbf{P}\left(\exists \mathbf{x}_0 \in \mathcal{M}_D, \|f(\mathbf{x}_0)\|_2 \leq (f_3\alpha/D)\sqrt{n}\right) &\leq (4C_0C_1)^{n-1}(f_3\alpha/D)^{n-1}|\mathcal{M}_D| \\
&\leq (O(1))^n \binom{n}{d-1} (f_3\alpha/D)^{n-1} \left(1 + d^{O(d)}D/\alpha\right)^d \left(1 + d^{O(d)}D/n^{d/2}\right)^{n-d+1} \\
&= (O(1))^n \binom{n}{d-1} (d^{O(d)})^d \left(\alpha d^{O(d)}/n^{d/2} + \alpha d^{O(d)}/n^{d/2}\right)^{n-d-1} \left(1 + d^{O(d)}D/n^{d/2}\right)^2 \\
&\leq (O(1))^n \binom{n}{d-1} d^{O(dn)} (n^{-d/16-1/4})^{n-d+1} (\varepsilon')^{-2}.
\end{aligned}$$

Now we use the assumption that $\gamma/2 \leq 19/36$. With this bound, $\varepsilon \geq n^{-(\gamma/2-17/36)n} \geq n^{-1/18}$, and hence $\varepsilon' = M_d\varepsilon^2 > n^{-n/8}$. Thus, as long as $2 \leq d \leq n^{\varepsilon_0}$ for sufficiently small ε_0 ,

$$\begin{aligned}
\mathbf{P}\left(\exists \mathbf{x}_0 \in \mathcal{M}_D, \|f(\mathbf{x}_0)\|_2, \|f(\mathbf{x}_0)\|_2 \leq (f_3\alpha/D)\sqrt{n}\right) &\leq d^{O(dn)} (n^{-d/16-1/4})^{n-d+1} (\varepsilon')^{-2} \\
&= O(n^{-n/16}). \tag{25}
\end{aligned}$$

In summary, we have shown that, conditioning on the the boundedness of the operator norm from Theorem 2.2,

$$\mathbf{P}\left(\exists \mathbf{x} : D < \mathbf{LCD}(\mathbf{y}_\mathbf{x}) \leq 2D \wedge \|f(\mathbf{x})\|_2 \leq \alpha\sqrt{n}/D\right) = n^{-n/16}.$$

Summing over the dyadic range $n^{d/2}/(O(d))^{O(d)} \leq D \leq \varepsilon'^{-1}$ for D , one thus obtains

$$\begin{aligned}
\mathbf{P}\left(\exists \mathbf{x} : \mathbf{LCD}(\mathbf{y}_\mathbf{x}) \leq \varepsilon'^{-1} \wedge \|f(\mathbf{x})\|_2 \leq n^{(d-1)/4}\sqrt{n}\varepsilon'^{-1}\right) &\leq O(n \log n) \times n^{-n/16}, \\
&\leq n^{-(1/16-o(1)n)},
\end{aligned}$$

completing the proof of Theorem 2.14.

Remark 3.6. Notice that in the last estimate of (25), if d is sufficiently large compared to γ , then $d^{O(dn)} (n^{-d/16-1/4})^{n-d+1} (\varepsilon')^{-2}$ is clearly at most $O(n^{-\Theta(n)})$ for any $\varepsilon \geq n^{-(\gamma/2-17/36)n}$.

Thus Theorem 1.16 holds for any $\gamma = O(1)$ provided that d is sufficiently large depending on γ .

4. CONTROL OF THE OPERATOR NORM: PROOF OF (9) OF THEOREM 2.2

We will first prove a general statement which will be useful for the next section.

Theorem 4.1. *Assume that ξ is a sub-gaussian random variable with zero mean and unit variance satisfying (3). Then there exists an absolute positive constant $C_0 = C_0(K_0)$ independent of d such that the following holds*

$$\mathbf{P}\left(\sup_{\mathbf{x}, \mathbf{y}, \dots, \mathbf{z} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d} \right)^2 \geq C_0 dn\right) \leq \exp(-dn).$$

Assuming this estimate for the moment, we now deduce Theorem 2.2.

Proof. (of Theorem 2.2) The bound on $\sup_{\mathbf{x} \in S^{n-1}} \|\mathbf{f}_{rand}(\mathbf{x})\|_2^2$ clearly follows from Theorem 4.1 by choosing $\mathbf{y}, \dots, \mathbf{z}$ to be \mathbf{x} . For the gradient, we have

$$D_{l, \mathbf{x}, rand}^{(1)}(\mathbf{y}) = \sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} (y_{i_1} x_{i_2} \dots x_{i_{d-1}} x_{i_d} + \dots + x_{i_1} x_{i_2} \dots x_{i_{d-1}} y_{i_d}). \quad (26)$$

Thus by Cauchy-Schwarz inequality

$$\begin{aligned} \sum_l (D_{l, \mathbf{x}, rand}^{(1)}(\mathbf{y}))^2 &\leq d \left[\sum_l \left(\sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} y_{i_1} x_{i_2} \dots x_{i_{d-1}} x_{i_d} \right)^2 + \dots + \right. \\ &\quad \left. + \sum_l \left(\sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} x_{i_1} x_{i_2} \dots x_{i_{d-1}} y_{i_d} \right)^2 \right]. \end{aligned}$$

By Theorem 4.1, each summand is bounded by $C_0 dn$ with probability at least $1 - \exp(-dn)$. Hence with probability at least $1 - d \exp(-dn) \geq 1 - \exp(-dn/2)$,

$$\sum_l (D_{l, \mathbf{x}, rand}^{(1)}(\mathbf{y}))^2 \leq C_0 d^3 n,$$

completing the bound for $\sup_{\mathbf{x}, \mathbf{y}} \|(D_{\mathbf{x}, rand}^{(1)}(\mathbf{y}))\|_2^2$.

The treatment for $\sup_{\mathbf{x}, \mathbf{y}, \mathbf{z} \in S^{n-1}} \|D_{\mathbf{x}, rand}^{(2)}(\mathbf{y}, \mathbf{z})\|_2^2$ is similar where in place of (26), we write $D_{\mathbf{x}, rand}^{(2)}(\mathbf{y}, \mathbf{z})$ as a sum of $O(d^2)$ summands. The upper bound $C_0 d^5 n$ can then be obtained again by applying Theorem 4.1 and Cauchy-Schwarz inequality. \square

What remains is to establish Theorem 4.1. We first prove it for the case of fixed $\mathbf{x}, \mathbf{y}, \dots, \mathbf{z}$.

Lemma 4.2. *Assume that $\mathbf{x}, \mathbf{y}, \dots, \mathbf{z} \in S^{n-1}$. Then there exists an absolute positive constant $C_0 = C_0(K_0)$ such that*

$$\mathbf{P}\left(\sup_{\mathbf{x}, \mathbf{y}, \dots, \mathbf{z} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq C_0 d n\right) \leq \exp(-16 d n).$$

Proof. (of Lemma 4.2) We observe that for any l , $(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d})^2$ is a sub-exponential random variable with mean one and bounded variance. Lemma 4.2 then follows by a standard deviation result. \square

We now extend the result above to the case $\mathbf{y}, \dots, \mathbf{z}$ are fixed.

Lemma 4.3. *Assume that $\mathbf{y}, \dots, \mathbf{z}$ are fixed unit vectors of S^{n-1} , then*

$$\mathbf{P}\left(\sup_{\mathbf{x} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq C_0 d n\right) \leq \exp(-(16d - 6)n).$$

Proof. (of Lemma 4.3) Consider an $1/2$ -net \mathcal{N} of S^{n-1} . We first claim that

$$\begin{aligned} & \mathbf{P}\left(\sup_{\mathbf{x} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq M^2\right) \\ & \leq \mathbf{P}\left(\sup_{\mathbf{x} \in \mathcal{N}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq (M/2)^2\right). \end{aligned} \quad (27)$$

For simplicity, consider the matrix $A_{\mathbf{y}, \dots, \mathbf{z}} := (a_{l i_1}(\mathbf{y}, \dots, \mathbf{z}))_{1 \leq l \leq n-1, 1 \leq i_1 \leq n}$, where $a_{l i_1}(\mathbf{y}, \dots, \mathbf{z}) := \sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} y_{i_2} \dots z_{i_d}$. It then follows that

$$\sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 = \|A_{\mathbf{y}, \dots, \mathbf{z}} \mathbf{x}\|_2^2.$$

Now assume that $\sup_{\mathbf{x} \in S^{n-1}} \|A_{\mathbf{y}, \dots, \mathbf{z}} \mathbf{x}\|_2 = \|A_{\mathbf{y}, \dots, \mathbf{z}}\|_{op}$ is attained at $\mathbf{x} = (x_1, \dots, x_n)$. Choose $\mathbf{x}' \in \mathcal{N}$ such that $\|\mathbf{x} - \mathbf{x}'\|_2 \leq 1/2$. By definition, as $A_{\mathbf{y}, \dots, \mathbf{z}}$ is a linear operator,

$$\|A_{\mathbf{y}, \dots, \mathbf{z}} \mathbf{x} - A_{\mathbf{y}, \dots, \mathbf{z}} \mathbf{x}'\|_2 = \|A_{\mathbf{y}, \dots, \mathbf{z}}(\mathbf{x} - \mathbf{x}')\|_2 \leq \|\mathbf{x} - \mathbf{x}'\|_2 \|A_{\mathbf{y}, \dots, \mathbf{z}}\|_{op} \leq \frac{1}{2} \|A_{\mathbf{y}, \dots, \mathbf{z}}\|_{op}.$$

By the triangle inequality, it is implied that

$$\|A_{\mathbf{y}, \dots, \mathbf{z}} \mathbf{x}'\|_2 \geq \frac{1}{2} \|\mathcal{A}\|_{op},$$

proving our claim.

To conclude the proof, notice that S^{n-1} has an $1/2$ -net \mathcal{N} of size at most $2n5^n$. We then apply Lemma 4.2 and the union bound

$$\mathbf{P}\left(\sup_{\mathbf{x} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq Cn\right) \leq 2n5^n \times \exp(-16dn) \leq \exp(-(16d-6)n).$$

□

Observe that one can also extend (27) to the case that \mathbf{x}, \mathbf{y} vary,

$$\begin{aligned} & \mathbf{P}\left(\sup_{\mathbf{x}, \mathbf{y} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq M^2\right) \\ & \leq \mathbf{P}\left(\bigvee_{\mathbf{y} \in \mathcal{N}} \sup_{\mathbf{x} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left|\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right|^2 \geq (M/2)^2\right). \end{aligned} \quad (28)$$

Thus one obtains the following analog of Lemma 4.3 when \mathbf{x} and \mathbf{y} are not fixed.

$$\begin{aligned} & \mathbf{P}\left(\sup_{\mathbf{x}, \mathbf{y} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq C_0 dn\right) \\ & \leq 2n5^n \times \mathbf{P}\left(\sup_{\mathbf{x} \in S^{n-1}} \sum_{1 \leq l \leq n-1} \left(\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 \geq (M/2)^2\right) \\ & \leq 2n5^n \exp(-(16d-6)n) \leq \exp(-(16d-12)n). \end{aligned}$$

To conclude the proof of Theorem 4.1, one just iterates the argument above d times. Finally, we remark that Theorem 4.1 yields the following more general looking version.

Theorem 4.4. *Assume that $1 \leq k \leq n$, and that $\mathcal{A} = \{a_{i_1, \dots, i_d}^{(l)}, 1 \leq i_1, \dots, i_d \leq k, 1 \leq l \leq n-1\}$ is an array of iid random copies of a subgaussian random variable ξ of zero mean and unit variance satisfying (3). Then there exists a positive constant C_0 such that the following holds*

$$\mathbf{P}\left(\sup_{\mathbf{x}, \mathbf{y}, \dots, \mathbf{z} \in S^{k-1}} \sum_{1 \leq l \leq n-1} \left|\sum_{i_1, \dots, i_d} a_{i_1 i_2 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right|^2 \geq C_0 dn\right) \leq \exp(-dn).$$

5. CONTROL OF COMPRESSIBLE VECTORS

We will prove a more general estimate as follows.

Theorem 5.1. *With sufficiently small constant c_{sparse} ,*

$$\mathbf{P}\left(\inf_{\mathbf{x} \in \text{Comp}(\delta, \rho)} \sum_{1 \leq l \leq n-1} |f_l(\mathbf{x})|^2 \leq c_{\text{sparse}} n\right) \leq c_0^n.$$

Recall from (19) that $\delta = \rho = \kappa_0/d^2$ for a sufficiently small absolute constant κ_0 . In order to prove Theorem 5.1, we will need to work with rectangular arrays.

Theorem 5.2. *Assume that $\mathcal{A} = \{a_{i_1 \dots i_d}^{(l)}, 1 \leq i_1, \dots, i_d \leq k, 1 \leq l \leq n-1\}$ is an array of iid random copies of ξ satisfying (3), with $k = \delta n$. Then there exist absolute constants c_1, c_2 such that the following holds*

$$\mathbf{P}\left(\inf_{\mathbf{x} \in S^k} \sum_{1 \leq l \leq n-1} |f_l(\mathbf{x})|^2 \leq c_1 n\right) \leq \exp(-c_2 n).$$

Indeed we shall prove a slightly stronger result as below.

Theorem 5.3 (Rectangular case for multilinear forms). *With the same assumption as in Theorem 5.2, there exist positive constants c_1, c_2 such that the following holds*

$$\mathbf{P}\left(\inf_{\mathbf{x}, \mathbf{y}, \dots, \mathbf{z} \in S^k} \sum_{1 \leq l \leq n} (a_{i_1 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d})^2 \leq c_1 n\right) \leq \exp(-c_2 n).$$

In order to prove Theorem 5.2, we first need the following easy result of non-concentration (see for instance [14, Lemma 2.6]).

Claim 5.4. *There exists $\mu \in (0, 1)$ such that for any $(a_1, \dots, a_N) \in S^{N-1}$, the random sum $S = \sum \xi_i a_i$, where ξ_1, \dots, ξ_N are independent copies of ξ from (3), satisfies*

$$\mathbf{P}(|S| \leq 1/2) \leq \mu.$$

We recall an analog of the tesorization lemma from Section 3.

Lemma 5.5. *Let η_1, \dots, η_n be independent non-negative random variable, and let $K, \delta \geq 0$.*

- *Assume that for each l , $\mathbf{P}(\eta_l < \varepsilon) \leq K\varepsilon$ for all $\varepsilon \geq \delta$. Then*

$$\mathbf{P}\left(\sum \eta_l^2 < \varepsilon^2 n\right) \leq (C_0 K \varepsilon)^n$$

for all $\varepsilon \geq \delta$.

- Consequently, assume that there exist λ and $\mu \in (0, 1)$ such that for each l , $\mathbf{P}(\eta_l < \lambda) \leq \mu$. Then there exist $\lambda_1 > 0$ and $\mu_1 \in (0, 1)$ depending on λ, μ such that

$$\mathbf{P}\left(\sum \eta_l^2 < \lambda_1 n\right) \leq \mu_1^n.$$

As $\sum_{1 \leq i_1, i_2, \dots, i_d \leq n} (x_{i_1} y_{i_2} \dots z_{i_d})^2 = 1$, it follows from Claim 5.4 and Lemma 5.5 the following analog of Theorem 5.2.

Lemma 5.6 (Estimate for fixed compressible vectors). *With the same assumption as in Theorem 5.2, and let $\mathbf{x}, \mathbf{y}, \dots, \mathbf{z}$ be fixed. Then there exist constants $\eta, \nu \in (0, 1)$ such that*

$$\mathbf{P}\left(\sum_{1 \leq l \leq n-1} \left(\sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 < \eta n\right) \leq \nu^n.$$

Similarly to our treatment of the operator norm in the previous section, we can improve the above as follows.

Theorem 5.7. *With the same assumption as in Theorem 5.2, and let $\mathbf{y}, \dots, \mathbf{z} \in S^k$ be fixed. Then there exist constants $\eta, \nu \in (0, 1)$ such that*

$$\mathbf{P}\left(\inf_{\mathbf{x} \in S^k} \sum_{1 \leq l \leq n-1} \left(\sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d}\right)^2 < 4\eta n\right) \leq \nu^{(1-o(1))n}.$$

For short, we denote $\sum_{1 \leq l \leq n-1} (\sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d})^2$ by $\|A_{\mathbf{y}, \dots, \mathbf{z}}(\mathbf{x})\|_2^2$, emphasizing that this operator depends on $\mathbf{y}, \dots, \mathbf{z}$.

Proof. (of Theorem 5.7) Let $\alpha_d = \alpha_0 d^{-3/2}$ with sufficiently small α_0 to be chosen. It is simple to show that (see for instance [10, Lemma 2.6] or [15, Proposition 2.1]) there exists an α_d -net \mathcal{N} in S^k of cardinality at most $|\mathcal{N}| \leq 2k(1 + \frac{2}{\alpha_d})^k$. Let η, ν be the numbers in Corollary 5.6, by the union bound,

$$\begin{aligned} \mathbf{P}\left(\exists \mathbf{x} \in \mathcal{N} : \|A_{\mathbf{y}, \dots, \mathbf{z}}(\mathbf{x})\|_2^2 < \eta n\right) &= \mathbf{P}\left(\exists \mathbf{x} \in \mathcal{N} : \sum_{1 \leq l \leq n-1} \left| \sum_{1 \leq i_1, \dots, i_d \leq n} a_{i_1 \dots i_d}^{(l)} x_{i_1} y_{i_2} \dots z_{i_d} \right|^2 < \eta n\right) \\ &\leq 2k\left(1 + \frac{2}{\alpha_d}\right)^k \nu^n \\ &\leq (\kappa_0 n / d^2) (1 + 2d^{3/2} / \alpha_0)^{\kappa_0 n / d^2} \nu^n \\ &\leq \nu^{(1-o(1))n}, \end{aligned}$$

where we used the fact that κ_0 is sufficiently small (compared to α_0) and $2 \leq d = o(n)$.

Within this event, let \mathbf{x} be any unit vector in S^k . Choose a point $\mathbf{x}' \in \mathcal{N}$ such that $\|\mathbf{x} - \mathbf{x}'\|_2 \leq \alpha_d$. By Theorem 4.4, with probability at least $1 - \exp(-dn)$ we have

$$\|A_{\mathbf{y}, \dots, \mathbf{z}}(\mathbf{x} - \mathbf{x}')\|_2 < \alpha_d \sqrt{C_0} d^{1/2} \sqrt{n} = \alpha_0 \sqrt{C_0} d^{-1} \sqrt{n} \leq \sqrt{\eta n},$$

where we chose α_0 so that $\alpha_0 \sqrt{C_0} \leq \sqrt{\eta}$. It thus follows that

$$\|A_{\mathbf{y}, \dots, \mathbf{z}} \mathbf{x}'\|_2 \leq \sqrt{\eta n} + \sqrt{\eta n} = 2\sqrt{\eta n},$$

completing the proof. \square

Proof. (of Theorem 5.3) Iterate the argument above d times by fixing lesser terms at each step, one arrives at the conclusion of Theorem 5.3, noting that the entropy loss is at most (taking into account the number of α_d -nets for all $\mathbf{x}, \mathbf{y}, \dots, \mathbf{z}$)

$$\left((\kappa_0 n / d^2) (1 + 2d^{3/2} / \alpha_0)^{\kappa_0 n / d^2} \right)^d = (\kappa_0 n / d^2)^d (1 + 2d^{3/2} / \alpha_0)^{\kappa_0 n / d} \nu^n \leq \nu^{(1-o(1))n},$$

again provided that κ_0 is sufficiently small compared to α_0 and $2 \leq d = o(n)$. \square

We now deduce Theorem 5.1 in the same manner.

Proof. (of Theorem 5.1) By Theorem 4.4, with probability at least $1 - \exp(-n)$ we have the following for any pair \mathbf{x}, \mathbf{x}' with $\|\mathbf{x} - \mathbf{x}'\|_2 \leq \rho$,

$$\begin{aligned} \|A_{\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}} \mathbf{x} - A_{\mathbf{x}', \mathbf{x}', \dots, \mathbf{x}'} \mathbf{x}'\|_2 &\leq \|A_{\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}} \mathbf{x} - A_{\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}} \mathbf{x}'\|_2 + \|A_{\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}} \mathbf{x}' - A_{\mathbf{x}', \mathbf{x}, \dots, \mathbf{x}} \mathbf{x}'\|_2 + \dots + \\ &\quad + \|A_{\mathbf{x}', \mathbf{x}', \dots, \mathbf{x}'} \mathbf{x}' - A_{\mathbf{x}', \mathbf{x}', \dots, \mathbf{x}'} \mathbf{x}'\|_2 \\ &\leq d\rho \sqrt{C_0} d^{1/2} \sqrt{n} \\ &\leq \sqrt{c_1 n}, \end{aligned}$$

where we used the fact that $\rho = \kappa_0 / d^2$ with sufficiently small κ_0 compared to c_1 .

As \mathbf{x}' ranges over vectors of S^{n-1} of support at most $k = \delta n$, an application of Theorem 5.2 implies that

$$\mathbf{P}\left(\inf_{\mathbf{x} \in \text{Comp}(\delta, \rho)} \sum_{1 \leq l \leq n-1} |f_l(\mathbf{x})|^2 \leq 2c_3 n \right) \leq \binom{n}{k} \exp(-c_2 n) \leq c_0^n,$$

for some $0 < c_0 < 1$, completing the proof of Theorem 5.1. \square

APPENDIX A. PROOF OF LEMMA 2.9

We restate the lemma.

Lemma A.1 (Lemma 2.9). *Let $K, \delta_0 \geq 0$ be given. Assume that $\mathbf{P}(|X_1| < \delta) \leq K\delta$ for all $\delta \geq \delta_0$. Then*

$$\mathbf{P}(X_1^2 + \cdots + X_n^2 < \delta n) \leq (C_0 K \delta)^n.$$

Proof. Assume that $\delta \geq \delta_0$. By Chebyshev's inequality

$$\mathbf{P}(X_1^2 + \cdots + X_n^2 \leq \delta n) \leq \mathbf{E} \exp(n - \sum_{i=1}^n X_i^2 / \delta) = \exp(n) \prod_{i=1}^n \mathbf{E} \exp(-X_i^2 / \delta).$$

On the other hand,

$$\mathbf{E} \exp(-X_i^2 / \delta) = \int_0^1 \mathbf{P}(\exp(-X_i^2 / \delta) > s) ds = \int_0^\infty 2u \exp(-u^2) \mathbf{P}(X_i < \delta u) du.$$

For $0 \leq u \leq 1$ we use $\mathbf{P}(X_i \leq \delta u) \leq \mathbf{P}(X_i \leq \delta) \leq K\delta$, while for $u \geq 1$ we have $\mathbf{P}(X_i \leq \delta u) \leq K\delta u$. Thus

$$\mathbf{E} \exp(-X_i^2 / \delta) = \int_0^1 2u \exp(-u^2) K\delta du + \int_1^\infty 2u \exp(-u^2) K\delta u du \leq C_0 K \delta.$$

□

Acknowledgement. The author would like to thank R. Vershynin and V. Vu for discussions at the early stage of this work. He is also grateful to the anonymous reviewer for helpful comments and suggestions that help correct and improve the presentation of the manuscript.

REFERENCES

- [1] L. Blum, F. Cucker, M. Shub and S. Smale, *Complexity and Real Computation*, Springer-Verlag, New York, 1998.
- [2] P. Burgisser and F. Cucker, *Condition, The Geometry of Numerical Algorithms*, Springer, Heidelberg, 2013.
- [3] F. Cucker, T. Krick, G. Malajovich and M. Wschebor, *A numerical algorithm for zero counting. I: Complexity and accuracy*, J. Complexity 24 (2008) 582-605.
- [4] F. Cucker, T. Krick, G. Malajovich and M. Wschebor, *A numerical algorithm for zero counting. II: Distance to Ill-posedness and smoothed analysis*, J. Fixed Point Theory Appl. 6 (2009) 285-294.
- [5] F. Cucker, T. Krick, G. Malajovich and M. Wschebor, *A Numerical Algorithm for Zero Counting. III: Randomization and Condition*, *Advances in Applied Mathematics* 48 (2012), 215-248.
- [6] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. 9 (1988), no. 4, 543-560.

- [7] H. Goldstine and J. von Neumann, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc. 53 (1947), 1021-1099.
- [8] J. Kahn, J. Komlós and E. Szemerédi, *On the probability that a random ± 1 matrix is singular*, J. Amer. Math. Soc. 8 (1995), 223-240.
- [9] E. Kostlan, *Random polynomials and the statistical fundamental theorem of algebra*, unpublished (1987).
- [10] V. D. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, Lecture Notes in Mathematics, 1200 Springer-Verlag, Berlin, 1986.
- [11] H. Nguyen and V. Vu, *Small ball probability, inverse theorems, and applications*, Erdos Centennial, Bolyai Society Mathematical Studies, Vol. 25 (2013).
- [12] H. Nguyen, O. Nguyen and V. Vu, *On the number of real roots of random polynomials*, submitted.
- [13] M. Rudelson, *Invertibility of random matrices: Norm of the inverse*, Annals of Mathematics, 168 (2008), no. 2, 575-600.
- [14] M. Rudelson and R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600-633.
- [15] M. Rudelson and R. Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics 62 (2009), 1707-1739.
- [16] M. Shub and S. Smale, *Complexity of Bezouts theorem I: geometric aspects*, J. Amer. Math. Soc. 6 (1993) 459-501.
- [17] M. Shub and S. Smale, *Complexity of Bezouts theorem II: volumes and probabilities*, Computational Algebraic Geometry, in: Progr. Math., vol. 109, Birkhuser, 1993, pp. 267-285.
- [18] M. Shub and S. Smale, *Complexity of Bezouts theorem III: condition number and packing*, J. Complexity 9 (1993) 4-14.
- [19] M. Shub and S. Smale, *Complexity of Bezouts theorem IV: polynomial time*, Theoret. Comput. Sci. 133 (1994) 141-164.
- [20] M. Shub and S. Smale, *Complexity of Bezouts theorem V: probability of success; extensions*, SIAM J. Numer. Anal. 33 (1996) 128-148.
- [21] D. A. Spielman and S. H. Teng, *Smoothed analysis of algorithms*, Proceedings of the International Congress of Mathematicians, Vol. I, 597-606, Higher Ed. Press, Beijing, 2002.
- [22] D. A. Spielman and S. H. Teng, *Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time*, J. ACM 51 (2004), no. 3, 385-463.
- [23] 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.
- [24] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, Mathematics of Computation 79 (2010), 2333-2352.
- [25] T. Tao and V. Vu, *Random matrices: the distribution of the smallest singular values*, Geom. Funct. Anal. 20 (2010), no. 1, 260-297.
- [26] M. Wschebor, *Smoothed analysis of $\kappa(A)$* , J. of Complexity 20 (2004), no. 1, 97-107.

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

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, 231 WEST 18TH AVENUE, COLUMBUS, OH 43210