

# On the Number of Faces and Radii of Cells Induced by Gaussian Spherical Tessellations

Eric Lybrand (elybrand@ucsd.edu)

Department of Mathematics, University of California, San Diego

Anna Ma (anna.ma@uci.edu)

Department of Mathematics, University of California, Irvine

Rayan Saab (rsaab@ucsd.edu)

Department of Mathematics and Halicioglu Data Science Institute,  
University of California, San Diego

## Abstract

We study a geometric property related to spherical hyperplane tessellations in  $\mathbb{R}^d$ . We first consider a fixed  $x$  on the Euclidean sphere and tessellations with  $M \gg d$  hyperplanes passing through the origin having normal vectors distributed according to a Gaussian distribution. We show that with high probability there exists a subset of the hyperplanes whose cardinality is on the order of  $d \log(d) \log(M)$  such that the radius of the cell containing  $x$  induced by these hyperplanes is bounded above by, up to constants,  $d \log(d) \log(M)/M$ . We extend this result to hold for all cells in the tessellation with high probability. Up to logarithmic terms, this upper bound matches the previously established lower bound of Goyal et al. (IEEE T. Inform. Theory 44(1):16–31, 1998).

## I. INTRODUCTION

Suppose we have a collection of non-zero vectors  $\varphi^{(1)}, \dots, \varphi^{(M)} \in \mathbb{R}^d$  and we consider the hyperplane tessellation these vectors induce on the unit sphere  $S^{d-1}$ . One might be tempted to believe that the number of cells formed by the hyperplane tessellation –that is, for  $x \in S^{d-1}$  with  $\langle \varphi^{(i)}, x \rangle \neq 0$  for all  $i \in [M]$ , sets of the form  $\{y \in S^{d-1} : \text{sign}(\langle x, \varphi^{(i)} \rangle) = \text{sign}(\langle y, \varphi^{(i)} \rangle) \text{ for all } i \in [M]\}$ – is exponential in  $M$ . However, it is not difficult to see that this is far from the truth when  $M$  is much larger than  $d$ , i.e.,  $M \gg d$ . In this setting and when the  $\varphi^{(i)}$  are in general position, or when any  $d$  of them are linearly independent, Schläfli proved, in the 1800s, that the number of cells is exactly

$$2 \sum_{i=0}^{d-1} \binom{M-1}{i}. \tag{1}$$

A.M. was partially supported by U.S. Air Force Award FA9550-18-1-0031 led by Roman Vershynin. R.S. was supported in part by NSF DMS-1517204 and DMS-2012546.

See [1] for a version of Schläfi's proof and for further references. To tease out what the asymptotic properties of this quantity are in terms of  $M$  and  $d$ , note that for any  $\frac{M+1}{2} > d$  the leading term in the sum is the final summand. Notice that

$$\frac{\binom{M}{d} + \binom{M}{d-1} + \dots + \binom{M}{0}}{\binom{M}{k}} \leq \sum_{j=0}^{\infty} \left( \frac{d}{M-d+1} \right)^j = \frac{M-d+1}{M-2d+1}. \quad (2)$$

Therefore, we see that  $\sum_{i=0}^d \binom{M}{i} \leq \binom{M}{d} \cdot \frac{M-d+1}{M-2d+1}$ .

When  $M \gg d$ , we can upper bound (1) by  $C_1 \binom{M-1}{d-1}$  which by a Sterling bound is bounded above by  $C_2 (eM/d)^d$ , where  $C_1, C_2 > 0$  are absolute constants. With so much redundancy a natural question to ask is how many faces<sup>1</sup> delimit a cell in such a tessellation? Further, how large are the radii of the cells? The first question has seen a recent resurgence of interest [2]–[4] and was initially addressed by Cover and Efron [5]. More specifically, it was shown in [5, Eq. (4.1)] that under the assumption that the  $\varphi^{(i)}$  are in general position, the expected number of faces for a cell drawn uniformly at random from the tessellation is

$$\frac{2 \cdot M \cdot \sum_{i=0}^{d-2} \binom{M-2}{i}}{\sum_{i=0}^{d-1} \binom{M-1}{i}}. \quad (3)$$

Using (2) to upper bound (3) gives us

$$\begin{aligned} \frac{2 \cdot M \cdot \sum_{i=0}^{d-2} \binom{M-2}{i}}{\sum_{i=0}^{d-1} \binom{M-1}{i}} &\leq \frac{2 \cdot M \cdot \binom{M-2}{d-2} \cdot \frac{M-d+1}{M-2d+3}}{\binom{M-1}{d-1}} \\ &\leq C_3 \cdot 2 \cdot M \cdot \frac{(M-2)!}{(d-2)!(M-d)!} \cdot \frac{(d-1)!(M-d)!}{(M-1)!} \\ &= C_3 \cdot 2 \cdot M \cdot \frac{d-1}{M-1} \leq C_4 d, \end{aligned} \quad (4)$$

where, again,  $C_3, C_4 > 0$  are absolute constants. In other words, we can expect there to be no more than on the order of  $d$  hyperplanes to form a cell drawn uniformly from our tessellation when there are  $M \gg d$  hyperplanes. In fact, we can find a matching lower bound by a similar argument which proves that the average cell has a number of faces that scales like  $d$ .

Unlike the setting of [5] we consider these two questions in the specific case when the  $\varphi^{(i)}$  are distributed according to a Gaussian distribution. We seek an upper bound (that holds with high probability on the draw of the  $\varphi^{(i)}$ ) on the radius of cells formed by a collection of hyperplanes whose cardinality, up to constants and logarithmic terms, matches the bound (4). Initially, we refer to a fixed cell by first fixing a point  $x \in S^{d-1}$ , drawing the vectors  $\varphi^{(i)}$  independently of  $x$ , and considering the cell containing  $x$ . We then pass to a uniform result where the result holds with high probability for all points  $x$  on the sphere.

## II. BACKGROUND AND CONNECTIONS

The geometric questions listed above play an important role in the theory of quantization for finite frame expansions (see, e.g., [6], [7]) and binary classification with linear separators [8]. These two contexts are dual to one another in the sense that for binary classifiers with linear separators the vector  $x$  is thought of as normal to a separating hyperplane which classifies the vectors  $\varphi^{(i)}$ , into one of two classes, under the mapping  $\varphi^{(i)} \mapsto \text{sign}(\langle x, \varphi^{(i)} \rangle)$ . Notice that the

<sup>1</sup>Here, we are referring to faces of maximal dimension, which are referred to as facets in other works.

mapping  $\varphi^{(i)} \mapsto \text{sign}(\langle x, \varphi^{(i)} \rangle)$  is invariant under positive scalings of  $x$  so that we may assume without loss of generality  $\|x\|_2 = 1$ . Frame quantization on the other hand views this problem in the setting as we have phrased it where the  $\varphi^{(i)}$  induce a spherical hyperplane tessellation and the goal is to recover an approximation of  $x$  from quantized measurements  $\text{sign}(\langle x, \varphi^{(i)} \rangle)$ . For example, such an approximation may take the form of a vector  $x^\sharp$  from the same cell as  $x$ .

Our geometric intuition concerning the radii of the cells in the tessellation is guided by a result from quantization of finite frame expansions. In that context, the so-called measurement vectors  $\varphi^{(i)}$  are independently and randomly distributed, exhibit a correlation structure arising from structured random matrix ensembles, or arise from some deterministic construction. It was in this setting that Goyal et al. proved that if the  $\varphi^{(i)}$  form a frame (i.e., a spanning set for  $\mathbb{R}^d$ ), not necessarily random, and  $x$  is taken to be random according to some distribution  $p$  over a set  $D$  then the root mean-squared error for recovering  $x$  using  $M$  quantized frame coefficients  $(\text{sign}(\langle x, \varphi^{(i)} \rangle))_{i=1}^M$  can decay *no faster than*  $C \frac{1}{M}$  where  $C$  is a constant that depends on  $d, p, D$ , but not  $M$  [6, Proposition 2]. The  $\varphi^{(i)}$  form a frame with probability 1 when they are i.i.d. Gaussian. Consequentially, this result essentially tells us that the largest cell in the induced tessellation has an inscribed ball of radius on the order of  $\frac{1}{M}$ . The fact that the quantization error can only decay linearly, rather than exponentially with  $M$ , has motivated the design of more sophisticated quantization schemes, such as  $\Sigma\Delta$  or noise shaping quantizers (see, e.g. [7], [9]), and non-adaptive universal one-bit quantizers [10]. Herein, we will demonstrate that for any  $x \in S^{d-1}$  there exists a sub-collection of  $C_5 d \log(d) \log(M)$  hyperplanes such that the radius of the cell containing  $x$  induced by this sub-collection is at most on the order of  $d \log(d) \log(M)/M$ . In other words, relatively few vectors  $\varphi^{(i)}$  identify the cell.

Binary classification with linear separators is often known under the moniker of the support vector machine. The support vector machine was first proposed in [8]. Various works have analyzed this related problem. Importantly, the results that we are aware of differ slightly from the geometric context that we have posed. Namely, most support vector machine results are concerned with the scenario where the spherical hyperplane tessellation is not fixed beforehand. In this dual setting, results are typically formulated in terms of how many point samples  $\varphi^{(i)}$  with labels  $\text{sign}(\langle \varphi^{(i)}, x \rangle)$  are required to learn a separating hyperplane with normal vector  $x^\sharp$  so that the probability of a new point  $\varphi_{M+1}$  being incorrectly classified according to  $\text{sign}(\langle \varphi_{M+1}, x^\sharp \rangle)$  is less than  $\varepsilon$ . In the primal setting, a separating hyperplane  $x^\sharp$  is any point in the same cell as  $x$  induced by the spherical tessellation from  $\varphi^{(1)}, \dots, \varphi^{(M)}$ . Learning an  $\varepsilon$ -accurate classifier amounts to a choice of  $x^\sharp$  so that with probability less than  $\varepsilon$  a newly drawn hyperplane with normal vector  $\varphi^{(M+1)}$  excludes  $x^\sharp$  from the new cell with  $x$  in it. For further reading, one may look at [11]–[16] for example.

### III. NOTATION

Henceforth,  $\langle \cdot, \cdot \rangle$  denotes the standard inner product over  $\mathbb{R}^d$ .  $S^{d-1}$  and  $B_2^d(x, \varepsilon)$  denote the unit sphere and the ball of radius  $\varepsilon$  centered at  $x$  in  $\mathbb{R}^d$ , respectively. The vector  $e_i \in S^{d-1}$  denotes the  $i^{\text{th}}$  standard basis vector in  $\mathbb{R}^d$  and the matrix  $I_d$  denotes the  $d \times d$  identity matrix. For any vector  $z \in \mathbb{R}^d$ , define the vector  $z_{[-1]} \in \mathbb{R}^{d-1}$  to be the projection of  $z$  which removes the first coordinate and let  $z_j \in \mathbb{R}$  denote the  $j^{\text{th}}$  entry of  $z$ . For any integer  $n \in \mathbb{N}$ , we

let  $[n] = \{1, 2, \dots, n\}$ . The relation  $a \gtrsim b$  denotes that there exists some universal constant  $C > 0$  so that  $a \geq Cb$ . Similarly, the relation  $a \gg b$  denotes that there exists some universal constant  $C > 1$  so that  $a \geq Cb$ . In order to avoid cumbersome notation, we reserve the symbols  $C, C', c, c'$  and so forth for universal constants. For a random variable  $X$ , we define the Orlicz norm (see [17, Section 2.7.1]) as

$$\|X\|_{\psi_2} = \inf \left\{ \omega > 0 : \mathbb{E} e^{X^2/\omega^2} \leq 2 \right\},$$

while for a random vector  $X \in \mathbb{R}^d$ , we define it as

$$\|X\|_{\psi_2} = \sup_{x \in S^{d-1}} \|\langle X, x \rangle\|_{\psi_2}.$$

With this notation, we define and approach our problem as follows. Let  $x \in S^{d-1}$  be some fixed unit vector in  $\mathbb{R}^d$ , and suppose we draw  $M \gg d$  i.i.d. vectors  $g^{(i)} \sim \mathcal{N}(0, d^{-1}I)$  at random and are given the sign pattern

$$y_i = \text{sign}(\langle g^{(i)}, x \rangle), \quad i \in [M].$$

The discussion in the introductory section tells us that the  $M$  measurements are highly redundant and that one should be able to infer which cell  $x$  is in from much fewer measurements. Given the derivation in (4) we might expect the number of hyperplanes forming the cell containing  $x$  to scale like  $d$ . In light of this, we consider sub-selecting a set  $S \subset [M]$  of size  $m \ll M$  from the i.i.d. vectors  $g^{(i)}$  so that with high probability

$$\text{sign}(\langle g^{(i)}, x \rangle) = \text{sign}(\langle g^{(i)}, y \rangle) \quad \forall i \in S \implies \|x - y\|_2 \text{ is small.}$$

More specifically, we consider sets of the form

$$T_{\tau, x} := \{i \in [M] : |\langle g^{(i)}, x \rangle| < \tau\}. \quad (5)$$

Our goal is to choose  $\tau$  so that  $T_{\tau, x}$  is large enough to define a spherical polytope containing  $x$  with a radius on the order of the smallest cell containing  $x$ . That is, we would like to choose  $\tau$  so that  $\{y \in S^{d-1} : \text{sign}(\langle \varphi^{(i)}, x \rangle) = \text{sign}(\langle \varphi^{(i)}, y \rangle) \text{ for all } i \in [M]\} \subset \{y \in S^{d-1} : \text{sign}(\langle \varphi^{(i)}, x \rangle) = \text{sign}(\langle \varphi^{(i)}, y \rangle) \text{ for all } i \in T_{\tau, x}\}$ . On the other hand, we also wish  $T_{\tau, x}$  to have cardinality roughly on the order of  $d$ . Our main results show that, indeed, this is possible with high probability.

#### IV. RESULTS

The main result of [6, Proposition 2] tells us that using  $M$  hyperplanes there exists a point  $x \in S^{d-1}$  and a point  $z$  in the same cell as  $x$  which satisfy  $\|x - z\|_2 \gtrsim \frac{d}{M}$ . Theorem IV.1 shows that if we were to have access to the sign patterns from the set  $T_{\tau, x}$ , then it is possible to find a matching upper bound up to an additional factor of  $\log(d) \log(M)$  with only  $Cd \log(d) \log(M)$  hyperplanes.

**Theorem IV.1.** *Fix  $x \in S^{d-1}$  and draw i.i.d. vectors  $g^{(i)} \sim \mathcal{N}(0, d^{-1}I_d)$  for  $i \in [M]$  with  $\frac{M}{\log(M)} \gtrsim d \log(d)$ . Select once and for all a collection of  $C_1 d \log(d)$  indices independent of  $x$  without replacement from  $[M]$  and put them in a set called  $V$ . Choose*

$$\tau = C_2 \frac{\sqrt{d} \log(d) \log(M)}{M},$$

and define the sets  $W_{\tau,x} := \{i \in [M] : -\tau < \langle g^{(i)}, x \rangle < 0\}$ ,  $S_{\tau,x} := W_{\tau,x} \cup V$ . Then, with probability at least  $1 - 3 \exp(-cd \log(d)) - \exp(-c'd \log(M))$  the following is true:

$$C_3 d \log(d) \log(M) \leq |S_{\tau,x}| \leq C_4 d \log(d) \log(M),$$

and if, for  $y \in S^{d-1}$ ,  $\text{sign}(\langle g^{(i)}, x \rangle) = \text{sign}(\langle g^{(i)}, y \rangle)$  for all  $i \in S_{\tau,x}$  then

$$\|x - y\|_2 \leq \frac{C_5 d \log(d) \log(M)}{\sqrt{M^2 + d^2 \log^2(d) \log^2(M)}}. \quad (6)$$

We can understand Theorem IV.1 using some simple geometric intuition. The vectors  $g^{(i)}$  whose inner products with  $x$  are the smallest in absolute value determine the normal hyperplanes which are closest to  $x$ . So, these hyperplanes are the most informative with respect to determining the faces of the cell in the full tessellation induced by all of the vectors  $g^{(1)}, \dots, g^{(M)}$ . However, because we only have finitely many  $g^{(i)}$  we need to choose  $\tau$  carefully so that the set  $W_{\tau,x} = \{i \in [M] : -\tau < \langle g^{(i)}, x \rangle < 0\}$  is non-empty.

Our proof of Theorem IV.1 relies on reducing the bound (6) to a lower bound on the minimal singular value of a particular random matrix. To see how, we provide the following quick sketch. By rotation invariance of the  $g^{(i)}$ , let's assume  $x = e_1$ . Notice we have  $\|x - y\|_2^2 = 2 - 2y_1$  for any  $y \in S^{d-1}$ . So, to control  $\max_y \|x - y\|_2^2$  where  $y$  is such that  $\text{sign}(\langle g^{(i)}, x \rangle) = \text{sign}(\langle g^{(i)}, y \rangle)$  for all  $i \in S_{\tau,x}$ , we just need a uniform lower bound on  $y_1$  for such  $y$ . This lower bound will come from the feasibility condition

$$\begin{aligned} \langle g^{(i)}, x \rangle \langle g^{(i)}, y \rangle &= g_1^{(i)} \langle g^{(i)}, y \rangle > 0 \text{ for all } i \in S_{\tau,x} \\ \iff (g_1^{(i)})^2 y_1 &> -g_1^{(i)} \langle g_{[-1]}^{(i)}, y_{[-1]} \rangle, \text{ for all } i \in S_{\tau,x}. \end{aligned} \quad (7)$$

Now, assuming  $g_1^{(i)} \neq 0$ , if we could preserve (7) by squaring both sides – for example if both sides were both positive – we would end up with an inequality of the form

$$g_1^{(i)2} y_1^2 > \|y_{[-1]}\|_2^2 \left\langle g_{[-1]}^{(i)}, \frac{y_{[-1]}}{\|y_{[-1]}\|_2} \right\rangle^2, \text{ for all } i \in S_{\tau,x}. \quad (8)$$

Using the fact that  $\|y_{[-1]}\|_2^2 = 1 - y_1^2$  and  $|g_1^{(i)}| < \tau$  on the subset of vectors  $W_{\tau,x} \subset S_{\tau,x}$  as defined in Theorem IV.1, we could collect like terms to get

$$y_1^2 \geq 1 - \frac{\tau^2}{\left\langle g_{[-1]}^{(i)}, \frac{y_{[-1]}}{\|y_{[-1]}\|_2} \right\rangle^2 + \tau^2}, \text{ for all } i \in W_{\tau,x} \subset S_{\tau,x}.$$

From here, all we would need to do is find a uniform lower bound for  $\left\langle g_{[-1]}^{(i)}, \frac{y_{[-1]}}{\|y_{[-1]}\|_2} \right\rangle^2$  which is where random matrix theory would come into play. The only way this strategy works is if we could show that the inequalities (7) and (8) were essentially equivalent. Of course, if both sides of (7) were positive we would be in great shape. The terms we have to worry about are  $y_1$  and  $\langle g_{[-1]}^{(i)}, y_{[-1]} \rangle$ . Our goal is to show that we can throw away a constant proportion of the vectors in  $S_{\tau,x}$  where the two conditions  $y_1 > 0$  and  $\langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > 0$  do not hold and still get the result of Theorem IV.1. Lemma IV.3 handles the condition  $y_1 > 0$  and Corollary IV.4 handles the condition  $\langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > 0$ . Besides these two results, proving Theorem IV.1 requires two main ingredients. First, Lemma IV.2 characterizes how  $\tau$  should be chosen to ensure that, with high probability,  $|S_{\tau,x}|$  is about  $d \log(d) \log(M)$ . This, along with Lemma IV.5

finish the heavy lifting needed to lower bound  $\left\langle g_{[-1]}^{(i)}, \frac{y_{[-1]}}{\|y_{[-1]}\|_2} \right\rangle^2$ . If not defined explicitly in the statements below, we note that variables are as in Theorem IV.1.

**Lemma IV.2.** *If  $\tau = C_2 \frac{\sqrt{d} \log(d) \log(M)}{M}$ , then with probability at least  $1 - \exp(-cd \log(d) \log(M))$  we have  $C_3 d \log(d) \log(M) \leq |S_{\tau,x}| \leq C_4 d \log(d) \log(M)$ .*

*Proof.* As in the statement of Theorem IV.1 let  $W_{\tau,x} := \{i \in [M] : -\tau < \langle g^{(i)}, x \rangle < 0\}$ . Note that  $W_{\tau,x} \subset S_{\tau,x}$  and

$$|W_{\tau,x}| = \sum_{i=1}^M \mathbb{1}(|\langle g^{(i)}, x \rangle| < \tau) \mathbb{1}(\langle g^{(i)}, x \rangle < 0). \quad (9)$$

We remark that the events defined in the indicator functions are independent. Further, each  $\mathbb{1}(|\langle g^{(i)}, x \rangle| < \tau)$  is a Bernoulli random variable with mean  $\mathbb{P}(|\langle g^{(i)}, x \rangle| < \tau)$ . By our choice of normalization,  $\langle g^{(i)}, x \rangle \sim \mathcal{N}(0, d^{-1})$ , so we have  $\mathbb{P}(|\langle g^{(i)}, x \rangle| < \tau) = \text{erf}\left(\frac{\tau\sqrt{d}}{\sqrt{2}}\right)$ , where erf denotes the Gauss error function

$$\text{erf}(t) = \frac{1}{\sqrt{\pi}} \int_{-t}^t e^{-x^2} dx.$$

Using a multiplicative Chernoff bound [17] for any  $s \in (0, 1]$

$$\begin{aligned} & \mathbb{P}\left(\left| |W_{\tau,x}| - \frac{M}{2} \text{erf}\left(\frac{\tau\sqrt{d}}{\sqrt{2}}\right) \right| \geq s \frac{M}{2} \text{erf}\left(\frac{\tau\sqrt{d}}{\sqrt{2}}\right)\right) \\ & \leq 2 \exp\left(-cs^2 M \text{erf}\left(\frac{\tau\sqrt{d}}{\sqrt{2}}\right)\right) \leq 2 \exp\left(-cs^2 M \tau \sqrt{d} e^{-\tau^2 d/2}\right). \end{aligned}$$

More specifically, this tells us that with high probability  $|W_{\tau,x}| \geq c' M \text{erf}\left(\frac{\tau\sqrt{d}}{\sqrt{2}}\right) \geq c'' M \tau \sqrt{d} \exp(-\tau^2 d/2) \geq c''' d \log(d) \log(M)$  using a lower Riemann sum estimate to lower bound erf in the penultimate inequality. Notice that the term  $e^{-\tau^2 d/2}$  does not pose any problems, since  $\tau^2 d/2 = C_2 d \log(d) \log(M)/2M$  and we assume that  $M \gtrsim d \log(d) \log(M)$ . Analogously, we also have using an upper Riemann sum estimate  $|W_{\tau,x}| \leq C M \text{erf}\left(\frac{\tau\sqrt{d}}{\sqrt{2}}\right) \leq C' M \tau \sqrt{d} = C'' d \log(d) \log(M)$ . Since  $S_{\tau,x} = W_{\tau,x} \cup V$  and  $|V| = C_1 d \log(d)$ , the claim then follows.  $\square$

**Lemma IV.3.** *Let  $A \in \mathbb{R}^{C_1 d \log(d) \times d}$  be a matrix whose rows are populated with the vectors in  $V$ , which is defined in the statement of Theorem IV.1. Then for any  $x, y \in S^{d-1}$ , the following is true with probability exceeding  $1 - e^{-cd \log(d)}$ :  $\text{sign}(Ax) = \text{sign}(Ay)$  implies  $\|x - y\|_2 < \sqrt{2}$ , and consequently  $\langle x, y \rangle > 0$ .*

*Proof.* This is a direct application of Theorem 2 in [18] considering the measurements formed using the matrix  $A$ .  $\square$

Recall that to achieve our goal to transition from (7) to (8), it would suffice to have  $y_1 > 0$  and  $\langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > 0$  for all  $i \in S_{\tau,x}$  and all  $y$  in the same cell as  $x$ . By nature of the latter condition being a “for all  $y$ ” statement, we will follow a relatively standard argument relying on a probabilistic union bound coupled with a continuity argument. To make the continuity argument work, we will use the slightly stronger condition  $\langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > \eta > 0$  for  $\eta$  to be defined below.

**Corollary IV.4.** *Let  $x = e_1$ , the first standard basis vector, fix  $y \in S^{d-1}$ , and suppose that  $\|y_{[-1]}\|_2^2 \geq \frac{1}{M^2}$ . Define  $W_{\tau,x} \subset S_{\tau,x}$  as in Theorem IV.1, and set  $\eta = C_5 \frac{\log(M)}{M^2}$ . Consider the subset*

$S_{\tau,x}^{\eta,y} := \{i \in W_{\tau,x} : \langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > \eta\} \subset S_{\tau,x}$ . Then with probability at least  $1 - e^{-cd\log(d)\log(M)}$ , we have  $\tilde{C}_3 d \log(d) \log(M) \leq |S_{\tau,x}^{\eta,y}| \leq \tilde{C}_4 d \log(d) \log(M)$ .

*Proof.* As before, we have

$$|S_{\tau,x}^{\eta,y}| := \sum_{i \in [M]} \mathbb{1}(-\tau \leq \langle g^{(i)}, x \rangle \leq 0) \mathbb{1}(\langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > \eta).$$

Notice that  $\mathbb{P}(\langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > \eta) \geq \mathbb{P}(\gamma \geq \eta M \sqrt{d})$ , where  $\gamma \sim \mathcal{N}(0, 1)$  and that, as in (9), the events defined in the indicator functions are independent. Using a Riemann sum to approximate this probability, we have that this aforementioned probability is lower bounded by  $\frac{1}{2} - c' \eta M \sqrt{d} = \frac{1}{2} - C' \frac{\sqrt{d} \log(M)}{M} = \tilde{c} > 0$ . So, we find  $\mathbb{E}|S_{\tau,x}^{\eta,y}| = \tilde{c} \mathbb{E}|W_{\tau,x}|$ , and using a Chernoff bound as in Lemma IV.2 the claim follows.  $\square$

The proof of the Lemma IV.5 depends on the following intermediate computation, whose statement and proof are provided in Lemma VII.1, Section VII.

**Lemma IV.5.** *Fix  $x = e_1$ . Then the following is true for all  $y \in S^{d-1}$  with  $\|y_{[-1]}\|_2^2 \geq \frac{1}{M^2}$ : define  $S_{\tau,x}^y := \{i \in W_{\tau,x} : \langle g_{[-1]}^{(i)}, y_{[-1]} \rangle > 0\}$ , and define the matrix  $G_y \in \mathbb{R}^{|S_{\tau,x}^y| \times (d-1)}$  by populating its rows with  $g_{[-1]}^{(i)}$ , where  $i \in S_{\tau,x}^y$ . Then with probability  $1 - e^{-c'd\log(d)\log(M)} - 2e^{-c''d\log(M)}$ , we have*

$$\min_{z \in S^{d-2}} \|G_y z\|_2^2 \gtrsim \log(d) \log(M).$$

*Proof.* To start, fix  $y$  and set  $\eta = C_5 \frac{\log(M)}{M^2}$ . By rotational invariance, we may assume  $y_{[-1]}$  is parallel to  $e_1 \in \mathbb{R}^{d-1}$ . Define the matrix  $\tilde{G}_y \in \mathbb{R}^{|S_{\tau,x}^y| \times (d-1)}$  by populating its rows with  $g_{[-1]}^{(i)}$ , where  $i \in S_{\tau,x}^y$  and  $S_{\tau,x}^y$  is defined in Corollary IV.4. By that very same corollary,  $\tilde{C}_3 d \log(d) \log(M) \leq |S_{\tau,x}^y| \leq \tilde{C}_4 d \log(d) \log(M)$  with probability at least  $1 - e^{-c_1 d \log(d) \log(M)}$ . Since any orthogonal projection of a standard Gaussian random vector is itself a Gaussian random vector, the rows of  $\tilde{G}_y$  are independent and identically distributed. Within each row, the first entry is distributed according to the conditional law  $g|g > \frac{\eta}{\|y_{[-1]}\|_2}$  and the remaining entries are i.i.d. Using Lemma VII.1 with  $t = \sqrt{c_2 d \log(M)}$ , we have with probability at least  $1 - e^{-c_2 d \log(M)}$ ,

$$\left\| \frac{1}{|S_{\tau,x}^y|} \tilde{G}_y^T \tilde{G}_y - \Sigma \right\|_{op} \leq CK^2 \left( \sqrt{\frac{d}{|S_{\tau,x}^y|}} + \sqrt{\frac{c_2 d \log(M)}{|S_{\tau,x}^y|}} \right),$$

where  $K = \max\{\|g_{[-1]}^{(i)}\|_{\psi_2} : i \in S_{\tau,x}^y\}$  is the maximal Orlicz 2-norm of the rows and  $\Sigma = \mathbb{E}g_{[-1]}^{(i)} g_{[-1]}^{(i)T} \in \mathbb{R}^{(d-1) \times (d-1)}$ . We will control both of these quantities shortly. For now though, notice that this immediately gives us the following lower bound on the minimal singular value of  $\tilde{G}_y$ , which holds with high probability, by the following algebra:

$$\begin{aligned} & \frac{1}{|S_{\tau,x}^y|} \inf_{\|z\|_2=1} \|\tilde{G}_y z\|_2^2 = \frac{1}{|S_{\tau,x}^y|} \inf_{\|z\|_2=1} \|\tilde{G}_y^T \tilde{G}_y z\|_2 \\ &= \inf_{\|z\|_2=1} \left\| \left( \frac{1}{|S_{\tau,x}^y|} \tilde{G}_y^T \tilde{G}_y - \Sigma + \Sigma \right) z \right\|_2 \\ &\geq \inf_{\|z\|_2=1} \|\Sigma z\|_2 - \sup_{\|z\|_2=1} \left\| \left( \frac{1}{|S_{\tau,x}^y|} \tilde{G}_y^T \tilde{G}_y - \Sigma \right) z \right\|_2 \\ &\geq \inf_{\|z\|_2=1} \|\Sigma z\|_2 - CK^2 \left( \sqrt{\frac{d}{|S_{\tau,x}^y|}} + \sqrt{\frac{c_2 d \log(M)}{|S_{\tau,x}^y|}} \right). \end{aligned}$$

Now, we quickly calculate what the minimal singular value of  $\Sigma$  is. In the following calculations, let  $\gamma \sim \mathcal{N}(0, 1)$ :

$$\Sigma = \mathbb{E} g_{[-1]} g_{[-1]}^T = \begin{bmatrix} \alpha & & & \\ & d^{-1} & & \\ & & \ddots & \\ & & & d^{-1} \end{bmatrix},$$

$$\alpha = \frac{1}{\mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} \frac{1}{\sqrt{2\pi d^{-1}}} \int_{\frac{\eta}{\|y_{[-1]}\|_2}}^{\infty} x^2 e^{-\frac{x^2}{2d^{-1}}} dx.$$

We can use integration by parts to expand this integral as

$$\int_{\frac{\eta}{\|y_{[-1]}\|_2}}^{\infty} x^2 e^{-\frac{x^2}{2d^{-1}}} dx = \frac{\eta e^{-\frac{\eta^2}{2\|y_{[-1]}\|_2^2}}}{\|y_{[-1]}\|_2 d^{-1}} + d^{-1} \int_{\frac{\eta}{\|y_{[-1]}\|_2}}^{\infty} e^{-\frac{x^2}{2d^{-1}}} dx. \quad (10)$$

Notice that when we multiply the second term in (10) by  $\frac{1}{\mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} \frac{1}{\sqrt{2\pi d^{-1}}}$  it simplifies to

$$\begin{aligned} & \frac{1}{\mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} \frac{1}{\sqrt{2\pi d^{-1}}} d^{-1} \int_{\frac{\eta}{\|y_{[-1]}\|_2}}^{\infty} e^{-\frac{x^2}{2d^{-1}}} dx \\ &= d^{-1} \frac{1}{\mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} \mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right) = d^{-1}. \end{aligned}$$

Therefore, we have

$$\alpha = d^{-1} + \frac{\sqrt{d^{-1}} \frac{\eta}{\|y_{[-1]}\|_2} e^{-\frac{1}{2d^{-1}} \left(\frac{\eta}{\|y_{[-1]}\|_2}\right)^2}}{\sqrt{2\pi} \mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} > d^{-1}.$$

As for  $K$ , by independence of the entries of  $g_{[-1]}^{(i)}$  we have using Lemma 2.6.8 and Lemma 3.4.2 from [17]

$$\|g_{[-1]}^{(i)}\|_{\psi_2} \leq C \max_{j \in \{2, 3, \dots, d\}} \|g_j^{(i)}\|_{\psi_2},$$

so it suffices to upper-bound the Orlicz 2-norm of the first entry  $g_{[-1],1}^{(i)}$ . By definition, we have

$$\|g_{[-1],1}^{(i)}\|_{\psi_2} = \inf \left\{ \omega > 0 : \mathbb{E} e^{g_{[-1],1}^{(i)2}/\omega^2} \leq 2 \right\}.$$

Note that when  $\omega = 2\sqrt{d^{-1}}$ , we have

$$\begin{aligned} \mathbb{E} e^{g_{[-1],1}^{(i)2}/\omega^2} &= \frac{1}{\mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} \frac{1}{\sqrt{2\pi d^{-1}}} \int_{\frac{\eta}{\|y_{[-1]}\|_2}}^{\infty} e^{\frac{x^2}{4d^{-1}}} e^{-\frac{x^2}{2d^{-1}}} dx \\ &= \frac{\sqrt{2}}{\mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} \frac{1}{\sqrt{2\pi(\sqrt{2}d^{-1/2})^2}} \int_{\frac{\eta}{\|y_{[-1]}\|_2}}^{\infty} e^{-\frac{x^2}{2(\sqrt{2}d^{-1/2})^2}} dx \\ &= \frac{\sqrt{2}\mathbb{P}\left(\sqrt{2d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)}{\mathbb{P}\left(\sqrt{d^{-1}}\gamma \geq \frac{\eta}{\|y_{[-1]}\|_2}\right)} = \frac{\sqrt{2}\mathbb{P}\left(\gamma \geq \frac{\eta\sqrt{d}}{\|y_{[-1]}\|_2\sqrt{2}}\right)}{\mathbb{P}\left(\gamma \geq \frac{\eta\sqrt{d}}{\|y_{[-1]}\|_2}\right)}. \end{aligned}$$

One may numerically check that the above quantity is less than 2 provided  $\frac{\eta\sqrt{d}}{\|y_{[-1]}\|_2} \leq \frac{C\sqrt{d}\log(M)}{M} < \frac{1}{2}$ . Since  $M \gtrsim d\log(d)\log(M)$ , the latter inequality will hold provided the constant in the lower bound of  $M$  is large enough. Hence,  $\|g_{[-1],1}^{(i)}\|_{\psi_2} \lesssim \sqrt{d^{-1}}$ . Consequentially, we have shown with probability at least  $1 - e^{-c_2 d \log(M)} - e^{-c_1 d \log(d) \log(M)}$  that

$$\begin{aligned} \frac{1}{|S_{\tau,x}^{\eta,y}|} \inf_{\|z\|_2=1} \|\tilde{G}_y z\|_2^2 &\geq d^{-1} - Cd^{-1} \left( \sqrt{\frac{d}{|S_{\tau,x}^{\eta,y}|}} + \sqrt{\frac{c_2 d \log(M)}{|S_{\tau,x}^{\eta,y}|}} \right) \\ &\geq d^{-1} \left( 1 - C' \left( \sqrt{\frac{1}{\log(d) \log(M)}} + \sqrt{\frac{1}{\log(d)}} \right) \right) \\ \implies \inf_{\|z\|_2=1} \|\tilde{G}_y z\|_2^2 &\gtrsim d^{-1} |S_{\tau,x}^{\eta,y}| \gtrsim \log(d) \log(M). \end{aligned} \quad (11)$$

To get the uniform result over all  $y$ , condition on the event  $\mathcal{E}$  where  $\max_{i \in [M]} \|g^{(i)}\|_2 \leq C' \sqrt{\log(M)}$ , which occurs with probability at least  $1 - \exp(-c_3 d \log(M))$  [17]. Let  $\varepsilon = \frac{1}{M^2}$  and  $\mathcal{N}$  be an  $\varepsilon$ -net of  $S^{d-1}$ . We know that the cardinality of this net is bounded by  $|\mathcal{N}| \leq (3M^2)^d \leq e^{c_4 d \log(M)}$  [17]. Union bounding over this net, we have that the event  $\mathcal{E}$ , the result of Corollary IV.4 and (11) hold for  $\tilde{G}_\xi$  for all  $\xi \in \mathcal{N}$  simultaneously with probability at least

$$1 - e^{-c_1 d \log(d) \log(M)} - e^{-c_5 d \log(M)} - e^{-c_3 d \log(M)}.$$

Here,  $c_5 = c_2 - c_4 > 0$  provided we choose  $c_2 > c_4$  in our choice of parameter  $t$  in Lemma VII.1. On this event, for any  $y \in S^{d-1} \setminus \mathcal{N}$  we have some  $\xi \in \mathcal{N}$  with  $\|y - \xi\|_2 \leq \varepsilon$  and therefore for any  $i \in \tilde{S}_{\tau,x}^\xi$ , we have

$$\begin{aligned} \langle g_{[-1]}^{(i)}, y_{[-1]} \rangle &\geq \langle g_{[-1]}^{(i)}, \xi_{[-1]} \rangle - \|g_{[-1]}^{(i)}\|_2 \varepsilon \\ &\geq \eta - C' \frac{\sqrt{\log(M)}}{M^2} > 0, \end{aligned}$$

where the above inequality holds, for example, provided  $C_5$  in  $\eta = C_5 \log(M)/M^2$  satisfies  $C_5 > C'$ . So, in other words, we have  $\tilde{S}_{\tau,x}^\xi \subset S_{\tau,x}^y$ . This immediately implies that  $\inf_{\|z\|_2=1} \|G_y z\|_2^2 \geq \inf_{\|z\|_2=1} \|\tilde{G}_\xi z\|_2^2 \gtrsim d^{-1} |\tilde{S}_{\tau,x}^\xi| \gtrsim \log(d) \log(M)$ .  $\square$

*Proof of Theorem IV.1.* We are now ready to make our heuristic proof sketch rigorous. We focus on bounding  $\max \|x - y\|_2$  where the maximum is taken over all  $y \in S^{d-1}$  which satisfy  $\text{sign}(\langle g^{(i)}, x \rangle) = \text{sign}(\langle g^{(i)}, y \rangle)$  for all  $i \in S_{\tau,x}$ . By rotational invariance, we may assume without loss of generality that  $x = e_1$ . Note that if  $\|y_{[-1]}\|_2^2 \leq M^{-2}$  then our result automatically holds, so we focus only on those  $y$  for which said inequality is violated. Since  $\max \|x - y\|_2^2 = 2 - 2 \min y_1$ , we turn to finding a lower bound on  $\min y_1$ . By feasibility, we know for any  $i \in S_{\tau,x}^y$

$$\begin{aligned} \langle g^{(i)}, x \rangle \langle g^{(i)}, y \rangle &= g_1^{(i)} \langle g^{(i)}, y \rangle > 0 \\ \iff (g_1^{(i)})^2 y_1 &> -g_1^{(i)} \langle g_{[-1]}^{(i)}, y_{[-1]} \rangle, \end{aligned}$$

where  $S_{\tau,x}^y$  is as defined in Lemma IV.5. By construction, both sides of this inequality are positive, as  $g_1^{(i)}$  is negative and  $y_1 > 0$  with probability at least  $1 - \exp(-c''' d \log(d))$  for all  $y$  in the same cell as  $x$  by Lemma IV.3. Squaring both sides of the inequality, using  $|g_1^{(i)}| < \tau$  for  $i \in S_{\tau,x}^y \subset W_{\tau,x}$ , and  $\|y_{[-1]}\|_2^2 = (1 - y_1^2)$  means the set of feasible  $y$ 's necessarily satisfy

$$y_1^2 \geq 1 - \frac{\tau^2}{\left\langle g_{[-1]}^{(i)}, \frac{y_{[-1]}}{\|y_{[-1]}\|_2} \right\rangle^2 + \tau^2}, \quad \text{for all } i \in S_{\tau,x}^y.$$

In other words, to control  $\max \|x - y\|_2$  for  $y$  in the same cell as  $x$  induced by  $S_{\tau,x}$ , it is sufficient to find a lower bound on the quantity

$$\min_{z \in S^{d-2}} \max_{i \in S_{\tau,x}^y} \left\langle g_{[-1]}^{(i)}, z \right\rangle^2.$$

Letting  $G_y$  be as defined in Lemma IV.5, notice that we may lower bound this quantity by

$$\begin{aligned} \min_{z \in S^{d-2}} \max_{i \in S_{\tau,x}^y} \left\langle g_{[-1]}^{(i)}, z \right\rangle^2 &\geq \min_{z \in S^{d-2}} |S_{\tau,x}^y|^{-1} \sum_{i \in S_{\tau,x}^y} \left\langle g_{[-1]}^{(i)}, z \right\rangle^2 \\ &= |S_{\tau,x}^y|^{-1} \min_{z \in S^{d-2}} \|G_y z\|_2^2. \end{aligned}$$

By Lemma IV.5, and Corollary IV.4 we know with probability at least  $1 - e^{-c'd\log(d)\log(M)} - 2e^{-c''d\log(M)} \min_{z \in S^{d-2}} \|G_y z\|_2^2 \gtrsim \log(d)\log(M)$  and  $|S_{\tau,x}^y| \lesssim d\log(d)\log(M)$  for all  $y$ . Therefore,

$$\min_{z \in S^{d-2}} \max_{i \in S_{\tau,x}^y} \left\langle g_{[-1]}^{(i)}, z \right\rangle^2 \gtrsim d^{-1}.$$

Combining our results from this section, we have with high probability

$$y_1^2 \geq 1 - \frac{\tau^2}{Cd^{-1} + \tau^2}.$$

With our choice of  $\tau = C_2 \frac{\sqrt{d}\log(d)\log(M)}{M}$ , this means

$$y_1^2 \geq 1 - C \frac{d\log^2(d)\log^2(M)}{\frac{M^2}{d} + d\log^2(d)\log^2(M)}.$$

This concludes the proof of Theorem IV.1.

## V. UNIFORM RESULTS

As a result of working so hard to get uniform control on  $\langle g_{[-1]}^{(i)}, y_{[-1]}\rangle$  for all  $y$  in the same cell as  $x$  using a union bound argument, we can get a uniform result for all  $x \in S^{d-1}$  nearly for free. The consequence of uniformly bounding the radius of the cell around  $x$  induced by the vectors selected according to  $S_{\tau,x}$  gives an upper bound on the radii for all cells in the tessellation induced by the entire collection of vectors  $g^{(1)}, \dots, g^{(M)}$ . This follows from the inclusion  $\{y \in S^{d-1} : \text{sign}(\langle g^{(i)}, y \rangle) = \text{sign}(\langle g^{(i)}, x \rangle)\}$ , for all  $i \in [M]\} \subset \{y \in S^{d-1} : \text{sign}(\langle g^{(i)}, y \rangle) = \text{sign}(\langle g^{(i)}, x \rangle)\}$ , for all  $i \in S_{\tau,x}$ .

**Corollary V.1.** *Let  $g^{(i)} \in \mathbb{R}^d$  be drawn from  $\mathcal{N}(0, d^{-1}I)$ , and set  $\tau = C_2 \frac{\sqrt{d}\log(d)\log(M)}{M}$ . Then with probability at least  $1 - 3\exp(-cd\log(d)) - 2\exp(-c'd\log(M))$  the result of Theorem IV.1 holds uniformly for all  $x \in S^{d-1}$ .*

*Proof.* As was the case in our previous lemmata that used a union bound argument and continuity, we will need to initially introduce some small modifications to our definitions to allow ourselves some wiggle room. As before, this will help us move from an argument that works for points on an  $\epsilon$ -net to arbitrary points. Consider for fixed  $x \in S^{d-1}$  the subset

$$\hat{S}_{\tau,x} := \{i \in [M] : -\tau < \langle g^{(i)}, x \rangle < -\tau/2\}.$$

Using a similar argument to that used in the proof of Lemma IV.2, we can show with high probability that  $C_6 d \log(d) \log(M) \leq |\hat{S}_{\tau,x}| \leq C_7 d \log(d) \log(M)$  for any fixed  $x$ . Indeed, the expected

value is  $\mathbb{E}|\hat{S}_{\tau,x}| = M\mathbb{P}\left(\sqrt{d^{-1}}g \in (-\tau, -\tau/2)\right)$  which means we can use upper and lower Riemann sum estimates to get analogous bounds as in Lemma IV.2. Lemma IV.3 already holds uniformly so there's no need to use a union bound for that event. The arguments in Corollary IV.4 and Lemma IV.5 hold without any modifications because they concern the projected random variables  $P_{x^\perp}g^{(i)}$ , the projections of  $g^{(i)}$  on the orthogonal complement of the span of  $x$ . These random variables are independent of the event associated with  $\hat{S}_{\tau,x}$ . Therefore, the statement of Theorem IV.1 also holds if  $S_{\tau,x}$  is replaced with  $\hat{S}_{\tau,x}$ .

With that matter settled, condition on the event that  $\max_{i \in [M]} \|g^{(i)}\|_2 \leq C'\sqrt{\log(M)}$  which occurs with probability at least  $1 - \exp(-c'd\log(M))$  [17]. Now, let  $\tau = \frac{C_2\sqrt{d}\log(d)\log(M)}{M}$ ,  $\tau' = \frac{\tau}{2C'\sqrt{\log(M)}}$ , and let  $\mathcal{N}$  be a  $\tau'$  covering of  $S^{d-1}$ . Note that  $|\mathcal{N}| \leq \left(\frac{3}{\tau'}\right)^d \leq M^{c''d}$ . By a union bound argument, we have the result of Theorem IV.1 using  $\hat{S}_{\tau,y}$  in place of  $S_{\tau,y}$  holding for all  $y \in \mathcal{N}$  with probability at least  $1 - 3\exp(-cd\log(d)) - \exp(-c'd\log(M))$ .

For an arbitrary  $x \in S^{d-1}$ , let  $y \in \mathcal{N}$  satisfy  $\|x - y\|_2 \leq \tau'$ . Then for any  $i \in \hat{S}_{\tau,y}$  we have

$$\begin{aligned} \langle g^{(i)}, x \rangle &= \langle g^{(i)}, y \rangle + \langle g^{(i)}, x - y \rangle < -\tau/2 + C'\sqrt{\log(M)}\|x - y\|_2 \\ &< -\tau/2 + \tau/2 = 0, \end{aligned}$$

and, additionally, using a similar argument

$$\langle g^{(i)}, x \rangle > -3\tau/2.$$

In other words, we've just shown that  $\hat{S}_{\tau,y} \subset S_{3\tau/2,x}$ , and furthermore that  $\text{sign}(\langle g^{(i)}, x \rangle) = \text{sign}(\langle g^{(i)}, y \rangle)$  for all  $i \in \hat{S}_{\tau,y}$ . The inclusion  $\hat{S}_{\tau,y} \subset S_{3\tau/2,x}$  tells us that the cell containing  $x$  induced by the hyperplanes in  $S_{3\tau/2,x}$  has a radius no larger than the radius of the cell containing  $x$  induced by the hyperplanes from  $\hat{S}_{\tau,y}$ . Denote the former cell by  $\mathcal{C}$  and the latter by  $\hat{\mathcal{C}}$ . By triangle inequality and Theorem IV.1

$$\begin{aligned} \max_{z \in \mathcal{C}} \|x - z\|_2 &\leq \max_{z \in \hat{\mathcal{C}}} \|x - z\|_2 \leq \|x - y\|_2 + \max_{z \in \hat{\mathcal{C}}} \|y - z\|_2 \\ &\lesssim \frac{d\log(d)\log(M)}{\sqrt{M^2 + d^2\log^2(d)\log^2(M)}}. \end{aligned}$$

□

**Remark 1.** One potential practical application of this result is an encoding algorithm for efficiently representing a vector  $x \in \mathbb{R}^d$  with  $M$  one-bit measurements of the form  $\text{sign}(\langle g^{(i)}, x \rangle)$ . Since the cardinality of  $S_{\tau,x}$  is  $k \approx d \text{ polylog}(M)$ , one could store or transmit the indices associated with  $S_{\tau,x}$  using  $\log_2 \binom{M}{k} \lesssim d \text{ polylog}(M)$  bits in addition to the  $k$  bits needed to encode the one bit measurements associated with  $S_{\tau,x}$ . Using standard reconstruction techniques (see, e.g., [6], [7]), one would then recover  $x$  from  $\text{sign}(\langle g^{(i)}, x \rangle)$ ,  $i \in S_{\tau,x}$ , with the error given in Corollary V.1. This encoding scheme results in a root-exponential decay of the error in the number of bits used.

## VI. ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for their thoughtful and useful suggestions which have significantly improved the manuscript.

## VII. APPENDIX

**Lemma VII.1.** Fix  $x = e_1$  and  $y \in S^{d-1}$ . Let  $\tilde{G} \in \mathbb{R}^{|S_{\tau,x}^{\eta,y}| \times (d-1)}$  be as in Lemma IV.5 and  $\Sigma := \mathbb{E} g_{[-1]}^{(i)} g_{[-1]}^{(i)T}$ . Then for any  $t > 0$  we have

$$\begin{aligned} \left\| |S_{\tau,x}^{\eta,y}|^{-1} \tilde{G}_y^T \tilde{G}_y - \Sigma \right\|_{op} &\leq K^2 \max\{\delta, \delta^2\} \\ \delta &:= C \left( \sqrt{\frac{d-1}{|S_{\tau,x}^{\eta,y}|}} + \frac{t}{\sqrt{|S_{\tau,x}^{\eta,y}|}} \right), \end{aligned}$$

with probability at least  $1 - 2 \exp(-t^2)$ . Here,  $K := \max_{j \in \{2, \dots, d\}} \|g_j^{(i)}\|_{\psi_2}$  where  $i \in S_{\tau,x}^{\eta,y}$ .

*Proof.* This proof is almost verbatim that of Theorem 4.6.1 in [17] and we only include it here for completeness. Let  $\mathcal{N} \subset S^{d-2}$  be a  $1/4$ -net. By definition

$$\begin{aligned} \left\| |S_{\tau,x}^{\eta,y}|^{-1} \tilde{G}_y^T \tilde{G}_y - \Sigma \right\|_{op}^2 &= \sup_{\|z\|_2=1} \left\| |S_{\tau,x}^{\eta,y}|^{-1} \tilde{G}_y z \right\|_2^2 - z^T \Sigma z \\ &\leq 2 \sup_{z \in \mathcal{N}} \left\| |S_{\tau,x}^{\eta,y}|^{-1} \tilde{G}_y z \right\|_2^2 - z^T \Sigma z. \end{aligned}$$

where the last line follows by Lemma 4.4.1 in [17]. We will use a union bound to control this quantity. To that end, fix  $z \in S^{d-2}$ . We can expand this random variable as

$$\left\| |S_{\tau,x}^{\eta,y}|^{-1} \tilde{G}_y z \right\|_2^2 - z^T \Sigma z = |S_{\tau,x}^{\eta,y}|^{-1} \sum_{i \in S_{\tau,x}^{\eta,y}} z^T g_{[-1]}^{(i)} g_{[-1]}^{(i)T} z - z^T \Sigma z.$$

The random variable  $g_{[-1]}^{(i)T} z$  is sub-gaussian with  $\|g_{[-1]}^{(i)T} z\|_{\psi_2} \leq \max_{j \in \{2, \dots, d\}} \|g_j^{(i)}\|_{\psi_2} = K$  since the entries of  $g^{(i)}$  are independent. Therefore, the random variables  $z^T g_{[-1]}^{(i)} g_{[-1]}^{(i)T} z - z^T \Sigma z$  are sub-exponential with  $\|z^T g_{[-1]}^{(i)} g_{[-1]}^{(i)T} z - z^T \Sigma z\|_{\psi_1} \leq CK^2$ . Using Bernstein's inequality [17] we get

$$\begin{aligned} &\mathbb{P} \left( \left| |S_{\tau,x}^{\eta,y}|^{-1} \sum_{i \in S_{\tau,x}^{\eta,y}} z^T g_{[-1]}^{(i)} g_{[-1]}^{(i)T} z - z^T \Sigma z \right| \geq \varepsilon/2 \right) \\ &\leq 2 \exp \left( -c \min \left\{ \frac{\varepsilon^2}{K^4}, \frac{\varepsilon}{K^2} \right\} |S_{\tau,x}^{\eta,y}| \right). \end{aligned}$$

Setting  $\varepsilon/K^2 = \max\{\delta, \delta^2\}$  and recalling the definition of  $\delta := C \left( \sqrt{\frac{d-1}{|S_{\tau,x}^{\eta,y}|}} + \frac{t}{\sqrt{|S_{\tau,x}^{\eta,y}|}} \right)$ , this bound reduces to

$$\begin{aligned} &\mathbb{P} \left( \left| |S_{\tau,x}^{\eta,y}|^{-1} \sum_{i \in S_{\tau,x}^{\eta,y}} z^T g_{[-1]}^{(i)} g_{[-1]}^{(i)T} z - z^T \Sigma z \right| \geq \frac{1}{2} K^2 \max\{\delta, \delta^2\} \right) \\ &\leq 2 \exp(-cC(d-1+t^2)). \end{aligned}$$

To get a uniform bound over the net, we remark that  $|\mathcal{N}| \leq 9^{d-2}$ , therefore

$$\begin{aligned} &\mathbb{P} \left( \max_{z \in \mathcal{N}} \left| |S_{\tau,x}^{\eta,y}|^{-1} \sum_{i \in S_{\tau,x}^{\eta,y}} z^T g_{[-1]}^{(i)} g_{[-1]}^{(i)T} z - z^T \Sigma z \right| \geq \frac{1}{2} K^2 \max\{\delta, \delta^2\} \right) \\ &\leq 9^{d-2} 2 \exp(-cC(d-1+t^2)) \leq 2 \exp(-ct^2), \end{aligned}$$

provided  $C$  in the definition of  $\delta$  is at least  $\log(9)/c$ .  $\square$

## REFERENCES

- [1] R. Schneider and W. Weil, *Stochastic and integral geometry*. Springer Science & Business Media, 2008.
- [2] D. Hug and R. Schneider, “Random conical tessellations,” *Discrete & Comput. Geom.*, vol. 56, no. 2, pp. 395–426, 2016.
- [3] ———, “Another look at threshold phenomena for random cones,” *arXiv preprint arXiv:2103.11394*, 2021.
- [4] T. Godland, Z. Kabluchko, and C. Thäle, “Random cones in high dimensions i: Donoho-tanner and cover-efron cones,” *arXiv preprint arXiv:2012.06189*, 2020.
- [5] T. M. Cover and B. Efron, “Geometrical probability and random points on a hypersphere,” *The Annals of Mathematical Statistics*, pp. 213–220, 1967.
- [6] V. K. Goyal, M. Vetterli, and N. T. Thao, “Quantized overcomplete expansions in  $\mathbb{R}^N$ : analysis, synthesis, and algorithms,” *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 16–31, 1998.
- [7] A. M. Powell, R. Saab, and Ö. Yilmaz, “Quantization and finite frames,” in *Finite frames*. Springer, 2013, pp. 267–302.
- [8] V. Vapnik and A. Chervonenkis, “Theory of pattern recognition,” 1974.
- [9] E. Chou, C. S. Güntürk, F. Krahmer, R. Saab, and Ö. Yilmaz, “Noise-shaping quantization methods for frame-based and compressive sampling systems,” *Sampling theory, a renaissance*, pp. 157–184, 2015.
- [10] P. T. Boufounos, “Universal rate-efficient scalar quantization,” *IEEE transactions on information theory*, vol. 58, no. 3, pp. 1861–1872, 2011.
- [11] P. M. Long, “On the sample complexity of PAC learning half-spaces against the uniform distribution,” *IEEE Trans. on Neural Networks*, vol. 6, no. 6, pp. 1556–1559, 1995.
- [12] D. Angluin, “Queries and concept learning,” *Mach. Learn.*, vol. 2, no. 4, pp. 319–342, 1988.
- [13] C. Campbell, N. Cristianini, and A. Smola, “Query learning with large margin classifiers,” in *ICML*, 2000, pp. 111–118.
- [14] S. Tong and D. Koller, “Support vector machine active learning with applications to text classification,” *J Mach. Learn. Res.*, vol. 2, pp. 45–66, 2001.
- [15] M.-F. Balcan, A. Broder, and T. Zhang, “Margin based active learning,” in *Proc. Learn. Theory*. Springer, 2007, pp. 35–50.
- [16] S. Dasgupta, D. J. Hsu, and C. Monteleoni, “A general agnostic active learning algorithm,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 353–360.
- [17] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018, vol. 47.
- [18] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, “Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors,” *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2082–2102, 2013.