ONE BIT SENSING, DISCREPANCY, AND STOLARSKY PRINCIPLE.

DMITRIY BILYK AND MICHAEL T. LACEY

ABSTRACT. A sign-linear one bit map from the $d$-dimensional sphere $S^d$ to the $n$-dimensional Hamming cube $H^n = \{-1, +1\}^n$ is given by

$$x \mapsto \{\text{sign}(x \cdot z_j) : 1 \leq j \leq n\}$$

where $\{z_j\} \subset S^d$. For $0 < \delta < 1$, we estimate $N(d, \delta)$, the smallest integer $n$ so that there is a sign-linear map which has the $\delta$-restricted isometric property, where we impose normalized geodesic distance on $S^d$, and Hamming metric on $H^n$. Up to a polylogarithmic factor, $N(d, \delta) \approx \delta^{-2+\frac{2}{d+1}}$, which has a dimensional correction in the power of $\delta$. This is a question that arises from the one bit sensing literature, and the method of proof follows from geometric discrepancy theory. We also obtain an analogue of the Stolarsky invariance principle for this situation, which implies that minimizing the $L^2$ average of the embedding error is equivalent to minimizing the discrete energy $\sum_{i,j} \left(\frac{1}{2} - d(z_i, z_j)\right)^2$, where $d$ is the normalized geodesic distance.

1. Introduction

Let $d \geq 2$ and let $S^d \subset \mathbb{R}^{d+1}$ denote the $d$-dimensional unit sphere. We approximate, above and below, the minimal number of hyperplanes required to guarantee that for any two points $x, y \in S^d$ the proportion of hyperplanes that separates these points approximates the distance between the points up to a given threshold. These questions have connections to different topics such as one-bit sensing, a non-linear variant of compressive sensing, and geometric functional analysis (almost isometric embeddings). Proof techniques are taken from geometric discrepancy theory.

On the sphere $S^d$ denote by $d(x,y)$ the geodesic distance between $x$ and $y$ on the sphere normalized so that the distance between antipodal points is 1, thus

$$d(x,y) = \frac{\cos^{-1}(x \cdot y)}{\pi}.$$ 

The $n$-dimensional Hamming cube $H^n = \{-1, +1\}^n$ has the Hamming metric

$$d_H(s,t) = \frac{1}{2n} \sum_{n=1}^n |s_j - t_j| = \frac{1}{n} \cdot \#\{1 \leq j \leq n : s_j \neq t_j\},$$

where $s = (s_1, \ldots, s_n) \in H^n$, and similarly for $t$, i.e. $d_H(s,t)$ measures the proportion of the coordinates in which $s$ and $t$ differ. We consider sign-linear maps from $S^d$ to $H^n$ given by

$$\varphi_Z(x) = \{\text{sgn}(z_j \cdot x) : 1 \leq j \leq n\}$$

1. Research supported in part by the NSF grant DMS 1101519.
2. Research supported in part by the NSF grant DMS 1265570.
where \( Z = \{z_1, z_2, \ldots, z_n\} \subset \mathbb{S}^d \). Note that, with abuse of notation, \( \varphi_Z(x) = \text{sgn}(Ax) \), where the rows of \( A \) consist of the vectors \( z_1, \ldots, z_n \).

Each coordinate of the map \( \varphi_Z \) divides \( \mathbb{S}^d \) into two hemispheres, and the Hamming distance

\[
d_H(\varphi_Z(x), \varphi_Z(y))
\]

is the proportion of of hyperplanes \( z_j^\perp \) that separate the points \( x \) and \( y \). It is easy to see that, if one chooses a hyperplane \( z_j^\perp \) uniformly at random, then

\[
\mathbb{P}\{\text{sgn}(x \cdot z) \neq \text{sgn}(y \cdot z)\} = d(x, y).
\]

This is the founding instance of the Crofton formula from integral geometry [35, p. 36-40]. Hence for a large number of random (or carefully chosen deterministic) hyperplanes, the Hamming distance \( d_H(x, y) \) should be close to the geodesic distance \( d(x, y) \).

The closeness is quantified by the following definition of the restricted isometric property (RIP), a basic concept in compressed sensing literature.

**Definition 1.2.** Let \( 0 < \delta < 1 \). We say that \( \varphi : \mathbb{S}^d \mapsto \mathbb{H}^n \) satisfies \( \delta \)-RIP if

\[
sup_{x, y \in \mathbb{S}^d} |d_H(\varphi(x), \varphi(y)) - d(x, y)| < \delta.
\]

The integer \( N(d, \delta) \) is the minimal integer \( n \) such that there exists an \( n \)-point set \( Z \subset \mathbb{S}^d \), such that \( \varphi_Z \) is a \( \delta \)-RIP map.

In the sign-linear case \( \varphi = \varphi_Z \), we set

\[
\Delta_Z(x, y) = d_H(\varphi_Z(x), \varphi_Z(y)) - d(x, y).
\]

Building intuition, we can set \( N_{\text{rdm}}(d, \delta) \) to be the smallest integer \( N \) so that drawing \( Z \) uniformly at random, the sign-linear map \( \varphi_Z \) is a \( \delta \)-RIP map, with chance at least \( 1/2 \). In a companion paper [10], we will show that

\[
N_{\text{rdm}}(d, \delta) \lesssim d\delta^{-2}.
\]

This is the best known bound for the linear embedding of the sphere, and the power of \( \delta^{-2} \) is sharp in the random case, as follows from the Central Limit Theorem.

In this paper we show that in general there is a dimensional correction to the power of \( \delta \). This is our first main result.

**Theorem 1.6.** For all \( d \in \mathbb{N} \) and \( 0 < \delta < 1 \), there holds,

\[
N(d, \delta) \approx \log \delta^{-2 + \frac{1}{d+1}}.
\]

where the equality holds up to a dimensional constant and a polylogarithmic factor in \( \delta \).

The upper bound in (1.7) is achieved by exhibiting a \( Z \) of small cardinality, for which \( \varphi_Z \) satisfies \( \delta \)-RIP. Jittered sampling, a cross between purely random and deterministic constructions, provides the example. Loosely speaking, first one divides the sphere \( \mathbb{S}^d \) into \( n \) roughly equal pieces, and then chooses a random point in each of them, see §2 for details. The lower bound is the universal statement that every \( Z \) of sufficiently small cardinality does not yield a \( \delta \)-RIP map. It is a deep fact from geometric discrepancy theory.

Most of the prior work concerns randomly selected \( Z \). Jacques and coauthors [24, Thm 2] proved (1.5), with an additional logarithmic term in \( \delta \). In fact, they also considered the sparse vectors in \( \mathbb{S}^d \). Plan and Vershynin [32] studied this question, looking for RIP’s for
general subsets $K \subset S^d$ into the Hamming cube, of which the sparse vectors are a prime example. They proved [32, Thm 1.2] that $N_{rdm}(d, \delta) \lesssim d\delta^6$, and conjectured (1.5), at least in the random case. Neither paper anticipates the dimensional correction in $\delta$ above.

Since in applications the dimension $d$ is often quite large, we considered a non-asymptotic version of the upper bound in (1.7) and computed an effective value of the constant $C_d$, proving that it grows roughly as $d^{5/2}$ (see Theorem 1.20 for a more precise statement):

$$N(d, \delta) \leq \max \left\{ C_d^{5/2} \delta^{-2+\frac{1}{10d}}, 100d \right\},$$

where $C > 0$ is an absolute constant.

Our second main result goes in a somewhat different direction, we also show in Theorem 1.21 that the $L^2$ norm of $\Delta_Z(x, y)$ given in (1.4) satisfies an analog of the Stolarsky principle [35], which implies that minimizing the $L^2$ average of $\Delta_Z(x, y)$ is equivalent to minimizing the discrete energy of the form $\frac{1}{n^2} \sum_{i,j} \left( \frac{1}{2} - d(z_i, z_j) \right)^2$. This suggests interesting connections to such objects as spherical codes, equiangular lines, and frames. See Theorem 1.21 and §3 for details.

**One-Bit Sensing.** The RIP property was formulated by Candes–Tao [16], and is a basic concept in the compressive sensing [21, Chap. 6]. It can be formulated as a property between two metric spaces, and so has many interesting variants.

One-bit sensing was initiated by Boufounos–Baraniuk [12]. The motivation for the one bit measurements $\text{sgn}(x \cdot y)$ are that (a) they form a canonical non-linearity on the measurement, as well as a canonical quantization of data, (b) there are striking technological advances which employ non-linear observations, and (c) it is therefore of interest to develop a comprehensive theory of non-linear signal processing.

The subsequent theory has been developed by [23, 24, 30, 32]. For upper bounds, random selection of points on a sphere is generally used. Note that [23, Thm 1] does contain a lower bound on the rate of recovery of a one-bit decoder. Plan–Vershynin [32] established results on one-bit RIP maps for arbitrary subsets of the unit sphere, and proposed some ambitious conjectures about bounds for these maps. In a companion paper [10] we will investigate some of these properties in the case of randomly selected hyperplanes. The results about one-bit sensing have been used in other interesting contexts, see the papers cited above as well as [4, 31].

Lower bounds, like the ones proved in Theorem 1.6, indicate the limits of what can be accomplished in compressive sensing. See for instance Larsen–Nelson [22] who prove a lower bound for dimension reduction in the Johnson–Lindenstrauss Lemma. This lemma is a foundational result in dimension reduction. In short, it states that for $X \subset S^d \subset \mathbb{R}^{d+1}$ of cardinality $k$, there is a linear map $A : \mathbb{R}^{d+1} \mapsto \mathbb{R}^n$, which, restricted to $X$, satisfies $\delta$-RIP, provided $n \gtrsim \delta^{-2} \log k$. This has many proofs, see for instance [34]. The connection of this lemma to compressed sensing is well known, see e.g. [3].

In our companion paper [10], we will show that the one-bit variant of the Johnson–Lindenstrauss bound holds, with the same bound $n \gtrsim \delta^{-2} \log k$. It would be interesting to know if this bound is also sharp. The clever techniques of [22] are essentially linear in
nature, so that a new technique is needed. Progress on this question could have consequences on lower bounds for non-linear Johnson–Lindenstrauss RIPs.

**Dvoretzky’s Theorem.** The results of this paper are also related to Dvoretzky’s Theorem [19], which states that for all \( \epsilon > 0 \) and all dimensions \( d \), there exists \( N = N(d, \epsilon) \) so that any Banach space \( X \) of dimension \( N \) contains a subspace \( Y \) of dimension \( d \) which embeds into Hilbert space with distortion at most \( 1 + \epsilon \). (Finite distortion must hold uniformly at all scales, in contrast to the RIP, which ignores sufficiently small scales.) This is a fundamental result in geometric functional analysis, and has sophisticated variants in metric spaces, [2,28].

It is interesting that the argument of Plan–Vershynin [19, §3.2] relies upon a variant of Dvoretzky’s Theorem and indeed ties improved bounds in Dvoretzky’s Theorem to improvements in one-bit RIP maps. In view of the connection between RIP properties in geometric discrepancy identified in this paper, there are new techniques that could be brought to bear on this question.

**Geometric Interpretation.** The results above can be interpreted as properties of about tessellations of the sphere \( S^d \) induced by the hyperplanes \( \{ z^+ : z \in Z \} \). The integer \( N(d, \delta) \) is the smallest size of \( Z \) so that for all \( x, y \in S^d \), the proportion of hyperplanes from \( Z \) that separate \( x, y \) is bounded above and below by \( d(x, y) \pm \delta \). This is the geometric language used by Plan–Vershynin [19], which indicates a connection with geometric discrepancy theory.

We point the reader to some recent papers which investigate integration on spheres and related geometrical questions: [1,15,33]

In this paper, we shall denote by \( \sigma \) the surface measure on the sphere, normalized so that \( \sigma(S^d) = 1 \). Unnormalized (Hausdorff) measure on \( S^d \) will be denoted by \( \sigma^* \). We shall use the notation \( \omega = \sigma^*_d(S^{d-1}) \) and \( \Omega = \sigma^*_d(S^d) \). In particular

\[
\Omega = \frac{2\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)},
\]

and the ratio between these two, which will appear often, satisfies (see [26])

\[
\frac{\omega}{\Omega} = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \leq \sqrt{\frac{d}{2\pi}}.
\]

The notation \( A \lesssim B \) means that \( A \leq CB \) for some fixed constant \( C > 0 \). Occasionally, the implicit constant may depend on the dimension \( d \) (this will be made clear in the context), but it is always independent of \( N \) and \( \delta \).

1.1. **Discrepancy.** We phrase the RIP property in the language of geometric discrepancy theory on the sphere \( S^d \). Let \( Z = \{ z_1, \ldots, z_N \} \) be an \( N \)-point subset of \( S^d \). The discrepancy of \( Z \) relative to a measurable subset \( S \subset S^d \) is

\[
D(Z, S) = \frac{1}{N} \cdot \# \{ Z \cap S \} - \sigma(S)
\]

We define the extremal \( (L^\infty) \) discrepancy of \( Z \) with respect to a family \( S \) of measurable subsets of \( S^d \) to be

\[
D_S(Z) = \sup_{S \in S} \left| D(Z, S) \right|.
\]
If the family $\mathcal{S}$ admits a natural measure then one may also replace the supremum above by an $L^2$ average, say. The main questions of discrepancy theory are: How small can discrepancy be? What are good or optimal point distributions? These questions have profound connections to approximation theory, probability, combinatorics, number theory, computer science, analysis, etc, see [7, 18, 27].

In this sense the quantity $\Delta_Z(x, y)$ defined in (1.4) clearly has a discrepancy flavor. In fact, (and this is perhaps the most important observation of the paper) the problem of uniform tessellations can actually be reformulated as a problem on geometric discrepancy with respect to spherical wedges.

Denote the set of normals of those hyperplanes that separate $x$ and $y$ by

$$W_{xy} = \{ z \in \mathbb{S}^d : \text{sgn}(x \cdot z) \neq \text{sgn}(y \cdot z) \}.$$  

The letter $W$ stands for wedge, since the set $W_{xy}$ does in fact look like a spherical wedge, i.e. the subset of the sphere lying between the hyperplanes $x \cdot z = 0$ and $y \cdot z = 0$, see Figure 1.

It follows from the Crofton formula (1.1) that

$$\sigma(W_{xy}) = \mathbb{P}(z^\perp \text{ separates } x \text{ and } y) = d(x, y).$$

Therefore we can rewrite the quantity (1.4) as

$$\Delta_Z(x, y) = \frac{\#(Z \cap W_{xy})}{N} - \sigma(W_{xy}) = \frac{1}{N} \sum_{k=1}^N 1_{W_{xy}}(z_k) - \sigma(W_{xy}) = D(Z, W_{xy}).$$

This is the discrepancy of the $N$-point distribution $Z$ with respect to the wedge $W_{xy}$.

The RIP property is the $L^\infty$ discrepancy with respect to wedges. Indeed, according to definitions (1.3) and (1.9), the map $\varphi_Z$ is a $\delta$-RIP exactly when the quantity below is at most $\delta$.

$$\|\Delta_Z(x, y)\|_\infty = \sup_{x,y \in \mathbb{S}^d} \left| \frac{\#(Z \cap W_{xy})}{N} - \sigma(W_{xy}) \right| = D_{\text{wedge}}(Z).$$

The problem of estimating $N(d, \delta)$ is thus simply inverse to obtaining discrepancy estimates in terms of $N$, and this is precisely the approach we shall take.
1.2. **A point of reference: spherical cap discrepancy.** We recall the classical results concerning the discrepancy for spherical caps. For \( x \in \mathbb{S}^d \), and \( t \in [-1, 1] \), let \( C(x, t) \) be the spherical cap of height \( t \) centered at \( x \), given by
\[
C(x, t) = \{ p \in \mathbb{S}^d : p \cdot x \geq t \}.
\]
Denote the set of all spherical caps by \( C \). For and \( N \)-point set \( Z \subset \mathbb{S}^d \) let
\[
D_{\text{cap}}(Z) = \sup_{C \in C} |D(Z, C)| = \sup_{C \in C} \left| \frac{\#(Z \cap C)}{N} - \sigma(C) \right|
\]
be the extremal discrepancy of \( Z \) with respect to spherical caps \( C \). The following classical results due to J. Beck \([5,6]\) yield almost precise information about the growth of this quantity in terms of \( N \).

**Beck’s Spherical Cap Discrepancy Theorem.** *For dimensions \( d \geq 2 \), there holds*

**Upper Bound:** There exists an \( N \)-point set \( Z \subset \mathbb{S}^d \) with spherical cap discrepancy
\[
D_{\text{cap}}(Z) \lesssim N^{-\frac{1}{2} - \frac{1}{d}} \sqrt{\log N}.
\]

**Lower Bound:** For any \( N \)-point set \( Z \subset \mathbb{S}^d \) the spherical cap discrepancy satisfies
\[
D_{\text{cap}}(Z) \gtrsim N^{-\frac{1}{2} - \frac{1}{d}}.
\]

We will elaborate on the upper bound (1.11). It is proved using a construction known as *jittered sampling*, which produces a semi-random point set. We describe this construction in much detail in §2. Stating this result in [6] Beck just states that “using probabilistic ideas it is not hard to show” and refers to his paper [5], where this fact is proved for rotated rectangles, not spherical caps. (It is well known that jittered sampling is applicable in many geometric settings.) In the book [7] the algorithm is described in some more detail, but one of the key steps, namely regular equal-area partition of the sphere, is only postulated. This construction was only recently rigorously formalized and effective values of the underlying constants have been found \([20,26]\). See §2.1 for further discussion.

It is generally believed that standard low-discrepancy sets, while providing good bounds with respect to the number of points \( N \), yield very bad, often exponential, dependence on the dimension. However, as we shall see, it appears that for jittered sampling this behavior is quite reasonable (see also [29] for a discussion of a similar effect). This is consistent with the fact that this construction is intermediate between purely random and deterministic sets.

Since we are interested both in asymptotic and non-asymptotic regimes, we shall explore this construction (in the case of spherical wedges) very scrupulously tracing the dependence of the constant on the dimension.

The proof of the lower bound (1.12) is Fourier-analytic in nature and holds with the smaller \( L^2 \) average in place of the supremum,
\[
D_{\text{cap},L^2}(Z) = \left( \int_{-1}^{1} \int_{\mathbb{S}^d} \left| \frac{\#(Z \cap C(x, t))}{N} - \sigma(C(x, t)) \right|^2 d\sigma(x) \, dt \right)^{1/2}.
\]

More precisely, a stronger lower bound holds.
\[
D_{\text{cap},L^2}(Z) \gtrsim N^{-\frac{1}{2} - \frac{1}{d'}}.
\]

(1.13)
This bound is sharp: the $L^2$ discrepancy of jittered sampling yields a bound akin to (1.11), but without $\sqrt{\log N}$.

Strikingly, minimizing the $L^2$ discrepancy is the same as maximizing the sum of pairwise distances between the vectors in $Z$, which is the main result of [35].

**Stolarsky Invariance Principle 1.14.** In all dimensions $d \geq 2$, for any $N$-point set $Z = \{z_1, \ldots, z_N\} \subset S^d$, the following holds

$$
(1.15) \quad \frac{1}{c_d} [D_{\text{cap},L^2}(Z)]^2 = \int_{S^d} \int_{S^d} \|x - y\| \, d\sigma(x) \, d\sigma(y) - \frac{1}{N^2} \sum_{i,j=1}^N \|z_i - z_j\|,
$$

where $c_d = \frac{1}{2} \int_{S^d} |p \cdot z| \, d\sigma(z)$.

The square of the $L^2$ discrepancy is exactly the difference between the continuous potential energy given by $\|x - y\|$ and the discrete energy induced by the points of $Z$. Alternate proofs of Stolarsky principle principle can be found in [9,14].

**1.3. Main results.** Analogues of Beck’s discrepancy estimates, Theorems 1.2 and Stolarsky invariance principle hold for spherical wedges $W_{xy}$. These in turn imply results for sign-linear RIP maps. Moreover, we shall explore the dependence of the upper estimates on the dimension $d$. Recall the definition of a wedge $W_{xy}$ in (1.10) and set

$$
D_{\text{wedge}}(Z) = \sup_{x,y \in S^d} |D(Z,W_{xy})| = \sup_{x,y \in S^d} \left| \frac{\#(Z \cap W_{xy})}{N} - \sigma(W_{xy}) \right|.
$$

**Theorem 1.16.** For all integers $d \geq 2$, there is a $B_d > 0$ so that for all integers $N \geq 1$,

**Upper Bound:** There exists a distribution of $N$ points $Z = \{z_1, \ldots, z_N\} \subset S^d$ such that

$$
(1.17) \quad D_{\text{wedge}}(Z) \leq C_d N^{-\frac{1}{d} + \frac{1}{4d}} \sqrt{\log N}.
$$

Provided $N \geq 100d$, we have $C_d \leq 20d^{\frac{1}{d} + \frac{1}{4d}}$.

**Lower Bound:** For any $Z \subset S^d$ with cardinality $N$,

$$
(1.18) \quad D_{\text{wedge}}(Z) \geq B_d N^{-\frac{1}{d} - \frac{1}{2d}}.
$$

A similar result (with different absolute constants, but the same exponents of $d$) holds for the implicit constants in Beck’s spherical cap discrepancy estimate (1.11).

Both inequalities are known in a very similar geometric situation. It was proved by Blümlinger [11] that (1.18) holds for the discrepancy with respect to spherical “slices”. For $x, y \in S^d$ denote

$$
S_{xy} = \{p \in S^d : p \cdot x > 0, p \cdot y < 0\}.
$$

In other words, the slice $S_{xy}$ is a half of the wedge $W_{xy}$. It should be noted that the discrepancy with respect to slices is in fact a better measure of equidistribution on the sphere than the wedge discrepancy (the wedge discrepancy doesn’t change if we move all points to the hemisphere $\{x : p \geq 0\}$ by changing some points $x$ to $-x$). Using jittering sampling in a manner almost identical to Beck’s, Blümlinger showed that there exists $Z \subset S^d$, $\#Z = N$, such that

$$
D_{\text{slice}}(Z) = \sup_{x,y \in S^d} |D(Z,S_{xy})| \lesssim N^{-\frac{1}{d} - \frac{1}{2d}} \sqrt{\log N}.
$$
Like Beck’s estimate, this bound did not say anything about the dependence of constants on the dimension. Without any regard for constants, the main estimate (1.17) of Theorem 1.16 follows immediately since

\[ D(Z, W_{xy}) = D(Z, S_{xy}) + D(Z, S_{x,-y}), \]

and hence \( D_{\text{wedge}}(Z) \leq 2D_{\text{slice}}(Z) \).

An effective value of the constant \( C_d \) in Theorem 1.16, which is important for uniform tesselation and one-bit compressed sensing problems, requires much more delicate considerations and constructions, some of which only became available recently, see §2.1.

Blümlinger also showed that the lower bound (1.12) holds for the slice discrepancy. The proof uses spherical harmonics and is quite involved (Matoušek [27] writes that ‘it would be interesting to find a simple proof’). In fact, it is proved that the \( L^2 \)-discrepancy for slices is bounded below by the \( L^2 \)-discrepancy for spherical caps, from which the result follows by Beck’s estimate (1.13):

\[
(1.19) \quad D_{\text{slice}}(Z) \gtrsim D_{\text{slice}, L^2}(Z) \gtrsim D_{\text{cap}, L^2}(Z) \gtrsim N^{-\frac{1}{d} - \frac{1}{d+1}}.
\]

The lower bound for spherical wedges can be deduced by the following symmetrization argument.

**Proof of (1.18).** For a point set \( Z = \{z_1, \ldots, z_N\} \subset S^d \), consider its symmetrization, i.e. a \( 2N \)-point set \( Z^* = Z \cup (-Z) \). It is easy to see that

\[
D(Z, W_{xy}) = D(Z, S_{xy}) + D(Z, S_{x,-y}) = D(Z, S_{xy}) + D(-Z, S_{xy}) = 2D(Z^*, S_{xy}).
\]

Therefore,

\[
D_{\text{wedge}}(Z) \geq 2D_{\text{slice}}(Z^*) \gtrsim (2N)^{-\frac{1}{d} - \frac{1}{d+1}},
\]

which proves (1.18).

Inverting the bounds of Theorems 1.16, one immediately obtains the first announced result (1.7), asymptotic bounds on the minimal dimension of a sign-linear \( \delta \)-RIP from \( S^d \) to the Hamming cube.

**Theorem 1.20.** There exists an absolute constant \( C > 0 \) (independent of the dimension) such that in every dimension \( d \geq 2 \) and for every \( \delta > 0 \), the the integer \( N(d, \delta) \) of Definition 1.2 satisfies

\[
b_d \delta^{-2 + \frac{2}{d+1}} \leq N(d, \delta) \leq \max \left\{ 100d, \ C d^\alpha \delta^{-2 + \frac{2}{d+1}} \left( 1 + \log d + \log \frac{1}{\delta} \right)^{\frac{d}{d+1}} \right\},
\]

where \( \alpha = \frac{5}{2} - \frac{2}{d+1} \), and \( b_d > 0 \).

The absolute constant \( C \) above can be taken to be e.g. \( C = 4000 \). Some details are given in the end of §2.3.

In a different vein, we also obtain a variant of the Stolarsky Invariance Principle (1.15).
Theorem 1.21 (Stolarsky principle for wedges). For any finite set \( Z = \{z_1, \ldots, z_N\} \subset S^d \), the following relation holds

\[
\|\Delta_Z(x, y)\|_2^2 = \frac{1}{N^2} \sum_{i,j=1}^N \left( \frac{1}{2} - d(z_i, z_j) \right)^2 - \int_{S^d} \int_{S^d} \left( \frac{1}{2} - d(x, y) \right)^2 d\sigma(x) d\sigma(y).
\]

Minimizing the \( L^2 \) average of the wedge discrepancy associated to the tessellation of the sphere is thus equivalent to minimizing the discrete potential energy of \( Z \) induced by the potential \( P(x, y) = \left( \frac{1}{2} - d(x, y) \right)^2 \). Intuitively, one would like to make the elements of \( Z \) ‘as orthogonal as possible’ on the average.

First of all, this suggests natural candidates for tessellations that are good or optimal on the average, e.g., spherical codes (sets \( X \subset S^d \) such that all \( x, y \in X \) satisfy \( x \cdot y < \mu \) for some parameter \( \mu \leq 1 \)), see [36, Chapter 5] and references therein, or equiangular lines (sets \( X \subset S^d \) such that all \( x, y \in X \) satisfy \( |x \cdot y| = \mu \) for some fixed \( \mu \in [0, 1) \)), see [17].

This also brings up connections to frame theory. Benedetto and Fickus [8] proved that a set \( Z = \{z_1, \ldots, z_N\} \subset S^d \) forms a normalized tight frame (i.e. there exists a constant \( A > 0 \) such that for every \( x \in \mathbb{R}^{d+1} \) an analog of Parseval’s identity holds: \( A\|x\|^2 = \sum_{i=1}^N |x \cdot z_i|^2 \)) if and only if \( Z \) is a minimizer of a discrete energy known as the total frame potential:

\[
TP(Z) = \sum_{i,j=1}^N |z_i \cdot z_j|^2,
\]

which looks unmistakably similar to the discrete energy on the right-hand side of (1.22).

It is not known yet whether the minimizers of (1.22) admit a similar geometric or functional-analytic characterization, or if some of the known distributions yield reasonable values for this energy. These are interesting questions to be addressed in future research. We prove Theorem 1.21 in §3.

2. Jittered sampling

Jittered (or stratified) sampling in discrepancy theory and statistics can be viewed as a semi-random construction, somewhat intermediate between the purely random Monte Carlo algorithms and the purely deterministic low discrepancy point sets. It is easy to describe the main idea in just a few words: initially, the ambient manifold (cube, torus, sphere etc) is subdivided into \( N \) regions of equal volume and (almost) equal diameter, then a point is chosen uniformly at random in each of these pieces, independently of others.

Intuitively, this construction guarantees that the arising point set is fairly well distributed (there are no clusters or large gaps). Amazingly, it turns out that in many situations this distribution yields nearly optimal discrepancy (while purely random constructions are far from optimal, and deterministic sets are hard to construct). As mentioned before, the construction in Theorem 1.2 is precisely the jittered sampling. It differs from the corresponding lower bound only by a factor of \( \sqrt{\log N} \), which is a result of the application of large deviation inequalities. If one replaces the \( L^\infty \) norm of the discrepancy by \( L^2 \), jittered sampling actually easily gives the sharp upper bound (without \( \sqrt{\log N} \)). Similar phenomena persist in other situations (discrepancy with respect to balls or rotated rectangles in the unit cube, slices on the sphere etc).
Jittered sampling is very well described in classical references on discrepancy theory, such as [7, 18, 27]. However, since the spherical case possesses certain subtleties, and, in addition, we want to trace the dependence of the constants on the dimension, we shall describe the construction in full detail. Besides, this procedure will also yield a quantitative bounds on the constant in the classical spherical caps discrepancy bound, Lemma 1.2.

In order to make the construction precise we need to introduce two notions: \textit{equal-area partitions with bounded diameters} and \textit{approximating families}.

\section*{2.1. Regular partitions of the sphere.} Let \( S_i \subset \mathbb{S}^d, \ i = 1, 2, \ldots, N \). We say that \( \{ S_i \}_{i=1}^N \) is a partition of the sphere if \( \mathbb{S}^d \) is a disjoint (up to measure zero) union of these sets, i.e. \( \mathbb{S}^d = \bigcup_{i=1}^N S_i \) and \( \sigma(S_i \cap S_j) = 0 \) for \( i \neq j \).

\begin{definition}
Let \( S = \{ S_i \}_{i=1}^N \) be a partition of \( \mathbb{S}^d \). We call it an equal-area partition if \( \sigma(S_i) = \frac{1}{N} \) for each \( i = 1, \ldots, N \).
\end{definition}

\begin{definition}
Let \( S = \{ S_i \}_{i=1}^N \) be an equal-area partition of \( \mathbb{S}^d \), and fix \( K_d > 0 \). We say that it is a regular partition (or, an equal area partition with bounded diameters) if for every \( i = 1, \ldots, N \),
\[
\text{diam}(S_i) \leq K_d N^{-\frac{1}{d}}.
\]
\end{definition}

In the case of the unit cube \([0,1)^d\), regular partitions are extremely easy to construct. Indeed, for \( N = M^d \), one can simply take disjoint squares of side length \( M^{-1} = N^{-1/d} \). The situation is much more complicated for the sphere \( \mathbb{S}^d \). Nevertheless, we have this explicit upper bound on the constant \( K_d \) above.

\begin{theorem}[Leopardi [26]]
For all \( N \in \mathbb{N} \) there exist regular partitions \( \{ S_i \}_{i=1}^N \) of \( \mathbb{S}^d \) with the constant \( K_d \) given by
\[
K_d = 8 \left( \frac{\Omega d}{\omega} \right)^{\frac{2}{d}},
\]
where as before \( \Omega \) is the \( d \)-dimensional Lebesgue surface measure of \( \mathbb{S}^d \), and \( \omega \) is the \( (d-1) \)-dimensional measure of \( \mathbb{S}^{d-1} \).
\end{theorem}

Admittedly, Leopardi states the result without the specific value of the constant \( K_d \). However, it can be easily extracted from the proof, see page 9 in [26].

The history of this issue (as described in [25]) is interesting: Stolarsky [35] asserts the existence of regular partitions of \( \mathbb{S}^d \) for all \( d \geq 2 \), but offers no construction or proof of this fact. Later, Beck and Chen [7] quote Stolarsky, and Bourgain and Lindenstrauss [13] quote Beck and Chen. A complete construction of a regular partition of the sphere in arbitrary dimension was given by Feige and Schechtman [20], and Leopardi [26] found an effective value of the constant in their construction, which we quote above.

\section*{2.2. Approximating families.} We approximate infinite family of sets (e.g., all spherical caps, or wedges) by finite families. This will facilitate the use of a union bound estimate in the next section.
Definition 2.5. Let $S$ and $Q$ be two collections of subsets of $\mathbb{S}^d$. We say that $Q$ is an $\varepsilon$-approximating family (also known as $\varepsilon$-bracketing) for $S$ if for each $S \in S$ there exist sets $A, B \in Q$ such that

$$A \subset S \subset B \quad \text{and} \quad \sigma(B \setminus A) < \varepsilon.$$ 

It is easy to see that for any $N$-point set $Z$ in $\mathbb{S}^d$, if $S, A, B$ are as in the definition above then the discrepancies of $Z$ with respect to these sets satisfy

$$|D(Z, S)| \leq \max\{|D(Z, A)|, |D(Z, B)|\} + \varepsilon. \quad (2.6)$$

Hence $D_S(Z) \leq D_Q(Z) + \varepsilon$. Thus for $\varepsilon \leq N^{-1}$, the discrepancy with respect to the original family is of the same order is the discrepancy with respect to the $\varepsilon$-approximating family.

Constructions of finite approximating families are obvious in some cases, e.g. axis-parallel boxes in the unit cube or spherical caps: just take the same sets with rational parameters with small denominators.

For the spherical wedges, which is our case of interest, we have the following Lemma.

Lemma 2.7. For any $0 < \varepsilon < 1$, and integers $d \geq 1$, there is an approximating family $Q$ for the the collection of spherical wedges $\{W_{x,y} : x, y \in \mathbb{S}^d\}$ with

$$\mathcal{N}(\varepsilon) \leq (Cd)^{d+1} \varepsilon^{-(d+1)}, \quad (2.8)$$

where $0 < C \leq 82$ is an absolute constant.

Proof. We construct two separate families: one for interior and one for exterior approximation of the spherical wedges. Let $\mathcal{N}(\varepsilon)$ be the covering number of $\mathbb{S}^d$ with respect to the Euclidean metric, in other words, the cardinality of the smallest set $\mathcal{H}_\varepsilon$ such that for each $x \in \mathbb{S}^d$ there exists $z \in \mathcal{H}_\varepsilon$ with $\|x - z\| \leq \varepsilon$.

A simple volume argument (see e.g. [38]) shows that

$$\mathcal{N}(\varepsilon) \leq \left(1 + \frac{2}{\varepsilon}\right)^{d+1} \leq \left(\frac{4}{\varepsilon}\right)^{d+1}.$$ 

(More precise estimates can be obtained, in particular by using $d$-dimensional, rather than $(d + 1)$-dimensional volume arguments, but this will suffice for our purposes).

We shall construct an $\varepsilon$-approximating family as follows. Start with an $\gamma$-net $\mathcal{H}_\gamma$ of size $\mathcal{N}(\gamma)$, where $\gamma > 0$ is to be specified. For $x, y \in \mathbb{S}^d$, define the exterior enlargement and interior reduction of $W_{xy}$ as

$$W_{xy}^\text{ext}(\gamma) = \{p \in \mathbb{S}^d : p \cdot x \geq -\gamma, p \cdot y \leq \gamma\} \cup \{p \in \mathbb{S}^d : p \cdot x \leq \gamma, p \cdot y \geq -\gamma\} \supseteq W_{xy}$$

$$W_{xy}^\text{int}(\gamma) = \{p \in \mathbb{S}^d : p \cdot x \geq \gamma, p \cdot y \leq -\gamma\} \cup \{p \in \mathbb{S}^d : p \cdot x \leq -\gamma, p \cdot y \geq \gamma\} \subseteq W_{xy}.$$ 

We claim that the collection

$$Q = \{W_{xy}^\text{int}(\gamma) : x, y \in \mathcal{H}_\gamma\} \cup \{W_{xy}^\text{ext}(\gamma) : x, y \in \mathcal{H}_\gamma\}$$

forms an approximating family for the set of all wedges $\{W_{xy} : x, y \in \mathbb{S}^d\}$. Indeed, let $x', y' \in \mathbb{S}^d$. Choose $x, y \in \mathcal{H}_\gamma$ so that $\|x - x'\| < \gamma, \|y - y'\| \leq \gamma$. Then it is easy to see that $W_{xy}^\text{int} \subseteq W_{x'y'} \subseteq W_{xy}^\text{ext}$. Indeed, e.g. if $p \cdot x \geq 0$, then $p \cdot x = p \cdot x' - p \cdot (x' - x) \geq -\gamma$, the rest is similar.
Moreover, it is easy to see that the normalized measure of a ‘tropical belt’ around the equator satisfies
\[
\sigma \left( \{ p \in \mathbb{S}^d : |p \cdot x| \leq \gamma \} \right) \leq \frac{2\gamma \omega}{\Omega}.
\]
Therefore we can estimate
\[
\sigma \left( W_{xy}^{\text{int}}(\gamma) \setminus W_{xy}^{\text{int}}(\gamma) \right) \leq \frac{4\omega \gamma}{\Omega},
\]
hence we have an \( \varepsilon \)-approximating family with \( \varepsilon = \frac{4\omega \gamma}{\Omega} \), i.e. \( \gamma = \frac{\Omega \varepsilon}{4\omega} \).

The cardinality of this family satisfies the bound
\[
\#Q = 2(\mathcal{N}(\gamma))^2 \leq 2 \left( \frac{4}{\gamma} \right)^{2(d+1)} = 2^{8d+9} \left( \frac{\omega}{\Omega} \right)^{2(d+1)} \varepsilon^{-2(d+1)} \leq (Cd)^{d+1} \varepsilon^{-2(d+1)},
\]
where \( C > 0 \) is an absolute constant which can be taken to be, e.g. \( C = 82 \), and we have used the standard fact (1.8).

\[ \square \]

2.3. The spherical wedge discrepancy of jittered sampling. Proof of Theorem 1.16. The algorithm, which we describe for the case of spherical wedges, is generic and applies to many other situations. We shall need the following version of the classical Chernoff–Hoeffding large deviation bound (see e.g. [18, 27])

**Lemma 2.9.** Let \( p_i \in [0, 1], i = 1, 2, \ldots, m \). Consider centered independent random variables \( X_i, i = 1, \ldots, m \) such that \( \mathbb{P}(X_i = -p_i) = 1 - p_i \) and \( \mathbb{P}(X_i = 1 - p_i) = p_i \). Let \( X = \sum_{i=1}^m X_i \). Then for any \( \lambda > 0 \)

\[
(2.10) \quad \mathbb{P}(|X| > \lambda) < 2 \exp \left( -\frac{2\lambda^2}{m} \right).
\]

We start with a regular partition \( \{ S_i \}_{i=1}^N \) of the sphere as described in §2.1, i.e. \( \mathbb{S}^d = \bigcup_{i=1}^N S_i \), \( \sigma(S_i \cap S_j) = 0 \) for \( i \neq j \), \( \sigma(S_i) = 1/N \), and \( \text{diam}(S_i) \leq K_d N^{-1/d} \) for all \( i = 1, \ldots, d \).

We now construct the set \( Z = \{ z_1, \ldots, z_N \} \) by choosing independent random points \( z_i \in S_i \) according to the uniform distribution on \( S_i \), i.e. \( N \cdot \sigma|_{S_i} \).

Let \( Q \) be a \( 1/N \)-approximating family for the family \( \mathcal{R} \) of interest, in our case the family of spherical wedges \( \{ W_{xy} : x, y \in \mathbb{S}^d \} \). The size of this family, as discussed in §2.2, satisfies \#\( Q \leq A_d N^{\alpha_d} \). According to (2.8) we may take \( A_d = (Cd)^{d+1} \) and \( \alpha_d = 2(d+1) \).

Consider a single set \( Q \in Q \). It is easy to see that for those \( i = 1, \ldots, N \), for which \( S_i \cap \partial Q = \emptyset \) (\( S_i \) lies completely inside or completely outside of \( Q \)), the input of \( z_i \) and \( S_i \) to the discrepancy of \( Z \) with respect to \( Q \) is zero. In other words,

\[
D(Z, Q) = \frac{1}{N} \sum_{i:S_i \cap \partial Q \neq \emptyset} 1_Q(z_i) - \sigma(S_i \cap Q) = \frac{1}{N} \sum_{i=1}^m X_i,
\]

where \( X_i \) are exactly as in Lemma 2.9 with \( p_i = N \cdot \sigma(S_i \cap Q) \), and \( m \leq M \), where \( M \) is the maximal number of sets \( S_i \) that may intersect the boundary of any element of \( Q \).

It is now straightforward to estimate \( M \) for spherical wedges. Let \( \varepsilon = K_d N^{-1/d} \). Since every \( S_i \) has diameter at most \( \varepsilon \), all the sets \( S_i \) which intersect \( \partial Q \) are contained in the
set \( \partial Q + \varepsilon B \), where \( B \) is the unit ball. Recall that \( \sigma^*_d \) denotes the unnormalized Lebesgue measure on the sphere, and that we defined \( \Omega = \sigma^*_d(S^d) \) and \( \omega = \sigma^*_{d-1}(S^{d-1}) \). We then have

\[
M \cdot \frac{\Omega}{N} \leq \sigma^*_d(\partial Q + \varepsilon B) \leq \sigma^*_{d-1}(\partial Q) \cdot 2\varepsilon \leq 8K_d N^{-1/d} \omega.
\]

Hence, invoking the diameter bounds for the regular partition (2.4), we find that

\[
M \leq \frac{8K_d \omega}{\Omega} \cdot N^{1-\frac{d}{2}} \leq 64d^{\frac{1}{2}} \left( \frac{\omega}{\Omega} \right)^{\frac{1}{2}-\frac{d}{2}} N^{1-\frac{d}{2}}.
\]

Choosing the parameter \( \lambda = (\alpha_d \cdot M)^{\frac{1}{2}} \sqrt{\log N} \), then invoking the representation (2.11) and the large deviation estimate (2.10), we find that for any given \( Q \in \mathcal{Q} \)

\[
\mathbb{P}(\lvert D(Z, Q) \rvert > \lambda/N) = \mathbb{P}(\lvert X \rvert > \lambda) \leq 2N^{-\alpha_d}.
\]

Since \( \#Q_d \leq A_d N^{\alpha_d} \), the union bound yields

\[
\mathbb{P}(\lvert D(Z, Q) \rvert > \lambda/N \text{ for at least one } Q \in \mathcal{Q}) \leq 2A_d N^{-\alpha_d} < 1,
\]

whenever \( N > (2A_d)^{1/\alpha_d} \). Therefore, for such \( N \), there exists \( Z \) such that

\[
\sup_{Q \in \mathcal{Q}} \lvert D(Z, Q) \rvert \leq N^{-1}(\alpha_d M)^{\frac{1}{2}} \sqrt{\log N} \leq 8 \sqrt{\alpha_d} d^{\frac{1}{2}} \left( \frac{\omega}{\Omega} \right)^{\frac{1}{2}-\frac{d}{2}} N^{-\frac{d}{2}-\frac{d}{2}} \sqrt{\log N},
\]

i.e. the discrepancy estimate of the form (1.17) holds for each member of the approximating family \( Q \in \mathcal{Q} \) with constant \( 8 \sqrt{\alpha_d} d^{\frac{1}{2}} \left( \frac{\omega}{\Omega} \right)^{\frac{1}{2}-\frac{d}{2}} \) for \( N > (2A_d)^{1/\alpha_d} \). Since this constant is greater then one, the right-hand side is greater than \( \frac{1}{N} \) for all \( N \). Thus according to (2.6), the discrepancy estimate (1.17) holds for all sets \( W_{xy}, x, y \in S^d \) with twice the constant.

Recalling that \( \alpha_d = 2(d+1) \) and \( A_d = (Cd)^{d+1} \) and the fact that \( \frac{\omega}{\Omega} \leq \sqrt{\frac{d}{2\pi}} \), we find that the constant is at most \( C_d = 20d^{\frac{1}{2}+\frac{d}{2}} \) whenever \( N \geq 100d \). This finishes the proof of Theorem 1.16. \( \square \)

**Proof of Theorem 1.20.** It is a straightforward, but tedious task to check that if \( N \geq 100d \) and

\[
N > 400d^\gamma \delta^{-\frac{2d}{d+1}} \left( (d+1) \log(400d^\gamma) + 2d \log \frac{1}{\delta} \right)^{\frac{d}{d+1}},
\]

where \( \gamma = \frac{3}{2} - \frac{1}{d+1} \), one has \( 20d^{\frac{d}{2}+\frac{d}{d+1}} N^{-\frac{1}{2}} \frac{1}{\delta} \sqrt{\log N} < \delta \). Therefore, with positive probability jittered sampling with \( N \) points yields a \( \delta \)-uniform tessellation. It is easy to see that the right-hand side in the equation above is bounded by \( 4000d^\alpha \delta^{-2+\frac{2d}{d+1}} \left( 1 + \log d + \log \frac{1}{\delta} \right)^{\frac{d}{d+1}} \), where \( \alpha = \frac{d}{2} - \frac{2}{d+1} \), which proves Theorem 1.20.
3. STOLARSKY PRINCIPLE FOR THE WEDGE DISCREPANCY.

We now turn to the proof of Stolarsky principle for tessellations, Theorem 1.21. Recall that the $L^2$ norm of the function $\Delta_Z(x,y)$ for a set $Z \subset \mathbb{S}^d$ is

\begin{equation}
\|\Delta_Z(x,y)\|_2^2 = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \left( \frac{1}{N} \sum_{k=1}^{N} 1_{W_{xy}}(z_k) - \sigma(W_{xy}) \right)^2 d\sigma(x) d\sigma(y).
\end{equation}

The proof is quite elementary in nature and conforms to a standard algorithm of many similar problems: we square out the expression above, and the cross terms yield the discrete potential energy of the interactions of points of $Z$. The idea is generally quite fruitful. Torquato \cite{37} applies this approach to many questions of discrete geometric optimization, such as packings, coverings etc (both theoretically and numerically) to recast them as energy-minimization problems.

**Proof of Theorem 1.21.** We recall that $\sigma(W_{xy}) = d(x,y)$ and notice that we can write (up to sets of measure zero)

\[ 1_{W_{xy}}(z_k) = 1_{\{\text{sgn}(x \cdot z_k) \neq \text{sgn}(y \cdot z_k)\}}(z_k) = \frac{1}{2} \left( 1 - \text{sgn}(x \cdot z_k) \cdot \text{sgn}(y \cdot z_k) \right), \]

therefore, using (3.1), we have

\begin{align}
\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \Delta_Z(x,y)^2 d\sigma(x) d\sigma(y) &= \\
&= \frac{1}{4N^2} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \sum_{i,j=1}^{N} \left( 1 - \text{sgn}(x \cdot z_i) \text{sgn}(y \cdot z_i) \right) \left( 1 - \text{sgn}(x \cdot z_j) \text{sgn}(y \cdot z_j) \right) d\sigma(x) d\sigma(y) \\
&\quad - \frac{2}{N} \sum_{k=1}^{N} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} 1_{W_{xy}}(z_k) \cdot d(x,y) \ d\sigma(x) d\sigma(y) \\
&\quad + \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x,y)^2 d\sigma(x) d\sigma(y).
\end{align}

The most interesting term is the first one (3.2). Using the obvious fact that the integral

\[ \int_{\mathbb{S}^d} \text{sgn}(p \cdot x) d\sigma(x) = 0 \]

for any $p \in \mathbb{S}^d$, we reduce this term to

\[ \frac{1}{4} + \frac{1}{4N^2} \sum_{i,j=1}^{N} \left( \int_{\mathbb{S}^d} \text{sgn}(x \cdot z_i) \cdot \text{sgn}(x \cdot z_j) d\sigma(x) \right)^2 = \frac{1}{4} + \frac{1}{N^2} \sum_{i,j=1}^{N} \left( \frac{1}{2} - d(z_i, z_j) \right)^2, \]

where the last line is obtained as follows. Rewrite the integrand as $\text{sgn}(x \cdot z_i)\text{sgn}(x \cdot z_j) = 1 - 21_{W_{z_izj}}(x)$, therefore

\[ \int_{\mathbb{S}^d} \text{sgn}(x \cdot z_i)\text{sgn}(x \cdot z_j) d\sigma(x) = 1 - 2 \int_{\mathbb{S}^d} 1_{W_{z_izj}}(x) d\sigma(x) = 1 - 2d(z_i, z_j). \]
We shall see that in the second term (3.3) one can easily replace the discrete average over $z_k \in Z$ over the continuous average over $p \in \mathbb{S}^d$, which is simpler to handle. Indeed, notice that by rotational invariance the integrand in (3.3) does not depend on the particular choice of $z_k \in \mathbb{S}^d$. Therefore, for an arbitrary pole $p \in \mathbb{S}^d$ we can write

$$\frac{2}{N} \sum_{k=1}^{N} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} 1_{W_{xy}}(z_k) \cdot d(x,y) \, d\sigma(x) \, d\sigma(y) = 2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} 1_{W_{xy}}(p) \cdot d(x,y) \, d\sigma(x) \, d\sigma(y).$$

Invoking rotational symmetry again, we see that the integral above may be replaced by the average over $p \in \mathbb{S}^d$:

$$2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} 1_{W_{xy}}(p) \cdot d(x,y) \, d\sigma(x) \, d\sigma(y) = 2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} 1_{W_{xy}}(p) \cdot d(x,y) \, d\sigma(x) \, d\sigma(y) \, d\sigma(p)$$

$$= 2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \left[ \int_{\mathbb{S}^d} 1_{W_{xy}}(p) \, d\sigma(p) \right] \cdot d(x,y) \, d\sigma(x) \, d\sigma(y) = 2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x,y)^2 \, d\sigma(x) \, d\sigma(y),$$

thus the term has the same form as the last one (3.4). Putting this together we find that

$$\|\Delta_Z(x,y)\|^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \left( \frac{1}{2} - d(z_i, z_j) \right)^2 + \frac{1}{4} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x,y)^2 \, d\sigma(x) \, d\sigma(y).$$

Observing that $\int_{\mathbb{S}^d} d(x,y) \, d\sigma(x) = \frac{1}{2}$ and hence

$$\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \left( \frac{1}{2} - d(x,y) \right)^2 \, d\sigma(x) \, d\sigma(y) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x,y)^2 \, d\sigma(x) \, d\sigma(y) - \frac{1}{4},$$

we arrive to the desired conclusion (1.22):

$$\|\Delta_Z(x,y)\|^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \left( \frac{1}{2} - d(z_i, z_j) \right)^2 - \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \left( \frac{1}{2} - d(x,y) \right)^2 \, d\sigma(x) \, d\sigma(y). \quad \square$$

3.1. $L^2$ discrepancy for random tessellations. Stolarsky principle provides a very simple way to compute exactly the expected value of the square of the $L^2$ discrepancy. Assume that the set $Z = \{z_1, \ldots, z_N\} \subset \mathbb{S}^d$ is random and compute the expectation of $\|\Delta_Z(x,y)\|^2$. Obviously, for a typical point set $Z$ and a typical wedge $W_{xy}$ the discrepancy is of the order $1/\sqrt{N}$, therefore this expected value naturally behaves as $O(1/N)$. We compute its value precisely.

**Lemma 3.6.** Let $Z = \{z_1, ..., z_N\} \subset \mathbb{S}^d$ consist of $N$ i.i.d. uniformly distributed points on the sphere. Then

$$\mathbb{E}_Z \|\Delta_Z(x,y)\|^2 = \frac{1}{N} \cdot \left( \frac{1}{2} - V_d \right).$$
Table 1. The values of $V_d$

| $d$ | $d = 2$ | $d = 3$ | $d = 4$ | $d = 5$ | $d = 6$ |
|-----|--------|--------|--------|--------|--------|
| $V_d$ | $\frac{1}{2} - \frac{2}{\pi^2}$ | $\frac{1}{3} - \frac{1}{2\pi^2}$ | $\frac{1}{2} - \frac{20}{9\pi^2}$ | $\frac{1}{3} - \frac{5}{8\pi^2}$ | $\frac{1}{2} - \frac{518}{225\pi^2}$ |

Proof. We shall need the quantity that already arose in the computations above, namely the second moment of the geodesic distance, i.e. the expected value of the square of the geodesic distance between two random points on the sphere

$$V_d = \mathbb{E}_{x,y} d(x,y)^2 = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x,y)^2 d\sigma(x) d\sigma(y).$$

It is obvious that $\mathbb{E}_{x,y} d(x,y) = \frac{1}{2}$ and hence $\mathbb{E}_{x,y} (\frac{1}{2} - d(x,y))^2 = V_d - \frac{1}{4}$. We use the final form of the Stolarsky principle (3.5) to find the value of $\mathbb{E}_Z \| \Delta Z(x,y) \|^2_2$. We separate the off-diagonal and diagonal terms in the discrete part of (3.5) to obtain

$$\mathbb{E}_Z \| \Delta Z(x,y) \|^2_2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j} \left( \frac{1}{2} - d(z_i,z_j) \right)^2 - \left( V_d - \frac{1}{4} \right)$$

$$= \frac{1}{N^2} \cdot (N^2 - N) \cdot \left( V_d - \frac{1}{4} \right) + \frac{1}{N^2} \cdot N \cdot \frac{1}{4} - \left( V_d - \frac{1}{4} \right)$$

$$= \frac{1}{N} \cdot \left( \frac{1}{2} - V_d \right),$$

which finishes the proof. \qed

Finally, we take a closer look at the expected value of the square of the geodesic distance $V_d$. We remark that it can be written as

$$V_d = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x,y)^2 d\sigma(x) d\sigma(y) = \frac{1}{\pi^2} \cdot \frac{\omega}{\Omega} \int_0^\pi \phi^2 (\sin \phi)^{d-1} d\phi,$$

where $\omega$ is the surface area of $\mathbb{S}^{d-1}$. For any given value of $d \geq 3$, the integrals above may be evaluated directly, although no simple closed form expression seems to be available. In Table 1 we list the values of $V_d$ in low dimensions.

In the case of spherical cap discrepancy, a computation similar to Lemma 3.6 is even simpler. It yields:

**Lemma 3.7.** Let $Z = \{z_1, ..., z_N\} \subset \mathbb{S}^d$ consist of $N$ i.i.d. uniformly distributed points on the sphere. Then

$$\mathbb{E}_Z D^2_{\text{cap},L^2} = \frac{U_d}{N},$$

where $U_d = \mathbb{E}_{x,y \in \mathbb{S}^d} \| x - y \|$. 

16
Indeed, using the original Stolarsky principle, Theorem 1.14, one gets

$$\mathbb{E}_Z D_{\text{cap}, L^2}^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}_z, z_j \|z_i - z_j\|^2 - U_d = \frac{N^2 - N}{N^2} U_d - U_d = \frac{U_d}{N}.$$  

3.2. $L^2$ wedge discrepancy for jittered sampling. Stolarsky principle (3.5) allows one to easily prove that jittered sampling yields optimal order of the $L^2$ wedge discrepancy.

**Lemma 3.8.** Let $Z = \{z_1, \ldots, z_N\} \subset \mathbb{S}^d$, $N \in \mathbb{N}$, be a point set constructed by jittered sampling corresponding to a regular partition of the sphere with constant $K_d$. Then

$$\mathbb{E}_Z \|\Delta_Z(x, y)\|^2_2 \leq K_d N^{-1 - \frac{1}{d}}.$$

A matching lower bound for arbitrary $N$-point sets is known for caps and slices and can be easily generalized to wedges, see (1.19) and the discussion immediately thereafter.

**Proof.** We notice that for $i \neq j$ we have $\mathbb{E}\left(\frac{1}{2} - d(z_i, z_j)\right)^2 = N^2 \int \int \left(\frac{1}{2} - d(x, y)\right)^2 d\sigma(x) d\sigma(y)$, while for $i = j$ one simply gets $\frac{1}{4}$. Therefore,

$$\mathbb{E}_Z \|\Delta_Z(x, y)\|_2^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}_z, z_j \left(\frac{1}{2} - d(z_i, z_j)\right)^2 - \int \int \left(\frac{1}{2} - d(x, y)\right)^2 d\sigma(x) d\sigma(y)$$

$$= \sum_{i=1}^{N} \int \int_{S_i} \left(d(x, y) - d^2(x, y)\right) d\sigma(x) d\sigma(y) \leq \sum_{i=1}^{N} K_d N^{-\frac{1}{d}} \cdot \frac{1}{N^2} = N^{-1 - \frac{1}{d}},$$

since $d(x, y) - d^2(x, y) \leq d(x, y) \leq \|x - y\| \leq K_d N^{-1/d}$ for $x, y \in S_i$. \hfill \qed

**References**

[1] C. Aistleitner, J. S. Brauchart, and J. Dick, *Point sets on the sphere $S^2$ with small spherical cap discrepancy*, Discrete Comput. Geom. 48 (2012), no. 4, 990–1024.

[2] Keith Ball, *The Ribe programme*, Astérisque 352 (2013), Exp. No. 1047, viii, 147–159. Séminaire Bourbaki. Vol. 2011/2012. Exposés 1043–1058.

[3] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wootters, *A simple proof of the restricted isometry property for random matrices*, Constr. Approx. 28 (2008), no. 3, 253–263.

[4] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, *Exponential decay of reconstruction error from binary measurements of sparse signals* (2014), available at [1407.8246](https://arxiv.org/abs/1407.8246).

[5] József Beck, *Some upper bounds in the theory of irregularities of distribution*, Acta Arith. 43 (1984), no. 2, 115–130.

[6] ______, *Sums of distances between points on a sphere—an application of the theory of irregularities of distribution to discrete geometry*, Mathematika 31 (1984), no. 1, 33–41.

[7] József Beck and William W. L. Chen, *Irregularities of distribution*, Cambridge Tracts in Mathematics, vol. 89, Cambridge University Press, Cambridge, 1987.

[8] John J. Benedetto and Matthew Fickus, *Finite normalized tight frames*, Adv. Comput. Math. 18 (2003), no. 2-4, 357–385.

[9] Dmitriy Bilyk, *Discrepancy problems as particle interactions*, preprint.

[10] Dmitriy Bilyk and Michael T. Lacey, *Random Tessellations, Restricted Isometric Embeddings and One Bit Sensing* (2015).

[11] Martin Blämlinger, *Slice discrepancy and irregularities of distribution on spheres*, Mathematika 38 (1991), no. 1, 105–116.
[12] P. Boufounos and R. Baraniuk, *1-bit compressive sensing*, Proc. 42nd Annu. Conf. Inf. Sci. Syst., Princeton, NJ, (Mar. 2008).

[13] J. Bourgain and J. Lindenstrauss, *Distribution of points on spheres and approximation by zonotopes*, Israel J. Math. **64** (1988), no. 1, 25–31.

[14] Johann S. Brauchart and Josef Dick, *A simple proof of Stolarsky’s invariance principle*, Proc. Amer. Math. Soc. **141** (2013), no. 6, 2085–2096.

[15] J. S. Brauchart, J. Dick, E. B. Saff, I. H. Sloan, Y. G. Wang, and R. S. Womersley, *Covering of spheres by spherical caps and worst-case error for equal weight cubature in Sobolev spaces*, J. Math. Anal. Appl. **431** (2015), no. 2, 782–811.

[16] Emmanuel J. Candes and Terence Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory **52** (2006), no. 12, 5406–5425.

[17] P. G. Casazza, D. Redmond, and J. C. Tremain, *Real equiangular frames*, Proceedings of CISS, Princeton, N.J. (2008).

[18] Bernard Chazelle, *The discrepancy method*, Cambridge University Press, Cambridge, 2000. Randomness and complexity.

[19] Aryeh Dvoretzky, *Some results on convex bodies and Banach spaces*, Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960), Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961, pp. 123–160.

[20] Uriel Feige and Gideon Schechtman, *On the optimality of the random hyperplane rounding technique for MAX CUT*, Random Structures Algorithms **20** (2002), no. 3, 403–440. Probabilistic methods in combinatorial optimization.

[21] Simon Foucart and Holger Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, New York, 2013.

[22] K. Green Larsen and J. Nelson, *The Johnson-Lindenstrauss lemma is optimal for linear dimensionality reduction*, available at 1411.2404.

[23] L. Jacques and K. Degraux, *Quantized Iterative Hard Thresholding: Bridging 1-bit and High-Resolution Quantized Compressed Sensing*, 2013.

[24] Laurent Jacques, Jason N. Laska, Petros T. Boufounos, and Richard G. Baraniuk, *Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors*, IEEE Trans. Inform. Theory **59** (2013), no. 4, 2082–2102.

[25] Paul Leopardi, *A partition of the unit sphere into regions of equal area and small diameter*, Electron. Trans. Numer. Anal. **25** (2006), 309–327 (electronic).

[26] , *Diameter bounds for equal area partitions of the unit sphere*, Electron. Trans. Numer. Anal. **35** (2009), 1–16.

[27] Jiří Matoušek, *Geometric discrepancy*, Algorithms and Combinatorics, vol. 18, Springer-Verlag, Berlin, 1999. An illustrated guide.

[28] Assaf Naor, *An introduction to the Ribe program*, Jpn. J. Math. **7** (2012), no. 2, 167–233.

[29] F. Pausinger and S. Steinerberger, *On the discrepancy of jittered sampling*, Journal of Complexity (to appear), available at 1510.00251.

[30] Yaniv Plan and Roman Vershynin, *One-bit compressed sensing by linear programming*, Comm. Pure Appl. Math. **66** (2013), no. 8, 1275–1297.

[31] , *Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach*, IEEE Trans. Inform. Theory **59** (2013), no. 1, 482–494.

[32] , *Dimension reduction by random hyperplane tessellations*, Discrete Comput. Geom. **51** (2014), no. 2, 438–461.

[33] A. Reznikov and E. B. Saff, *The covering radius of randomly distributed points on a manifold*, available at 1504.03029.

[34] Gideon Schechtman, *Two observations regarding embedding subsets of Euclidean spaces in normed spaces*, Adv. Math. **200** (2006), no. 1, 125–135.

[35] Kenneth B. Stolarsky, *Sums of distances between points on a sphere. II*, Proc. Amer. Math. Soc. **41** (1973), 575–582.

[36] Vladimir Temlyakov, *Greedy approximation*, Cambridge Monographs on Applied and Computational Mathematics, vol. 20, Cambridge University Press, Cambridge, 2011.
[37] S. Torquato, Reformulation of the covering and quantizer problems as ground states of interacting particles, Phys. Rev. E 82 (2010), 056109.

[38] Roman Vershynin, Introduction to the non-asymptotic analysis of random matrices, Compressed sensing, Cambridge Univ. Press, Cambridge, 2012, pp. 210–268.

School of Mathematics, University of Minnesota, Minneapolis MN 55455, USA
E-mail address: dbilyk@math.umn.edu

School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA
E-mail address: lacey@math.gatech.edu
UNIFORM TESSELLATIONS OF THE SPHERE, DISCREPANCY, AND STOLARSKY PRINCIPLE.

DMITRIY BILYK AND MICHAEL T. LACEY

ABSTRACT. In this paper we apply the ideas and methods of discrepancy theory to the problems of tessellations of spheres by hyperplanes, which yields several interesting results. In particular, we prove that for the cardinality $N(\delta)$ of the smallest $\delta$-uniform tessellation of $\mathbb{S}^d$ satisfies the bounds $c_d \delta^{-2 + \frac{d}{2+d}} \leq N(\delta) \leq \delta^{-2 + \frac{d}{2+d}} \log \delta \left( \frac{1}{\delta} + \frac{d}{\delta} \right)^2$, which disproves a prior conjecture of Plan and Vershynin. We also obtain an analogue of the Stolarsky principle for this situation, which implies that minimizing the $L^2$ average of the tessellation error is equivalent to minimizing the discrete energy $\sum_{i,j} \left( \frac{1}{2} - d(z_i, z_j) \right)^2$, where $d$ is the normalized geodesic distance.

1. Introduction

Let $d \geq 2$ and let $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ denote the $d$-dimensional unit sphere. We consider the following question: how uniformly can one tessellate $\mathbb{S}^d$ (i.e. cut it into pieces) with $N$ hyperplanes? What is the minimal number of hyperplanes required to guarantee that for any two points $x, y \in \mathbb{S}^d$ the proportion of hyperplanes that separates these points approximates the distance between the points up to a given threshold?

These questions were first considered by Plan and Vershynin [MR3164174] and have connections to several different topics such as compressed sensing ("one-bit sensing"), geometric functional analysis (almost isometric embeddings), and information theory.

To set the stage, consider a finite set of vectors $Z = \{z_1, z_2, ..., z_N\}$ on the sphere $\mathbb{S}^d$, and consider the map $\varphi_Z$ from the sphere $\mathbb{S}^d$ into the Hamming cube $\{-1, +1\}^N$ of dimension $N$ given by

$$\varphi_Z(x) = \{ \text{sgn} (\langle z_j, x \rangle) : 1 \leq j \leq N \}.$$  

For any two points $x, y \in \mathbb{S}^d$ we define the Hamming distance $d_H(x, y)$ generated by $Z$ between these points (the letter $H$ stands for "Hamming") as

$$d_H(x, y) := \frac{1}{N} \cdot \# \{ z_k \in Z : \text{sgn}(x \cdot z_k) \neq \text{sgn}(y \cdot z_k) \},$$

in other words we compute the proportion of those hyperplanes $z_k^\perp$ that separate the points $x$ and $y$ among all hyperplanes given by $Z$.

Research supported in part by the NSF grant DMS 1101519.
Research supported in part by the NSF grant DMS 1265570.
We shall denote by $d(x, y)$ the geodesic distance between $x$ and $y$ on the sphere normalized so that the distance between the opposite poles is 1, i.e.

$$d(x, y) = \frac{\cos^{-1}(x \cdot y)}{\pi}.$$ 

It is easy to see that, if one chooses a hyperplane $z^\perp$ uniformly at random, then

$$P\{\text{sgn}(x \cdot z) \neq \text{sgn}(y \cdot z)\} = d(x, y).$$

In fact this relation is a simple instance of the Crofton formula from integral geometry [MR0333995, p. 36-40]. Hence for a large number of random (or carefully chosen deterministic) hyperplanes, the Hamming distance $d_H(x, y)$ should be close to the geodesic distance $d(x, y)$. Exactly how close can be measured by the quantity

$$\Delta_Z(x, y) := d_H(x, y) - d(x, y).$$

The following definition was introduced in [MR3164174].

**Definition 1.1.** Let $\delta > 0$. We say that the set $Z = \{z_1, \ldots, z_N\} \subset S^d$ induces a $\delta$-uniform tessellation of the sphere if

$$\Delta(Z) := \|\Delta_Z\|_{\infty} = \sup_{x, y \in S^d} |\Delta_Z(x, y)| \leq \delta,$$

i.e. $|\Delta_Z(x, y)| \leq \delta$ for all pairs of points on $S^d$.

We define $N(\delta)$ to be the minimal integer $N$ such that there exists an $N$-point set $Z \subset S^d$, which induces a $\delta$-uniform tessellation.

Even when they wrote their paper, the term 'RIP' would have more consistent with the literature.

Our main questions then become the following: can we determine the behavior of $N(\delta)$? Can we identify constructions that yield optimal or at least reasonable uniform tessellations?

Partial answers to these questions were given by Plan and Vershynin in [MR3164174]. To be fair, we need to mention that their Indeed, motivated by the considerations from compressive sensing, they considered arbitrary subsets of the sphere, i.e. in the definition above the supremum is taken over $x, y \in K \subset S^d$. They provided sufficient lower bounds on the size of random $Z$ in in terms of geometric characteristics of $K$. In numerous applications various specific choices of the set $K$ are dictated by the nature of the problem: e.g. sparse vectors for compressed sensing, finite or lower dimensional sets for almost isometric embeddings akin to the famous Johnson–Lindenstrauss lemma. However already the case of the whole sphere $K = S^d$ is non-trivial, and we concentrate on this situation in the present paper, leaving the more general case for future work.

Restricted to the current setup, the result of Plan and Vershynin [MR3164174, Theorem 1.2] reads as follows: there exists an absolute constant $C > 0$ such that in all dimensions $d \geq 2$, for any $\delta > 0$ there exists a $\delta$-uniform tessellation of the sphere $S^d$ by $N$ hyperplanes with

$$N \leq Cd\delta^{-6}.$$
In fact, they prove that a random selection of \( N \approx d\delta^{-6} \) hyperplanes achieves a \( \delta \)-uniform tessellation with high probability. They also conjecture that in all dimensions \( d \geq 2 \) the optimal estimate is

\[
(4) \quad cd \leq \sup_{\delta > 0} \delta^2 N(\delta) \leq Cd,
\]

for some absolute constants \( c, C > 0 \). In a companion paper [?BL], we will show that this conjecture is correct for uniformly at random selection of hyperplanes. (The lower bound follows from the Central Limit Theorem, while the upper bound requires argument.)

In the present paper, by bridging this circle of questions to the geometric discrepancy theory, we show that this conjecture is not true in general. Instead, there is a dimensional correction on the power of \( \delta \). We shall prove (see Theorem 1.7) that there exist constants \( c_d, C_d > 0 \) (dependent on the dimension) such that

\[
(5) \quad c_d \delta^{-2+\frac{2}{d+1}} \leq N(\delta) \leq C_d \delta^{-2+\frac{2}{d+1}} \left( \log \frac{1}{\delta} \right)^{\frac{d}{d+1}}.
\]

Probability theory and empirical processes suggest that the conjectured bound (4) may indeed be the right relation for random tessellations by hyperplanes, i.e., it is feasible that in the purely random regime \( N(\delta) \approx \delta^{-2} \). However, discrepancy theory dictates that in the general case there is a dimensional correction in the exponent of \( \delta \) which vanishes as \( d \to \infty \).

The upper bound in (5) is achieved by a classical construction of discrepancy theory, known as jittered sampling, which is a cross between purely random and deterministic constructions. Loosely speaking, first one divides the sphere \( S^d \) into \( N \) roughly equal pieces, and then chooses a random point in each of them, see §2 for details.

Since in applications the dimension \( d \) is often quite large, we considered a non-asymptotic version of the upper bound in (5) and computed an effective value of the constant \( C_d \), proving that it grows roughly as \( d^{5/2} \) (see Theorem 1.8 for a more precise statement):

\[
N(\delta) \leq \max \left\{ C d^{5/2} \delta^{-2+\frac{2}{d+1}} \left( 1 + \log d + \log \frac{1}{\delta} \right)^{\frac{d}{d+1}}, \quad 100d \right\},
\]

where \( C > 0 \) is an absolute constant.

In a somewhat different direction, we also show in Theorem 1.9 that the \( L^2 \) norm of \( \Delta(x, y) \) satisfies an analog of the Stolarsky principle (see Theorem 1.4), which implies that minimizing the \( L^2 \) average of \( \Delta(x, y) \) is equivalent to minimizing the discrete energy of the form \( \frac{1}{N^2} \sum_{i,j} \left( \frac{1}{2} - d(z_i, z_j) \right)^2 \) and suggests interesting connections to such objects as spherical codes, equiangular lines, and frames. See Theorem 1.9 and §3 for details.

In this paper, we shall denote by \( \sigma \) the surface measure on the sphere, normalized so that \( \sigma(S^d) = 1 \). Unnormalized (Hausdorff) measure on \( S^d \) will be denoted by \( \sigma^*_d \). We shall use the notation \( \omega = \sigma^*_{d-1}(S^{d-1}) \) and \( \Omega = \sigma^*_d(S^d) \). The notation \( A \lesssim B \) means that \( A \leq CB \) for some fixed constant \( C > 0 \). Occasionally, the implicit constant may depend on the dimension \( d \) (this will be made clear in the context), but it is always independent of \( N \) and \( \delta \).

1.1. Discrepancy. Geometric discrepancy theory deals with various ways to versions of the following problem. We shall resort just to the case when the ambient manifold is the sphere \( S^d \), although other setups are of course possible. Consider a collection \( S \) of measurable
Figure 1. The spherical wedge $W_{xy}$.

subsets of $\mathbb{S}^d$. Let $Z = \{z_1, \ldots, z_N\}$ be an $N$-point subset of $\mathbb{S}^d$. Equidistribution of $Z$ may be quantified by means of discrepancy. Consider the discrepancy

$$D(Z, S) = \frac{1}{N} \cdot \# \{Z \cap S\} - \sigma(S)$$

of the point set $Z$ with respect to $S \in \mathcal{S}$, i.e. error of approximation of surface measure of $S$ by an average of point masses. We define the extremal discrepancy of $Z$ with respect to the family $\mathcal{S}$

$$D_{\mathcal{S}}(Z) = \sup_{S \in \mathcal{S}} |D(Z, S)|.$$  

If the family $\mathcal{S}$ admits a natural measure then one may also replace the supremum above by an $L^2$ average (this is known as $L^2$-discrepancy) or other averages. The main questions of discrepancy theory are: How small can discrepancy be? What are good or optimal point distributions? What are the limitations? These questions have profound connections to approximate, probability, combinatorics, number theory, computer science, analysis, etc, see [MR903025, MR1779341, MR1697825]. In short, geometric discrepancy deals with approximations of continuous objects by discrete ones, or in other words, with differences between observed and expected values.

In this sense the quantity $\Delta_Z(x, y)$ defined in (2) clearly has a discrepancy flavor. In fact, (and this is perhaps the most important observation of the paper) the problem of uniform tessellations can actually be reformulated as a problem on geometric discrepancy with respect to “spherical wedges”.

Indeed, denote by

$$W_{xy} = \{z \in S^{d-1} : \text{sgn}(x \cdot z) \neq \text{sgn}(y \cdot z)\}$$

the set of normals of those hyperplanes that separate $x$ and $y$. The letter $W$ stands for “wedge”, since the set $W_{xy}$ does in fact look like a spherical wedge, i.e. the subset of the sphere lying between the hyperplanes $x \cdot z = 0$ and $y \cdot z = 0$, see Figure 1.

It follows from (1) that

$$\sigma(W_{xy}) = \mathbb{P}(z^\perp \text{ separates } x \text{ and } y) = d(x, y).$$
Therefore we can rewrite the quantity (2) as
\[ \Delta_Z(x, y) = \frac{\#(Z \cap W_{xy})}{N} - \sigma(W_{xy}) = \frac{1}{N} \sum_{k=1}^{N} 1_{W_{xy}}(z_k) - \sigma(W_{xy}) = D(Z, W_{xy}), \]
i.e. it is precisely the discrepancy of the \( N \)-point distribution \( Z \) with respect to the wedge \( W_{xy} \).

Under this correspondence the problem of uniform tessellation corresponds to the extremal \((L^\infty)\) discrepancy with respect to wedges. Indeed, according to definitions (3) and (6)
\[ \Delta(Z) = \|\Delta(Z, x, y)\|_\infty = \sup_{x, y \in S^{d-1}} \left| \frac{\#(Z \cap W_{xy})}{N} - \sigma(W_{xy}) \right| = D_{\text{wedge}}(Z). \]
The problem of estimating \( N(\delta) \) is thus simply inverse to obtaining discrepancy estimates in terms of \( N \), and this is precisely the approach we shall take.

1.2. A point of reference: spherical cap discrepancy. While we shall be mostly interested in discrepancy with respect to spherical wedges, we shall start by recalling a much better studied problem about discrepancies of spherical point sets, namely, discrepancy with respect to spherical caps. We shall use this situation as an important point of reference for our case.

For \( x \in \mathbb{S}^d \), and \( t \in [-1, 1] \), let \( C(x, t) \) be the spherical cap of height \( t \) centered at \( x \), i.e.
\[ C(x, t) = \{ p \in \mathbb{S}^d : p \cdot x \geq t \}. \]
Denote the set of all spherical caps by \( \mathcal{C} \). For and \( N \)-point set \( Z \subset \mathbb{S}^d \) let
\[ D_{\text{cap}}(Z) = \sup_{C(x, t) \in \mathcal{C}} \left| D(Z, C(x, t)) \right| = \sup_{C(x, t) \in \mathcal{C}} \left| \frac{\#(Z \cap C(x, t))}{N} - \sigma(C(x, t)) \right| \]
be the extremal discrepancy of \( Z \) with respect to spherical caps \( C(x, t) \). The following classical results due to J. Beck [MR736726] yield almost precise information about the growth of this quantity in terms of \( N \).

**Theorem 1.2.** There exists an \( N \)-point set \( Z \subset \mathbb{S}^d \) with spherical cap discrepancy
\[ D_{\text{cap}}(Z) \lesssim N^{-\frac{1}{2} - \frac{d}{2}} \sqrt{\log N}. \]

**Theorem 1.3.** For any \( N \)-point set \( Z \subset \mathbb{S}^d \) the spherical cap discrepancy satisfies
\[ D_{\text{cap}}(Z) \gtrsim N^{-\frac{1}{2} - \frac{d}{2}}. \]

Some remarks are in order. First we would like to note that the implicit constants in inequalities above naturally depend on the dimension. However, this dependence on \( d \) doesn’t seem to have been explored before.

The upper bound, Theorem 1.2, is proved using a construction known as “jittered sampling”, which produces a semi-random point set. We describe this construction in much detail in §2. Stating this result in [MR762175] Beck just states that “using probabilistic ideas it is not hard to show” and refers to his paper [MR736726], where this fact is proved for rotated rectangles, not spherical caps (jittered sampling is a sufficiently generic algorithm applicable to many geometric setups). In the book [MR903025] the algorithm is described in some more detail, but one of the key steps, namely regular equal-area partition of the
sphere, was only postulated. This construction was only recently rigorously formalized with and effective values of the underlying constants have been found [MR1900615, MR2582801]. See §2.1 for further discussion.

It is generally believed that standard low-discrepancy sets, while providing good bounds with respect to the number of points \(N\), yield very bad, often exponential dependence on the dimension. However, as we shall see it appears that for jittered sampling this behavior is quite reasonable (see also [PaSt] for a discussion of the similar effect). This is consistent with the fact that this construction is intermediate between purely random and deterministic sets.

Since we are interested both in asymptotic and non-asymptotic regimes, we shall explore this construction (in the case of spherical wedges) very scrupulously tracing the dependence of the constant on the dimension.

The proof of Lemma 1.3 in [MR762175] is Fourier-analytic in nature, and it actually holds for the \(L^2\) average in place of the supremum:

\[
D_{\text{cap},L^2}(Z) = \left( \int_{-1}^1 \int_{S^d} \left| \frac{\#(Z \cap C(x,t))}{N} - \sigma(C(x,t)) \right|^2 \, d\sigma(x) \, dt \right)^{1/2},
\]

in other words

\[
(8) \quad D_{\text{cap},L^2}(Z) \gtrsim N^{-\frac{1}{2} - \frac{1}{2d}}.
\]

In fact the estimate is sharp in \(N\) for this quantity: for the \(L^2\) discrepancy jittered sampling yields a bound akin to (7), but without \(\sqrt{\log N}\). Another famous result about the \(L^2\)-discrepancy of point sets with respect to spherical caps relates this notion to another famous problem about uniform distributions of points on the sphere: maximizing the sum of pairwise distances. This fact was proved by Stolarski in [MR0333995] and became known as Stolarsky principle:

**Theorem 1.4.** In all dimensions \(d \geq 2\), for any \(N\)-point set \(Z = \{z_1, \ldots, z_N\} \subset S^d\), the following holds

\[
\frac{1}{c_d} \left[ D_{\text{cap},L^2}(Z) \right]^2 = \int_{S^d} \int_{S^d} \|x - y\| \, d\sigma(x) \, d\sigma(y) - \frac{1}{N^2} \sum_{i,j=1}^N \|z_i - z_j\|,
\]

where \(c_d = \frac{1}{2} \int_{S^{d-1}} |p \cdot z| \, d\sigma(z)\).

This relation means that the square of the \(L^2\) discrepancy can be realized as the integration error of the cubature formula given by the points of \(Z\) for the potential \(\|x - y\|\) on \(S^d \times S^d\) (i.e. a Riesz s-potential with \(s = -1\)), or, in other words, the difference between the continuous potential energy given by this potential on the sphere and the discrete energy induced by the points of \(Z\). Minimizing this discrepancy is thus equivalent to maximizing the sum of distances. Simpler proofs of Stolarsky principle can be found in [MR3034434, B].

1.3. **Main results.** We shall demonstrate that analogues of Beck’s discrepancy estimates, Theorems 1.2 and 1.3, hold for spherical wedges \(W_{xy}\). Moreover, we shall explore the dependence of the upper estimates on the dimension \(d\).
Theorem 1.5. For all $N \geq 1$, there exists a distribution of $N$ points $Z = \{z_1, \ldots, z_N\} \subset S^d$ such that
\begin{equation}
\Delta(Z) \leq C_d N^{-\frac{1}{2} - \frac{1}{2d}} \sqrt{\log N}.
\end{equation}
The constant $C_d$ may be chosen as
\[ C_d = 20d^{\frac{3}{4} + \frac{1}{4d}} \]
whenever $N \geq 100d$.

The proof of this theorem is taken up in §2.3. We would like to make a remark that a similar result (with different absolute constants, but the same exponents of $d$) holds for the implicit constants in Beck’s spherical cap discrepancy estimate, Theorem 1.2.

We also have the complementing lower bound:

Theorem 1.6. In all dimensions $d \geq 2$, there exists a constant $B_d > 0$ such that any point distribution satisfies
\begin{equation}
\Delta(Z) \geq B_d N^{-\frac{1}{2} - \frac{1}{2d}}.
\end{equation}

Both of these theorems are known in a very similar geometric situation. It was proved by Blümlinger [MR1116689] that (10) holds for the discrepancy with respect to spherical “slices”. For $x, y \in S^d$ denote
\[ S_{xy} = \{p \in S^d : p \cdot x > 0, p \cdot y < 0\}. \]
In other words, the slice $S_{xy}$ is a half of the wedge $W_{xy}$. It should be noted that the discrepancy with respect to slices is in fact a better measure of equidistribution on the sphere than the wedge discrepancy (as mentioned earlier, wedge discrepancy doesn’t change if we move all points to the hemisphere $\{x \cdot p \geq 0\}$ by changing some points $x$ to $-x$). Using jittered sampling in a manner almost identical to Beck’s, Blümlinger showed that there exists $Z \subset S^d, \#Z = N$, such that
\[ D_{\text{slice}}(Z) = \sup_{x,y \in S^d} |D(Z, S_{xy})| \lesssim N^{-\frac{1}{2} - \frac{1}{2d}} \sqrt{\log N}. \]

Like Beck’s estimate, this bound did not say anything about the dependence of constants on the dimension. Without any regard for constants, the main estimate (9) of Theorem 1.5 follows immediately since
\[ D(Z, W_{xy}) = D(Z, S_{xy}) + D(Z, S_{x,-y}), \text{ and hence } \Delta(Z) = D_{\text{wedge}}(Z) \leq 2D_{\text{slice}}(Z). \]

An effective value of the constant $C_d$ in Theorem 1.5, which is important for uniform tesselation and one-bit compressed sensing problems, requires much more delicate considerations and constructions, some of which only became available recently, see §2.1.

Blümlinger also showed that the lower bound (1.6) holds for the slice discrepancy. The proof uses spherical harmonics and is quite involved (Matoušek [MR1697825] writes that ‘it would be interesting to find a simple proof’). In fact, it is proved that the $L^2$-discrepancy for slices is bounded below by the $L^2$-discrepancy for spherical caps, from which the result follows by Beck’s estimate (8):
\[ D_{\text{slice}}(Z) \geq D_{\text{slice}, L^2}(Z) \geq D_{\text{cap}, L^2}(Z) \geq N^{-\frac{1}{2} - \frac{1}{2d}}. \]
The result for spherical wedges can be deduced by the following symmetrization argument.

**Proof of Theorem 1.6.** For a point set \( Z = \{z_1, \ldots, z_N\} \subset S^d \), consider its symmetrization, i.e. a \( 2N \)-point set \( Z^* = Z \cup (-Z) \). It is easy to see that

\[
D(Z, W_{xy}) = D(Z, S_{xy}) + D(Z, S_{x,-y}) = D(Z, S_{xy}) + D(-Z, S_{xy}) = 2D(Z^*, S_{xy}).
\]

Therefore,

\[
\Delta(Z) = D_{\text{wedge}}(Z) \geq D_{\text{slice}}(Z^*) \gtrsim (2N)^{-\frac{1}{2} - \frac{1}{2d}},
\]

which proves (10). \qed

Inverting the bounds of Theorems 1.5 and 1.6, one immediately obtains the first announced result (5): asymptotic bounds on the size of the minimal \( \delta \)-uniform tessellation of \( S^d \), which are sharp up to a logarithmic factor.

**Theorem 1.7.** The following bounds hold for \( \delta \)-uniform tessellations of the sphere \( S^d \):  
- There exists an \( N \)-point set \( Z = \{z_1, \ldots, z_N\} \subset S^d \) with
  \[
  N \lesssim \delta^{-2 + \frac{2}{d+1}} \left( \log \frac{1}{\delta} \right)^{\frac{d+1}{d}},
  \]
  which induces a \( \delta \)-uniform tessellation of \( S^d \).
- On the other hand, any \( N \)-point set \( Z = \{z_1, \ldots, z_N\} \subset S^d \), which induces a \( \delta \)-uniform tessellation of \( S^d \), satisfies
  \[
  N \gtrsim \delta^{-2 + \frac{2}{d+1}}.
  \]

In other words, the size \( N(\delta) \) of the smallest \( \delta \)-uniform tessellation of the sphere \( S^d \) satisfies

\[
\delta^{-2 + \frac{2}{d+1}} \lesssim N(\delta) \lesssim \delta^{-2 + \frac{2}{d+1}} \left( \log \frac{1}{\delta} \right)^{\frac{d+1}{d}},
\]

where all the implicit constants depend on the dimension \( d \).

If we assume that \( \delta^{-2} \) is indeed the correct order of growth of the size of a \( \delta \)-uniform tessellation induced by random hyperplanes, this result shows that the optimal value of \( N \) is asymptotically better by a factor of \( \delta^{\frac{2}{d+1}} \). The optimal tessellation is almost achieved by the semi-random jittered sampling, see §2. As \( d \to \infty \), the situation starts looking more and more like the random case. As \( d \) tends to infinity, the bounds approach those of the random case, however.

Invoking the effective value of the constant \( C_d \), we obtain a more precise upper bound on \( N(\delta) \) in terms of both \( \delta \) and \( d \), i.e. a non-asymptotic version of the upper bound in the theorem above.

**Theorem 1.8.** There exists an absolute constant \( C > 0 \) (independent of the dimension) such that in every dimension \( d \geq 2 \) and for every \( \delta > 0 \), the cardinality \( N(\delta) \) of the smallest \( \delta \)-uniform tessellation of \( S^d \) satisfies

\[
N(\delta) \leq \max \left\{ 100d, \ C d^\alpha \delta^{-2 + \frac{2}{d+1}} \left( 1 + \log d + \log \frac{1}{\delta} \right)^{\frac{d+1}{d}} \right\},
\]

\( 8 \)
where \( \alpha = \frac{5}{2} - \frac{2}{d+1} \).

Hence, for large dimensions the implicit constant in the upper bound in (11) is close to \( d^{5/2} \) in order of magnitude. The absolute constant \( C \) above can be taken to be e.g. \( C = 4000 \). Some details are given in the end of §2.3.

**Remark:** It is an easy observation that a \( \delta \)-uniform tessellation has cells of diameter at most \( \delta \). Indeed, for \( x, y \) in the same cell \( d_H(x,y) = 0 \), therefore \( d(x,y) \leq \delta \). Therefore our results imply that there exists a tessellation of \( S^d \) by \( N \) hyperplanes with \( \delta \)-small cells with
\[
N \lesssim \delta^{-2+\frac{d}{2}} \left( \log \frac{1}{\delta} \right)^{\frac{d+1}{d}}.
\]
This result is better, in terms of the power of \( \delta \), than Theorem 3.1 in [MR3164174], where the bound is \( Cd\delta^{-4} \). However, it is much weaker than the more general proved by the authors of this paper in a simultaneous work [BL, Theorem 1.4] which shows that a random tessellation by only \( C d \delta^{-1} \log \frac{1}{\delta} \) hyperplanes has \( \delta \)-small cells with high probability.

The remark above confuses the RIP property and the small-cell property. Should be removed.

In a different vein, we also obtain an analog of the Stolarsky principle, Theorem 1.4. It is very similar in spirit to the original Stolarsky principle in the sense that it represents the \( L^2 \) discrepancy as the difference of a discrete and a continuous energies, but the underlying potential is quite different from the Euclidean distance in the case of spherical caps.

**Theorem 1.9** (Stolarsky principle for the tessellation of the sphere). For any finite set \( Z = \{z_1, \ldots, z_N\} \subset S^{d-1} \), the following relation holds
\[
\| \Delta_Z(x,y) \|_2^2 = \frac{1}{N^2} \sum_{i,j=1}^N \left( \frac{1}{2} - d(z_i,z_j) \right)^2 - \int_{S^{d-1}} \int_{S^{d-1}} \left( \frac{1}{2} - d(x,y) \right)^2 d\sigma(x) d\sigma(y).
\]

Minimizing the \( L^2 \) average of the wedge discrepancy associated to the tessellation of the sphere is thus equivalent to minimizing the discrete potential energy of \( Z \) induced by the potential that arises in this setting \( P(x,y) = \left( \frac{1}{2} - d(x,y) \right)^2 \). Intuitively, one would like to make the elements of \( Z \) ‘as orthogonal as possible’ on the average.

First of all, this suggests natural candidates for tessellations that are good or optimal on the average, e.g., spherical codes (sets \( X \subset S^d \) such that all \( x, y \in X \) satisfy \( x \cdot y < \mu \) for some parameter \( \mu \leq 1 \), see [Tem, Chapter 5] and references therein, or equiangular lines (sets \( X \subset S^d \) such that all \( x, y \in X \) satisfy \( |x \cdot y| = \mu \) for some fixed \( \mu \in [0,1) \)), see [CasTrem].

This also brings up connections to frame theory. Benedetto and Fickus [MR1968126] proved that a set \( Z = \{z_1, \ldots, z_N\} \subset S^d \) forms a normalized tight frame (i.e. there exists a constant \( A > 0 \) such that for every \( x \in \mathbb{R}^{d+1} \) an analog of Parseval’s identity holds: \( A \|x\|^2 = \sum_{i=1}^N |x \cdot z_i|^2 \) if and only if \( Z \) is a minimizer of a discrete energy known as the total frame potential:
\[
TP(Z) = \sum_{i,j=1}^N |z_i \cdot z_j|^2,
\]
which looks unmistakably similar to the discrete energy on the right-hand side of (12).

It is not known yet whether the minimizers of (12) admit a similar geometric or functional-analytic characterization, or if some of the known distributions yield reasonable values for this energy. These are interesting questions to be addressed in future research. We prove Theorem 1.9 in §3.

2. Jittered sampling

Jittered (or stratified) sampling in discrepancy theory and statistics can be viewed as a semi-random construction, somewhat intermediate between the purely random Monte Carlo algorithms and the purely deterministic low discrepancy point sets. It is easy to describe the main idea in just a few words: initially, the ambient manifold (cube, torus, sphere etc) is subdivided into \( N \) regions of equal volume and (almost) equal diameter, then a point is chosen uniformly at random in each of these pieces, independently of others.

Intuitively, this construction guarantees that the arising point set is fairly well distributed (there are no clusters or large gaps). Amazingly, it turns out that in many situations this distribution yields nearly optimal discrepancy (while purely random constructions are far from optimal, and deterministic sets are hard to construct). As mentioned before, the construction in Theorem 1.2 is precisely the jittered sampling. It differs from the corresponding lower bound only by a factor of \( \sqrt{\log N} \), which is a result of the application of large deviation inequalities. If one replaces the \( L^\infty \) norm of the discrepancy by \( L^2 \), jittered sampling actually easily gives the sharp upper bound (without \( \sqrt{\log N} \)). Similar phenomena persist in other situations (discrepancy with respect to balls or rotated rectangles in the unit cube, slices on the sphere etc).

Jittered sampling is very well described in classical references on discrepancy theory, such as [MR903025, MR1779341, MR1697825]. However, since the spherical case possesses certain subtleties, and, in addition, we want to trace the dependence of the constants on the dimension, we shall describe the construction in full detail. Besides, this procedure will also yield a quantitative bounds on the constant in the classical spherical caps discrepancy bound, Lemma 1.2.

In order to make the construction precise we need to introduce two notions: equal-area partitions with bounded diameters and approximating families.

2.1. Regular partitions of the sphere. Let \( S_i \subset \mathbb{S}^d, i = 1, 2, \ldots, N \). We say that \( \{S_i\}_{i=1}^N \) is a partition of the sphere if \( \mathbb{S}^d \) is a disjoint (up to measure zero) union of these sets, i.e.

\[
\mathbb{S}^d = \bigcup_{i=1}^N S_i \quad \text{and} \quad \sigma(S_i \cap S_j) = 0 \quad \text{for} \quad i \neq j.
\]

**Definition 2.1.** Let \( \mathcal{S} = \{S_i\}_{i=1}^N \) be a partition of \( \mathbb{S}^d \). We call it an equal-area partition if \( \sigma(S_i) = \frac{1}{N} \) for each \( i = 1, \ldots, N \).

**Definition 2.2.** Let \( \mathcal{S} = \{S_i\}_{i=1}^N \) be an equal-area partition of \( \mathbb{S}^d \), and fix \( K_d > 0 \). We say that it is a regular partition (or, an equal area partition with bounded diameters) if for every \( i = 1, \ldots, N \),

\[
diam(S_i) \leq K_d N^{-\frac{1}{d}}.
\]
In the case of the unit cube $[0, 1]^d$, regular partitions are extremely easy to construct. Indeed, for $N = M^d$, one can simply take disjoint squares of side length $M^{-1} = N^{-1/d}$. The situation is much more complicated for the sphere $\mathbb{S}^d$. Nevertheless, we have this explicit upper bound on the constant $K_d$ above.

**Theorem 2.3. [Leopardi]** For the constant $K_d$ as below, there are regular partitions $\{S_i\}_{i=1}^N$, for all $N \in \mathbb{N}$.

$$K_d = 8 \left( \frac{\Omega d}{\omega} \right)^{\frac{1}{d}},$$

where as before $\Omega$ is the $d$-dimensional Lebesgue surface measure of $\mathbb{S}^d$, and $\omega$ is the $(d-1)$-dimensional measure of $\mathbb{S}^{d-1}$.

The history of this issue (as described in [Le2]) is interesting: Stolarsky [MR0333395] asserts the existence of regular partitions of $\mathbb{S}^d$ for all $d \geq 2$, but offers no construction or proof of this fact. Later, Beck and Chen [MR903025] quote Stolarsky, and Bourgain and Lindenstrauss [MR981745] quote Beck and Chen. A complete construction of a regular partition of the sphere in arbitrary dimension was given by Feige and Schechtman [MR1900615], and Leopardi [MR2582801] found an effective value of the constant in their construction, which we quote above.

### 2.2. Approximating families

We approximate infinite family of sets (e.g., all spherical caps, or wedges) by finite families. This will facilitate the use of a union bound estimate in the next section.

**Definition 2.4.** Let $S$ and $Q$ be two collections of subsets of $\mathbb{S}^d$. We say that $Q$ is an $\varepsilon$-approximating family for $S$ if for each $S \in S$ there exist sets $A, B \in Q$ such that

$$A \subset S \subset B \quad \text{and} \quad \sigma(B \setminus A) < \varepsilon.$$

It is easy to see that for any $N$-point set $Z$ in $\mathbb{S}^d$, if $S, A, B$ are as in the definition above then the discrepancies of $Z$ with respect to these sets satisfy

$$|D(Z, S)| \leq \max\{|D(Z, A)|, |D(Z, B)|\} + \varepsilon.$$  

Hence $D_S(Z) \leq D_Q(Z) + \varepsilon$. Thus for $\varepsilon \leq N^{-1}$, the discrepancy with respect to the original family is of the same order is the discrepancy with respect to the $\varepsilon$-approximating family.

Constructions of finite approximating families are obvious in some cases, e.g. axis-parallel boxes in the unit cube or spherical caps: just take the same sets with rational parameters with small denominators.

For the spherical wedges, which is our case of interest, we have the following Lemma.

**Lemma 2.5.** For any $0 < \varepsilon < 1$, and integers $d \geq 1$, there is an approximating family $Q$ for the the collection of spherical wedges $\{W_{x,y} : x, y \in \mathbb{S}^d\}$, so that

$$|\mathcal{Q}| \leq (Cd)^{d+1} \varepsilon^{-2(d+1)},$$

where $0 < C \leq 82$ is an absolute constant.

**Proof.** We construct two separate families: one for interior and one for exterior approximation of the spherical wedges. Let $N(\varepsilon)$ be the covering number of $\mathbb{S}^d$ with respect to the
Euclidean metric, in other words, the cardinality of the smallest \( \varepsilon \)-net, i.e. set \( \mathcal{H}_\varepsilon \) such that for each \( x \in \mathbb{S}^d \) there exists \( z \in \mathcal{H}_\varepsilon \) with \( \| x - z \| \leq \varepsilon \).

A simple volume argument (see e.g. [?Vex]) shows that

\[
\mathcal{N}(\varepsilon) \leq \left( 1 + \frac{2}{\varepsilon} \right)^{d+1} \leq \left( \frac{4}{\varepsilon} \right)^{d+1}.
\]

(More precise estimates can be obtained, in particular by using \( d \)-dimensional, rather than \( (d+1) \)-dimensional volume arguments, but this will suffice for our purposes).

We shall construct an \( \varepsilon \)-approximating family as follows. Start with a \( \gamma \)-net \( \mathcal{H}_\gamma \) of size \( \mathcal{N}(\gamma) \), where \( \gamma > 0 \) is to be specified. For \( x, y \in \mathbb{S}^d \), define the exterior enlargement and interior reduction of \( W_{xy} \) as

\[
W_{xy}^\text{ext}(\gamma) = \{ p \in \mathbb{S}^d : p \cdot x \geq -\gamma, p \cdot y \leq \gamma \} \cup \{ p \in \mathbb{S}^d : p \cdot x \leq \gamma, p \cdot y \geq -\gamma \} \supset W_{xy}
\]

\[
W_{xy}^\text{int}(\gamma) = \{ p \in \mathbb{S}^d : p \cdot x \geq \gamma, p \cdot y \leq -\gamma \} \cup \{ p \in \mathbb{S}^d : p \cdot x \leq -\gamma, p \cdot y \geq \gamma \} \subset W_{xy}.
\]

We claim that the collection

\[
\mathcal{Q} = \{ W_{xy}^\text{int}(\gamma) : x, y \in \mathcal{H}_\gamma \} \cup \{ W_{xy}^\text{ext}(\gamma) : x, y \in \mathcal{H}_\gamma \}
\]

forms an approximating family for the set of all wedges \( \{ W_{xy} : x, y \in \mathbb{S}^d \} \). Indeed, let \( x', y' \in \mathbb{S}^d \). Choose \( x, y \in \mathcal{H}_\gamma \) so that \( \| x - x'\| < \gamma \), \( \| y - y'\| \leq \gamma \). Then it is easy to see that \( W_{xy}^\text{int} \subset W_{x'y'} \subset W_{xy}^\text{ext} \). Indeed, e.g. if \( p \cdot x' \geq 0 \), then \( p \cdot x = p \cdot x' - p \cdot (x' - x) \geq -\gamma \), the rest is similar.

Moreover, it is easy to see that the normalized measure of a ‘tropical belt’ around the equator satisfies

\[
\sigma(\{ p \in \mathbb{S}^d : |p \cdot x| \leq \gamma \}) \leq \frac{2\gamma \omega}{\Omega}.
\]

Therefore we can estimate

\[
\sigma(W_{xy}^\text{int}(\gamma) \setminus W_{xy}^\text{int}(\gamma)) \leq \frac{4\omega \gamma}{\Omega},
\]

hence we have an \( \varepsilon \)-approximating family with \( \varepsilon = \frac{4\omega \gamma}{\Omega} \), i.e. \( \gamma = \frac{\Omega \varepsilon}{4\omega} \).

The cardinality of this family satisfies the bound

\[
\# \mathcal{Q} = 2(\mathcal{N}(\gamma))^2 \leq 2 \left( \frac{4}{\gamma} \right)^{2(d+1)} = 2^{8d+9} \left( \frac{\omega}{\Omega} \right)^{2(d+1)} \varepsilon^{-2(d+1)} \leq (Cd)^{d+1} \varepsilon^{-2(d+1)},
\]

where \( C > 0 \) is an absolute constant which can be taken to be, e.g. \( C = 82 \), and we have used the standard fact that \( \frac{\omega}{\Omega} \leq \sqrt{\frac{d}{2\pi}} \), see e.g. [?MR2582801].

\[ \square \]

### 2.3. The spherical wedge discrepancy of jittered sampling. Proof of Theorem 1.5.

The algorithm, which we describe for the case of spherical wedges, is generic and applies to many other situations. We shall need the following version of the classical Chernoff–Hoefding large deviation bound (see e.g. [?MR1779341, ?MR1697825])
Lemma 2.6. Let \( p_i \in [0, 1], i = 1, 2, \ldots, m \). Consider centered independent random variables \( X_i, i = 1, \ldots, m \) such that \( \mathbb{P}(X_i = -p_i) = 1 - p_i \) and \( \mathbb{P}(X_i = 1 - p_i) = p_i \). Let \( X = \sum_{i=1}^{m} X_i \). Then for any \( \lambda > 0 \)

\[
\mathbb{P}(|X| > \lambda) < 2 \exp \left( - \frac{2\lambda^2}{m} \right).
\]

We start with a regular partition \( \{ S_i \}_{i=1}^{N} \) of the sphere as described in §2.1, i.e., \( S^d = \bigcup_{i=1}^{N} S_i \), \( \sigma(S_i \cap S_j) = 0 \) for \( i \neq j \), \( \sigma(S_i) = 1/N \), and \( \text{diam}(S_i) \leq K_d N^{-1/d} \) for all \( i = 1, \ldots, d \).

We now construct the set \( Z = \{ z_1, \ldots, z_N \} \) by choosing independent random points \( z_i \in S_i \) according to the uniform distribution on \( S_i \), i.e. \( N \cdot \sigma(S_i) \).

Let \( Q \) be a \( 1/N \)-approximating family for the family \( \mathcal{R} \) of interest, in our case the family of spherical wedges \( \{ W_{xy} : x, y \in S^d \} \). The size of this family, as discussed in §2.2, satisfies \( \# \mathcal{Q} \leq A_d N^{\alpha_d} \). According to (15) we may take \( A_d = (Cd)^{d+1} \) and \( \alpha_d = 2(d+1) \).

Consider a single set \( Q \in \mathcal{Q} \). It is easy to see that for those \( i = 1, \ldots, N \), for which \( S_i \cap \partial Q = \emptyset \) (\( S_i \) lies completely inside or completely outside of \( Q \)), the input of \( z_i \) and \( S_i \) to the discrepancy of \( Z \) with respect to \( Q \) is zero. In other words,

\[
D(Z, Q) = \sum_{i:S_i \cap \partial Q \neq \emptyset} \left( \frac{1}{N} \cdot 1\{ z_i \in Q \} - \sigma(S_i \cap Q) \right) = \frac{1}{N} \sum_{i=1}^{m} X_i,
\]

where \( X_i \) are exactly as in Lemma 2.6 with \( p_i = N \cdot \sigma(S_i \cap Q) \), and \( m \leq M \), where \( M \) is the maximal number of sets \( S_i \) that may intersect the boundary of any element of \( \mathcal{Q} \).

It is now straightforward to estimate \( M \) for spherical wedges. Let \( \varepsilon = K_d N^{-1/d} \). Since every \( S_i \) has diameter at most \( \varepsilon \), all the sets \( S_i \) which intersect \( \partial Q \) are contained in the set \( \partial Q + \varepsilon B \), where \( B \) is the unit ball. Recall that \( \sigma_d^* \) denotes the unnormalized Lebesgue measure on the sphere, and that we defined \( \Omega = \sigma_d^*(S^d) \) and \( \omega = \sigma_{d-1}^*(S^{d-1}) \). We then have

\[
M \cdot \frac{\Omega}{N} \leq \sigma_d^*(\partial Q + \varepsilon B) \leq \sigma_{d-1}^*(\partial Q) \cdot 2\varepsilon \leq 8K_d N^{-1/d} \omega.
\]

Hence, invoking the diameter bounds for the regular partition (13), we find that

\[
M \leq \frac{8K_d \omega}{\Omega} \cdot N^{1-\frac{1}{d}} \leq 64d^2 \left( \frac{\omega}{\Omega} \right)^{1-\frac{1}{d}} N^{1-\frac{1}{d}}.
\]

Choosing the parameter \( \lambda = (\alpha_d \cdot M)^{\frac{1}{d}} \sqrt{\log N} \), then invoking the representation (17) and the large deviation estimate (16), we find that for any given \( Q \in \mathcal{Q} \)

\[
\mathbb{P}(|D(Z, Q)| > \lambda/N) = \mathbb{P}(|X| > \lambda) \leq 2N^{-2\alpha_d}.
\]

Since \( \#Q_d \leq A_d N^{\alpha_d} \), the union bound yields

\[
\mathbb{P}(|D(Z, Q)| > \lambda/N \text{ for at least one } Q \in \mathcal{Q}) \leq 2A_d N^{-\alpha_d} < 1,
\]

whenever \( N > (2A_d)^{1/\alpha_d} \). Therefore, for such \( N \), there exists \( Z \) such that

\[
\sup_{Q \in \mathcal{Q}} |D(Z, Q)| \leq N^{-1} (\alpha_d M)^{\frac{1}{d}} \sqrt{\log N} \leq 8\sqrt{\alpha_d d^{\frac{1}{d}}} \left( \frac{\omega}{\Omega} \right)^{\frac{1}{2}-\frac{1}{2d}} N^{-\frac{1}{2}-\frac{1}{2d}} \sqrt{\log N},
\]

13
i.e. the discrepancy estimate of the form (9) holds for each member of the approximating family $Q \in Q$ with constant $8 \sqrt{\alpha_d d \frac{1}{2}} (\frac{\alpha}{\pi})^{\frac{1}{2} - \frac{2d}{d+1}}$ for $N > (2A_d)^{1/\alpha_d}$. Since this constant is greater than one, the right-hand side is greater than $\frac{1}{N}$ for all $N$. Thus according to (14), the discrepancy estimate (9) holds for all sets $W_{xy}$, $x, y \in S^d$ with twice the constant.

Recalling that $\alpha_d = 2(d+1)$ and $A_d = (Cd)^{d+1}$ and the fact that $\frac{\alpha}{\pi} \leq \sqrt{\frac{d}{2\pi}}$, we find that the constant is at most $C_d = 20d^{\frac{3}{4}} + \frac{1}{4}d^{\frac{1}{2}}$ whenever $N \geq 100$. This finishes the proof of Theorem 1.5. □

**Proof of Theorem 1.8.** It is a straightforward, but tedious task to check that if $N \geq 100d$ and

$$N > 400d^\gamma \delta^{-\frac{2d}{d+1}} \left((d+1) \log(400d^\gamma) + 2d \log \frac{1}{\delta}\right) d^{\frac{1}{d+1}},$$

where $\gamma = \frac{3}{2} - \frac{1}{d+1}$, one has $20d^{\frac{3}{4}} + \frac{1}{4}d^{\frac{1}{2}} \sqrt{\log N} < \delta$. Therefore, with positive probability jittered sampling with $N$ points yields a $\delta$-uniform tessellation. It is easy to see that the right-hand side in the equation above is bounded by $4000d^\alpha \delta^{-2+\frac{2d}{d+1}} (1 + \log d + \log \frac{1}{\delta}) d^{\frac{1}{d+1}}$, where $\alpha = \frac{5}{2} - \frac{2}{d+1}$, which proves Theorem 1.8.

3. **Stolarsky principle for the wedge discrepancy.**

We now turn to the proof of Stolarsky principle for tessellations, Theorem 1.9. Recall that the $L^2$ norm of the function $\Delta_Z(x, y)$ for a set $Z \subset S^d$

$$\|\Delta_Z(x, y)\|_2^2 = \int_{S^d} \int_{S^d} \left(\frac{1}{N} \sum_{k=1}^{N} 1_{W_{xy}}(z_k) - \sigma(W_{xy})\right)^2 d\sigma(x) d\sigma(y).$$

The proof is quite elementary in nature and conforms to a standard algorithm of many similar problems: we square out the expression above, and the cross terms yield the discrete potential energy of the interactions of points of $Z$. The idea is generally quite fruitful. Torquato [To] applies this approach to many questions of discrete geometric optimization, such as packings, coverings etc (both theoretically and numerically) to recast them as energy-minimization problems.

**Proof of Theorem 1.9.** We recall that $\sigma(W_{xy}) = d(x, y)$ and notice that we can write (up to sets of measure zero)

$$1_{W_{xy}}(z_k) = 1_{\{\text{sgn}(x \cdot z_k) \neq \text{sgn}(y \cdot z_k)\}}(z_k) = \frac{1}{2} \left(1 - \text{sgn}(x \cdot z_k) \cdot \text{sgn}(y \cdot z_k)\right),$$

where
therefore, using (18), we have

\[
\int \int \Delta_z(x, y)^2 d\sigma(x) d\sigma(y) =
\]

\[
= \frac{1}{4N^2} \int \int \sum_{i,j=1}^{N} \left(1 - \text{sgn}(x \cdot z_i) \text{sgn}(y \cdot z_i)\right) \left(1 - \text{sgn}(x \cdot z_j) \text{sgn}(y \cdot z_j)\right) d\sigma(x) d\sigma(y)
\]  

\[
= \frac{1}{N} \sum_{k=1}^{N} \int \int 1_{W_{xy}}(z_k) \cdot d(x, y) d\sigma(x) d\sigma(y)
\]

\[
+ \int \int d(x, y)^2 d\sigma(x) d\sigma(y).
\]

The most interesting term is the first one (19). Using the obvious fact that the integral \(\int sgn(p \cdot x) d\sigma(x) = 0\) for any \(p \in S^d\), we reduce this term to

\[
\int sgn(x \cdot z_i) sgn(x \cdot z_j) d\sigma(x) = 1 - 2 \int 1_{W_{zizj}}(x) d\sigma(x) = 1 - 2d(z_i, z_j).
\]

We shall see that in the second term (20) one can easily replace the discrete average over \(z_k \in Z\) over the continuous average over \(p \in S^d\), which is simpler to handle. Indeed, notice that by rotational invariance the integrand in (20) does not depend on the particular choice of \(z_k \in S^d\). Therefore, for an arbitrary pole \(p \in S^d\) we can write

\[
\int sgn(x \cdot z_i) sgn(x \cdot z_j) d\sigma(x) = 1 - 2 \int 1_{W_{zizj}}(x) d\sigma(x) = 1 - 2d(z_i, z_j).
\]
thus the term has the same form as the last one (21). Putting this together we find that
\[
\| \Delta_Z(x, y) \|_2^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \left( \frac{1}{2} - d(z_i, z_j) \right)^2 + \frac{1}{4} - \int_{S^d} \int_{S^d} d(x, y)^2 d\sigma(x) d\sigma(y).
\]
Observing that \( \int_{S^d} d(x, y) d\sigma(x) = \frac{1}{2} \) and hence
\[
\int_{S^d} \int_{S^d} \left( \frac{1}{2} - d(x, y) \right)^2 d\sigma(x) d\sigma(y) = \int_{S^d} \int_{S^d} d(x, y)^2 d\sigma(x) d\sigma(y) - \frac{1}{4},
\]
we arrive to the desired conclusion (12):
\[
\| \Delta_Z(x, y) \|_2^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \left( \frac{1}{2} - d(z_i, z_j) \right)^2 - \int_{S^d} \int_{S^d} \left( \frac{1}{2} - d(x, y) \right)^2 d\sigma(x) d\sigma(y). \quad \square
\]

3.1. \( L^2 \) discrepancy for random tesselations. Stolarsky principle provides a very simple way to compute exactly the expected value of the square of the \( L^2 \) discrepancy. Assume that the set \( Z = \{z_1, \ldots, z_N\} \subset S^d \) is random and compute the expectation of \( \| \Delta_Z(x, y) \|_2^2 \).

Obviously, for a typical point set \( Z \) and a typical wedge \( W_{xy} \) the discrepancy is of the order \( 1/\sqrt{N} \), therefore this expected value naturally behaves as \( O(1/N) \). We compute its value precisely.

**Lemma 3.1.** Let \( Z = \{z_1, \ldots, z_N\} \subset S^d \) consist of \( N \) i.i.d. uniformly distributed points on the sphere. Then
\[
\mathbb{E}_Z \| \Delta_Z(x, y) \|_2^2 = \frac{1}{N} \cdot \left( \frac{1}{2} - V_d \right).
\]

**Proof.** We shall need the quantity that already arose in the computations above, namely the second moment of the geodesic distance, i.e. the expected value of the square of the geodesic distance between two random points on the sphere
\[
V_d = \mathbb{E}_{xy} d(x, y)^2 = \int_{S^d} \int_{S^d} d(x, y)^2 d\sigma(x) d\sigma(y).
\]
It is obvious that \( \mathbb{E}_{xy} d(x, y) = \frac{1}{2} \) and hence \( \mathbb{E}_{xy} \left( \frac{1}{2} - d(x, y) \right)^2 = V_d - \frac{1}{4} \). We use the final form of the Stolarsky principle (22) to find the value of \( \mathbb{E}_Z \| \Delta_Z(x, y) \|_2^2 \). We separate the off-diagonal and diagonal terms in the discrete part of (22) to obtain
\[
\mathbb{E}_Z \| \Delta_Z(x, y) \|_2^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}_{z_i, z_j} \left( \frac{1}{2} - d(z_i, z_j) \right)^2 - \left( V_d - \frac{1}{4} \right)
\]
\[
= \frac{1}{N^2} \cdot (N^2 - N) \cdot \left( V_d - \frac{1}{4} \right) + \frac{1}{N^2} \cdot N \cdot \frac{1}{4} - \left( V_d - \frac{1}{4} \right)
\]
\[
= \frac{1}{N} \cdot \left( \frac{1}{2} - V_d \right),
\]
which finishes the proof. \( \square \)
Finally, we take a closer look at the expected value of the square of the geodesic distance $V_d$. We remark that it can be written as

$$V_d = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x, y)^2 d\sigma(x) d\sigma(y) = \frac{1}{\pi^2} \cdot \frac{\omega_{d-1}}{\Omega} \int_0^\pi \phi^2 (\sin \phi)^{d-1} d\phi,$$

where $\omega_{d-1}$ is the surface area of $\mathbb{S}^d$. For any given value of $d \geq 3$, the integrals above may be evaluated directly, although no simple closed form expression seems to be available. In Table 1 we list the values of $V_d$ in low dimensions.

| $d$ | $d = 2$ | $d = 3$ | $d = 4$ | $d = 5$ | $d = 6$ |
|-----|---------|---------|---------|---------|---------|
| $V_d$ | $\frac{1}{2} - \frac{2}{\pi^2}$ | $\frac{1}{3} - \frac{1}{2\pi^2}$ | $\frac{1}{2} - \frac{20}{9\pi^2}$ | $\frac{1}{3} - \frac{5}{8\pi^2}$ | $\frac{1}{2} - \frac{518}{225\pi^2}$ |

Are the powers above correct? There were a number of $d - 1$'s above. And I changed $(\sin \phi)^{d-2}$ to $(\sin \phi)^{d-1}$ above.

In the case of spherical cap discrepancy, a computation similar to Lemma 3.1 is even simpler. It yields:

**Lemma 3.2.** Let $Z = \{z_1, ..., z_N\} \subset \mathbb{S}^d$ consist of $N$ i.i.d. uniformly distributed points on the sphere. Then

$$\mathbb{E}_Z D_{\text{cap},L^2}^2 = \frac{U_d}{N},$$

where $U_d = \mathbb{E}_{x,y \in \mathbb{S}^d} \|x - y\|$.

Indeed, using the original Stolarsky principle, Theorem 1.4, one gets

$$\mathbb{E}_Z D_{\text{cap},L^2}^2 = \frac{1}{N^2} \sum_{i,j=1}^N \mathbb{E}_{z_i,z_j} \|z_i - z_j\| - U_d = \frac{N^2 - N}{N^2} U_d - U_d = \frac{U_d}{N}.$$