PAPER

One-bit sensing, discrepancy and Stolarsky's principle

To cite this article: D. Bilyk and M. T. Lacey 2017 *Sb. Math.* **208** 744

DOI: https://doi.org/10.1070/SM8656

One-bit sensing, discrepancy and Stolarsky's principle

D. Bilyk and M. T. Lacey

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

$$x \mapsto \{\operatorname{sign}(x \cdot z_j) \colon 1 \leqslant j \leqslant N\},$$

where $\{z_j\} \subset \mathbb{S}^d$. For $0 < \delta < 1$, we estimate $N(d, \delta)$, the smallest integer N so that there is a sign-linear map which has the δ-restricted isometric property, where we impose the normalized geodesic distance on $\mathbb{S}^d$ and the Hamming metric on H^N. Up to a polylogarithmic factor, $N(d, \delta) \approx \delta^{-2+2/(d+1)}$, which has a dimensional correction in the power of δ. 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 $\sum_{i,j}\bigl(\frac{1}{2} - d(z_i, z_j)\bigr)^2$, where d is the normalized geodesic distance.

Bibliography: 39 titles.

Keywords: discrepancy, one-bit sensing, restricted isometry property, Stolarsky principle.

§ 1. Introduction

The present paper is concerned with the following question: what is the minimal number of hyperplanes such that, for any two points on the unit sphere, the geodesic distance between them is well-approximated by the proportion of hyperplanes which separate these points. This question has connections to different topics, such as one-bit sensing (a nonlinear variant of compressive sensing), geometric functional analysis (almost isometric embeddings), and combinatorial geometry (tessellations of the sphere), while our proof techniques are taken from geometric discrepancy theory.

We now introduce the notation and make this question more precise.

Let $d \geqslant 2$ and let $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ denote the d-dimensional unit sphere. We denote by $d(x, y)$ the *geodesic* distance between x and y on $\mathbb{S}^d$ normalized so that the distance between antipodal points is 1, that is,

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

The research was partially supported by NSF (grants nos. DMS 1101519 and DMS 1265570).

AMS 2010 *Mathematics Subject Classification*. Primary 11K38, 94A12, 94A20; Secondary 52C99.

where $x{\cdot}y$ is the scalar product of the vectors x and y. The N-dimensional Hamming cube $H^N = \{-1,+1\}^N$ has the Hamming metric

$$d_H(s,t) = \frac{1}{2N}\sum_{j=1}^{N}|s_j - t_j| = \frac{1}{N}\#\{1 \leqslant j \leqslant N\colon s_j \neq t_j\},$$

where $s = (s_1,\dots,s_N) \in H^N$, and similarly for t, that is, $d_H(s,t)$ measures the proportion of the coordinates in which s and t differ. We consider *sign-linear* maps from $\mathbb{S}^d$ to H^N given by

$$\varphi_Z(x) = \{\operatorname{sgn}(z_j \cdot x)\colon 1 \leqslant j \leqslant N\},$$

where $Z = \{z_1, z_2, \dots, z_N\} \subset \mathbb{S}^d$. Note that, with an abuse of notation, $\varphi_Z(x) = \operatorname{sgn}(Ax)$, where the rows of A consist of the vectors $z_1,\dots,z_N$.

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

$$d_H(\varphi_Z(x), \varphi_Z(y))$$

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

$$\mathbb{P}\big\{\operatorname{sgn}(x\cdot z) \neq \operatorname{sgn}(y \cdot z)\big\} = d(x,y). \tag{1.1}$$

This is the original instance of the Crofton formula from integral geometry (see [36], pp. 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. Let $0 < \delta < 1$. A map $\varphi\colon \mathbb{S}^d \mapsto H^N$ *satisfies* the δ-RIP if

$$\sup_{x,y\in\mathbb{S}^d} |d_H(\varphi(x),\varphi(y)) - d(x,y)| < \delta. \tag{1.2}$$

We set $N(d,\delta)$ to be the minimal integer N for which there exists an N-point set $Z \subset \mathbb{S}^d$, such that φ_Z is a δ-RIP map.

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

$$\Delta_Z(x,y) = d_H(\varphi_Z(x),\varphi_Z(y)) - d(x,y). \tag{1.3}$$

Building on intuition, we can set $N_{\mathrm{rdm}}(d,\delta)$ to be the smallest integer N so that drawing Z uniformly at random, the sign-linear map φ_Z is a δ-RIP map, with probability at least 1/2. In a companion paper [10], we conjecture, following [32], that

$$N_{\mathrm{rdm}}(d,\delta) \lesssim d\delta^{-2}. \tag{1.4}$$

Such bounds are known for the *linear* embedding of the sphere into $\mathbb{R}^N$ (Dvoretzky's theorem). The power of δ^{-2} is sharp in the random case, as follows from the Central Limit Theorem. In [10] we prove that a δ-RIP map from $\mathbb{S}^d$ to H^N exists for N

as in (1.4), although our map is not sign-linear, but rather a composition of the 'nearest neighbour' map and a sign-linear map. We also prove an analogue of (1.4) for sparse vectors.

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

Theorem 1. *For all $d \in \mathbb{N}$ and $0 < \delta < 1$,*

$$N(d,\delta) \approx_{\log} \delta^{-2+\frac{2}{d+1}}, \tag{1.5}$$

where the equality holds up to a dimensional constant and a polylogarithmic factor in d and δ.

The upper bound in (1.5) is achieved by exhibiting a Z of small cardinality, for which φ_Z satisfies δ-RIP. *Jittered* (or *stratified*) *sampling*, a cross between purely random and deterministic constructions, provides the example. Loosely speaking, first we divide the sphere $\mathbb{S}^d$ into N roughly equal pieces, and then we choose 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 δ-RIP map. It is a deep fact from geometric discrepancy theory.

Most of the prior work concerns randomly selected Z. Jacques and coauthors [24], Theorem 2, proved an analogue of (1.4) for sparse vectors in $\mathbb{S}^d$ with an additional logarithmic term in δ. Plan and Vershynin [32] studied this question, looking for RIP for general subsets $K \subset \mathbb{S}^d$ mapped into the Hamming cube, of which the sparse vectors are a prime example. They proved [32], Theorem 1.2, that $N_{\mathrm{rdm}}(d,\delta) \lesssim d\delta^{-6}$, and conjectured (1.4), at least in the random case. Neither paper anticipates the dimensional correction in δ above.

Since in applications the dimension d is often quite large, we considered a non-asymptotic version of the upper bound in (1.5) and computed an effective value of the constant C_d, proving that it grows roughly as $d^{5/2}$ (see Theorem 4 for a more precise statement):

$$N(d,\delta) \leqslant \max\left\{ C d^{\frac{5}{2}} \delta^{-2+\frac{2}{d+1}} \left(1 + \log d + \log \frac{1}{\delta}\right)^{\frac{d}{d+1}},\ 100d \right\},$$

where $C > 0$ is an absolute constant.

Our second main result goes in a somewhat different direction. In Theorem 5 we show that the L^2-norm of $\Delta_Z(x,y)$ given in (1.3) satisfies an analogue 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 $\frac{1}{n^2}\sum_{i,j}\bigl(\frac{1}{2} - d(z_i,z_j)\bigr)^2$. This suggests interesting connections to such objects as spherical codes, equiangular lines and frames. See Theorem 5 and § 3 for details.

One-bit sensing. The restricted isometry property (RIP) was formulated by Candes and Tao [16] and is a basic concept in compressive sensing [21], Ch. 6. It can be studied in various metric spaces, and thus has many interesting variants.

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

The subsequent theory was then developed in [23], [24], [30] and [32]. For upper bounds, random selection of points on a sphere is generally used. Note that [23], Theorem 1, does contain a lower bound on the rate of recovery of a one-bit decoder. Plan and 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] and [31].

Lower bounds, like the ones proved in Theorem 1, 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 fundamental result in dimension reduction. In short, it states that for $X \subset \mathbb{S}^d \subset \mathbb{R}^{d+1}$ of cardinality k, there is a linear map $A\colon \mathbb{R}^{d+1} \to \mathbb{R}^N$, which, restricted to X, satisfies δ-RIP, provided that $N \gtrsim \delta^{-2} \log k$. This has many proofs, see [34], for instance. The connection of this lemma to compressed sensing is well known, see, for example, [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 in [22] are essentially linear in nature, so that new techniques are needed. Progress on this question might have consequences on lower bounds for *nonlinear* Johnson-Lindenstrauss RIPs.

Dvoretzky's Theorem. The results in this paper are also related to Dvoretzky's Theorem [19], which states that for all $\varepsilon > 0$ and all dimensions d there exists $N = N(d, \varepsilon)$ such that any Banach space X of dimension N contains a subspace Y of dimension d which embeds into Hilbert space with distortion at most $1 + \varepsilon$. (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 and 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 tessellations of the sphere $\mathbb{S}^d$ induced by the hyperplanes $\{z^\perp \colon z \in Z\}$. The integer $N(d, \delta)$ is the smallest size of Z so that for all $x, y \in \mathbb{S}^d$ the proportion of hyperplanes from Z that separate x and y is bounded above and below by $d(x, y) \pm \delta$. This is the geometric language used in 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] and [33].

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

$$\Omega = \frac{2\pi^{\frac{d+1}{2}}}{\Gamma(\frac{d+1}{2})},$$

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

$$\frac{\omega}{\Omega} = \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})\sqrt{\pi}} \leqslant \sqrt{\frac{d}{2\pi}}. \tag{1.6}$$

The notation $A \lesssim B$ means that $A \leqslant CB$ 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 δ.

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

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

We define the extremal (L^∞) discrepancy of Z with respect to a family $\mathscr{S}$ of measurable subsets of $\mathbb{S}^d$ to be

$$D_{\mathscr{S}}(Z) = \sup_{S \in \mathscr{S}} |D(Z, S)|. \tag{1.7}$$

If the family $\mathscr{S}$ admits a natural measure then the supremum above can also be replaced by an L^2-average. The main questions in discrepancy theory are the following. 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 and so on, see [7], [18] and [27].

In this sense the quantity $\Delta_Z(x, y)$ defined in (1.3) clearly has a discrepancy flavour. 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

$$W_{xy} = \big\{z \in \mathbb{S}^d \colon \operatorname{sgn}(x \cdot z) \neq \operatorname{sgn}(y \cdot z)\big\}. \tag{1.8}$$

The letter W stands for *wedge*, since the set W_{xy} does in fact look like a spherical wedge, that is, 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

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

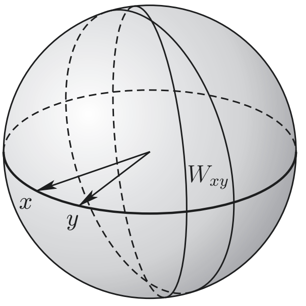

Figure 1. The spherical wedge W_{xy}.

Therefore we can rewrite the quantity (1.3) as

$$\Delta_Z(x,y) = \frac{\#(Z \cap W_{xy})}{N} - \sigma(W_{xy}) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{W_{xy}}(z_k) - \sigma(W_{xy}) =: D(Z, W_{xy}), \tag{1.9}$$

that is, the discrepancy of the N-point distribution Z with respect to the wedge W_{xy}, see § 1.3.

The RIP property can now be reformulated in terms of the L^∞-discrepancy with respect to wedges. Indeed, according to definitions (1.2) and (1.7), the map φ_Z is δ-RIP exactly when the quantity

$$\|\Delta_Z\|_\infty = \sup_{x,y\in\mathbb{S}^d} \left| \frac{\#(Z \cap W_{xy})}{N} - \sigma(W_{xy}) \right| =: D_{\mathrm{wedge}}(Z) \tag{1.10}$$

is at most δ. 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 centred at x, given by

$$C(x,t) = \{y \in \mathbb{S}^d \colon y \cdot x \geqslant t\}.$$

Denote the set of all spherical caps by $\mathscr{C}$. For an N-point set $Z \subset \mathbb{S}^d$ let

$$D_{\mathrm{cap}}(Z) = \sup_{C\in\mathscr{C}} |D(Z,C)| = \sup_{C\in\mathscr{C}} \left| \frac{\#(Z\cap C)}{N} - \sigma(C) \right|$$

be the extremal discrepancy of Z with respect to spherical caps $\mathscr{C}$. The following classical results due to Beck [5], [6] yield almost precise information about the growth of this quantity in terms of N.

Beck's Theorem (on spherical cap discrepancy). *For dimensions $d \geqslant 2$, the following hold.*

Upper bound. There exists an N-point set $Z \subset \mathbb{S}^d$ with spherical cap discrepancy

$$D_{\mathrm{cap}}(Z) \lesssim N^{-\frac{1}{2}-\frac{1}{2d}} \sqrt{\log N}. \tag{1.11}$$

Lower bound. For any N-point set $Z \subset \mathbb{S}^d$ the spherical cap discrepancy satisfies

$$D_{\mathrm{cap}}(Z) \gtrsim N^{-\frac{1}{2}-\frac{1}{2d}}. \tag{1.12}$$

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. The proof of (1.11) in [6] states that '... using probabilistic ideas it is not hard to show...' and refers to [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 behaviour 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 in both asymptotic and nonasymptotic regimes, we shall explore this construction (in the case of spherical wedges) tracing the dependence of the constant on the dimension very scrupulously.

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,

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

More precisely, a lower bound stronger than (1.12) holds:

$$D_{\mathrm{cap},L^2}(Z) \gtrsim N^{-\frac{1}{2}-\frac{1}{2d}}. \tag{1.13}$$

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 in [35].

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

$$\frac{1}{c_d} [D_{\mathrm{cap},L^2}(Z)]^2 = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \|x - y\|\, d\sigma(x)\, d\sigma(y) - \frac{1}{N^2} \sum_{i,j=1}^{N} \|z_i - z_j\|, \tag{1.14}$$

where $\|\cdot\|$ *is the Euclidean norm and*

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

for an arbitrary pole $p \in \mathbb{S}^d$.

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 the Stolarsky principle can be found in [9] and [14].

1.3. Main results. Analogues of Beck's discrepancy estimates (1.11) and (1.12), as well as of the 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 (1.8) and of the wedge discrepancy (1.9), (1.10).

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

Theorem 3. *For all integers* $d \geqslant 2$ *there are* B_d *and* $C_d > 0$ *such that for all integers* $N \geqslant 1$ *the following hold.*

Upper bound. There exists a distribution of N *points* $Z \subset \mathbb{S}^d$ *with*

$$D_{\text{wedge}}(Z) \leqslant C_d N^{-\frac{1}{2}-\frac{1}{2d}}\sqrt{\log N}. \tag{1.15}$$

Provided $N \geqslant 100d$, *we have* $C_d \leqslant 20 d^{\frac{3}{4}+\frac{1}{4d}}$.

Lower bound. For any $Z \subset \mathbb{S}^d$ *with cardinality* N,

$$D_{\text{wedge}}(Z) \geqslant B_d N^{-\frac{1}{2}-\frac{1}{2d}}. \tag{1.16}$$

Both inequalities are known in a very similar geometric situation. It was proved by Blümlinger [11] that the upper bound (1.15) holds for the discrepancy with respect to spherical 'slices'. For $x, y \in \mathbb{S}^d$ denote

$$S_{xy} = \{z \in \mathbb{S}^d\colon z\cdot x > 0,\ z\cdot y < 0\}.$$

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 does not change if we move all points to the hemisphere $\{x\cdot p \geqslant 0\}$ by changing some points x to $-x$). Using jittered sampling in a manner almost identical to Beck's, Blümlinger showed that there exists $Z \subset \mathbb{S}^d$, $\#Z = N$, such that

$$D_{\text{slice}}(Z) = \sup_{x,y\in\mathbb{S}^d} |D(Z, S_{xy})| \lesssim N^{-\frac{1}{2}-\frac{1}{2d}}\sqrt{\log N}.$$

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 (1.15) of Theorem 3 follows immediately since

$$D(Z, W_{xy}) = D(Z, S_{xy}) + D(Z, S_{-x,-y}),$$

and hence

$$D_{\text{wedge}}(Z) \leqslant 2D_{\text{slice}}(Z).$$

An effective value of the constant C_d in Theorem 3, which is important for uniform tessellation 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 (1.16) 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 was shown 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):

$$D_{\text{slice}}(Z) \gtrsim D_{\text{slice},L^2}(Z) \gtrsim D_{\text{cap},L^2}(Z) \gtrsim N^{-\frac{1}{2}-\frac{1}{2d}}. \tag{1.17}$$

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

Proof of (1.16). For a point set $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$, consider its symmetrization, that is, 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,

$$D_{\text{wedge}}(Z) \geqslant 2D_{\text{slice}}(Z^*) \gtrsim (2N)^{-\frac{1}{2}-\frac{1}{2d}},$$

which proves (1.16).

Inverting the bounds of Theorems 3, one immediately obtains the result announced first (1.5), which gives asymptotic bounds on the minimal dimension of a sign-linear δ-RIP from $\mathbb{S}^d$ to the Hamming cube.

Theorem 4. *There exists an absolute constant $C > 0$ (independent of the dimension) such that in every dimension $d \geqslant 2$ and for every $\delta > 0$, the integer $N(d, \delta)$ in Definition 1 satisfies*

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

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

The absolute constant C above can be taken to be $C = 4000$, for instance. Some details are given in the end of § 2.3.

In a different vein, we also obtain a variant of the Stolarsky Invariance Principle (1.14).

Theorem 5 (Stolarsky principle for wedges). *For any finite set $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$, the following relation holds:*

$$\|\Delta_Z(x,y)\|_2^2 = \frac{1}{N^2} \sum_{i,j=1}^{N} \left(\frac{1}{2} - d(z_i, z_j)\right)^2 - \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} \left(\frac{1}{2} - d(x,y)\right)^2 d\sigma(x)\, d\sigma(y). \tag{1.18}$$

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, we 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, for example, spherical codes (sets $X \subset \mathbb{S}^d$ such that all $x, y \in X$ satisfy $x \cdot y < \mu$ for some parameter $\mu \leqslant 1$, see [37], Ch. 5, and references there), or equiangular lines (sets $X \subset \mathbb{S}^d$ 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 $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$ forms a normalized tight frame (that is, there exists a constant $A > 0$ such that for every $x \in \mathbb{R}^{d+1}$ an analogue 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*:

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

which looks unmistakably similar to the discrete energy on the right-hand side of (1.18).

It is not yet known whether the minimizers of (1.18) 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 5 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, and so on) 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 the others.

Intuitively, this construction guarantees that the point set which results 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 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 we replace the L^∞-norm of the discrepancy by L^2, jittered sampling actually gives the sharp upper bound easily (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 and so on).

Jittered sampling has been very well described in the classical references on discrepancy theory, such as [7], [18] and [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 bound on the constant in the classical spherical caps discrepancy estimate (1.11).

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, that is, $\mathbb{S}^d = \bigcup_{i=1}^N S_i$ and $\sigma(S_i \cap S_j) = 0$ for $i \neq j$.

Definition 2. Let $\mathscr{S} = \{S_i\}_{i=1}^N$ be a partition of $\mathbb{S}^d$. We call it an *equal-area partition* if $\sigma(S_i) = 1/N$ for each $i = 1, \ldots, N$.

Definition 3. Let $\mathscr{S} = \{S_i\}_{i=1}^N$ be an equal-area partition of $\mathbb{S}^d$. We say that it is a *regular partition* (or an equal-area partition with bounded diameters) with constant $K_d > 0$ if, for every $i = 1, \ldots, N$,

$$\operatorname{diam}(S_i) \leqslant K_d N^{-\frac{1}{d}}.$$

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

Theorem 6 (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*

$$K_d = 8\left(\frac{\Omega d}{\omega}\right)^{\frac{1}{d}}, \tag{2.1}$$

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

Note that Leopardi states the result without the specific value of the constant K_d. However, it can easily be extracted from the proof, see p. 9 in [26].

The history of this issue (as described in [25]) is interesting. In [35], Stolarsky asserts the existence of regular partitions of $\mathbb{S}^d$ for all $d \geqslant 2$, 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 for the constant in their construction, which we have given above.

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

Definition 4. Let $\mathscr{S}$ and $\mathscr{Q}$ be two collections of subsets of $\mathbb{S}^d$. We say that $\mathscr{Q}$ is an *ε-approximating family* (also known as ε-bracketing) for $\mathscr{S}$ if, for each $S \in \mathscr{S}$, there exist sets $A, B \in \mathscr{Q}$ 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 and B are as in the definition above then the discrepancies of Z with respect to these sets satisfy

$$|D(Z,S)| \leqslant \max\{|D(Z,A)|, |D(Z,B)|\} + \varepsilon. \tag{2.2}$$

Hence $D_{\mathscr{S}}(Z) \leqslant D_{\mathscr{Q}}(Z) + \varepsilon$. Thus, for $\varepsilon \leqslant N^{-1}$ the discrepancy with respect to the original family is of the same order as the discrepancy with respect to the ε-approximating family.

Constructions of finite approximating families are obvious in some cases, for example 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 1. *For any $0 < \varepsilon < 1$ and integer $d \geqslant 1$ there is an approximating family $\mathscr{Q}$ for the the collection of spherical wedges $\{W_{xy} \colon x, y \in \mathbb{S}^d\}$ with*

$$\#\mathscr{Q} \leqslant (Cd)^{d+1} \varepsilon^{-2(d+1)}, \tag{2.3}$$

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

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

A simple volume argument (see [39], for example) shows that

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

(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 construct an ε-approximating family as follows. Start with an γ-net $\mathscr{H}_\gamma$ of size $\mathscr{N}(\gamma)$, 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

$$W_{xy}^{\text{ext}}(\gamma) = \{p \in \mathbb{S}^d \colon p \cdot x \geqslant -\gamma, p \cdot y \leqslant \gamma\} \cup \{p \in \mathbb{S}^d \colon p \cdot x \leqslant \gamma, p \cdot y \geqslant -\gamma\} \supset W_{xy},$$
$$W_{xy}^{\text{int}}(\gamma) = \{p \in \mathbb{S}^d \colon p \cdot x \geqslant \gamma, p \cdot y \leqslant -\gamma\} \cup \{p \in \mathbb{S}^d \colon p \cdot x \leqslant -\gamma, p \cdot y \geqslant \gamma\} \subset W_{xy}.$$

We claim that the collection

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

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

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

$$\sigma\big(\{p \in \mathbb{S}^d \colon |p \cdot x| \leqslant \gamma\}\big) \leqslant \frac{2\gamma\omega}{\Omega}.$$

Therefore we can estimate

$$\sigma\big(W_{xy}^{\text{ext}}(\gamma) \setminus W_{xy}^{\text{int}}(\gamma)\big) \leqslant \frac{4\omega\gamma}{\Omega},$$

hence we have an ε-approximating family with $\varepsilon = 4\omega\gamma/\Omega$, that is, $\gamma = \Omega\varepsilon/(4\omega)$.

The cardinality of this family satisfies the bound

$$\#\mathscr{Q} = 2(\mathscr{N}(\gamma))^2 \leqslant 2\bigg(\frac{4}{\gamma}\bigg)^{2(d+1)} = 2^{8d+9}\bigg(\frac{\omega}{\Omega}\bigg)^{2(d+1)} \varepsilon^{-2(d+1)} \leqslant (Cd)^{d+1}\varepsilon^{-2(d+1)},$$

where $C > 0$ is an absolute constant which can be taken to be $C = 82$, for example. Here we have used the standard fact (1.6).

Lemma 1 is proved.

2.3. The spherical wedge discrepancy of jittered sampling. Proof of Theorem 3. 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, for example, [18] and [27]).

Lemma 2. *Let $p_i \in [0,1]$, $i = 1, 2, \dots, m$. Consider centred independent random variables X_i, $i = 1, \dots, m$, such that*

$$\mathbb{P}(X_i = -p_i) = 1 - p_i \quad \textit{and} \quad \mathbb{P}(X_i = 1 - p_i) = p_i.$$

Let $X = \sum_{i=1}^m X_i$. Then for any $\lambda > 0$

$$\mathbb{P}(|X| > \lambda) < 2\exp\bigg(-\frac{2\lambda^2}{m}\bigg). \tag{2.4}$$

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

We now construct the set $Z = \{z_1, \dots, z_N\}$ by choosing independent random points $z_i \in S_i$ according to the uniform distribution on S_i, that is, $N \cdot \sigma|_{S_i}$.

Let $\mathscr{Q}$ be a $1/N$-approximating family for the family $\mathscr{R}$ of interest, in our case the family of spherical wedges $\{W_{xy} \colon x, y \in \mathbb{S}^d\}$. The size of this family, as discussed in § 2.2, satisfies $\#\mathscr{Q} \leqslant A_d N^{\alpha_d}$. According to (2.3) we can take $A_d = (Cd)^{d+1}$ and $\alpha_d = 2(d+1)$.

Consider a single set $Q \in \mathscr{Q}$. It is easy to see that for those $i = 1, \dots, N$ for which $S_i \cap \partial Q = \varnothing$ (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,

$$D(Z, Q) = \frac{1}{N} \sum_{i\colon S_i \cap \partial Q \neq \varnothing} \big(\mathbf{1}_Q(z_i) - N\sigma(S_i \cap Q)\big) = \frac{1}{N} \sum_{i=1}^m X_i, \tag{2.5}$$

where the X_i are exactly as in Lemma 2 with $p_i = N \cdot \sigma(S_i \cap Q)$, and $m \leqslant M$, where M is the maximal number of sets S_i that may intersect the boundary of any element of $\mathscr{Q}$.

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 ε, all the sets S_i which intersect ∂Q are contained in the set $\partial Q + \varepsilon B$, where B is the unit ball. Recall that σ_d^* denotes the unnormalized Lebesgue measure on the sphere, and that we defined $\Omega = \sigma_d^*(\mathbb{S}^d)$ and $\omega = \sigma_{d-1}^*(\mathbb{S}^{d-1})$. We then have

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

Hence, invoking the diameter bounds for the regular partition (2.1) we find that

$$M \leqslant \frac{8K_d\omega}{\Omega} N^{1-\frac{1}{d}} \leqslant 64 d^{\frac{1}{d}} \left(\frac{\omega}{\Omega}\right)^{1-\frac{1}{d}} N^{1-\frac{1}{d}}.$$

Choosing the parameter $\lambda = (\alpha_d \cdot M)^{\frac{1}{2}}\sqrt{\log N}$, then invoking the representation (2.5) and the large deviation estimate (2.4), we find that for any given $Q \in \mathscr{Q}$

$$\mathbb{P}\left(|D(Z,Q)| > \frac{\lambda}{N}\right) = \mathbb{P}(|X| > \lambda) \leqslant 2N^{-2\alpha_d}.$$

Since $\#Q_d \leqslant A_d N^{\alpha_d}$, the union bound yields

$$\mathbb{P}(|D(Z,Q)| > \frac{\lambda}{N} \quad \text{for at least one } Q \in \mathscr{Q}) \leqslant 2A_d N^{-\alpha_d} < 1,$$

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

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

that is, the discrepancy estimate of the form (1.15) holds for each member of the approximating family $Q \in \mathscr{Q}$ with constant $8\sqrt{\alpha_d}\, d^{\frac{1}{2d}}(\omega/\Omega)^{\frac{1}{2}-\frac{1}{2d}}$ for $N > (2A_d)^{1/\alpha_d}$. Since this constant is greater then one, the right-hand side is greater than $1/N$ for all N. Thus according to (2.2), the discrepancy estimate (1.15) holds for all sets W_{xy}, $x, y \in \mathbb{S}^d$ with twice the constant.

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

Proof of Theorem 4. It is a straightforward, but tedious task to check that if $N \geqslant 100d$ and

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

where $\gamma = \frac{3}{2} - \frac{1}{d+1}$, then $20d^{\frac{3}{4}+\frac{1}{4d}} N^{-\frac{1}{2}-\frac{1}{2d}}\sqrt{\log N} < \delta$. Therefore, with positive probability, jittered sampling with N points yields a δ-uniform tessellation. It is easy to see that the right-hand side in the above equation is bounded by

$$4000d^{\alpha}\delta^{-2+\frac{2}{d+1}}\left(1+\log d+\log\frac{1}{\delta}\right)^{\frac{d}{d+1}},$$

where $\alpha=\frac{5}{2}-\frac{2}{d+1}$, which proves Theorem 4.

§ 3. Stolarsky principle for the wedge discrepancy

We now turn to the proof of Theorem 5, the Stolarsky principle for tessellations. Recall that the L^2-norm of the function $\Delta_Z(x,y)$ for a set $Z\subset\mathbb{S}^d$ is

$$\|\Delta_Z(x,y)\|_2^2=\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\left(\frac{1}{N}\sum_{k=1}^{N}\mathbf{1}_{W_{xy}}(z_k)-\sigma(W_{xy})\right)^2 d\sigma(x)\,d\sigma(y). \tag{3.1}$$

The proof is quite elementary in nature and conforms to a standard algorithm for 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 [38] applies this approach (both theoretically and numerically) to many questions in discrete geometric optimization, such as packings, coverings, number variance, to recast them as energy-minimization problems.

Proof of Theorem 5. We recall that $\sigma(W_{xy})=d(x,y)$ and notice that we can write (up to sets of measure zero)

$$\mathbf{1}_{W_{xy}}(z_k)=\mathbf{1}_{\{\mathrm{sgn}(x\cdot z_k)\neq\mathrm{sgn}(y\cdot z_k)\}}(z_k)=\frac{1}{2}\big(1-\mathrm{sgn}(x\cdot z_k)\mathrm{sgn}(y\cdot z_k)\big).$$

Therefore, using (3.1), we have

$$\begin{aligned}
&\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\Delta_Z(x,y)^2\,d\sigma(x)\,d\sigma(y)\\
&=\frac{1}{4N^2}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\sum_{i,j=1}^{N}\big(1-\mathrm{sgn}(x\cdot z_i)\mathrm{sgn}(y\cdot z_i)\big)\big(1-\mathrm{sgn}(x\cdot z_j)\mathrm{sgn}(y\cdot z_j)\big)\,d\sigma(x)\,d\sigma(y)\\
&\qquad-\frac{2}{N}\sum_{k=1}^{N}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\mathbf{1}_{W_{xy}}(z_k)d(x,y)\,d\sigma(x)\,d\sigma(y)\\
&\qquad+\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}d(x,y)^2\,d\sigma(x)\,d\sigma(y).
\end{aligned} \tag{3.2}$$

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

$$\int_{\mathbb{S}^d}\mathrm{sgn}(p\cdot x)\,d\sigma(x)=0\quad\forall\,p\in\mathbb{S}^d,$$

we reduce this term to

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

The last line is obtained by rewriting the integrand as $\mathrm{sgn}(x \cdot z_i)\mathrm{sgn}(x \cdot z_j) = 1 - 2 \cdot \mathbf{1}_{W_{z_i z_j}}(x)$, so that

$$\int_{\mathbb{S}^d} \mathrm{sgn}(x \cdot z_i)\mathrm{sgn}(x \cdot z_j)\, d\sigma(x) = 1 - 2\int_{\mathbb{S}^d} \mathbf{1}_{W_{z_i z_j}}(x)\, d\sigma(x) = 1 - 2d(z_i, z_j).$$

We shall see that in the second term in (3.2) we can easily replace the discrete average over $z_k \in Z$ by the continuous average over $p \in \mathbb{S}^d$, which is simpler to handle. Indeed, notice that by rotational invariance the integrand in (3.2) 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

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

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

$$\begin{aligned} &2\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} \mathbf{1}_{W_{xy}}(p) d(x,y)\, d\sigma(x)\, d\sigma(y) \\ &\quad = 2\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} \mathbf{1}_{W_{xy}}(p) d(x,y)\, d\sigma(x)\, d\sigma(y)\, d\sigma(p) \\ &\quad = 2\int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\left[\int_{\mathbb{S}^d} \mathbf{1}_{W_{xy}}(p)\, d\sigma(p)\right] d(x,y)\, d\sigma(x)\, d\sigma(y) \\ &\quad = 2\int_{\mathbb{S}^d}\int_{\mathbb{S}^d} d(x,y)^2\, d\sigma(x)\, d\sigma(y), \end{aligned}$$

thus the term has the same form as the last one in (3.2). Putting these together we find that

$$\|\Delta_Z(x,y)\|_2^2 = \frac{1}{N^2}\sum_{i,j=1}^{N}\left(\frac{1}{2} - d(z_i,z_j)\right)^2 + \frac{1}{4} - \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} d(x,y)^2\, d\sigma(x)\, d\sigma(y).$$

Observing that $\displaystyle\int_{\mathbb{S}^d} d(x,y)\, d\sigma(x) = \frac{1}{2}$ 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 at the desired conclusion (1.18):

$$\|\Delta_Z(x,y)\|_2^2 = \frac{1}{N^2}\sum_{i,j=1}^{N}\left(\frac{1}{2} - d(z_i,z_j)\right)^2 - \int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\left(\frac{1}{2} - d(x,y)\right)^2 d\sigma(x)\, d\sigma(y). \quad (3.3)$$

3.1. L^2-discrepancy for random tessellations. The Stolarsky principle provides a very simple way of computing the expected value of the square of the L^2-discrepancy. Assume that the set $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$ 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 $\mathscr{O}(1/N)$. To compute its value precisely we shall need the quantity that has already arisen in the computations above, namely the second moment of the geodesic distance, or in other words, the expected value of the square of the geodesic distance between two random points on the sphere:

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

Lemma 3. *Let $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$ consist of N i.i.d. (uniformly distributed) points on the sphere. Then*

$$\mathbb{E}_Z \|\Delta_Z(x,y)\|_2^2 = \frac{1}{N}\left(\frac{1}{2} - V_d\right).$$

Proof. It is obvious that $\mathbb{E}_{xy} d(x,y) = 1/2$ and hence $\mathbb{E}_{xy}(1/2 - d(x,y))^2 = V_d - 1/4$. To find the value of $\mathbb{E}_Z\|\Delta_Z(x,y)\|_2^2$ we use the final form of the Stolarsky principle (3.3). We separate the off-diagonal and diagonal terms in the discrete part of (3.3) to obtain

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

which completes the proof.

In the case of spherical cap discrepancy this computation is even simpler.

Lemma 4. *Let $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$ consist of N i.i.d. (uniformly distributed) points on the sphere. Then*

$$\mathbb{E}_Z D^2_{\mathrm{cap},L^2} = \frac{c_d U_d}{N}, \tag{3.4}$$

where $U_d = \mathbb{E}_{x,y\in\mathbb{S}^d}\|x-y\|$ and c_d is the constant from Theorem 2.

Using the original Stolarsky principle, Theorem 2, gives

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

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

$$V_d = \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} d(x,y)^2\, d\sigma(x)\, d\sigma(y) = \frac{1}{\pi^2}\frac{\omega}{\Omega}\int_0^{\pi} \phi^2 (\sin\phi)^{d-1} d\phi, \tag{3.5}$$

where ω is the surface area of $\mathbb{S}^{d-1}$. In Table 1 we list the values of V_d in low dimension.

Table 1. The values of V_d.

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

3.2. L^2 wedge discrepancy for jittered sampling. The Stolarsky principle (3.3) allows us to prove that jittered sampling yields optimal order of the L^2 wedge discrepancy quite easily.

Lemma 5. *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*

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

A matching lower bound for arbitrary N-point sets is known for caps and slices and can be easily generalized to wedges, see (1.17) and the discussion immediately following it.

Proof of Lemma 5. We notice that for $i \neq j$ we have

$$\mathbb{E}\left(\frac{1}{2} - d(z_i,z_j)\right)^2 = N^2 \int_{S_i}\int_{S_j}\left(\frac{1}{2} - d(x,y)\right)^2 d\sigma(x)\, d\sigma(y),$$

while for $i = j$ we simply get $\frac{1}{4}$. Therefore,

$$\begin{aligned}
&\mathbb{E}_Z \|\Delta_Z(x,y)\|_2^2 \\
&\quad = \frac{1}{N^2}\sum_{i,j=1}^{N} \mathbb{E}_{z_i,z_j}\left(\frac{1}{2} - d(z_i,z_j)\right)^2 - \int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\left(\frac{1}{2} - d(x,y)\right)^2 d\sigma(x)\, d\sigma(y) \\
&\quad = \sum_{i=1}^{N}\int_{S_i}\int_{S_i}\big(d(x,y) - d^2(x,y)\big) d\sigma(x)\, d\sigma(y) \\
&\quad \leqslant \sum_{i=1}^{N} K_d N^{-\frac{1}{d}} \frac{1}{N^2} = K_d N^{-1-\frac{1}{d}},
\end{aligned}$$

since $d(x,y) - d^2(x,y) \leqslant d(x,y) \leqslant \|x-y\| \leqslant K_d N^{-\frac{1}{d}}$ for $x, y \in S_i$.

Bibliography

[1] C. Aistleitner, J. S. Brauchart and J. Dick, "Point sets on the sphere $\mathbb{S}^2$ with small spherical cap discrepancy", *Discrete Comput. Geom.* **48**:4 (2012), 990–1024.

[2] K. Ball, "The Ribe programme", *Séminaire Bourbaki*, vol. 2011/2012, Exposés 1043–1058, Astérisque, vol. 352, Soc. Math. France, Paris 2013, pp. viii, 147–159, Exp. No. 1047.

[3] R. Baraniuk, M. Davenport, R. DeVore and M. Wakin, "A simple proof of the restricted isometry property for random matrices", *Constr. Approx.* **28**:3 (2008), 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*, arXiv: 1407.8246.

[5] J. Beck, "Some upper bounds in the theory of irregularities of distribution", *Acta Arith.* **43**:2 (1984), 115–130.

[6] J. Beck, "Sums of distances between points on a sphere – an application of the theory of irregularities of distribution to discrete geometry", *Mathematika* **31**:1 (1984), 33–41.

[7] J. Beck and W. W. L. Chen, *Irregularities of distribution*, Cambridge Tracts in Math., vol. 89, Cambridge Univ. Press, Cambridge 1987, xiv+294 pp.

[8] J. J. Benedetto and M. Fickus, "Finite normalized tight frames", *Adv. Comput. Math.* **18**:2–4 (2003), 357–385.

[9] D. Bilyk, F. Dai and R. Matzke, *Stolarsky principle and energy optimization on the sphere*, arXiv: 1611.04420.

[10] D. Bilyk and M. T. Lacey, *Random tessellations, restricted isometric embeddings, and one bit sensing*, arXiv: 1512.06697.

[11] M. Blümlinger, "Slice discrepancy and irregularities of distribution on spheres", *Mathematika* **38**:1 (1991), 105–116.

[12] P. Boufounos and R. Baraniuk, "1-bit compressive sensing", 2008 42*nd annual conference on information sciences and systems CISS*, IEEE, Princeton, NJ 2008, pp. 16–21.

[13] J. Bourgain and J. Lindenstrauss, "Distribution of points on spheres and approximation by zonotopes", *Israel J. Math.* **64**:1 (1988), 25–31.

[14] J. S. Brauchart and J. Dick, "A simple proof of Stolarsky's invariance principle", *Proc. Amer. Math. Soc.* **141**:6 (2013), 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**:2 (2015), 782–811.

[16] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?", *IEEE Trans. Inform. Theory* **52**:12 (2006), 5406–5425.

[17] P. G. Casazza, D. Redmond and J. C. Tremain, "Real equiangular frames", 2008 42*nd annual conference on information sciences and systems*, IEEE, Princeton, NJ 2008, pp. 715–720.

[18] B. Chazelle, *The discrepancy method. Randomness and complexity*, Cambridge Univ. Press, Cambridge 2000, xviii+463 pp.

[19] A. 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] U. Feige and G. Schechtman, "On the optimality of the random hyperplane rounding technique for MAX CUT", *Random Structures Algorithms* **20**:3 (2002), 403–440.

[21] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Appl. Numer. Harmon. Anal., Birkhäuser/Springer, New York 2013, xviii+625 pp.

[22] K. G. Larsen and J. Nelson, "The Johnson-Lindenstrauss lemma is optimal for linear dimensionality reduction", 43*rd international colloquium on automata, languages, and programming* (*ICALP* 2016) (Rome 2016), LIPIcs, vol. 55, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl 2016, 82, 11 pp.

[23] L. Jacques, K. Degraux and C. De Vleeschouwer, "Quantized iterative hard thresholding: bridging 1-bit and high-resolution quantized compressed sensing", *10th international conference on sampling theory and applications* (*SampTA* 2013), 2013, pp. 105–108, http://www.eurasip.org/Proceedings/Ext/SampTA2013/papers/p105-jacques.pdf.
[24] L. Jacques, J. N. Laska, P. T. Boufounos and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors", *IEEE Trans. Inform. Theory* **59**:4 (2013), 2082–2102.
[25] P. Leopardi, "A partition of the unit sphere into regions of equal area and small diameter", *Electron. Trans. Numer. Anal.* **25** (2006), 309–327.
[26] P. Leopardi, "Diameter bounds for equal area partitions of the unit sphere", *Electron. Trans. Numer. Anal.* **35** (2009), 1–16.
[27] J. Matoušek, *Geometric discrepancy. An illustrated guide*, Algorithms Combin., vol. 18, Springer-Verlag, Berlin 1999, xii+288 pp.
[28] A. Naor, "An introduction to the Ribe program", *Jpn. J. Math.* **7**:2 (2012), 167–233.
[29] F. Pausinger and S. Steinerberger, "On the discrepancy of jittered sampling", *J. Complexity* **33** (2016), 199–216; arXiv: 1510.00251.
[30] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming", *Comm. Pure Appl. Math.* **66**:8 (2013), 1275–1297.
[31] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach", *IEEE Trans. Inform. Theory* **59**:1 (2013), 482–494.
[32] Y. Plan and R. Vershynin, "Dimension reduction by random hyperplane tessellations", *Discrete Comput. Geom.* **51**:2 (2014), 438–461.
[33] A. Reznikov and E. B. Saff, "The covering radius of randomly distributed points on a manifold", *Int. Math. Res. Not. IMRN* **2016**:19 (2016), 6065–6094.
[34] G. Schechtman, "Two observations regarding embedding subsets of Euclidean spaces in normed spaces", *Adv. Math.* **200**:1 (2006), 125–135.
[35] K. B. Stolarsky, "Sums of distances between points on a sphere. II", *Proc. Amer. Math. Soc.* **41**:2 (1973), 575–582.
[36] S. Tabachnikov, *Geometry and billiards*, Stud. Math. Libr., vol. 30, Amer. Math. Soc., Providence, RI; Mathematics Advanced Study Semesters, University Park, PA 2005, xii+176 pp.
[37] V. Temlyakov, *Greedy approximation*, Cambridge Monogr. Appl. Comput. Math., vol. 20, Cambridge Univ. Press, Cambridge 2011, xiv+418 pp.
[38] S. Torquato, "Reformulation of the covering and quantizer problems as ground states of interacting particles", *Phys. Rev. E* (3) **82**:5 (2010), 056109, 52 pp.; arXiv: 1009.1443.
[39] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices", *Compressed sensing*, Cambridge Univ. Press, Cambridge 2012, pp. 210–268.

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

Received 21/DEC/15 and 13/DEC/16

Michael T. Lacey
School of Mathematics,
Georgia Institute of Technology,
Atlanta, GA, USA
E-mail: lacey@math.gatech.edu