CONSTRUCTIVE
APPROXIMATION

CrossMark

The Stolarsky Principle and Energy Optimization on the Sphere

Dmitriy Bilyk[1] · Feng Dai[2] · Ryan Matzke[1]

Abstract The classical *Stolarsky invariance principle* connects the spherical cap L^2 discrepancy of a finite point set on the sphere to the pairwise sum of Euclidean distances between the points. In this paper, we further explore and extend this phenomenon. In addition to a new elementary proof of this fact, we establish several new analogs, which relate various notions of discrepancy to different discrete energies. In particular, we find that the hemisphere discrepancy is related to the sum of geodesic distances. We also extend these results to arbitrary measures on the sphere and arbitrary notions of discrepancy and apply them to problems of energy optimization and combinatorial geometry and find that, surprisingly, the geodesic distance energy behaves differently than its Euclidean counterpart.

Communicated by Doug Hardin.

✉ Dmitriy Bilyk
   dbilyk@math.umn.edu

   Feng Dai
   fdai@ualberta.ca

   Ryan Matzke
   matzk053@umn.edu

[1] School of Mathematics, University of Minnesota, Minneapolis, MN 55408, USA

[2] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada

🌐 Springer

1 Introduction and Main Results

In numerous areas of mathematics and other sciences, one is faced with the problem of distributing large finite sets of points on the sphere as uniformly as possible. There exist various quantitative measures of uniformity of spherical point distributions. Among the most popular ones are discrepancy and energy.

Let $Z = \{z_1, ..., z_N\} \subset \mathbb{S}^d$ be an N-point set in the d-dimensional sphere, and let σ denote the normalized Lebesgue surface measure on $\mathbb{S}^d$; i.e., $\sigma(\mathbb{S}^d) = 1$. For a given subset of the sphere, $A \subset \mathbb{S}^d$, the discrepancy of Z with respect to A is defined as

$$D(Z, A) = \frac{1}{N} \sum_{k=1}^{N} \mathbf{1}_A(z_k) - \sigma(A); \tag{1.1}$$

in other words, $D(Z, A)$ indicates how well the Lebesgue measure of A is approximated by the counting measure $\frac{1}{N} \sum_{k=1}^{N} \delta_{z_k}$. To obtain good finite distributions Z, one normally evaluates and strives to minimize the supremum (extremal discrepancy) or average (e.g., L^2 discrepancy) of $|D(Z, A)|$ over some rich and well-structured collection of sets A. Typical examples of such collections include spherical caps, slices, convex sets, etc. – the specific choice depends on the problem at hand. For a good exposition of discrepancy and numerical integrations on the sphere, as well as discrepancy theory in general, the reader is referred to, e.g., [14,26].

On the other hand, the energy of Z with respect to a function $F : \mathbb{S}^d \times \mathbb{S}^d \to \mathbb{R}$ is defined as

$$E_F(Z) = \frac{1}{N^2} \sum_{i,j=1}^{N} F(z_i, z_j);$$

i.e., points of Z are viewed as interacting particles on the sphere that repel according to the potential given by F. If $F(x, x)$ is undefined, which is the case for the most common example arising in electrostatics, the Riesz potential $F(x, y) = \|x - y\|^{-s}$, the diagonal terms are omitted in the sum above. In many situations, minimizing (or maximizing, depending on the structure of F) the energy $E_F(Z)$ yields well-distributed point-sets on the sphere, and the quality of this distribution may be measured by the difference of the discrete energy $E_F(Z)$ and the energy of the continuous uniform distribution $I_F(\sigma) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x, y) d\sigma(x) d\sigma(y)$. Vast literature exists on problems of this nature: we refer the reader, e.g., to the upcoming book [10].

It is known that in some cases, these two ways of quantifying equidistribution are closely connected. One of the first instances of such a connection was obtained in 1973 by Stolarsky [32], who proved that minimizing the L^2 discrepancy with respect to spherical caps is equivalent to maximizing the pairwise sum of Euclidean distances; i.e., $E_F(Z)$ with $F(x, y) = \|x - y\|$. More precisely, he established the identity

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

which came to be known as the *Stolarsky invariance principle* (see Sect. 2 for more details).

In this paper, we further explore these connections and obtain various versions of the Stolarsky principle in different settings, yielding new relations between discrepancy and energy optimization.

In Sect. 2, we revisit the classical Stolarsky invariance principle and give a very simple elementary proof of this identity.

In Sect. 3, we observe that by replacing *all* spherical caps with *hemispheres* (i.e., spherical caps with aperture $\pi/2$), one obtains a variant of the Stolarsky principle with the geodesic distance $d(x, y)$ in place of the Euclidean distance, Theorem 3.1. This allows one to easily characterize finite point sets on $\mathbb{S}^d$ which maximize the sum of geodesic distances in all dimensions $d \geq 1$ (for even N these are just symmetric sets), and points to a drastic difference with the case of Euclidean distances, see Theorem 3.2.

In Sect. 4, we take this idea one step further and show that an analog of the Stolarsky principle holds for general probability measures μ in place of the counting measure (Theorem 4.1). This provides a way to characterize the maximizers of the geodesic distance energy integral

$$I_g(\mu) = \int_{\mathbb{S}^d}\int_{\mathbb{S}^d} \big(d(x,y)\big)^\delta d\mu(x)d\mu(y),$$

with $\delta > 0$, over all probability measures on the sphere, Theorem 4.7. In particular,

- for $0 < \delta < 1$, the unique maximizer is σ;
- for $\delta = 1$, maximizers are centrally symmetric measures;
- for $\delta > 1$, maximizers are measures of the form $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$.

The second part of this statement is a consequence of the Stolarsky principle. The case $\delta \in (0, 1)$ is actually proved in the companion paper of the authors [5], which also studies the cases of $\delta < 0$ (geodesic Riesz energy) and $\delta = 0$ (logarithmic energy) by means of analyzing ultraspherical expansions. This brings up a surprising difference between the geodesic and Euclidean settings. In the latter case, a result of Björck [8] from 1955, see Theorem 4.6, states that the critical value is $\delta = 2$ rather than $\delta = 1$. This effect in dimension $d = 1$ has been previously noticed in [12].

In Sect. 5, we explore the connections between energy optimization and discrepancy on a more general level. We show that for positive definite functions F, one can define a natural notion of discrepancy, so that the analog of the Stolarsky principle holds for general measures μ, Theorem 5.6, and make connections between positive definiteness and refinements of this property and minimization of energy integrals

$$I_F(\mu) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) d\mu(x) d\mu(y),$$

in particular, the question whether σ is a unique minimizer of $I_F(\mu)$, in other words, whether attaining equilibrium under the potential F imposes uniform distribution. These results are further developed and applied in the parallel paper of the authors [5], yielding sharp asymptotic estimates for the difference $E_F(Z) - I_F(\sigma)$ as well as new proofs of classical discrepancy bounds (2.2).

In Sect. 6, we consider discrepancy with respect to spherical "slices" (i.e., intersections of half-spaces, previously studied in [9]) and spherical "wedges" (i.e., symmetrized slices – this notion of discrepancy came up recently [7] in connection with problems of uniform tessellation of the sphere by hyperplanes and one-bit compressed sensing). In these cases, one also obtains Stolarsky-type identities with potentials $F(x, y) = \left(1 - d(x, y)\right)^2$ and $F(x, y) = \left(\frac{1}{2} - d(x, y)\right)^2$, respectively.

Finally, in the appendix, Sect. 7, we compute the values of some of the spherical integrals that arise in the exposition.

We note that the basic strategy behind most versions of the Stolarsky principle, at a very low level, is straightforward. Computing the L^2 discrepancy, one squares out the expression in (1.1), thus pairwise interactions between points of Z arise from cross terms of the form $\mathbf{1}_A(z_i) \cdot \mathbf{1}_A(z_j)$. When integrated over the test sets A in a given class, this yields the interaction potential $F(z_i, z_j)$, which is often represented as the volume of intersection of test sets "centered" at z_i and z_j, see, e.g., (2.5); however, the details in some settings get rather technical. This approach is employed in Sects. 2–4 and 6. A similar idea has been used by Torquato [34] for "number variance," a quantity very similar to L^2 discrepancy. In Sect. 5 we go in the opposite direction and show that for any positive definite interaction potential, one can construct an appropriate notion of discrepancy, so that the Stolarsky principle holds.

We would also like to mention that the interest in the Stolarsky principle in different settings has recently spiked: [13] studied it from the point of view of numerical integration on the sphere, [21] uses it in applications to genomics, [30] explores the Stolarsky principle in general metric spaces, [7] connects it to tessellations of the sphere, while the present paper and [5] deal with it in the context of energy optimization.

In the text, the dimension of the sphere is $d \geq 1$, i.e., $\mathbb{S}^d \subset \mathbb{R}^{d+1}$; σ is the *normalized* Lebesgue (d-dimensional Hausdorff) measure on $\mathbb{S}^d$, i.e., $\sigma(\mathbb{S}^d) = 1$; $\|x\|$ denotes the Euclidean norm of $x \in \mathbb{R}^{d+1}$; and $d(x, y)$ denotes the geodesic distance between x and $y \in \mathbb{S}^d$, *normalized* so that the distance between antipodal poles is equal to 1, i.e., $d(x, y) = \frac{1}{\pi} \cos^{-1}(x \cdot y)$. The cardinality of a finite set Z is denoted by $\#Z$. The set of all finite signed Borel measures on $\mathbb{S}^d$ is denoted by $\mathcal{B}$, and $\mathcal{M}$ stands for the set of probability measures on $\mathbb{S}^d$ (positive Borel measures with total mass one). Further explanations, background information, and references will be given in each section.

2 Classical Stolarsky Invariance Principle for Spherical Caps

We consider "spherical caps" $C(x, t)$ with center $x \in \mathbb{S}^d$ and "height" $t \in [-1, 1]$; i.e.,

$$C(x, t) = \{z \in \mathbb{S}^d : z \cdot x > t\}.$$

We define the L^2 discrepancy of Z with respect to spherical caps:

$$[D_{L^2,\text{cap}}(Z)]^2 = \int\limits_{-1}^{1} \int\limits_{\mathbb{S}^d} \left| \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{C(x,t)}(z_j) - \sigma\big(C(x,t)\big) \right|^2 d\sigma(x)\, dt$$

$$= \int\limits_{-1}^{1} [D_{L^2,\text{cap}}^{(t)}(Z)]^2\, dt.$$

The following result was proved by Stolarsky in 1973 [32]:

Theorem 2.1 *(Stolarsky invariance principle) Let $Z = \{z_1, z_2, \ldots, z_N\} \subset \mathbb{S}^d$. Then the following relation holds:*

$$[D_{L^2,\text{cap}}(Z)]^2 = C_d \left(\int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} \|x - y\|\, d\sigma(x)\, d\sigma(y) - \frac{1}{N^2} \sum_{i,j=1}^{N} \|z_i - z_j\| \right). \quad (2.1)$$

The constant C_d satisfies

$$C_d = \frac{1}{2} \int_{\mathbb{S}^d} |p \cdot z|\, d\sigma(z) = \frac{1}{d} \frac{\omega_{d-1}}{\omega_d} = \frac{v_d}{\omega_d}$$

$$= \frac{1}{d} \frac{\Gamma\big((d+1)/2\big)}{\sqrt{\pi}\, \Gamma(d/2)} \sim \frac{1}{\sqrt{2\pi d}} \quad \text{as} \ \ d \to \infty,$$

where ω_d is the surface area of $\mathbb{S}^d$, v_d is the volume of the unit ball in $\mathbb{R}^d$, and p is an arbitrary point on the sphere $\mathbb{S}^d$.

This theorem states that:

- minimizing the L^2 spherical cap discrepancy of Z is equivalent to maximizing the sum of Euclidean distances between the points of Z;
- the L^2 spherical cap discrepancy can be realized as the difference between the continuous and discrete energies or, equivalently, the error of numerical integration of the distance integral $\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \|x - y\|\, d\sigma(x)\, d\sigma(y)$ by the cubature formula with knots at the points of Z.

It is well known [2,3] that the optimal order of the L^2 spherical cap discrepancy is $N^{-\frac{1}{2} - \frac{1}{2d}}$; i.e.,

$$c_d N^{-\frac{1}{2} - \frac{1}{2d}} \leq \inf_{\#Z = N} D_{L^2,\text{cap}}(Z) \leq c_d' N^{-\frac{1}{2} - \frac{1}{2d}}, \quad (2.2)$$

which in turn bounds the difference of continuous and discrete energies in (2.1).

In addition to the original proof in [32], an alternative proof has been given in [13]. Here we present a new short and simple proof of the Stolarsky invariance principle (2.1). It strongly resonates with the proof in [13], but is completely elementary in nature. A similar proof in a probabilistic interpretation has been independently given

in [21] (compare Lemmas 2.2 and 2.3 below to Proposition 1 of [21]), and analogous ideas are used in [30] on general metric spaces.

The proof of (2.1) follows the aforementioned strategy: one squares out the integrand, and the discrete part (pairwise interactions) arises naturally from the cross terms. The important ingredient is the following relation between intersections of spherical caps and the Euclidean distance between their centers:

Lemma 2.2 *For arbitrary $x, y \in \mathbb{S}^d$, we have*

$$\int_{-1}^{1} \sigma\big(C(x,t) \cap C(y,t)\big)\, dt = 1 - C_d \|x - y\|, \tag{2.3}$$

where the constant C_d is given by $C_d = \frac{1}{2} \int_{\mathbb{S}^d} |p \cdot z|\, d\sigma(z)$ for any fixed point $p \in \mathbb{S}^d$.

Proof Recall that σ is normalized so that $\sigma(\mathbb{S}^d) = 1$. We have

$$\int_{-1}^{1} \sigma\big(C(x,t) \cap C(y,t)\big)\, dt = \int_{-1}^{1} \int_{\mathbb{S}^d} \mathbf{1}_{C(x,t)}(z) \cdot \mathbf{1}_{C(y,t)}(z) d\sigma(z)\, dt$$

$$= \int_{\mathbb{S}^d} \int_{-1}^{1} \mathbf{1}_{C(z,t)}(x) \cdot \mathbf{1}_{C(z,t)}(y)\, dt\, d\sigma(z)$$

$$= \int_{\mathbb{S}^d} \int_{-1}^{\min\{x \cdot z, y \cdot z\}} dt\, d\sigma(z)$$

$$= \int_{\mathbb{S}^d} \big(\min\{x \cdot z, y \cdot z\} + 1\big)\, d\sigma(z).$$

We now write $\min\{x \cdot z, y \cdot z\} = \frac{1}{2}\big(x \cdot z + y \cdot z - |(x - y) \cdot z|\big)$. Obviously $\int_{\mathbb{S}^d} x \cdot z\, d\sigma(z) = \int_{\mathbb{S}^d} y \cdot z\, d\sigma(z) = 0$. By rotational invariance, we observe that

$$\int_{\mathbb{S}^d} |(x - y) \cdot z|\, d\sigma(z) = \|x - y\| \cdot \int_{\mathbb{S}^d} \left| \frac{x - y}{\|x - y\|} \cdot z \right| d\sigma(z) = 2C_d \|x - y\|,$$

where $C_d = \frac{1}{2} \int_{\mathbb{S}^d} |p \cdot z|\, d\sigma(z)$, and this finishes the proof. $\square$

Here we essentially repeated the proof from [13], but the proof of the next lemma, which gives the quadratic mean value of the size of the spherical caps, is simpler (does not use reproducing kernels).

Lemma 2.3 *For any $p \in \mathbb{S}^d$, we have*

$$\int\limits_{-1}^{1} \left(\sigma\big(C(p,t)\big)\right)^2 dt = 1 - C_d \int\limits_{\mathbb{S}^d} \|x - p\| \, d\sigma(x). \tag{2.4}$$

Proof It is clear that

$$\int\limits_{\mathbb{S}^d} \|x - p\| \, d\sigma(x) = \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} \|x - y\| \, d\sigma(x) \, d\sigma(y).$$

We use the result of the previous lemma, i.e., relation (2.3), to compute

$$1 - C_d \int\limits_{\mathbb{S}^d} \|x - p\| \, d\sigma(x) = \int\limits_{\mathbb{S}^d} \big(1 - C_d \|x - p\|\big) \, d\sigma(x)$$

$$= \int\limits_{\mathbb{S}^d} \int\limits_{-1}^{1} \int\limits_{\mathbb{S}^d} \mathbf{1}_{C(x,t)}(z) \cdot \mathbf{1}_{C(p,t)}(z) d\sigma(z) \, dt \, d\sigma(x)$$

$$= \int\limits_{-1}^{1} \int\limits_{\mathbb{S}^d} \mathbf{1}_{C(p,t)}(z) \left(\int\limits_{\mathbb{S}^d} \mathbf{1}_{C(z,t)}(x) \, d\sigma(x) \right) d\sigma(z) \, dt$$

$$= \int\limits_{-1}^{1} \left(\sigma\big(C(p,t)\big)\right)^2 dt.$$

$\square$

This is one of numerous examples of a situation in which averaging over scales simplifies things. For the L^2 discrepancy for spherical caps of fixed height t:

$$D_{L^2,\mathrm{cap}}^{(t)}(Z) := \left(\int_{\mathbb{S}^d} \left| \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{C(x,t)}(z_j) - \sigma\big(C(x,t)\big) \right|^2 d\sigma(x) \right)^{1/2},$$

one would have to deal with $\sigma\big(C(x,t) \cap C(y,t)\big)$, which has complicated structure, and no short relation akin to (2.3) is available, see, e.g., [21]. Hence in this case, there is no formula as succinct and explicit as the Stolarsky principle; however, one can still write down a generic relation where the interactions between z_j's would involve $\sigma\big(C(z_i,t) \cap C(z_j,t)\big)$.

Proposition 2.4 *For any $Z = \{z_1, z_2, \ldots, z_N\} \subset \mathbb{S}^d$ and a fixed $t \in [-1,1]$, the following relation holds:*

$$\left[D_{L^2,\text{cap}}^{(t)}(Z)\right]^2 = \frac{1}{N^2}\sum_{i,j=1}^{N}\sigma\big(C(z_i,t)\cap C(z_j,t)\big) - \big(\sigma\big(C(p,t)\big)\big)^2, \qquad (2.5)$$

where $p \in \mathbb{S}^d$ is arbitrary.

Proof We note that $\sigma\big(C(x,t)\big)$ is independent of $x \in \mathbb{S}^d$; hence

$$\left[D_{L^2}^{(t)}(Z)\right]^2 = \int_{\mathbb{S}^d}\left|\frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_{C(x,t)}(z_j) - \sigma\big(C(x,t)\big)\right|^2 d\sigma(x)$$

$$= \frac{1}{N^2}\sum_{i,j=1}^{N}\int_{\mathbb{S}^d}\mathbf{1}_{C(x,t)}(z_i)\cdot\mathbf{1}_{C(x,t)}(z_j)d\sigma(x)$$

$$-\frac{2}{N}\int_{\mathbb{S}^d}\sum_{j=1}^{N}\mathbf{1}_{C(z_j,t)}(x)\cdot\sigma\big(C(x,t)\big)\,d\sigma(x) + \big(\sigma\big(C(p,t)\big)\big)^2$$

$$= \frac{1}{N^2}\sum_{i,j=1}^{N}\sigma\big(C(z_i,t)\cap C(z_j,t)\big) - \big(\sigma\big(C(p,t)\big)\big)^2.$$

□

Integrating identity (2.5) with respect to t and applying relations (2.3) and (2.4), we finish the proof of the Stolarsky principle (2.1).

3 Stolarsky Principle for Hemispheres

An (open) hemisphere in the direction of $x \in \mathbb{S}^d$ is simply a spherical cap of height $t = 0$:
$$H(x) = \{z \in \mathbb{S}^d : z \cdot x > 0\} = C(x,0).$$
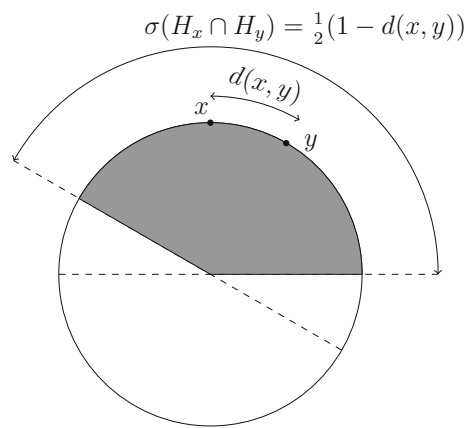Since $\sigma\big(H(x)\big) = \frac{1}{2}$, the natural L^2 discrepancy for this set system is

$$D_{L^2,\text{hem}}(Z) := \left(\int_{\mathbb{S}^d}\left|\frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_{H(x)}(z_j) - \frac{1}{2}\right|^2 d\sigma(x)\right)^{1/2} = D_{L^2,\text{cap}}^{(0)}(Z). \qquad (3.1)$$

While, as mentioned above, generally the quantity $\sigma\big(C(x,t)\cap C(y,t)\big)$ is complicated, in the case $t = 0$ (hemispheres) it has a very simple representation: for $x, y \in \mathbb{S}^d$,

$$\sigma\big(H(x)\cap H(y)\big) = \sigma\big(C(x,0)\cap C(y,0)\big) = \frac{1}{2}\cdot\big(1 - d(x,y)\big), \qquad (3.2)$$

where $d(x,y)$ is the normalized geodesic distance on the sphere between x and y. This can be very easily seen from Figure 1.

Fig. 1 The size $\sigma(H_x \cap H_y)$ of the intersection of two hemispheres depends linearly on the geodesic distance $d(x, y)$

$$\sigma(H_x \cap H_y) = \tfrac{1}{2}(1 - d(x, y))$$

Combining (3.2), Proposition 2.4, and the obvious fact that $\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x, y)\, d\sigma(x)\, d\sigma(y)$ $= \frac{1}{2}$, one arrives at the following result.

Theorem 3.1 *(Stolarsky principle for hemispheres) For any N-point set $Z \in \mathbb{S}^d$, the following relation holds:*

$$[D_{L^2,\mathrm{hem}}(Z)]^2 = \frac{1}{2}\left(\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x, y)\, d\sigma(x)\, d\sigma(y) - \frac{1}{N^2} \sum_{i,j=1}^{N} d(z_i, z_j) \right). \quad (3.3)$$

The statement looks strikingly similar to the original Stolarsky principle (2.1). One can say that the Euclidean distance corresponds to the *mean* over $t \in [-1, 1]$, while the geodesic distance corresponds to the *median* ($t = 0$) of the heights of the spherical caps. Despite the fact that the original Stolarsky principle was proved in 1973, this version is new and has not been observed before. (At the time of preparation of this manuscript, we have learned that this version of the Stolarsky principle has been independently and simultaneously proved by Skriganov [30].) Relation (3.3) has several interesting features and consequences.

First of all, the principle of *irregularities of distribution* does not hold in this situation; that is, the hemisphere discrepancy can be very small, even zero, for large N. Indeed, for any symmetric distribution Z, it is easy to see that the L^2 hemisphere discrepancy is equal to zero. Moreover, (3.3) allows us to characterize finite point distributions in $\mathbb{S}^d$, which maximize the sum of geodesic distances.

Theorem 3.2 *Let $d \geq 1$. Then the following holds:*

(i) For any point distribution $Z = \{z_1, \ldots, z_N\} \subset \mathbb{S}^d$,

$$\frac{1}{N^2} \sum_{i,j=1}^{N} d(z_i, z_j) \leq \frac{1}{2}.$$

(ii) *For a given $N \in \mathbb{N}$, the sum above is maximized if and only if the following condition holds: for any $x \in \mathbb{S}^d$, such that the hyperplane $x^\perp$ contains no points of Z, the numbers of points of Z on either side of $x^\perp$ differ by at most one; i.e.,*

$$\left| \#\big(Z \cap H(x)\big) - \#\big(Z \cap H(-x)\big) \right| \le 1.$$

(iii) *If N is even,*

$$\max_{\#Z=N} \frac{1}{N^2} \sum_{i,j=1}^{N} d(z_i, z_j) = \frac{1}{2},$$

and this maximum is achieved if and only if Z is a centrally symmetric set.

(iv) *If N is odd,*

$$\max_{\#Z=N} \frac{1}{N^2} \sum_{i,j=1}^{N} d(z_i, z_j) = \frac{1}{2} - \frac{1}{2N^2},$$

and this maximum is achieved if and only if Z can be represented as a union $Z = Z_1 \cup Z_2$, where Z_1 is symmetric, while Z_2 lies on a two-dimensional hyperplane (i.e. on the same great circle) and satisfies $\frac{1}{M^2} \sum_{z_i, z_j \in Z_2} d(z_i, z_j) = \frac{1}{2} - \frac{1}{2M^2}$ with $M = \#Z_2$; i.e., Z_2 is a maximizer of the sum of geodesic distances on $\mathbb{S}^1$.

Before we turn to the proof of the theorem, we briefly discuss the history of these questions. These results have been previously known in dimensions $d = 1$ and $d = 2$. Parts (i) and (ii) in $d = 1$ were proved by Fejes-Tóth [18] (and reproved later in [23] in relation to musical rhythms). Fejes-Tóth also conjectured that the same holds for $d \ge 2$. In dimension $d = 2$ for even N, part (iii) of the theorem above was proved by Sperling [31]. Our Stolarsky principle (3.3) makes the proof of this case very simple in all dimensions $d \ge 2$. For odd N, Larcher [25] proved part (iv) of Theorem 3.2 in dimension $d = 2$; however, we believe that there is a mistake in his proof (statement (ii) at the bottom of page 48). We use a different geometric approach to prove (iv) in *all dimensions $d \ge 2$*, based on part (ii) of Theorem 3.2 and an interesting fact from combinatorial geometry known as the Sylvester–Gallai theorem.

Thus, our Stolarsky-type formula (3.3) greatly simplifies the proof of these facts in dimensions one and two and, moreover, allows us to extend them to all dimensions $d \ge 2$. We now turn to the proof of Theorem 3.2.

Proof Part (i) of the theorem is now obvious since the left-hand side of (3.3) is non-negative and $\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x, y) \, d\sigma(x) \, d\sigma(y) = \frac{1}{2}$.

Part (ii) also follows easily from (3.3). Indeed, for every $x \in \mathbb{S}^d$ such that $x^\perp$ does not contain any points of Z, the minimal value of the integrand in the left-hand side of (3.3), i.e., the integrand in (3.1), equals 0 for even N (if exactly half the points lie on either side of $x^\perp$) and is $\frac{1}{4N^2}$ for odd N (if the numbers of points on both sides of $x^\perp$

differ exactly by 1). Obviously, configurations Z for which this is achieved for each such $x \in \mathbb{S}^d$ are possible: e.g., $\lfloor N/2 \rfloor$ and $\lceil N/2 \rceil$ points in antipodal poles. Moreover, if for some $x \in \mathbb{S}^d$ with $x^\perp \cap Z = \emptyset$ this condition is not satisfied, then it also fails on a small set of positive measure around x; hence $D_{L^2,\text{hem}}(Z)$ is not minimal, and therefore $\frac{1}{N^2} \sum d(z_i, z_j)$ is not maximized.

To prove part (iii), first observe that symmetric sets Z trivially satisfy the condition of part (ii). Now assume that for some $x \in \mathbb{S}^d$, the number of points of Z located at x and $-x$ is not the same. Consider a hyperplane passing through x, which contains no other points of Z. Perturbing it in opposite directions, we find that the difference of number of points on either side changes by at least 2, i.e., cannot stay equal to zero. Thus nonsymmetric sets Z with even number of points don't satisfy the condition of part (ii), i.e., cannot maximize the sum of geodesic distances.

We now turn to part (iv). We shall rely on the Sylvester–Gallai theorem. In the Euclidean case, it states the following: *if a finite set Z in $\mathbb{R}^d$ has the property that for every two points of Z, the straight line passing through them contains at least one other point of Z, then all points of Z lie on the same straight line*. A spherical version of this theorem also holds.

Theorem 3.3 *(Spherical Sylvester–Gallai Theorem) Assume that a set Z of N points on the sphere $\mathbb{S}^d$ contains no antipodal points and satisfies the following condition: for every two points of Z, the great circle passing through them contains at least one more point of Z. Then all points of Z lie on the same great circle.*

For the history and several proofs of these theorems, we refer the reader to the book [1, pages 73 and 88]. Normally, these theorems are stated in dimension $d = 2$, but higher dimensional extensions are simple. Indeed, for $Z \subset \mathbb{S}^d$, consider a copy of $\mathbb{S}^2$ that contains $z_1, z_2, z_3 \in Z$. The two-dimensional version of Theorem 3.3 applies, and hence z_1, z_2, z_3 lie on the same great circle. In the same manner, considering a copy of $\mathbb{S}^2$ containing this great circle and any other point $z_i \in Z$, we find that z_i has to lie on the same great circle.

We are now ready to prove part (iv). Assume that N is odd. It follows from (3.3) and the proof of part (ii) that the maximal value of $\frac{1}{N^2} \sum d(z_i, z_j)$ is $\frac{1}{2} - \frac{1}{2N^2}$. Observe that adding a pair of antipodal points to Z does not change maximality of Z; i.e., Z is a maximizer if and only if $Z \cup \{p, -p\}$ is a maximizer (with N replaced by $N + 2$). Indeed, since $d(x, p) + d(x, -p) = 1$, it is easy to check that

$$\sum_{x,y \in Z \cup \{-p,p\}} d(x, y) = \sum_{x,y \in Z} d(x, y) + 2(N + 1);$$

thus the second sum equals $\frac{N^2}{2} - \frac{1}{2}$ if and only if the first sum is $\frac{(N+2)^2}{2} - \frac{1}{2}$. This immediately proves sufficiency of the condition in (iv). Moreover, it shows that, in order to prove necessity, it is enough to consider maximizers without antipodal points and to prove that they have to be contained in some great circle.

Assume that Z with $\#Z = N$ maximizes $\frac{1}{N^2} \sum d(z_i, z_j)$ and contains no pair of antipodal points. Consider two arbitrary points $z_1, z_2 \in Z$, and assume that no other point of Z lies on the great circle defined by z_1 and z_2. Since Z is finite, there exists a

hyperplane containing z_1 and z_2, which does not contain any other points of Z. Since z_1 and z_2 are not antipodal, one can perturb the hyperplane in such a way that it does not touch other points of Z and both points z_1 and z_2 end up on the same side of the hyperplane. Perturbing in the opposite direction, we observe that the difference between the numbers of points on opposite sides of the hyperplane changes by 4, i.e., cannot stay equal to ± 1; i.e., by part (ii), Z cannot be a maximizer.

We thus conclude that, for any two points of Z, at least one other point of Z has to lie on the same great circle; i.e., the spherical Sylvester–Gallai theorem, Theorem 3.3, applies. Hence Z is contained in a great circle. □

Remark Observe that the one-dimensional maximizers of odd cardinality N, which arise in part (iv) of Theorem 3.2, are characterized by the condition that the sum of any $\lceil N/2 \rceil$ consecutive central angles defined by the points is at least π. In particular, any acute triangle is a maximizer for $d = 1$ and $N = 3$.

Remark While the paper was in preparation, we learned that on $\mathbb{S}^2$, Theorem 3.2 has been proved by Nielsen [27] (the paper is written in Danish). In the case of odd N, the argument in [27] also exploits the spherical version of Sylvester–Gallai theorem, but the main methods, which come from projective geometry, are different.

Theorem 3.2 demonstrates that the situation is drastically different from the spherical cap discrepancy and the sum of Euclidean distances. In the latter case, minimizing the L^2 spherical cap discrepancy (equivalently, maximizing the sum of Euclidean distances) leads to a rather uniform distribution of Z. In particular, for $d = 1$, the sum is maximized by the vertices of a regular N-gon [17], and in higher dimensions, maximizing distributions have to be well separated [33]. The sum of *geodesic* distances, however, may be maximized by very nonuniform sets, e.g., $N/2$ points in two antipodal poles.

4 Geodesic Distance Energy Integral

The results of the previous section naturally suggest a more general problem of finding equilibrium distributions for the geodesic energy integral. Let μ, ν be Borel measures on $\mathbb{S}^d$. Define the geodesic distance energy integrals as

$$I(\mu, \nu) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x, y) \, d\mu(x) d\nu(y); \quad I(\mu) = I(\mu, \mu)$$
$$= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} d(x, y) \, d\mu(x) d\mu(y).$$

Let $\mathcal{M}$ denote the set of non-negative Borel probability measures on $\mathbb{S}^d$ (i.e., $\mu(\mathbb{S}^d) = 1$). We shall be interested in the quantity

$$M = \sup_{\mu \in \mathcal{M}} I(\mu),$$

as well as the maximizers of this expression, i.e. the measures $\mu \in \mathcal{M}$ for which $I(\mu) = M$ (the existence of maximizers follows easily from the weak-star-compactness of $\mathcal{M}$). The hemisphere Stolarsky principle, Theorem 3.1, may be extended to more general measures than the counting measure $\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i}$.

Theorem 4.1 *(Hemisphere Stolarsky principle for general measures) Let μ be a Borel measure on $\mathbb{S}^d$ with $\mu(\mathbb{S}^d) = 1$. Then the following relation holds:*

$$\int_{\mathbb{S}^d} \left(\mu\big(H(x)\big) - \frac{1}{2} \right)^2 d\sigma(x) = \frac{1}{2} \cdot \left(\frac{1}{2} - I(\mu) \right). \tag{4.1}$$

Proof Notice that

$$\int_{\mathbb{S}^d} \int_{H(x)} d\mu(y) d\sigma(x) = \int_{\mathbb{S}^d} \int_{H(y)} d\sigma(x) d\mu(y) = \frac{1}{2} \cdot \int_{\mathbb{S}^d} d\mu(y) = \frac{1}{2},$$

and, according to (3.2),

$$\int_{\mathbb{S}^d} \int_{H(x)} \int_{H(x)} d\mu(y) d\mu(z) d\sigma(x) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \sigma\big(H(y) \cap H(z)\big) d\mu(y) d\mu(z)$$

$$= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \frac{1}{2} \cdot \big(1 - d(y, z)\big) d\mu(y) d\mu(z) = \frac{1}{2} - \frac{1}{2} I(\mu).$$

Using the two relations above, we obtain

$$\int_{\mathbb{S}^d} \left(\mu\big(H(x)\big) - \frac{1}{2} \right)^2 d\sigma(x) = \int_{\mathbb{S}^d} \left(\int_{H(x)} \int_{H(x)} d\mu(y) d\mu(z) - \int_{H(x)} d\mu(y) + \frac{1}{4} \right) d\sigma(x)$$

$$= \frac{1}{2} - \frac{1}{2} I(\mu) - \frac{1}{2} + \frac{1}{4} = \frac{1}{2} \cdot \left(\frac{1}{2} - I(\mu) \right),$$

which proves (4.1). □

Since the left-hand side of identity (4.1) is non-negative, Theorem 4.1 immediately yields a corollary about the maximizers of $I(\mu)$:

Corollary 4.2 *For any $\mu \in \mathcal{M}$, $I(\mu) \leq \frac{1}{2}$. Measures $\mu \in \mathcal{M}$, for which $I(\mu) = \frac{1}{2}$, are exactly the measures that satisfy the following condition:*

$$\mu\big(H(x)\big) = \frac{1}{2} \quad \text{for } \sigma\text{-a.e. } x \in \mathbb{S}^d. \tag{4.2}$$

It is very easy to see that if the measure μ is symmetric, it is a maximizer of $I(\mu)$; i.e., $I(\mu) = \frac{1}{2}$. Indeed, let μ^* be the reflection of μ; i.e., $\mu^*(E) = \mu(-E)$. It is easy to see that

$$I(\mu, \mu^*) = \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} d(x, y) d\mu(x) d\mu^*(y) = \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} d(x, -y) d\mu(x) d\mu(y)$$

$$= \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} \left(1 - d(x, y)\right) d\mu(x) d\mu(y) = 1 - I(\mu).$$

If, moreover, μ is symmetric, i.e., $\mu^* = \mu$, then

$$I(\mu) = I(\mu, \mu^*) = 1 - I(\mu) \implies I(\mu) = \frac{1}{2}.$$

Therefore, in particular, every symmetric measure $\mu \in \mathcal{M}$ satisfies (4.2). The converse of this fact is less obvious.

Proposition 4.3 *Assume that the measure* $\mu \in \mathcal{M}$ *satisfies the condition*

$$\mu\big(H(x)\big) = \frac{1}{2} \quad \text{for } \sigma\text{-a.e. } x \in \mathbb{S}^d. \tag{4.3}$$

Then the measure μ *is symmetric, i.e.,* $\mu(E) = \mu(-E)$ *for every Borel set* $E \subset \mathbb{S}^d$.

We are unaware of an elementary proof of this seemingly simple statement. Our approach is based on spherical harmonics and Gegenbauer polynomials. We refer the reader to [15,20] for background information. Let $w_\lambda(t) = (1 - t^2)^{\lambda - \frac{1}{2}}$ with $\lambda > 0$. Given $1 \le p < \infty$, we denote by $L_{w_\lambda}^p[-1, 1]$ the space of all real integrable functions f on $[-1, 1]$ with $\|f\|_{p,\lambda} := \left(\int_{-1}^1 |f(t)|^p w_\lambda(t) \, dt\right)^{1/p} < \infty$. Every function $f \in L_{w_\lambda}^1[-1, 1]$ has an expansion in terms of Gegenbauer (ultraspherical) polynomials C_n^λ:

$$f(t) \sim \sum_{n=0}^\infty \widehat{f}(n, \lambda) \frac{n + \lambda}{\lambda} C_n^\lambda(t), \quad t \in [-1, 1].$$

Let $\mathcal{H}_n(\mathbb{S}^d)$ denote the space of spherical harmonics of degree n on $\mathbb{S}^d$, i.e., homogeneous harmonic polynomials of degree n in $d + 1$ variables. We start with an auxiliary lemma, which will also be useful in Sect. 5.

Lemma 4.4 *Let* γ *be a signed Borel measure on* $\mathbb{S}^d$ *and* $f \in L_{w_\lambda}^2[-1, 1]$ *with* $\lambda = \frac{d-1}{2}$. *Assume that*

$$\int_{\mathbb{S}^d} f(x \cdot y) \, d\gamma(y) = 0 \text{ for } \sigma\text{-almost every } x \in \mathbb{S}^d. \tag{4.4}$$

Assume also that $\widehat{f}(n, \lambda) \ne 0$ *for some* $n \ge 1$. *Then for every spherical harmonic* $Y_n \in \mathcal{H}_n(\mathbb{S}^d)$,

$$\int_{\mathbb{S}^d} Y_n(y) d\gamma(y) = 0. \tag{4.5}$$

Proof The Funk–Hecke formula (see, e.g., Theorem 1.2.9 in [15]) states that

$$\int_{\mathbb{S}^d} f(x \cdot y) Y_n(x) d\sigma(x) = \widehat{f}(n, \lambda) Y_n(y). \tag{4.6}$$

Therefore,

$$\int_{\mathbb{S}^d} Y_n(y) d\gamma(y) = \frac{1}{\widehat{f}(n, \lambda)} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} f(x \cdot y) Y_n(x) d\sigma(x) d\gamma(y)$$

$$= \frac{1}{\widehat{f}(n, \lambda)} \int_{\mathbb{S}^d} \left(\int_{\mathbb{S}^d} f(x \cdot y) d\gamma(y) \right) Y_n(x) d\sigma(x) = 0.$$

$\square$

Proof of Proposition 4.3 Let μ^* be the reflection of μ, defined by $\mu^*(E) = \mu(-E)$, and set $\gamma = \mu - \mu^*$. Condition (4.3) then implies that

$$\gamma\big(H(x)\big) = \int_{\mathbb{S}^d} \mathbf{1}_{(0,1]}(x \cdot y) \, d\gamma(y) = 0 \quad \text{for } \sigma\text{-a.e. } x \in \mathbb{S}^d.$$

The Gegenbauer coefficients $\widehat{f}(n, \lambda)$ of the function $f(t) = \mathbf{1}_{(0,1]}(t)$ are nonzero for odd n (Lemma 3.4.6 in [20]). Therefore, according to Lemma 4.4, relation (4.5) holds for all odd n. But for even values of n it obviously holds, since in this case $Y_n \in \mathcal{H}_n(\mathbb{S}^d)$ is an even function, and γ is antisymmetric. Therefore, $\int_{\mathbb{S}^d} f(y) d\gamma(y) = 0$ for every polynomial f and hence for each $f \in C(\mathbb{S}^d)$, which implies that $\gamma = 0$. Hence $\mu = \mu^*$; i.e., μ is symmetric. $\square$

From the above discussion, we obtain the following characterization of the maximizers of $I(\mu)$:

Theorem 4.5 *For a measure $\mu \in \mathcal{M}$,*

$$I(\mu) = \sup_{\gamma \in \mathcal{M}} I(\gamma) = \frac{1}{2}$$

if and only if μ is centrally symmetric.

This behavior of $I(\mu)$ is in sharp contrast to the behavior of the seemingly similar energy integral $\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \|x - y\| \, d\mu(x) d\mu(y)$. It is known [8] that the unique maximizer of this energy integral is $\mu = \sigma$, the uniform distribution on $\mathbb{S}^d$. In this sense, the behavior of $I(\mu)$ is more similar to (albeit still different from) that of $\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \|x - y\|^2 \, d\mu(x) d\mu(y)$ which is maximized by any measure with center of mass at the origin, which may be easily seen from the relation

$$\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \|x - y\|^2 \, d\mu(x) d\mu(y) = 2 - 2 \cdot \left\| \int_{\mathbb{S}^d} x d\mu(x) \right\|^2. \tag{4.7}$$

It is thus natural to analyze energy integrals with general powers $\delta > 0$.

4.1 Geodesic Distance Energy Integrals with Exponent $\delta > 0$

We would like to understand which measures $\mu \in \mathcal{M}$ maximize the energy

$$I(\mu) = \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} \left(d(x, y)\right)^\delta d\mu(x) d\mu(y)$$

for $\delta > 0$, and how maximizers depend on δ.

While the geodesic distance energy integral is a novel object, such integrals with Euclidean distances are well investigated. For an extensive study of the energy integrals $\int_F \int_F \|x - y\|^\delta d\mu(x) d\mu(y), \delta > 0$, see [8]. Specialized to the case $F = \mathbb{S}^d$, these results are formulated below.

Theorem 4.6 *(Bjorck, [8]) For $\delta > 0$, define the energy integral*

$$I_E(\mu) = \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} \|x - y\|^\delta d\mu(x) d\mu(y).$$

The maximizers of this energy integral over $\mu \in \mathcal{M}$ (Borel probability measures on $\mathbb{S}^d$) can be characterized as follows:

  (i) $0 < \delta < 2$: the unique maximizer of $I_E(\mu)$ is $\mu = \sigma$ (the normalized surface measure).
 (ii) $\delta = 2$: $I_E(\mu)$ is maximized if and only if the center of mass of μ is at the origin.
(iii) $\delta > 2$: $I_E(\mu)$ is maximized if and only if $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$; i.e., the mass is equally concentrated at two antipodal poles.

The proof of part (i) uses potential analysis, in particular, the semigroup property of the Riesz potentials; part (ii) is explained in (4.7); and part (iii) is almost trivial.

We observe that there is a "breaking point" $\delta = 2$ in the behavior of maximizers of the Euclidean energy integral. Surprisingly, for the seemingly similar geodesic distance integral, this critical value is different: $\delta = 1$. We have the following theorem:

Theorem 4.7 *For $\delta > 0$, let $I(\mu)$ be the geodesic distance energy integral*

$$I(\mu) = \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} \left(d(x, y)\right)^\delta d\mu(x) d\mu(y).$$

The maximizers of this energy integral over $\mu \in \mathcal{M}$ can be characterized as follows:

  (i) $0 < \delta < 1$: the unique maximizer of $I(\mu)$ is $\mu = \sigma$ (the normalized surface measure).
 (ii) $\delta = 1$: $I(\mu)$ is maximized if and only if μ is centrally symmetric.
(iii) $\delta > 1$: $I(\mu)$ is maximized if and only if $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$; i.e., the mass is equally concentrated at two antipodal poles.

Part (i) is proved in the companion paper of the authors [5] through extensive analysis of spherical harmonics expansions. Part (ii) is the result of Theorem 4.5 above, which is a consequence of the hemisphere Stolarsky principle (4.1). The proof of part (iii) is quite simple: since $d(x, y) \leq 1$, we have for $\delta > 1$,

$$I(\mu) \leq \int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} d(x, y) \, d\mu(x) d\mu(y) \leq \frac{1}{2}.$$

The first inequality turns into an equality when $\mu \times \mu \{(x, y) : d(x, y) = 0 \text{ or } 1\} = 1$, while the second bound becomes exact when μ is symmetric, according to part (ii). This readily implies that $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$.

This peculiar effect (that geodesic distance energy behaves differently from its Euclidean counterpart) has been noticed in dimension $d = 1$, i.e., on the circle, in [12], where the one-dimensional case of parts (i) and (ii) of the above theorem have been proved. In [5], the follow-up to the present paper, we conduct a more detailed analysis of the geodesic distance energy (including negative powers and logarithmic energy).

4.2 Average Case Integration Error

Finally, the right-hand side of the hemisphere Stolarsky principle (3.3) also yields the average-case squared integration error on $C(\mathbb{S}^d)$ with respect to the law of the *hemisphere Gaussian process* introduced in [6]. This is a mean-zero Gaussian process G on $\mathbb{S}^d$, which is defined by $\mathbb{E}G_x^2 = \frac{1}{4}$, $\mathbb{E}(G_x - G_y)^2 = \|\mathbf{1}_{H(x)} - \mathbf{1}_{H(x)}\|_2^2 = d(x, y)$; i.e., its covariance is given by $\mathbb{E}G_x G_y = \frac{1}{4} - \frac{1}{2}d(x, y)$. It induces a Gaussian measure on the space of continuous functions $C(\mathbb{S}^d)$, which we also denote by G. Then the average-case integration error with respect to G is equal to the L^2 hemisphere discrepancy.

Theorem 4.8 *Let $Z = \{z_1, z_2, \ldots, z_N\} \subset \mathbb{S}^d$. The following holds:*

$$\int\limits_{C(\mathbb{S}^d)} \left| \frac{1}{N} \sum_{i=1}^N f(z_i) - \int\limits_{\mathbb{S}^d} f(x) d\sigma(x) \right|^2 dG(f) = \frac{1}{2} \cdot \left(\frac{1}{2} - \frac{1}{N^2} \sum_{i,j=1}^N d(z_i, z_j) \right)$$

$$= [D_{L^2, \text{hem}}(Z)]^2.$$

Proof We note that

$$\int\limits_{C(\mathbb{S}^d)} f(x) f(y) dG(f) = \mathbb{E}G_x G_y = \frac{1}{4} - \frac{1}{2} d(x, y),$$

and thus

$$\int\limits_{C(\mathbb{S}^d)} \left(\int\limits_{\mathbb{S}^d} f(x)d\sigma(x) \right)^2 dG(f) = \int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \left(\frac{1}{4} - \frac{1}{2}d(x,y) \right) d\sigma(x)d\sigma(y) = 0;$$

i.e., $\int\limits_{\mathbb{S}^d} f(x)d\sigma(x) = 0$ for G-a.e. $f \in C(\mathbb{S}^d)$. Therefore, for the average case integration error, we obtain

$$\int\limits_{C(\mathbb{S}^d)} \left| \frac{1}{N} \sum_{i=1}^N f(z_i) - \int\limits_{\mathbb{S}^d} f(x)d\sigma(x) \right|^2 dG(f) = \int\limits_{C(\mathbb{S}^d)} \left| \frac{1}{N} \sum_{i=1}^N f(z_i) \right|^2 dG(f)$$

$$= \frac{1}{N^2} \sum_{i,j=1}^N \left(\frac{1}{4} - \frac{1}{2}d(z_i,z_j) \right)$$

$$= \frac{1}{2} \cdot \left(\frac{1}{2} - \frac{1}{N^2} \sum_{i,j=1}^N d(z_i,z_j) \right),$$

which is exactly the right-hand side of the hemisphere Stolarsky principle (3.3). □

We remark that the first result of this type has been obtained in [36] for the anchored L^2 discrepancy on $[0, 1]^d$ and the average-case integration error with respect to the Wiener sheet measure.

5 Positive Definite Functions and Generalized Stolarsky Principle

In this section, we take a more general look at energy minimization and the Stolarsky principle. For any Borel measure μ on $\mathbb{S}^d$ and a bounded or non-negative Borel measurable function F on $[-1, 1]$, we define the energy integral

$$I_F(\mu) := \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) \, d\mu(x) \, d\mu(y).$$

As before, let $\mathcal{B}$ denote the class of finite Borel signed measures on $\mathbb{S}^d$.

We start with a simple observation that shows that, while $I_F(\mu)$ is quadratic in μ, it behaves linearly near σ.

Lemma 5.1 *For any bounded or non-negative Borel measurable function F on $[-1, 1]$ and any signed measure $\mu \in \mathcal{B}$ with $\mu(\mathbb{S}^d) = 1$, the following relation holds:*

$$I_F(\mu) - I_F(\sigma) = I_F(\mu - \sigma). \tag{5.1}$$

Proof The result follows from the simple observation that, due to rotational invariance, for any $x \in \mathbb{S}^d$,

$$\int_{\mathbb{S}^d} F(x \cdot y) d\sigma(y) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) d\sigma(x) d\sigma(y) = I_F(\sigma);$$

i.e., the left-hand side is independent of $x \in \mathbb{S}^d$. Therefore,

$$I_F(\mu - \sigma) = I_F(\mu) + I_F(\sigma) - 2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) d\sigma(x) d\mu(y)$$

$$= I_F(\mu) + I_F(\sigma) - 2 I_F(\sigma) \int_{\mathbb{S}^d} d\mu(y) = I_F(\mu) - I_F(\sigma).$$

$\square$

Next, we recall the concept of positive definite functions on the sphere. A function $F \in C[-1, 1]$ is called *positive definite* on the sphere $\mathbb{S}^d$ if for any set of points $Z = \{z_1, ..., z_N\} \subset \mathbb{S}^d$, the matrix $\left[F(z_i \cdot z_j) \right]_{i,j=1}^{N}$ is positive semidefinite; i.e.,

$$\sum_{i,j} F(z_i \cdot z_j) c_i c_j \geq 0 \tag{5.2}$$

for all $c_i \in \mathbb{R}$. We denote the class of positive definite functions by Φ_d. This class admits several different characterizations.

Proposition 5.2 *For a function $F \in C[-1, 1]$, the following conditions are equivalent:*

(i) F is positive definite on $\mathbb{S}^d$; i.e., $F \in \Phi_d$.
(ii) For $\lambda = \frac{d-1}{2}$, all Gegenbauer coefficients of F are non-negative; i.e.,

$$\widehat{F}(n, \lambda) \geq 0 \ \text{ for all } n \geq 0.$$

(iii) For any signed measure $\mu \in \mathcal{B}$ the energy integral is non-negative: $I_F(\mu) \geq 0$.
(iv) There exists a function $f \in L^2_{w_\lambda}[-1, 1]$ such that

$$F(x \cdot y) = \int_{\mathbb{S}^d} f(x \cdot z) f(z \cdot y) \, d\sigma(z), \ \ x, y \in \mathbb{S}^d; \tag{5.3}$$

i.e., F is the spherical convolution of f with itself.

We shall briefly outline the proof of this proposition. The equivalence of (i) and (ii) is a celebrated theorem of Schoenberg [29]. In addition, it is known that the Gegenbauer expansion of $F \in \Phi_d$ is absolutely summable. Since (5.2) states that $I_F\left(\sum c_i \delta_{z_i} \right) \geq 0$, obviously (iii) implies (i). The converse implication is proved by a standard argument based on the compactness of $\mathbb{S}^d$ and the weak-star density of the linear span of Dirac masses in $\mathcal{B}$. Finally, the equivalence of (ii) and (iv) can be established by defining f

through the identity $(\widehat{f}(n, \lambda))^2 = \widehat{F}(n, \lambda)$. Absolute summability of the Gegenbauer series of $F \in \Phi_d$ will guarantee that $f \in L^2_{w_\lambda}[-1, 1]$. For more details on positive definite functions, see Chapter 14 in [15].

Condition (iii) above suggests that the property of being positive definite is related to energy minimization. We show that this is indeed the case. We shall need to make a technical assumption that $I_\sigma(F) \geq 0$ (in view of (iii), it is necessary for $F \in \Phi_d$). But, since adding a constant to F does not affect minimizing energy over $\mathcal{M}$, this assumption is easily removable. We first prove the following theorem.

Theorem 5.3 *Assume that $F \in C[-1, 1]$ and $I_\sigma(F) \geq 0$. Then σ is a minimizer of $I_F(\mu)$ over $\mathcal{M}$ (probability Borel measures on $\mathbb{S}^d$) if and only if $F \in \Phi_d$.*

Proof The sufficiency follows easily from (5.1) and condition (iii) of Proposition 5.2. Indeed, if $F \in \Phi_d$, then for any $\mu \in \mathcal{M}$,

$$I_F(\mu) - I_F(\sigma) = I_F(\mu - \sigma) \geq 0; \quad \text{i.e.,} \quad I_F(\mu) \geq I_F(\sigma).$$

$\square$

We now prove the necessity. Assume that σ is a minimizer of $I_F(\mu)$ over $\mu \in \mathcal{M}$. We first state an auxiliary lemma:

Lemma 5.4 *Let $F \in C[-1, 1]$ with $I_F(\sigma) \geq 0$. Assume that σ is a minimizer of $I_F(\mu)$ over $\mu \in \mathcal{M}$. Then for any measure $\gamma \in \mathcal{B}$ with total mass zero, $I_F(\gamma) \geq 0$.*

Assuming the lemma, it is easy to finish the proof of the theorem. Indeed, let $\mu \in \mathcal{B}$ be an arbitrary signed measure with total mass one. Then according to (5.1) and the lemma above,

$$I_F(\mu) - I_F(\sigma) = I_F(\mu - \sigma) \geq 0,$$

since $(\mu - \sigma)(\mathbb{S}^d) = 0$. Therefore, $I_F(\mu) \geq I_F(\sigma) \geq 0$, and by part (iii) of Proposition 5.2, F is positive definite.

Remark We would like to observe that along the way we have proved that if σ is a minimizer of I_F over $\mathcal{M}$, the set of positive measures of mass one, it is also a minimizer over the class of all *signed* measures of total mass one. This is not necessarily the case in other settings. In particular, for the integral over the ball

$$\int_{\mathbb{B}^{d+1}} \int_{\mathbb{B}^{d+1}} \|x - y\| d\mu(x) d\mu(y),$$

according to [8] the unique maximizer over $\mathcal{M}$ is σ, while in the case of signed measures the maximizer does not exist [22].

It remains to prove Lemma 5.4

Proof of Lemma 5.4 Assume, on the contrary, that for some $\gamma \in \mathcal{B}$ with $\gamma(\mathbb{S}^d) = 0$, we have $I_F(\gamma) < 0$. We shall smooth γ out by considering, for $\varepsilon > 0$, the function $\rho_\varepsilon(x) = \dfrac{\gamma(C(x, 1-\varepsilon))}{\sigma(C(x, 1-\varepsilon))}$ and defining $d\gamma_\varepsilon(x) = \rho_\varepsilon(x)d\sigma(x)$, where as before $C(x, t) = \{z \in \mathbb{S}^d : x \cdot z > t\}$ is the spherical cap. The measure γ_ε has total mass zero, since, letting $m_\varepsilon = \sigma(C(x, 1-\varepsilon))$, we have

$$\int_{\mathbb{S}^d} d\gamma_\varepsilon(x) = \frac{1}{m_\varepsilon} \int_{\mathbb{S}^d} \int_{C(x,1-\varepsilon)} d\gamma(y)d\sigma(x)$$

$$= \frac{1}{m_\varepsilon} \int_{\mathbb{S}^d} \left(\int_{C(y,1-\varepsilon)} d\sigma(x) \right) d\gamma(y) = 0.$$

Next, we claim that, for ε small enough, $I_F(\gamma_\varepsilon) < 0$. This follows from

$$I_F(\gamma_\varepsilon) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \frac{1}{m_\varepsilon^2} \int_{C(x,1-\varepsilon)} \int_{C(y,1-\varepsilon)} F(x \cdot y)d\gamma(u)d\gamma(v)d\sigma(x)d\sigma(y)$$

$$= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \left[\frac{1}{m_\varepsilon^2} \int_{C(u,1-\varepsilon)} \int_{C(v,1-\varepsilon)} F(x \cdot y)d\sigma(x)d\sigma(y) \right] d\gamma(u)d\gamma(v)$$

$$\longrightarrow I_F(\gamma) < 0$$

as $\varepsilon \to 0$, since the expression inside the brackets converges to $F(u \cdot v)$ uniformly in u and v. This proves the claim.

It is also easy to see that the density ρ_ε is a bounded function. Therefore, there exists a constant $c > 0$ such that the measure $\mu = c\gamma_\varepsilon + \sigma$; i.e., $d\mu(x) = (1 + c\rho_\varepsilon(x))d\sigma(x)$ is non-negative. Hence $\mu \in \mathcal{M}$. Since σ minimizes I_F over $\mathcal{M}$, by (5.1) we have

$$0 \leq I_F(\mu) - I_F(\sigma) = I_F(\mu - \sigma) = c^2 I_F(\gamma_\varepsilon),$$

which contradicts the fact that $I_F(\gamma_\varepsilon) < 0$.

Since $I_F(\sigma) = \widehat{F}(0, \lambda)$ for $\lambda = \frac{d-1}{2}$, we can easily remove the assumption $I_F(\sigma) \geq 0$ in Theorem 5.3. $\qquad\blacksquare$

Corollary 5.5 *Assume that $F \in C[-1, 1]$. Then σ is a minimizer of $I_F(\mu)$ over $\mathcal{M}$ if and only if $F + C \in \Phi_d$ for some constant $C \in \mathbb{R}$ or, equivalently, if $\widehat{F}(n, \lambda) \geq 0$ for all $n \geq 1$.*

We now turn to the generalization of the Stolarsky principle for positive definite functions. Assume that $F \in \Phi_d$, $\lambda = \frac{d-1}{2}$, and the function $f \in L^2_{w_\lambda}[-1, 1]$ is such that (5.3) is satisfied.

For a non-negative Borel probability measure μ on $\mathbb{S}^d$, we define the L^2 discrepancy of μ with respect to f as

$$D_{L^2,f}(\mu) = \left(\int_{\mathbb{S}^d} \left| \int_{\mathbb{S}^d} f(x \cdot y) \, d\mu(y) - \int_{\mathbb{S}^d} f(x \cdot y) \, d\sigma(y) \right|^2 d\sigma(x) \right)^{\frac{1}{2}}.$$

The L^2 discrepancy of a finite point-set $Z \subset \mathbb{S}^d$ is simply

$$D_{L^2,f}(Z) = D_{L^2,f}\left(\frac{1}{N} \sum_{j=1}^{N} \delta_{z_j} \right)$$

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

Notice that various choices of f recover different geometric notions of discrepancy, although this object is more general. We now prove a general version of the Stolarsky principle, which connects the energies with respect to F to the L^2 discrepancy built upon f.

Theorem 5.6 (*Generalized Stolarsky principle*) *Let $\mu \in \mathcal{B}$ be a signed Borel probability measure on $\mathbb{S}^d$ with total mass $\mu(\mathbb{S}^d) = 1$, and let $F \in \Phi_d$ with f as in* (5.3). *Then*

$$I_F(\mu) - I_F(\sigma) = D_{L^2,f}^2(\mu). \tag{5.4}$$

In particular, in the case of $\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i}$, this relation becomes

$$\frac{1}{N^2} \sum_{i,j=1}^{N} F(z_i \cdot z_j) - \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) d\sigma(x) d\sigma(y) = D_{L^2,f}^2(Z). \tag{5.5}$$

Proof According to the definition of $D_{L^2,f}(\mu)$, (5.3), and (5.1), we have

$$D_{L^2,f}^2(\mu) = \int_{\mathbb{S}^d} \left(\int_{\mathbb{S}^d} f(x \cdot y) \, d(\mu - \sigma)(y) \right)^2 d\sigma(x)$$

$$= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} f(x \cdot y) f(x \cdot z) d(\mu - \sigma)(y) d(\mu - \sigma)(z) d\sigma(x)$$

$$= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(y \cdot z) d(\mu - \sigma)(y) d(\mu - \sigma)(z) = I_F(\mu - \sigma) = I_F(\mu)$$

$$- I_F(\sigma).$$

$\square$

This approach brings up several novel points. First of all, in most contexts, the Stolarsky identity arises from the notion of the L^2 discrepancy, which in turn dictates the specific form of the interaction potential F. Theorem 5.6, on the other hand, allows one to go in the opposite direction: starting with the potential $F \in \Phi_d$, one can produce

a natural notion of discrepancy, for which the Stolarsky principle holds. The precise form of the function f, defined through the identity $\left(\widehat{f}(n,\lambda)\right)^2 = \widehat{F}(n,\lambda)$, cannot be made explicit in most cases (in fact, many different choices of f corresponding to the same F can be constructed by changing the signs of the coefficients $\widehat{f}(n,\lambda)$). However, this does not prevent one from being able to obtain estimates for $D_{L^2,f}(\mu)$. In [5] (Theorem 4.2, part (ii)), we prove that

$$C_d \min_{1 \leq k \leq c_d N^{1/d}} \widehat{F}(k,\lambda) \leq \inf_{\#Z=N} D^2_{L^2,f}(Z) \leq N^{-1} \max_{0 \leq \theta \leq c'_d N^{-\frac{1}{d}}} \left(F(1) - F(\cos\theta)\right).$$

Hence, e.g., lower bounds can be proved using information about either F or f. In [5] we use these estimates to give an alternative proof of the spherical cap discrepancy bounds (2.2), and employ (5.5) to obtain sharp asymptotic behavior of the difference between discrete and continuous energies, $E_F(Z) - I_F(\sigma)$, as the number of points $N \to \infty$ both in the case of Riesz energy, $F(x \cdot y) = \|x - y\|^\delta$ (recovering results of [11,24,35]), and the geodesic distance energies, $F(x \cdot y) = \left(d(x,y)\right)^\delta$, introduced in this paper.

Here we concentrate on the applications of the Stolarsky principle (5.4) to characterizing minimizers of I_F. Since $D^2_{L^2,f}(\mu) \geq 0$, identity (5.4) gives yet another proof that for $F \in \Phi_d$, the uniform measure σ is a minimizer of $I_F(\mu)$ over $\mathcal{M}$ (in fact, over all *signed* Borel probability measures on $\mathbb{S}^d$ with total mass $\mu(\mathbb{S}^d) = 1$). Furthermore, the generalized Stolarsky identity (5.4) also allows one to characterize those $F \in \Phi_d$ for which σ is the *unique* minimizer of $I_\mu(F)$.

Theorem 5.7 *Let $F \in C[-1,1]$. Then σ is the unique minimizer of $I_F(\mu)$ if and only if $\widehat{F}(n,\lambda) > 0$ for each $n \geq 1$.*

We shall need a lemma that is a simple corollary of Lemma 4.4 and the density of polynomials in $C(\mathbb{S}^d)$ – compare it to the proof of Proposition 4.3.

Lemma 5.8 *Let γ be a signed Borel measure on $\mathbb{S}^d$ with $\gamma(\mathbb{S}^d) = 0$, and let $f \in L^2_{w_\lambda}[-1,1]$. Assume that condition (4.4) of Lemma 4.4 is satisfied; i.e.,*

$$\int_{\mathbb{S}^d} f(x \cdot y)\,d\gamma(y) = 0 \text{ for } \sigma\text{-almost every } x \in \mathbb{S}^d,$$

and $\widehat{f}(n,\lambda) \neq 0$ for all $n \geq 1$. Then $\gamma = 0$.

We are now ready to prove Theorem 5.7:

Proof of Theorem 5.7 We start with the proof of sufficiency. Without loss of generality, we may assume that $\widehat{F}(0;\lambda) > 0$. Assume that for each $n \geq 1$, we have $\widehat{f}(n;\lambda) = (\widehat{F}(n;\lambda))^{1/2} \neq 0$. Let μ be a minimizer of $I_\mu(F)$; i.e., $I_\mu(F) = I_\sigma(F)$. Therefore, the Stolarsky principle (5.4) implies that $D^2_{L^2,f}(\mu) = 0$; i.e., $\int_{\mathbb{S}^d} f(x \cdot y)\,d(\sigma - \mu)(y) = 0$ for σ-almost every x. Then by Lemma 5.8, $\sigma - \mu = 0$. Hence σ is the unique minimizer of $I_\mu(F)$.

Conversely, assume that $\widehat{F}(n_0, \lambda) \leq 0$ for some $n_0 \geq 1$. Let Y_{n_0} be a spherical harmonic of degree n_0 with $\|Y_{n_0}\|_2 = 1$. Then, for $\varepsilon > 0$ small enough, the measure $d\mu = (1 + \varepsilon Y_{n_0}(x))d\sigma \in \mathcal{M}$, and by the Funk–Hecke formula (4.6), we have

$$
\begin{aligned}
I_\mu(F) &= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y)\Big(1 + \varepsilon Y_{n_0}(x)\Big)\Big(1 + \varepsilon Y_{n_0}(y)\Big)\, d\sigma(x)\, d\sigma(y) \\
&= I_\sigma(F) + \varepsilon^2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) Y_{n_0}(x) Y_{n_0}(y)\, d\sigma(x) d\sigma(y) \\
&\quad + 2\varepsilon \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) Y_{n_0}(x)\, d\sigma(x)\, d\sigma(y) \\
&= I_\sigma(F) + \varepsilon^2 \widehat{F}(n_0, \lambda) \leq I_\sigma(F),
\end{aligned}
$$

with equality being valid only if $\widehat{F}(n_0, \lambda) = 0$. This is impossible since $d\sigma$ is the unique minimizer of $I_\mu(\text{F})$.   □

The relations between positive definite functions and energy minimization on the sphere are well known [24,29]. Here we have attempted to give an essentially self-contained exposition with minimal references to ultraspherical expansions and a special emphasis on the novel role of the Stolarsky principle. It has come to our attention that similar ideas have been explored also in [16].

6 Stolarsky Principle for Spherical Wedges and Slices

The spherical wedge as defined in [7] is the subregion of $\mathbb{S}^d$ between two hyperplanes: for $x, y \in \mathbb{S}^d$,
$$
W_{xy} = \{z \in \mathbb{S}^d : \text{sign}(x \cdot z) \neq \text{sign}(y \cdot z)\};
$$

in other words, it is the collection of all points $z \in \mathbb{S}^d$ such that the hyperplane $z^\perp$ separates x and y. It is easy to see that (compare to (3.2))

$$
\sigma(W_{xy}) = \sigma\{\text{sign}(x \cdot z) \neq \text{sign}(y \cdot z)\} = \sigma\big(H(x)\Delta H(y)\big) = d(x, y)
$$

(this is also a simple instance of the Crofton formula in integral geometry). For a finite set of vectors $Z = \{z_1, z_2, ..., z_N\}$ on the sphere $\mathbb{S}^d$, we define the Hamming distance between the points $x, y \in \mathbb{S}^d$ as

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

i.e., the proportion of those hyperplanes $z_k^\perp$ that *separate* the points x and y. Therefore, the quantity

$$
\Delta_Z(x, y) := d_H(x, y) - d(x, y) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{W_{xy}}(z_i) - \sigma(W_{xy})
$$

is precisely the discrepancy of Z with respect to W_{xy}. This quantity arises naturally in *one-bit compressed sensing* and in uniform tessellations of the sphere, as well as dimension reduction and almost isometric embedding results (e.g., one-bit analogs of the Johnson–Lindenstrauss lemma), see [6,7,28] for more details.

We define the L^2 wedge discrepancy

$$D_{L^2,\text{wedge}}(Z) = \left\| \Delta_Z(x, y) \right\|_2$$

$$= \left(\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{1}_{W_{xy}}(z_i) - \sigma(W_{xy}) \right)^2 d\sigma(x) d\sigma(y) \right)^{\frac{1}{2}}.$$

The analog of the Stolarsky principle for wedges has been proved by the first author and M. Lacey in [7]:

Theorem 6.1 *(Stolarsky principle for wedges)*

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

$$(6.1)$$

This theorem implies that, in order to minimize $D_{L^2,\text{wedge}}(Z)$, one should minimize the discrete energy with the potential $\left(\frac{1}{2} - d(x, y) \right)^2$, i.e. make the vectors z_k as orthogonal as possible on the average. We would like to point out the strong similarity between this discrete energy and another similar quantity, the *frame potential* introduced in [4]:

$$FP(Z) = \frac{1}{N^2} \sum_{i,j=1}^N (z_i \cdot z_j)^2.$$

A finite set $Z = \{z_1, \ldots, z_N\} \subset \mathbb{S}^d$ is called a *tight frame* if and only if there exists a constant $C > 0$ such that for any vector $x \in \mathbb{R}^{d+1}$,

$$\|x\|^2 = C \sum_{i=1}^N (x \cdot z_j)^2.$$

It was proved in [4] that Z is a tight frame if and only if Z is a minimizer of the frame potential $FP(Z)$.

Similarly, we define the spherical slices: for $x, y \in \mathbb{S}^d$,

$$S_{xy} = \{z \in \mathbb{S}^d : x \cdot z > 0, \ y \cdot z < 0\};$$
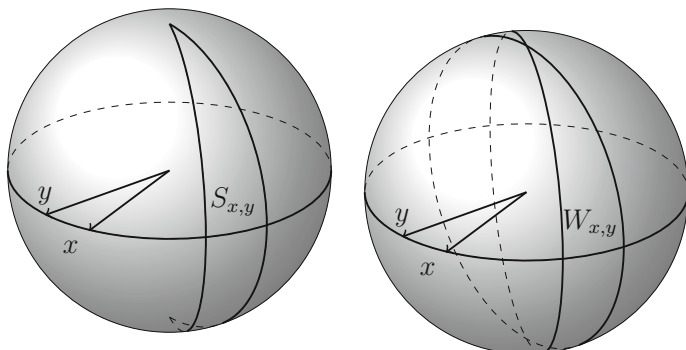
Fig. 2 The slice $S_{x,y} = H_x \cap H_{-y}$ and the wedge $W_{x,y} = H_x \triangle H_y$

i.e., a slice is half of a wedge (Fig. 2). The L^2 slice discrepancy naturally is

$$D_{L^2,\text{slice}}(Z) = \left(\iint_{\mathbb{S}^d \ \mathbb{S}^d} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{1}_{S_{xy}}(z_i) - \sigma(S_{xy}) \right)^2 d\sigma(x) d\sigma(y) \right)^{\frac{1}{2}}. \quad (6.2)$$

This discrepancy has been previously considered in [9]; however, the Stolarsky principle in this setting is new:

Theorem 6.2 *(Stolarsky principle for slices)*

$$4[D_{L^2,\text{slice}}(Z)]^2 = \frac{1}{N^2} \sum_{i,j=1}^N \left(1 - d(z_i, z_j)\right)^2 - \iint_{\mathbb{S}^d \ \mathbb{S}^d} \left(1 - d(x, y)\right)^2 d\sigma(x) d\sigma(y). \tag{6.3}$$

Proof This proof is very similar to the proof of (6.1), see [7, Theorem 1.21]. Recall that $S_{xy} = \{z \in \mathbb{S}^d : x \cdot z > 0, \ y \cdot z < 0\}$ and $\sigma(S_{xy}) = \frac{1}{2}d(x, y)$. Using the definition (6.2), we obtain

$$\begin{aligned}
[D_{L^2,\text{slice}}(Z)]^2 = {} & \frac{1}{N^2} \sum_{i,j=1}^N \iint_{\mathbb{S}^d \ \mathbb{S}^d} \mathbf{1}_{S_{xy}}(z_i) \cdot \mathbf{1}_{S_{xy}}(z_j) \, d\sigma(x) \, d\sigma(y) \\
& - \frac{2}{N} \sum_{k=1}^N \iint_{\mathbb{S}^d \ \mathbb{S}^d} \mathbf{1}_{S_{xy}}(z_k) \cdot \sigma(S_{xy}) \, d\sigma(x) \, d\sigma(y) \\
& + \iint_{\mathbb{S}^d \ \mathbb{S}^d} \sigma(S_{xy})^2 \, d\sigma(x) \, d\sigma(y).
\end{aligned}$$

It is easy to see that $z_i, z_j \in S_{xy}$ if and only if $x \in S_{z_i,-z_j}$ and $y \in S_{-z_i,z_j}$. Since

$$\sigma\left(S_{\pm z_i, \mp z_j}\right) = \frac{1}{2}d(\pm z_i, \mp z_j) = \frac{1}{2}\left(1 - d(z_i, z_j)\right),$$

we find that

$$\int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \mathbf{1}_{S_{xy}}(z_j) \cdot \mathbf{1}_{S_{xy}}(z_i) d\sigma(x) d\sigma(y) = \sigma\left(S_{z_i,-z_j}\right) \cdot \sigma\left(S_{-z_i,z_j}\right) = \frac{1}{4} \cdot \left(1 - d(z_i, z_j)\right)^2.$$

Notice that by rotational invariance, the double integral in the second term does not depend on the choice of $z_k \in \mathbb{S}^d$, and therefore it can be replaced by the average over $z \in \mathbb{S}^d$:

$$\int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \mathbf{1}_{S_{xy}}(z_k) \cdot \sigma(S_{xy}) \, d\sigma(x) \, d\sigma(y) = \int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \mathbf{1}_{S_{xy}}(z) \cdot \sigma(S_{xy}) \, d\sigma(x) \, d\sigma(y) \, d\sigma(z)$$

$$= \int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \left[\int\limits_{\mathbb{S}^d} \mathbf{1}_{S_{xy}}(z) \, d\sigma(z)\right] \sigma(S_{xy}) \, d\sigma(x) \, d\sigma(y) = \int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \sigma(S_{xy})^2 \, d\sigma(x) \, d\sigma(y).$$

Since $\sigma(S_{xy}) = \frac{1}{2}d(x, y)$, it follows that

$$[D_{L^2,\text{slice}}(Z)]^2 = \frac{1}{4N^2} \sum_{i,j=1}^{N} \left(1 - d(z_i, z_j)\right)^2 - \frac{1}{4}\int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \left(d(x, y)\right)^2 d\sigma(x) d\sigma(y)$$

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

which proves Theorem 6.2.   □

The value of the integral $\int\limits_{\mathbb{S}^d} \int\limits_{\mathbb{S}^d} \left(d(x, y)\right)^2 d\sigma(x) d\sigma(y)$, which arises in this theorem, will be computed in the next section.

We also note that the optimal order of magnitude both for $D_{L^2,\text{slice}}(Z)$ [9] and $D_{L^2,\text{wedge}}(Z)$ [7] satisfy the same bounds as the spherical cap discrepancy (2.2).

7 Appendix: Mean-Square Geodesic Distance

The following integral arises in the formulations of Stolarsky principles for wedges (6.1) and slices (6.3):

$$V_d = \int\limits_{\mathbb{S}^d}\int\limits_{\mathbb{S}^d} \left(d(x, y)\right)^2 d\sigma(x) d\sigma(y);$$

hence we compute it and examine its properties. A standard calculation yields that

$$V_d = \frac{1}{\pi^2} \cdot \frac{\omega_{d-1}}{\omega_d} \int_0^\pi \phi^2 \sin^{d-1} \phi \, d\phi.$$

Applying the recursive relation [19]

$$\int x^m \sin^n x \, dx = \frac{x^{m-1} \sin^{n-1} x}{n^2} \left(m \sin x - n x \cos x \right) +$$

$$+ \frac{n-1}{n} \int x^m \sin^{n-2} x \, dx - \frac{m(m-1)}{n^2} \int x^{m-2} \sin^n x \, dx$$

with $m = 2$ and $n = d - 1$, as well as the facts that

$$\frac{\omega_{d-1}}{\omega_d} = \frac{d-1}{d-2} \cdot \frac{\omega_{d-3}}{\omega_{d-2}}$$

and

$$\int_0^\pi \sin^{d-1} \phi \, d\phi = \frac{\sqrt{\pi} \, \Gamma(d/2)}{\Gamma\big((d+1)/2\big)},$$

we obtain the recursive relation

$$V_d = V_{d-2} - \frac{2}{\pi^2 (d-1)^2}.$$

Together with simple identities $V_1 = \frac{1}{3}$ and $V_2 = \frac{1}{2} - \frac{2}{\pi^2}$ (or even $V_0 = \frac{1}{2}$), this yields:

Lemma 7.1 *For odd values of $d \geq 1$,*

$$V_d = \frac{1}{3} - \frac{2}{\pi^2} \sum_{k=1}^{(d-1)/2} \frac{1}{(2k)^2},$$

while for even values of $d \geq 2$,

$$V_d = \frac{1}{2} - \frac{2}{\pi^2} \sum_{k=1}^{d/2} \frac{1}{(2k-1)^2}.$$

Since $\displaystyle\sum_{k=1}^\infty \frac{1}{(2k)^2} = \frac{\pi^2}{24}$ and $\displaystyle\sum_{k=1}^\infty \frac{1}{(2k-1)^2} = \frac{\pi^2}{8}$, we find that

$$\lim_{d \to \infty} V_d = \frac{1}{4},$$

which is consistent with the concentration of measure phenomenon ("most points" on the high-dimensional sphere are nearly orthogonal).

Notice that this confirms the result of Theorem 4.7 that, unless $d = 0$, the uniform distribution σ is *not* a maximizer of $I(\mu) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \big(d(x, y) \big)^2 \, d\mu(x) \, d\mu(y)$, since for $\mu = \frac{1}{2}\delta_p + \frac{1}{2}\delta_{-p}$ we have $I(\mu) = \frac{1}{2}$.

References

1. Aigner, M., Ziegler, G.: Proofs from The Book, 5th edn. Springer, Berlin (2014)
2. Beck, J.: Some upper bounds in the theory of irregularities of distribution. Acta Arith. **43**(2), 115–130 (1984)
3. Beck, J.: Sums of distances between points on a sphere-an application of the theory of irregularities of distribution to discrete geometry. Mathematika **31**(1), 33–41 (1984)
4. Benedetto, J., Fickus, M.: Finite normalized tight frames. Adv. Comput. Math. **18**(2–4), 357–385 (2003)
5. Bilyk, D., Dai, F.: Geodesic distance Riesz energy on the sphere, 2017, to appear, available at arXiv: 1612.08442
6. Bilyk, D., Lacey, M.: Random tessellations, restricted isometric embeddings, and one bit sensing, 2017, to appear, available at arXiv: 1512.06697
7. Bilyk, D., Lacey, M.: One bit sensing, discrepancy, and Stolarsky principle. Sb. Math. **208**(6), 744–763 (2017)
8. Björck, G.: Distributions of positive mass, which maximize a certain generalized energy integral. Ark. Mat. **3**, 255–269 (1956)
9. Blümlinger, M.: Slice discrepancy and irregularities of distribution on spheres. Mathematika **38**(1), 105–116 (1991)
10. Borodachov, S., Hardin, D., Saff, E.: Minimal Discrete Energy on Rectifiable Sets, Springer, Monographs in Math. (to appear)
11. Brauchart, J.: About the second term of the asymptotics for optimal Riesz energy on the sphere in the potential-theoretical case. Integral Transforms Spec. Funct. **17**(5), 321–328 (2006)
12. Brauchart, J., Hardin, D., Saff, E.: Discrete energy asymptotics on a Riemannian circle. Unif. Distrib. Theory **7**(2), 77–108 (2012)
13. Brauchart, J.S., Dick, J.: A simple proof of Stolarsky's invariance principle. Proc. Am. Math. Soc. **141**, 2085–2096 (2013)
14. Drmota, M., Tichy, R.: Sequences, Discrepancies and Applications. Lecture Notes in Mathematics, vol. 1651. Springer, Berlin (1997)
15. Dai, F., Xu, Y.: Approximation Theory and Harmonic Analysis on Spheres and Balls. Springer Monographs in Mathematics. Springer, New York (2013)
16. Damelin, S., Hickernell, F., Ragozin, D., Zeng, X.: On energy, discrepancy and group invariant measures on measurable subsets of Euclidean space. J. Fourier Anal. Appl. **16**(6), 813–839 (2010)
17. Fejes Tóth, L.: On the sum of distances determined by a pointset. Acta Math. Acad. Sci. Hungar. **7**, 397–401 (1956)
18. Fejes Tóth, L.: Über eine Punktverteilung auf der Kugel (German). Acta Math. Acad. Sci. Hung. **10**, 13–19 (1959)
19. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series, and Products. Academic Press Inc, San Diego, CA (2000)
20. Groemer, H.: Geometric Applications of Fourier Series and Spherical Harmonics. Encyclopedia of Mathematics and Its Applications, vol. 61. Cambridge University Press, Cambridge (1996)
21. He, H., Basu, K., Zhao, Q., Owen, A.: Permutation p-value approximation via generalized Stolarsky invariance, to appear, available at arXiv: 1603.02757
22. Hinrichs, A., Nickolas, P., Wolf, R.: A note on the metric geometry of the unit ball. Math. Z. **268**(3–4), 887–896 (2011)
23. Jiang, M.: On the sum of distances along a circle. Discrete Math. **308**(10), 2038–2045 (2008)
24. Kuijlaars, A., Saff, E.: Asymptotics for minimal discrete energy on the sphere. Trans. Am. Math. Soc. **350**(2), 523–538 (1998)

25. Larcher, H.: Solution of a geometric problem by Fejes Tóth. Mich. Math. J. **9**, 45–51 (1962)
26. Matoušek, J.: Geometric Discrepancy. An Illustrated Guide. Springer, Berlin (1999)
27. Nielsen, F.: On the sum of the distances between n points on a sphere. Nordisk Mat. Tidskr. **13**, 45–50 (1965). (Danish)
28. Plan, Y., Vershynin, R.: Dimension reduction by random uniform tessellation. Discrete Comput. Geom. to appear
29. Schoenberg, I.: Some extremal problems for positive definite sequences and related extremal convex conformal maps of the circle. Nederl. Akad. Wetensch. Proc. Ser A **20**, 28–37 (1958)
30. Skriganov, M.: Point distributions in compact metric spaces, to appear, available at arXiv: 1512.00364
31. Sperling, G.: Lösung einer elementargeometrischen Frage von Fejes Tóth. (German). Arch. Math. **11**, 69–71 (1960)
32. Stolarsky, K.B.: Sums of distances between points on a sphere. II. Proc. Am. Math. Soc. **41**, 575–582 (1973)
33. Stolarsky, K.B.: Spherical distributions of N points with maximal distance sums are well spaced. Proc. Am. Math. Soc. **48**, 203–206 (1975)
34. Torquato, S.: Reformulation of the covering and quantizer problems as ground states of interacting particles. Phys. Rev. E **82**(5), 1–52 (2010)
35. Wagner, G.: On means of distances on the surface of a sphere. II. Upper bounds. Pac. J. Math. **154**(2), 381–396 (1992)
36. Woźniakowski, H.: Average case complexity of multivariate integration. Bull. Am. Math. Soc. **24**(1), 185–194 (1991)