

GEODESIC DISTANCE RIESZ ENERGY ON THE SPHERE

DMITRIY BILYK AND FENG DAI

ABSTRACT. We study energy integrals and discrete energies on the sphere, in particular, analogs of the Riesz energy with the geodesic distance in place of Euclidean, and determine that the range of exponents for which uniform distribution optimizes such energies is different from the classical case. We also obtain a very general form of the Stolarsky principle, which relates discrete energies to certain L^2 discrepancies, and prove optimal asymptotic estimates for both objects. This leads to sharp asymptotics of the difference between optimal discrete and continuous energies in the geodesic case, as well as new proofs of discrepancy estimates.

1. INTRODUCTION

In the present paper we study optimization properties of some energy integrals and discrete energies on the unit sphere $\mathbb{S}^d \subset \mathbb{R}^{d+1}$, particularly those related to the geodesic distance on the sphere.

Let $\mathcal{B}$ denote the collection of all Borel probability measures on $\mathbb{S}^d$. Given a measure $\mu \in \mathcal{B}$, define the energy integral $I_F(\mu)$ of a measurable (non-negative or bounded) function $F : [-1, 1] \rightarrow \mathbb{R}$ by

$$(1.1) \quad I_F(\mu) := \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) d\mu(x) d\mu(y).$$

We are interested in finding the optimal (maximal or minimal, depending on F) values of $I_F(\mu)$ over $\mu \in \mathcal{B}$, as well as extremal measures μ for which these values are achieved, i.e. equilibrium distributions with respect to F . In particular, it is natural to investigate whether the Lebesgue surface measure σ , normalized by $\sigma(\mathbb{S}^d) = 1$, is a minimizer (maximizer), and if so, whether it is unique; in other words, whether optimizing the energy with potential F induces uniform distribution.

For a finite set of N points $Z = \{z_1, \dots, z_N\}$ in $\mathbb{S}^d$, its discrete energy with respect to F is defined as

$$(1.2) \quad E_F(Z) = \sum_{1 \leq i < j \leq N} F(z_i \cdot z_j).$$

Note that in the case when $F(1) = 0$, we have

$$E_F(Z) = \frac{1}{2} N^2 I_F(\mu) \quad \text{with} \quad \mu = \frac{1}{N} \sum_{i=1}^N \delta_{z_i},$$

where δ_x is the Dirac mass at $x \in \mathbb{S}^d$. Appropriate modifications accounting for the diagonal terms should be made in the general case. One is interested in optimizing the discrete energy for a given N , analyzing extremal N -point configurations, comparing optimal values of the discrete energy to the optimal energy integral, and finding the asymptotic behavior of this difference.

Such problems arise naturally in various fields, e.g., in electrostatics, in the study of equilibrium distributions of charges which repel according to the law given by F . One of the most natural choices of the potential is the so-called Riesz potential $F(x \cdot y) = |x - y|^{-s}$, where $|x - y|$ is the Euclidean distance between x and y in $\mathbb{R}^{d+1}$. In particular, for $d = 2$ and $s = 1$, minimization of the energy (1.2) is known as Thomson's problem and amounts to finding the equilibrium (according to Coulomb's Law) distributions of N electrons on the sphere. This situation has been studied extensively and numerous questions posed above are well understood in this case [8, 10, 12, 19, 24, 34, 35], although precise optimal discrete distributions are still elusive for most values of N (the problem for $N = 5$ has only been solved recently [25]).

While our present work establishes many general facts and relations, we primarily concentrate on the case, seemingly similar to the classical Riesz energy, that uses geodesic, rather than Euclidean, distance in the definition of energy. This object naturally arose in the companion paper of the authors [5] in relation to discrepancy theory and Stolarsky principle. It has also been considered previously [12, 15, 20, 29] in various contexts. To make things precise, let $\rho(x, y)$ denote the geodesic distance between x and y on $\mathbb{S}^d$, i.e.

$$\rho(x, y) = \arccos(x \cdot y).$$

We shall consider energies defined by the function

$$F_\delta(x \cdot y) = (\rho(x, y))^\delta, \quad \text{i.e.} \quad F_\delta(t) = (\arccos t)^\delta,$$

for an arbitrary $\delta \in \mathbb{R} \setminus \{0\}$; for $\delta = 0$, the standard modification is the logarithmic potential $F_0(t) = -\log(\frac{1}{\pi} \arccos t)$. We would like to characterize extremizers of the energy integral

$$(1.3) \quad I_{d,\delta}(\mu) = I_{F_\delta}(\mu) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y))^\delta d\mu(x) d\mu(y),$$

which we shall refer to as the *geodesic distance (Riesz) energy*. Naturally one is interested in minimizers when $\delta \leq 0$ and maximizers for $\delta > 0$.

One may expect that the behavior of the geodesic distance energy should be similar to its Euclidean counterpart, i.e. the standard Riesz energy. Perhaps surprisingly, this is not quite the case. In dimension $d = 1$ (on the circle) this phenomenon has been previously observed in [13]: in the geodesic case the uniform distribution σ ceases to be the unique extremizer of $I_{d,\delta}$ when $\delta \geq 1$, while in the case of Riesz energy the analogous critical value is $\delta = 2$.

In the present work together with our companion paper [5] we prove this fact in all dimensions $d \geq 1$. More precisely, we prove the following theorem:

Theorem 1.1. *Let $I_{d,\delta}(\mu)$ be the geodesic distance energy integral on $\mathbb{S}^d$, with $\delta \in \mathbb{R}$, as defined in (1.3). The extremizers of this energy integral over $\mu \in \mathcal{B}$ can be characterized as follows:*

- (i) $-d < \delta \leq 0$: the unique minimizer of $I_{d,\delta}(\mu)$ is $\mu = \sigma$ (the normalized surface measure).
- (ii) $0 < \delta < 1$: the unique maximizer of $I_{d,\delta}(\mu)$ is $\mu = \sigma$.
- (iii) $\delta = 1$: $I_{d,\delta}(\mu)$ is maximized if and only if μ is centrally symmetric.
- (iv) $\delta > 1$: $I_{d,\delta}(\mu)$ is maximized if and only if $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$ for some $p \in \mathbb{S}^d$, i.e. the mass is equally concentrated at two antipodal poles.

We shall present the proof of the first two parts, (i) and (ii), in which optimal energy leads to uniform distribution (see Theorem 3.1). The last two parts of Theorem 1.1 can also be obtained by our methods (see Lemma 2.4), besides, they have been proved by the authors and R. Matzke in [5] via a different approach: the critical case (iii) is obtained as a consequence of the geodesic distance Stolarsky principle, relating the hemisphere discrepancy and the geodesic distance energy (see also §4 in the present paper), while the degenerate case (iv) easily follows from the critical case.

In the one-dimensional case parts (i) and (ii) of Theorem 1.1 have been previously established by Brauchart, Hardin, and Saff [13], along with the precise asymptotic energy of the discrete geodesic energy of N equally spaced points. Immediately after our result, Tan [33] gave an alternative proof of parts (ii)-(iv) of Theorem 1.1. See the remark at the very end of §3.

We note that in the case of Riesz energy, i.e. for $F(x \cdot y) = |x - y|^\delta$, part (i) also holds, see e.g. [19], while for $\delta > 0$ the phase transition is different, as was established by Björck [8]. Here σ is the unique maximizer when $\delta \in (0, 2)$, while for $\delta > 2$ the maximizers collapse to symmetric two-point measures as in (iv). At the critical value $\delta = 2$ the maximizers are precisely those measures whose center of mass is at the origin. Intuitively, for small values of δ , in particular, when $\delta < 0$, small scale interactions contribute the most to the energy, therefore (since $\rho(x, y) \approx |x - y|$ when x and y are close) both energies exhibit similar behavior, while for larger values of δ mid-range and long-range interactions come into play and the difference between geodesic and Euclidean distances manifests itself in the energy integrals. The restriction that $\delta > -d$ is natural for the classical Riesz energy, since on a d -dimensional manifold, the corresponding energy integral with $\delta \leq -d$ would be infinite for any $\mu \in \mathcal{B}$.

Our proofs rely on *spherical harmonic* expansions. We briefly review the basic notions in §2, but for a detailed and extensive exposition the reader is directed to, e.g., [14]. In §2 we discuss connections between the extremizers of energy integrals and properties of the potential F (signs of the Gegenbauer coefficients, positive definiteness). Some of these connections are well known in the theory and go back to Schoenberg [24], while some formulations are new. We would like to point out that some of these properties are also discussed in our parallel paper [5] without resorting to (or with minimal use of) spherical harmonics.

In §3 we apply the general results presented in §2 to the specific case of the geodesic distance energy integral (1.3) and prove parts (i) and (ii) of Theorem 1.1: these results are contained in Theorem 3.1. Up to some technical details, the proofs boil down to demonstrating that the Gegenbauer coefficients of the potential are all positive (negative). Essential computations are carried out in Lemma 3.2.

In §4 we connect two different objects, which quantify equidistribution: energy and discrepancy. We show that for all positive definite functions F the difference between the discrete and continuous energies may be represented as the L^2 norm of a certain discrepancy function (part (i) of Theorem 4.2, see also §5 in [5] for a more detailed discussion): this is a generalization of the *Stolarsky principle* in discrepancy theory [31]. Furthermore, this discrepancy may be estimated using the function F . In particular, lower bounds involve Gegenbauer coefficients of F (see part (ii) of Theorem 4.2). We apply these results to give an alternative proof of a classical bound on the spherical cap discrepancy due to Beck [2], see Theorem 4.3.

Results of Theorem 4.2 can also be used in the opposite direction. In §5 we turn to the problem of estimating the asymptotic difference between the geodesic distance energy of the uniform distribution $I_{d,\delta}(\sigma)$ and the corresponding optimal energy of discrete N -point distributions, as $N \rightarrow \infty$. Setting $\mathcal{E}_{d,\delta}(N) = \inf_{\#Z=N} E_{F_\delta}(Z)$, where as before $F_\delta(t) = (\arccos t)^\delta$, and using the results of §4, in Theorem 5.2 we establish that the asymptotic estimate

$$I_{d,\delta}(\sigma) - \frac{2}{N^2} \mathcal{E}_{d,\delta}(N) \sim N^{-1-\frac{\delta}{d}}$$

holds for $-d < \delta < 1$ (with a logarithmic correction for $\delta = 0$). Note that here and throughout the paper the symbol $\sim$ stands for “is of the same order as”; in other words, $f(N) \sim g(N)$ means that for some $A, B > 0$ the inequality $Af(N) \leq g(N) \leq Bf(N)$ holds for all $N \in \mathbb{N}$. The implicit constants may depend on the dimension and some other parameters, but are independent of the number of points N .

This result closely mirrors the case of the classical Riesz energy, but for the Euclidean distance this estimate is valid for $-d < \delta < 2$ (this has been established in a series of papers: [34, 35], [19], and [10]).

In order to prove Theorem 5.2, one needs sharp asymptotic estimates of the Gegenbauer coefficients of F_δ (while to establish the optimality of $I_{d,\delta}(\sigma)$ in §3, it suffices to show just that these coefficients are positive). These bounds, which in the geodesic case are much more complicated than for the Euclidean distance, are stated in Lemma 5.4 and their rather technical proof is presented in §6.

Acknowledgment. The authors are extremely grateful to CRM (Barcelona): their collaboration has started while both of them participated in the research program on “Constructive Approximation and Harmonic Analysis” in 2016. The stay of the first author at CRM has been sponsored by NSF grant DMS 1613790. This work is partially supported by NSERC Canada under grant RGPIN 04702 (Dai) and by the Simons foundation collaboration grant and NSF grant DMS 1665007 (Bilyk).

2. PRELIMINARIES

Let $w_\lambda(t) = (1 - t^2)^{\lambda - \frac{1}{2}}$ with $\lambda > 0$. Given $1 \leq p < \infty$, 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 a Gegenbauer

(ultraspherical) polynomial expansion:

$$(2.1) \quad F(t) \sim \sum_{n=0}^{\infty} \widehat{F}(n; \lambda) \frac{n + \lambda}{\lambda} C_n^\lambda(t), \quad t \in [-1, 1],$$

where C_n^λ are Gegenbauer polynomials (see [14] for an extensive discussion) and

$$\widehat{F}(n; \lambda) = \frac{\Gamma(\lambda + 1)}{\Gamma(\lambda + \frac{1}{2})\Gamma(\frac{1}{2})} \int_{-1}^1 F(t) R_n^\lambda(t) (1 - t^2)^{\lambda - \frac{1}{2}} dt, \quad n = 0, 1, \dots,$$

where $R_n^\lambda(t) = \frac{C_n^\lambda(t)}{C_n^\lambda(1)}$. From now on in this text we shall set the value of λ to

$$\lambda = \frac{d - 1}{2}.$$

In the special case $\lambda = 0$ (which corresponds to the circle $\mathbb{S}^1$) one obtains Chebyshev polynomials of the first kind $T_n(t)$, which satisfy

$$\frac{1}{2} \lim_{\lambda \rightarrow 0} \frac{n + \lambda}{\lambda} C_n^\lambda(t) = T_n(t) = \cos(n \arccos t).$$

Denote by σ the Lebesgue surface measure on $\mathbb{S}^d$ normalized by $\sigma(\mathbb{S}^d) = 1$, and let $\mathcal{H}_n$ be the space of all spherical harmonics of degree n on $\mathbb{S}^d$. Let $\{Y_{n,1}, \dots, Y_{n,a_n^d}\}$ denote a real orthonormal basis of the space $\mathcal{H}_n$. The addition formula for spherical harmonics states that (see, for instance, [14, 1.2.8])

$$(2.2) \quad \sum_{j=1}^{a_n^d} Y_{n,j}(x) Y_{n,j}(y) = \frac{n + \lambda}{\lambda} C_n^\lambda(x \cdot y) \quad \text{for all } x, y \in \mathbb{S}^d,$$

where

$$a_n^d = \frac{n + \lambda}{\lambda} C_n^\lambda(1) = \dim \mathcal{H}_n \sim n^{d-1}, \quad \lambda = \frac{d - 1}{2}.$$

As before we denote the collection of all Borel probability measures on $\mathbb{S}^d$ by $\mathcal{B}$, and δ_{x_0} is the Dirac Borel probability measure supported at $x_0 \in \mathbb{S}^d$. Given $\mu \in \mathcal{B}$, define the energy integral $I_F(\mu)$ of a (bounded or nonnegative) measurable function $F : [-1, 1] \rightarrow \mathbb{R}$ as in (1.1) by $I_F(\mu) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) d\mu(x) d\mu(y)$. We have the following proposition on extremizers of $I_F(\mu)$ over $\mathcal{B}$:

Proposition 2.1. *Let $\lambda = \frac{d-1}{2}$ and let F be a continuous function on $[-1, 1]$. The following conditions are equivalent:*

- (a) $\widehat{F}(n; \lambda) \geq 0$ for all $n \geq 1$;
- (b) the surface Lebesgue measure $\mu = \sigma$ on $\mathbb{S}^d$ is a minimizer of the energy integral $I_F(\mu)$;

If the above conditions hold, then every Dirac mass $\mu = \delta_e$, $e \in \mathbb{S}^d$ is a maximizer of the energy integral $I_F(\mu)$.

The case when some coefficient $\widehat{F}(n; \lambda) = 0$ corresponds to σ not being the only minimizer. This is made precise in the following statement concerning uniqueness of extremizers:

Proposition 2.2. *Let $\lambda = \frac{d-1}{2}$ and let $F \in C[-1, 1]$. The following conditions are equivalent:*

- (a) $\widehat{F}(n; \lambda) > 0$ for all $n \geq 1$;
(b) the normalized surface measure σ is the unique minimizer of $I_F(\mu)$;

In this case, every maximizer of $I_F(\mu)$ is a Dirac mass.

Obviously changing the inequality signs in the above propositions reverts the roles of maximizers and minimizers. In addition, since adding constants to F does not change the extremizers of $I_F(\mu)$, it is natural that $\widehat{F}(0; \lambda)$ does not play a role in these statements. In order to prove these statements we shall need the following technical lemma:

Lemma 2.3. *Let $F \in C[-1, 1]$ and assume that $\widehat{F}(n; \lambda) \geq 0$ for all $n \geq 1$. Then*

$$(2.3) \quad \sum_{n=1}^{\infty} n^{2\lambda} \widehat{F}(n; \lambda) < \infty,$$

which, in particular, will imply that the series on the right hand side of (2.1) converges uniformly and absolutely to the function F on $[-1, 1]$.

Proof. Let $\sigma_n^\delta F$ denote the Cesàro (C, δ) -means of the Gegenbauer polynomial expansion of F (see [14, Section 2.4] for details), i.e.

$$(2.4) \quad \sigma_n^\delta F(t) = \sum_{k=0}^n \frac{A_{n-k}^\delta}{A_n^\delta} \frac{k+\lambda}{\lambda} \widehat{F}(k; \lambda) C_k^\lambda(t), \quad A_j^\delta = \frac{\Gamma(j+\delta+1)}{\Gamma(j+1)\Gamma(\delta+1)}.$$

It is known (e.g. Theorem 2.4.3 in [14]) that for $\delta > \lambda$,

$$\lim_{n \rightarrow \infty} \|\sigma_n^\delta F - F\|_{L^\infty[-1,1]} = 0, \quad \forall F \in C[-1, 1].$$

On the other hand, since for each fixed j , the sequence $\{\frac{A_{n-j}^\delta}{A_n^\delta}\}_{n=j}^\infty$ increases to 1 as $n \rightarrow \infty$, it follows by Levi's monotone convergence theorem that

$$F(1) = \lim_{n \rightarrow \infty} \sigma_n^\delta(F)(1) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{A_{n-k}^\delta}{A_n^\delta} \widehat{F}(k; \lambda) \frac{k+\lambda}{\lambda} C_k^\lambda(1) = \sum_{k=0}^{\infty} \widehat{F}(k; \lambda) \frac{k+\lambda}{\lambda} C_k^\lambda(1).$$

This yields (2.3) since $(k+\lambda)C_k^\lambda(1) \sim k^{2\lambda}$ as $k \rightarrow \infty$. $\square$

Remark: The self-improving property (2.3) (positivity of coefficients implies their decay) has various manifestations in harmonic analysis: e.g., if a function $f \in L^1(\mathbb{T})$ has Fourier series $\sum ic_n e^{2\pi i n x}$ with $c_n = -c_{-n} \geq 0$ (i.e. sine series with non-negative coefficients), then necessarily $\sum_{n>0} \frac{c_n}{n} < \infty$ (see e.g. [18, page 24]), which is a direct analog of (2.3).

We now prove Proposition 2.1:

Proof of Proposition 2.1. We first prove that, when (a) holds, σ is a minimizer and δ_e is a maximizer of $I_F(\mu)$ for any $e \in \mathbb{S}^d$. Indeed, by (2.3), (2.2) and the dominated

convergence theorem, it follows that

$$\begin{aligned}
 I_F(\mu) &= \sum_{n=0}^{\infty} \widehat{F}(n; \lambda) \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \frac{n+\lambda}{\lambda} C_n^\lambda(x \cdot y) d\mu(x) d\mu(y) \\
 &= \sum_{n=0}^{\infty} \widehat{F}(n; \lambda) \sum_{j=1}^{a_n^d} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} Y_{n,j}(x) Y_{n,j}(y) d\mu(x) d\mu(y) \\
 (2.5) \quad &= \widehat{F}(0; \lambda) + \sum_{n=1}^{\infty} \widehat{F}(n; \lambda) b_{n,\mu},
 \end{aligned}$$

where, by the addition formula (2.2),

$$(2.6) \quad b_{n,\mu} = \frac{n+\lambda}{\lambda} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} C_n^\lambda(x \cdot y) d\mu(x) d\mu(y) = \sum_{j=1}^{a_n^d} \left(\int_{\mathbb{S}^d} Y_{n,j}(x) d\mu(x) \right)^2 \geq 0.$$

Using the Cauchy–Schwarz inequality and (2.2) with $x = y$, we obtain

$$(2.7) \quad 0 \leq b_{n,\mu} \leq \int_{\mathbb{S}^d} \sum_{j=1}^{a_n^d} |Y_{n,j}(x)|^2 d\mu(x) = a_n^d.$$

Thus, if $\widehat{F}(n; \lambda) \geq 0$ for all $n \geq 1$, then by (2.5) and (2.7), one has

$$\begin{aligned}
 I_F(\mu) &\geq \widehat{F}(0; \lambda) = I_F(\sigma) \quad \text{and} \\
 I_F(\mu) &\leq \widehat{F}(0; \lambda) + \sum_{n=1}^{\infty} \widehat{F}(n; \lambda) a_n^d = F(1) = I_F(\delta_e),
 \end{aligned}$$

i.e. σ and δ_e are a minimizer and a maximizer of the integral $I_F(\mu)$, respectively.

Conversely, if $\widehat{F}(n; \lambda) < 0$ for some $n \geq 1$, then define μ by setting $d\mu(x) = (1 + \varepsilon Y_{n,1}(x)) d\sigma(x)$, where $\varepsilon > 0$ is chosen small enough so that $1 + \varepsilon Y_{n,1}(x) \geq 0$ on $\mathbb{S}^d$. The Funk–Hecke formula (see, e.g., Theorem 1.2.9 in [14]) states that for any spherical harmonic $Y \in \mathcal{H}_n$

$$\int_{\mathbb{S}^d} F(x \cdot y) Y(x) d\sigma(x) = \widehat{F}(n; \lambda) Y(y).$$

Thus, using the fact that $\int_{\mathbb{S}^d} Y_{n,1}(x) d\sigma(x) = 0$, we find that $\mu \in \mathcal{B}$ and

$$\begin{aligned}
 (2.8) \quad I_F(\mu) &= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) (1 + \varepsilon Y_{n,1}(x)) (1 + \varepsilon Y_{n,1}(y)) d\sigma(x) d\sigma(y) \\
 &= I_F(\sigma) + \varepsilon^2 \widehat{F}(n; \lambda) \int_{\mathbb{S}^d} Y_{n,1}^2(y) d\sigma(y) < I_F(\sigma),
 \end{aligned}$$

i.e. σ is not a minimizer of I_F . $\square$

Remark: We observe that the fact that δ_e maximizes I_F is not equivalent to conditions (a)-(b) of Proposition 2.1. Indeed, a sufficient condition for this is that $\max_{t \in [-1,1]} F(t) = F(1)$, since then for each $\mu \in \mathcal{B}$, we have $I_F(\mu) \leq \|F\|_\infty = F(1) = I_F(\delta_e)$. For example, for $F(t) = -(\arccos t)^2$, the maximizer is obviously

δ_e , while the minimizer is not σ : according to part (iv) of Theorem 1.1, minimizers are measures of the form $\frac{1}{2}(\delta_e + \delta_{-e})$.

We now prove Proposition 2.2 about the uniqueness of minimizers.

Proof of Proposition 2.2. Suppose that (a) holds, i.e. $\widehat{F}(n; \lambda) > 0$ for all $n \geq 1$. In this case, by (2.5), the equality $I_F(\mu) = \widehat{F}(0; \lambda) = I_F(\sigma)$ holds if and only if $b_{n, \mu} = 0$ for all $n \geq 1$, i.e. if and only if $\int_{\mathbb{S}^d} g(x) d\mu(x) = 0$ for all $g \in \mathcal{H}_n$ and $n \geq 1$. This last condition implies that for each spherical polynomial P on $\mathbb{S}^d$,

$$\int_{\mathbb{S}^d} P(x) d\mu(x) = \int_{\mathbb{S}^d} P(x) d\sigma(x).$$

By the density of spherical polynomials in the space $C(\mathbb{S}^d)$, we then conclude that $d\mu = d\sigma$.

Next, we show that if $\mu_0 \in \mathcal{B}$ is a maximizer of $I_F(\mu)$, then $\mu_0 = \delta_e$ for some $e \in \mathbb{S}^d$. To see this, we first note that according to (2.5), (2.6) and (2.7), in order that

$$I_F(\mu_0) = \max_{\mu \in \mathcal{B}} I_F(\mu) = I_F(\delta_e) = \widehat{F}(0; \lambda) + \sum_{n=1}^{\infty} \widehat{F}(n; \lambda) a_n^d,$$

it is necessary that

$$\left(\int_{\mathbb{S}^d} Y_{n,j}(x) d\mu_0(x) \right)^2 = \int_{\mathbb{S}^d} |Y_{n,j}(x)|^2 d\mu_0(x), \quad \forall n \geq 1, \quad \forall 1 \leq j \leq a_n^d,$$

or equivalently,

$$Y_{n,j}(x) \equiv \text{constant} \quad \mu_0\text{-a.e. on } \mathbb{S}^d, \quad \forall n \geq 1, \quad \forall 1 \leq j \leq a_n^d,$$

which in turn implies that each spherical polynomial is constant μ_0 -a.e. on $\mathbb{S}^d$. Since the space of spherical polynomials is dense in $C(\mathbb{S}^d)$, we further conclude that every continuous function on $\mathbb{S}^d$ is constant μ_0 -a.e. on $\mathbb{S}^d$.

Assume that the support of μ_0 contains at least two distinct points $e_0 \neq e_1$. (Recall that $\text{supp } \mu_0$ is the complement of the union of all open sets U with $\mu_0(U) = 0$.) Choose open neighborhoods U_i of e_i such that $\mu_0(U_i) > 0$ and $U_0 \cap U_1 = \emptyset$, and construct a function $f \in C(\mathbb{S}^d)$ with $f|_{U_i} = i$. Then f is not constant μ_0 -a.e. on $\mathbb{S}^d$. Thus $\text{supp } \mu_0$ contains only one point, and hence μ_0 is a Dirac mass.

If we assume that (a) fails, i.e. $\widehat{F}(n; \lambda) = 0$ for some $n \geq 1$ (if $\widehat{F}(n; \lambda) < 0$, then σ is not a minimizer by Proposition 2.1), then the argument of (2.8) shows that for $d\mu(x) = (1 + \varepsilon Y_{n,1}(x)) d\sigma(x)$ we have $I_F(\mu) = I_F(\sigma)$, i.e. σ is not a unique minimizer. $\square$

We would like to note that functions with $\widehat{F}(n; \lambda) \geq 0$ for all $n \geq 1$ are, up to constant terms, *positive definite functions on the sphere*. This class was introduced by Schoenberg [24] also in the context of energy optimization. See the discussion in the beginning of §4.

In the end of this section we state some additional results about the extremizers of I_F in terms of the signs of the Gegenbauer coefficients $\widehat{F}(n; \lambda)$, which may be proved by identical arguments. These statements could be used to prove parts (iii) and (iv) of Theorem 1.1, which were proved in [5] by other means.

Lemma 2.4. *Let $F \in C[-1, 1]$ and $\lambda = \frac{d-1}{2}$.*

(i) If $(-1)^n \widehat{F}(n; \lambda) \geq 0$ for all $n \geq 1$, then

$$\max_{\mu \in \mathcal{B}} I_F(\mu) = \widehat{F}(0; \lambda) + \sum_{n=1}^{\infty} \widehat{F}(2n; \lambda) a_{2n}^d = \frac{F(1) + F(-1)}{2} = I_F\left(\frac{\delta_e + \delta_{-e}}{2}\right).$$

Moreover, if $(-1)^n \widehat{F}(n; \lambda) > 0$ for all $n \geq 1$, then any maximizer of I_F is of the form $\mu = \frac{1}{2}(\delta_e + \delta_{-e})$ for some $e \in \mathbb{S}^d$.

(ii) If $\widehat{F}(2n; \lambda) = 0$ and $\widehat{F}(2n-1; \lambda) \leq 0$ for all $n \in \mathbb{N}$, then

$$\max_{\mu \in \mathcal{B}} I_F(\mu) = \widehat{F}(0; \lambda),$$

and, in addition, any symmetric measure μ (i.e. such that $\mu(E) = \mu(-E)$ for all measurable $E \subset \mathbb{S}^d$) is a maximizer of I_F .

If $\widehat{F}(2n; \lambda) = 0$ and $\widehat{F}(2n-1; \lambda) < 0$ for all $n \in \mathbb{N}$, then μ is a maximizer of I_F if and only if μ is symmetric.

3. GEODESIC DISTANCE RIESZ ENERGY INTEGRALS

We now apply the results of the previous section to the specific case of the geodesic distance energy integral (1.3). In order to avoid singularities, we introduce standard modifications of the potentials. For $t \in [-1, 1]$ and $0 \leq \varepsilon < 1$, we define as in (1.3)

$$F_{\delta, \varepsilon}(t) = \begin{cases} (\varepsilon + \arccos t)^\delta, & \text{if } \delta \neq 0; \\ \log\left(\frac{\pi}{\varepsilon + \arccos t}\right), & \text{if } \delta = 0. \end{cases}$$

We write $F_\delta(t) = F_{\delta, 0}(t)$. For $\mu \in \mathcal{B}$, define

$$I_{d, \delta}(\mu) := I_{F_\delta}(\mu) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F_\delta(x \cdot y) d\mu(x) d\mu(y).$$

The main goal in this section is to show the following theorem, which constitutes parts (i) and (ii) of Theorem 1.1.

Theorem 3.1. *The normalized Lebesgue measure σ on $\mathbb{S}^d$ is the unique maximizer for the integral $I_{d, \delta}(\mu)$, when $\delta \in (0, 1)$, and is the unique minimizer of $I_{d, \delta}(\mu)$, when $-d < \delta \leq 0$.*

The following lemma plays a crucial role in the proof of Theorem 3.1.

Lemma 3.2. *Let $\varepsilon \in [0, 1)$ and $\lambda > 0$. For $\delta > -(2\lambda + 1)$, define*

$$\mathcal{I}_{n, \varepsilon}^\delta := \int_{-1}^1 F_{\delta, \varepsilon}(t) C_n^\lambda(t) (1 - t^2)^{\lambda - \frac{1}{2}} dt, \quad n = 0, 1, \dots$$

Then the following statements hold:

- (i) *If $\delta \in (0, 1)$, then $\mathcal{I}_{n, \varepsilon}^\delta < 0$ for all $n = 1, 2, \dots$.*
- (ii) *If $-(2\lambda + 1) < \delta \leq 0$, then $\mathcal{I}_{n, \varepsilon}^\delta > 0$ for $n = 0, 1, 2, \dots$.*

Proof. (i) By Rodrigues' formula for ultraspherical polynomials (see, for instance, [32, 4.1.72]),

$$C_n^\lambda(t) = \frac{(-1)^n 2^n}{n!} \frac{\Gamma(n + \lambda) \Gamma(n + 2\lambda)}{\Gamma(\lambda) \Gamma(2n + 2\lambda)} (1 - t^2)^{-(\lambda - \frac{1}{2})} \left(\frac{d}{dt}\right)^n (1 - t^2)^{n + \lambda - \frac{1}{2}},$$

and the fact that $C_n^\lambda(-t) = (-1)^n C_n^\lambda(t)$ for all $t \in [-1, 1]$, it is easily seen from integration by parts that for $k, n = 0, 1, \dots$,

$$(3.1) \quad \int_{-1}^1 t^k C_n^\lambda(t) (1-t^2)^{\lambda-\frac{1}{2}} dt \begin{cases} > 0, & \text{if } k \geq n \text{ and } k-n \text{ is even,} \\ = 0, & \text{otherwise.} \end{cases}$$

Thus, for the proof of assertion (i), it is sufficient to show that for $\delta \in (0, 1)$,

$$(3.2) \quad F_{\delta, \varepsilon}(t) = a_0(\varepsilon) + \sum_{k=1}^{\infty} a_k(\varepsilon) t^k, \quad |t| < 1 \quad \text{with } a_k(\varepsilon) < 0 \text{ for } k = 1, 2, \dots.$$

Indeed, once (3.2) is proved, we obtain from (3.1) that for $n \geq 1$,

$$\begin{aligned} \mathcal{I}_{n, \varepsilon}^\delta &= \sum_{k=0}^{\infty} a_k(\varepsilon) \int_{-1}^1 t^k C_n^\lambda(t) (1-t^2)^{\lambda-\frac{1}{2}} dt \\ &= - \sum_{j=0}^{\infty} |a_{n+2j}(\varepsilon)| \int_{-1}^1 t^{n+2j} C_n^\lambda(t) (1-t^2)^{\lambda-\frac{1}{2}} dt < 0. \end{aligned}$$

To show (3.2), we use the Maclaurin series of the function $\arccos t$ on the interval $[-1, 1]$:

$$(3.3) \quad \arccos t = \frac{\pi}{2} - \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{4^n(2n+1)} t^{2n+1} =: \frac{\pi}{2} - A(t), \quad |t| \leq 1.$$

The main point in (3.3) lies in the fact that

$$(3.4) \quad A(t) := \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{4^n(2n+1)} t^{2n+1}$$

is an odd power series with positive coefficients. Clearly,

$$(3.5) \quad |A(t)| = \left| \frac{\pi}{2} - \arccos t \right| < \frac{\pi}{2}, \quad t \in (-1, 1).$$

Thus, using (3.3) and (3.5), we obtain that for $t \in (-1, 1)$ and $\varepsilon \in [0, 1)$,

$$F_{\delta, \varepsilon}(t) = \left(\frac{\pi}{2} + \varepsilon \right)^\delta \left(1 - \frac{2A(t)}{\pi + 2\varepsilon} \right)^\delta = \left(\frac{\pi}{2} + \varepsilon \right)^\delta + \left(\frac{\pi}{2 + \varepsilon} \right)^\delta \sum_{j=1}^{\infty} b_j^\delta \left(\frac{2A(t)}{\pi + 2\varepsilon} \right)^j,$$

where

$$b_j^\delta = \frac{(-1)^j \delta(\delta-1) \cdots (\delta-j+1)}{j!} = - \frac{\delta(1-\delta) \cdots (j-1-\delta)}{j!}.$$

Clearly, each b_j^δ ($j \geq 1$) is negative for $\delta \in (0, 1)$. Then (3.2) follows from (3.4).

(ii) As in the proof of assertion (i), it suffices to show that $F_{\delta, \varepsilon}(t)$ has a Maclaurin series representation with positive coefficients on the interval $[-1, 1]$. For $\delta = 0$, we use the Maclaurin series of the function $\log(1-t)$ on the interval $(-1, 1)$:

$$\log(1-x) = - \sum_{n=1}^{\infty} \frac{x^n}{n}, \quad |x| < 1.$$

We then obtain from (3.5) that for $t \in (-1, 1)$,

$$\begin{aligned} F_{0,\varepsilon}(t) &= \log \frac{2\pi}{\pi + 2\varepsilon} - \log \left(1 - \frac{2A(t)}{\pi + 2\varepsilon} \right) \\ (3.6) \quad &= \log \frac{2\pi}{\pi + 2\varepsilon} + \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{2A(t)}{\pi + 2\varepsilon} \right)^n. \end{aligned}$$

For $\delta = -s < 0$, we have

$$\begin{aligned} F_{\delta,\varepsilon}(t) &= \left(\frac{\pi}{2} + \varepsilon - A(t) \right)^{-s} \\ (3.7) \quad &= \left(\frac{2}{\pi + 2\varepsilon} \right)^{-s} \sum_{j=0}^{\infty} \frac{s(s+1) \cdots (s+j-1)}{j!} \left(\frac{2A(t)}{\pi + 2\varepsilon} \right)^j. \end{aligned}$$

Combining (3.6), (3.7) with (3.4), we conclude that if $\delta \in (-2\lambda - 1, 0]$ and $\varepsilon \in [0, 1)$, then all the coefficients of the Maclaurin series of the function $F_{\delta,\varepsilon}(t)$ are positive. Assertion (ii) then follows by (3.1). $\square$

We are now in a position to show Theorem 3.1.

Proof of Theorem 3.1. For $\delta \in (0, 1)$, the function $F_{\delta}(t)$ is continuous on $[-1, 1]$, and hence the stated assertion follows directly from Proposition 2.2 and part (i) of Lemma 3.2.

For $-d < \delta \leq 0$, the function F_{δ} is not continuous at $t = 1$ and, therefore, we need a slight modification of the proof. For the moment, we assume that $\delta \neq 0$ and write $\delta = -s$ with $0 < s < d$. Recall that for each $\varepsilon \in (0, 1)$,

$$F_{\delta,\varepsilon}(t) = \left(\arccos t + \varepsilon \right)^{-s}, \quad t \in [-1, 1],$$

and by part (ii) of Lemma 3.2,

$$\widehat{F_{\delta,\varepsilon}}(n; \lambda) = c_n \int_{-1}^1 F_{\delta,\varepsilon}(t) C_n^{\lambda}(t) (1 - t^2)^{\lambda - \frac{1}{2}} dt > 0, \quad n = 0, 1, \dots$$

Using Proposition 2.1, we conclude that $I_{\mu}(F_{\delta,\varepsilon})$ has a unique minimizer $d\sigma$. Hence, for any $\mu \in \mathcal{B}$ and any $\varepsilon > 0$,

$$\begin{aligned} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y))^{-s} d\mu(x) d\mu(y) &\geq \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\varepsilon + \rho(x, y))^{-s} d\mu(x) d\mu(y) \\ &\geq \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\varepsilon + \rho(x, y))^{-s} d\sigma(x) d\sigma(y). \end{aligned}$$

Letting $\varepsilon \rightarrow 0$, and using the monotone convergence theorem, we get

$$\int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y))^{-s} d\mu(x) d\mu(y) \geq \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y))^{-s} d\sigma(x) d\sigma(y), \quad \forall \mu \in \mathcal{B}.$$

This shows that σ is a minimizer of $I_{F_{\delta}}(\mu)$ for $0 < s = -\delta < d$.

Next, we show the minimizer is unique. Let $\mu_0 \in \mathcal{B}$. If $d\mu_0 \neq d\sigma$, then there must exist a spherical harmonic P of degree $n_0 \geq 1$ with $\|P\|_2 = 1$ such that

$\int_{\mathbb{S}^d} P(x) d\mu_0(x) \neq 0$. By (2.6), this implies that

$$\begin{aligned} b_{n_0, \mu_0} &= \frac{n_0 + \lambda}{\lambda} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} C_{n_0}^\lambda(x \cdot y) d\mu_0(x) d\mu_0(y) \\ &\geq \left(\int_{\mathbb{S}^d} P(x) d\mu_0(x) \right)^2 \geq c > 0. \end{aligned}$$

However, according to (2.5), we have that for any $\varepsilon \in (0, 1)$,

$$I_{F_\delta}(\mu_0) \geq I_{F_{\delta, \varepsilon}}(\mu_0) \geq \widehat{F_{\delta, \varepsilon}}(0; \lambda) + \widehat{F_{\delta, \varepsilon}}(n_0; \lambda) b_{n_0, \mu_0}.$$

Letting $\varepsilon \rightarrow 0$, we obtain from part (ii) of Lemma 3.2 (ii) that

$$I_{F_\delta}(\mu_0) \geq I_{F_\delta}(\sigma) + \widehat{F_\delta}(n_0; \lambda) b_{n_0, \mu_0} > I_{F_\delta}(\sigma).$$

Since μ_0 is an arbitrary Borel probability measure on $\mathbb{S}^d$, this shows the uniqueness of the minimizer. Finally, we point out that the above proof with a slight modification works equally well for the case $\delta = 0$. $\square$

The methods employed here are quite standard in the context of energy optimization on the sphere (see, e.g., [19] for the classical Riesz energy). While Theorem 3.1 covers parts (i) and (ii) of Theorem 1.1, the remaining two cases (proved in [5]) could be proved in the same way, using the above computations and results of Lemma 2.4.

Notice that we have, in fact, used Taylor expansions of the underlying function F in order to obtain information about the signs of the ultraspherical coefficients. Recently, after the first author's presentation of the results of this paper and [5], Y.S. Tan [33] found a beautiful alternative proof of parts (ii)-(iv) of Theorem 1.1, which involves only Taylor series and uses an interesting “tensorization trick”. It does *not* resort to the use of spherical harmonics. This method is somewhat less general than the one presented in §2, since it requires F to be analytic.

4. DISCREPANCY AND THE STOLARSKY PRINCIPLE

Let us denote by Φ_d the set of all continuous functions F on $[-1, 1]$ for which $\widehat{F}(n; \lambda) \geq 0$ for all $n \in \mathbb{N} \cup \{0\}$, where, as before, $\lambda = \frac{d-1}{2}$. Functions in the class Φ_d are known as *positive definite* functions on the sphere and, up to constants, are precisely the functions discussed in §2 (Proposition 2.1). Their connection to energy optimization is well known [24]. There is a variety of characterizations of the class Φ_d , but we shall particularly use the following.

Lemma 4.1. *A function $F \in \Phi_d$ if and only if there exists a function $f \in L^2_{w_\lambda}[-1, 1]$ such that*

$$(4.1) \quad F(x \cdot y) = \int_{\mathbb{S}^d} f(x \cdot z) f(z \cdot y) d\sigma(z), \quad x, y \in \mathbb{S}^d.$$

Proof. The *sufficiency* part is obvious. Indeed, if (4.1) holds for some $f \in L^2_{w_\lambda}[-1, 1]$, then F is continuous and $\widehat{F}(n; \lambda) = |\widehat{f}(n, \lambda)|^2 \geq 0$ for all $n \in \mathbb{N} \cup \{0\}$.

It remains to show the *necessity*. Let F be a continuous function on $[-1, 1]$ such that $\widehat{F}(n; \lambda) \geq 0$ for all $n \geq 0$. Define

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

f is a well defined function in $L^2_{w_\lambda}[-1, 1]$ since, by Plancherel's formula and (2.3),

$$c_d \int_{-1}^1 |f(t)|^2 (1 - t^2)^{\lambda - \frac{1}{2}} dt = \sum_{n=0}^{\infty} \widehat{F}(n; \lambda) \frac{n + \lambda}{\lambda} C_n^\lambda(1) < \infty.$$

Furthermore, (4.1) holds since $\widehat{F}(n; \lambda) = |\widehat{f}(n, \lambda)|^2$ for all $n \geq 0$. $\square$

Observe that the relation $\widehat{f}(n, \lambda) = \pm \sqrt{\widehat{F}(n; \lambda)}$ implies that there are infinitely many choices of the function f .

For the rest of this section, we will assume that $F \in \Phi_d$, and $f \in L^2_{w_\lambda}[-1, 1]$ is chosen so that (4.1) is satisfied.

Given a finite set of points $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$, we define its L^2 discrepancy with respect to a function $f : [-1, 1] \rightarrow \mathbb{R}$ by

$$(4.2) \quad D_{L^2, f}(Z) = \left(\int_{\mathbb{S}^d} \left| \int_{\mathbb{S}^d} f(x \cdot y) d\sigma(y) - \frac{1}{N} \sum_{j=1}^N f(x \cdot z_j) \right|^2 d\sigma(x) \right)^{\frac{1}{2}}.$$

We define the optimal L^2 discrepancy by setting

$$\mathcal{D}_{L^2, f, N} = \inf_Z D_{L^2, f}(Z),$$

where the infimum is taken over all $Z \subset \mathbb{S}^d$ with $\#Z = N$. The discrepancy $D_{L^2, f}(Z)$ measures the uniformity of the finite distribution of points Z with respect to the function f . Choosing, for example, $f_t(\tau) = \mathbf{1}_{[1-t, 1]}(\tau)$ and averaging over $t \in [-1, 1]$, one obtains the well-studied spherical cap discrepancy, see (4.12).

The link between discrepancy and energy on the sphere has been first established by Stolarsky [31] who established an identity relating the spherical cap L^2 discrepancy and the sum of pairwise Euclidean distances between the points of Z . Identities of this type came to be known as instances of *the Stolarsky invariance principle*. There has been an increase of activity on this subject in the recent years [7, 9, 11, 22, 26, 27, 28]. In our companion paper [5] we explore a number of variations of this principle and its applications to energy optimization, in particular, part (iii) of Theorem 1.1.

Here we present a general form of the Stolarsky principle, which relates the discrepancy $D_{L^2, f}(Z)$, the discrete energy $E_F(Z)$, and the energy integral $I_F(\sigma)$. We also apply this principle to estimating the optimal discrepancy $\mathcal{D}_{L^2, f, N}$.

Theorem 4.2. *Let $\lambda = \frac{d-1}{2}$. Assume that $F \in \Phi_d$ and $f \in L^2_{w_\lambda}[-1, 1]$ satisfy relation (4.1).*

(i) (*Stolarsky principle*) *Given a set of N -points $Z = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$,*

$$(4.3) \quad N^{-2} \sum_{i=1}^N \sum_{j=1}^N F(z_i \cdot z_j) = D_{L^2, f}^2(Z) + \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} F(x \cdot y) d\sigma(x) d\sigma(y).$$

(ii) *There exist constants $c_d, C_d > 0$, such that for any $N \in \mathbb{N}$,*

$$(4.4) \quad C_d \min_{1 \leq k \leq c_d N^{1/d}} \widehat{F}(k, \lambda) \leq \mathcal{D}_{L^2, f, N}^2 \leq N^{-1} \max_{0 \leq t \leq c_d N^{-1/d}} (F(1) - F(\cos t)).$$

We make a few remarks before proceeding to the proof of this theorem. First of all, notice that (4.3) implies that minimizing the discrete energy $E_F(Z)$ is equivalent to minimizing the discrepancy $D_{L^2, f}(Z)$. Moreover, the square of this discrepancy yields the difference between the discrete energy and the optimal energy integral, a quantity which will be investigated more deeply in the following section, §5. We also observe that, while this approach is rather general, it is somewhat indirect, since it is not easy to explicitly find the function f for a given F (and vice versa). However, since $\widehat{F}(n; \lambda) = |\widehat{f}(n, \lambda)|^2$, the lower bound in (4.4) may be used to estimate both discrepancy and energy.

We also would like to point out that the lower bound in (4.4) is very similar in spirit to Montgomery's lower bound on the L^2 discrepancy with respect to disks in $\mathbb{T}^d$, where an analogous lower bound is obtained in terms of the Fourier coefficients (see [21]). This idea has been further explored in [6, 30].

Finally, the lower estimate in (4.4) is in certain sense sharp. In particular, the minimum cannot be replaced by some averages. For example, let $F(t) = t^{2m}$, $m \in \mathbb{N}$. According to (3.1), $F \in \Phi_d$. The lower bound trivially yields zero, since the odd coefficients $\widehat{F}(2k+1, \lambda)$ vanish. And there indeed exist finite point distributions for which this estimate is exact: these are the so-called spherical $2k$ -designs, i.e. cubature formulas which integrate polynomials of degree up to $2k$ exactly (for recent sharp results on the existence of spherical designs of optimal order see [4]). Since in this case $F(x \cdot y)$ is a polynomial of degree $2m$ in both x and y , it follows that the discrete sum in (4.3) is equal to the continuous energy integral, i.e. the L^2 discrepancy is zero.

Proof. (i) This identity can be verified by a direct computation involving definition (4.2) and relation (4.1). Indeed, we have

$$\begin{aligned} D_{L^2, f}^2(Z) &= \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} f(x \cdot y) f(x \cdot y') d\sigma(y) d\sigma(y') d\sigma(x) \\ &\quad - \frac{2}{N} \sum_{i=1}^N \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} f(x \cdot y) f(x \cdot z_i) d\sigma(y) d\sigma(x) \\ &\quad + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int_{\mathbb{S}^d} f(x \cdot z_i) f(x \cdot z_j) d\sigma(x) \\ &= I_F(\sigma) - 2 \int_{\mathbb{S}^d} F(y \cdot z_1) d\sigma(y) + \frac{1}{N^2} \sum_{i,j=1}^N F(z_i \cdot z_j) \\ &= \frac{1}{N^2} \sum_{i,j=1}^N F(z_i \cdot z_j) - I_F(\sigma), \end{aligned}$$

where we have used the fact that, due to rotational invariance, for any $p \in \mathbb{S}^d$ we have $\int_{\mathbb{S}^d} F(y \cdot p) d\sigma(y) = I_F(d\sigma)$. An even more general form of the Stolarsky principle is proved in Theorem 5.10 of our parallel paper [5], which, in particular,

gives an alternative proof of the fact that σ minimizes I_F for $F \in \Phi_d$.

(ii) We start with the proof of the upper estimate:

$$(4.5) \quad \mathcal{D}_{L^2, f, N}^2 \leq N^{-1} \max_{0 \leq t \leq c_d N^{-\frac{1}{d}}} (F(1) - F(\cos t)).$$

The proof follows along the same lines as that of Theorem 1 of [19]. Let $\{R_1, \dots, R_N\}$ be a partition of $\mathbb{S}^d$ such that (see, for instance, [14, Sec. 6.4, p. 140])

$$\sigma(R_j) = \frac{1}{N} \quad \text{and} \quad \text{diam}(R_j) \leq c_d N^{-\frac{1}{d}}, \quad j = 1, 2, \dots, N.$$

Denote by σ_j^* the restriction of the measure $N\sigma$ to R_j , and let $\Omega^N = R_1 \times \dots \times R_N$ denote the product measure space with probability measure $d\sigma_1^* \times \dots \times d\sigma_N^*$. Then

$$\begin{aligned} \mathcal{D}_{L^2, f, N}^2 &\leq N^{-2} \int_{\mathbb{S}^d} \int_{\Omega^N} \left| \sum_{j=1}^N \left[f(x \cdot z_j) - \int_{R_j} f(x \cdot z) d\sigma_j^*(z) \right] \right|^2 \times \\ &\quad \times d\sigma_1^*(z_1) \cdots d\sigma_N^*(z_N) d\sigma(x) \\ &= N^{-2} \int_{\mathbb{S}^d} \sum_{j=1}^N \left[\int_{R_j} |f(x \cdot z)|^2 d\sigma_j^*(z) - \left(\int_{R_j} f(x \cdot z) d\sigma_j^*(z) \right)^2 \right] d\sigma(x) \\ (4.6) \quad &= N^{-1} F(1) - \sum_{j=1}^N \int_{\mathbb{S}^d} \left(\int_{R_j} f(x \cdot z) d\sigma(z) \right)^2 d\sigma(x). \end{aligned}$$

Note, however, that for each $1 \leq j \leq N$,

$$\begin{aligned} \int_{\mathbb{S}^d} \left(\int_{R_j} f(x \cdot z) d\sigma(z) \right)^2 d\sigma(x) &= \int_{\mathbb{S}^d} \int_{R_j} \int_{R_j} f(x \cdot z) f(x \cdot y) d\sigma(z) d\sigma(y) d\sigma(x) \\ (4.7) \quad &= \int_{R_j} \int_{R_j} F(y \cdot z) d\sigma(y) d\sigma(z) \geq N^{-2} \min_{0 \leq t \leq c_d N^{-\frac{1}{d}}} F(\cos t). \end{aligned}$$

Combining (4.6) with (4.7), we deduce estimate (4.5).

Next, we prove the lower estimate:

$$(4.8) \quad \mathcal{D}_{L^2, f, N}^2 \geq C_d \min_{1 \leq k \leq c_d N^{1/d}} \widehat{F}(k, \lambda).$$

Let $a > 1$ be a large parameter depending only on d such that $n := aN^{1/d}$ is an integer. Let $K_n(t)$ denote the Cesàro kernel of order $d+1$ for the spherical harmonic expansions on $\mathbb{S}^d$; that is,

$$K_n(t) = \sum_{k=0}^n \frac{A_{n-k}^{d+1}}{A_n^{d+1}} \frac{k + \lambda}{\lambda} C_k^\lambda(t), \quad t \in [-1, 1],$$

where the coefficients A_n^j are defined in (2.4). It is known that (see, for instance, [3] and [1, Theorem 7.6.1, p. 389])

$$0 \leq K_n(\cos \theta) \leq C n^d (1 + n\theta)^{-d-1}, \quad \forall \theta \in [0, \pi].$$

We claim that for each $Z_N = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$,

$$(4.9) \quad \int_{\mathbb{S}^d} \left| 1 - N^{-1} \sum_{j=1}^N K_n(x \cdot z_j) \right|^2 d\sigma(x) \geq c_d > 0.$$

To see this, we first note that Bernstein's inequality for trigonometric polynomials implies that

$$(4.10) \quad \min_{0 \leq \theta \leq \frac{1}{2n}} K_n(\cos \theta) \geq \frac{1}{2} K_n(1) = \frac{1}{2} \|K_n\|_\infty \geq c_d n^d.$$

Next, let $\{R_1, \dots, R_{N_1}\}$ be an area-regular partition of $\mathbb{S}^d$ such that $N_1 = C_{d,a} N$, $\sigma(R_j) = \frac{1}{N_1}$ and $\text{diam}(R_j) \leq \frac{1}{2n}$ for $1 \leq j \leq N_1$. Set

$$\Lambda := \left\{ j : 1 \leq j \leq N_1, \quad R_j \cap Z_N \neq \emptyset \right\}.$$

Using (4.10) and positivity of the kernel K_n , we have that for each $x \in R_j$ with $j \in \Lambda$,

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N K_n(x \cdot z_i) &\geq \frac{1}{N} \sum_{z \in Z_N \cap R_j} K_n(x \cdot z) \geq c_d \frac{n^d}{N} \cdot \# \{R_j \cap Z_N\} \\ &= c_d a^d \cdot \# \{R_j \cap Z_N\} > 2 \# \{R_j \cap Z_N\}, \end{aligned}$$

provided that the parameter a is large enough. It then follows that

$$\begin{aligned} \int_{\mathbb{S}^d} \left| 1 - N^{-1} \sum_{i=1}^N K_n(x \cdot z_i) \right|^2 d\sigma(x) &\geq \sum_{j \in \Lambda} \int_{R_j} \left| 1 - \frac{1}{N} \sum_{i=1}^N K_n(x \cdot z_i) \right|^2 d\sigma(x) \\ &\geq \frac{1}{N_1} \sum_{j \in \Lambda} |\#(R_j \cap Z_N)|^2 \geq \frac{1}{N_1} \sum_{j \in \Lambda} \#(R_j \cap Z_N) = \frac{N}{N_1} \geq \frac{1}{C_{d,a}} > 0. \end{aligned}$$

This proves the claim (4.9). We are now ready to show the lower estimate (4.8). Recall that

$$f(x \cdot e) = \sum_{k=0}^{\infty} \widehat{f}(k; \lambda) \frac{k + \lambda}{\lambda} C_k^\lambda(x \cdot e), \quad \forall e \in \mathbb{S}^d,$$

where the series converges in the norm of $L^2(\mathbb{S}^d)$, and $\widehat{f}(0; \lambda) = \int_{\mathbb{S}^d} f(x \cdot z) d\sigma(z)$ for any $x \in \mathbb{S}^d$. Thus, by orthogonality of spherical harmonics, we obtain

$$\begin{aligned} &\int_{\mathbb{S}^d} \left| \int_{\mathbb{S}^d} f(x \cdot z) d\sigma(z) - N^{-1} \sum_{j=1}^N f(x \cdot z_j) \right|^2 d\sigma(x) \\ &= \int_{\mathbb{S}^d} \left| \sum_{k=1}^{\infty} \widehat{f}(k, \lambda) \frac{k + \lambda}{\lambda} N^{-1} \sum_{j=1}^N C_k^\lambda(x \cdot z_j) \right|^2 d\sigma(x) \\ &= \sum_{k=1}^{\infty} \left\| \widehat{F}(k, \lambda) \right\|^2 \left\| \frac{k + \lambda}{\lambda} N^{-1} \sum_{j=1}^N C_k^\lambda(\langle z_j, \cdot \rangle) \right\|_2^2. \end{aligned}$$

Since $\frac{A_{n-j}^{d+1}}{A_n^{d+1}} \leq 1$ for all $0 \leq j \leq n$, it follows that

$$(4.11) \quad \begin{aligned} & \int_{\mathbb{S}^d} \left| \int_{\mathbb{S}^d} f(x \cdot z) d\sigma(z) - N^{-1} \sum_{j=1}^N f(x \cdot z_j) \right|^2 d\sigma(x) \\ & \geq \sum_{k=1}^n \widehat{F}(k, \lambda) \left| \frac{A_{n-k}^{d+1}}{A_n^{d+1}} \right|^2 \left\| \frac{k + \lambda}{\lambda} N^{-1} \sum_{j=1}^N C_k^\lambda(\langle z_j, \cdot \rangle) \right\|_2^2, \end{aligned}$$

which, by (4.9), is bounded below by

$$\begin{aligned} & \left(\min_{1 \leq k \leq n} \widehat{F}(k, \lambda) \right) \int_{\mathbb{S}^d} \left| 1 - N^{-1} \sum_{j=1}^N K_n(x \cdot z_j) \right|^2 d\sigma(x) \\ & \geq C_d \min_{1 \leq k \leq n} \widehat{F}(k, \lambda). \end{aligned}$$

This yields the desired lower estimate (4.8). $\square$

Using Theorem 4.2, one can give a new simpler proof of a well-known result of Beck [2] regarding the lower estimate of the spherical cap discrepancy:

$$(4.12) \quad D_{L^2, \text{cap}}(Z_N)^2 := \int_{-1}^1 \int_{\mathbb{S}^d} \left| \frac{\#(Z_N \cap B(x, t))}{N} - \sigma(B(x, t)) \right|^2 d\sigma(x) dt,$$

where $Z_N := \{z_1, \dots, z_N\}$ is a set of N -distinct points on $\mathbb{S}^d$ and

$$B(x, t) := \{y \in \mathbb{S}^d : x \cdot y \geq t\}, \quad x \in \mathbb{S}^d, \quad t \in [-1, 1].$$

Corollary 4.3. [2, J. Beck, 1984] *Given an arbitrary set Z_N of N -distinct points on the sphere $\mathbb{S}^d$,*

$$D_{L^2, \text{cap}}(Z_N) \geq C_d N^{-\frac{1}{2} - \frac{1}{2d}}.$$

Proof. Let $f_t(s) = \chi_{[t, 1]}(s)$ for $t, s \in [-1, 1]$. Using the formula,

$$\frac{d}{dx} \left(C_{n-1}^{\lambda+1}(x) (1-x^2)^{\lambda+\frac{1}{2}} \right) = -\frac{n(n+2\lambda)}{2\lambda} C_n^\lambda(x) (1-x^2)^{\lambda-\frac{1}{2}},$$

we have

$$\widehat{f}_t(n; \lambda) = c_\lambda \frac{\Gamma(n+1)}{\Gamma(n+2\lambda)} \int_t^1 C_n^\lambda(x) (1-x^2)^{\lambda-\frac{1}{2}} dx = c_d (1-t^2)^{\lambda+\frac{1}{2}} R_{n-1}^{\lambda+1}(t).$$

This implies that

$$\int_{-1}^1 |\widehat{f}_t(n; \lambda)|^2 dt \sim \int_{-1}^1 |R_{n-1}^{\lambda+1}(t)|^2 (1-t^2)^{2\lambda+1} dt \sim n^{-2\lambda-2} = n^{-d-1},$$

where the second step uses the known estimates on integrals of Jacobi polynomials (see, for instance, [32, Ex. 91, p. 391]).

On the other hand, using (4.11) and (4.9), with $n \sim N^{1/d}$, we have

$$\begin{aligned}
D_{L^2, \text{cap}}(A)^2 &= \int_{-1}^1 D_{L^2, f_t}(Z_N)^2 dt \\
&\geq \sum_{k=1}^n \int_{-1}^1 |\widehat{f}_t(k, \lambda)|^2 dt \left| \frac{A_{n-k}^{d+1}}{A_n^{d+1}} \right|^2 \left\| \frac{k + \lambda}{\lambda} N^{-1} \sum_{j=1}^N C_k^\lambda(\langle z_j, \cdot \rangle) \right\|_2^2 \\
&\geq \left(\min_{1 \leq k \leq n} \int_{-1}^1 |\widehat{f}_t(k, \lambda)|^2 dt \right) \int_{\mathbb{S}^d} \left| 1 - N^{-1} \sum_{j=1}^N K_n(x \cdot z_j) \right|^2 d\sigma(x) \\
&\geq C_d \min_{1 \leq k \leq n} \int_{-1}^1 |\widehat{f}_t(k, \lambda)|^2 dt \geq c_d n^{-d-1} \sim N^{-1-\frac{1}{d}}.
\end{aligned}$$

□

5. DISCRETE RIESZ ENERGY

Now that we understand that the energy $I_{d,\delta}(\sigma)$ is optimal for $-d < \delta < 1$, it is natural to investigate how well it can be approximated by discrete distributions. We define the appropriate discrete energy in accordance with (1.2).

Definition 5.1. For $\delta > -d$ and $\delta \neq 0$, define the discrete δ -energy of a finite subset $Z_N = \{z_1, \dots, z_N\}$ of N distinct points on $\mathbb{S}^d$ by

$$E_{d,\delta}(Z_N) := \sum_{1 \leq i < j \leq N} \rho(z_i, z_j)^\delta,$$

where $\rho(x, y) = \arccos(x \cdot y)$ is the geodesic distance between x and y on $\mathbb{S}^d$. The discrete N -point δ -energy of $\mathbb{S}^d$ is defined by

$$\mathcal{E}_{d,\delta}(N) := \inf_{Z_N} E_{d,\delta}(Z_N),$$

where the infimum is taken over all N -point subsets of $\mathbb{S}^d$ (and infimum is replaced by supremum for $\delta > 0$).

Notice that, when $\delta > 0$, we have

$$I_{d,\delta}(N^{-1} \sum_{i=1}^N \delta_{z_i}) = \frac{2}{N^2} E_{d,\delta}(Z_N),$$

while this energy integral is infinite for $\delta \leq 0$ because of diagonal terms.

Our main result in this section can be stated as follows:

Theorem 5.2. Let $d \geq 2$. If $-d < \delta < 1$ and $\delta \neq 0$, then

$$I_{d,\delta}(\sigma) - \frac{2}{N^2} \mathcal{E}_{d,\delta}(N) \sim N^{-1-\frac{\delta}{d}}.$$

In the logarithmic case, $\delta = 0$, we have the estimate

$$I_{d,0}(\sigma) - \frac{2}{N^2} \mathcal{E}_{d,0}(N) \sim N^{-1} \log N.$$

Remark 5.3. For the discrete δ -Riesz energy defined with respect to the Euclidean distance on $\mathbb{S}^d$, similar results were previously established for $-d < \delta < 2$ in a series of papers (see [12, Proposition 2] and also [10, 19, 23, 34, 35]).

Recall that for $\delta > -d$ and $\delta \neq 0$,

$$I_{d,\delta}(\sigma) = \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \rho(x, y)^\delta d\sigma(x) d\sigma(y) = \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})\Gamma(\frac{1}{2})} \int_0^\pi \theta^\delta \sin^{d-1} \theta d\theta.$$

The following lemma is needed in the proof of the lower estimates in Theorem 5.2:

Lemma 5.4. *If $\lambda > 0$ and $0 < \delta < 1$, then*

$$- \int_0^\pi \theta^\delta R_n^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \sim n^{-2\lambda-1-\delta}, \quad n = 1, 2, \dots,$$

where $R_n^\lambda(t) = \frac{C_n^\lambda(t)}{C_n^\lambda(1)}$.

We postpone the proof of Lemma 5.4 to the next section. For the moment, we take it for granted, and proceed with the proof of Theorem 5.2.

5.1. Proof of Theorem 5.2 for $0 < \delta < 1$. By part (i) of Lemma 3.2, the function $F(t) = \left(\frac{\pi}{2}\right)^\delta - (\arccos t)^\delta$ belongs to the class Φ_d for $\delta \in (0, 1)$. Furthermore, by Lemma 5.4, we have that

$$(5.1) \quad \widehat{F}(k; \lambda) = -c_\lambda \int_0^\pi \theta^\delta R_k^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \sim k^{-d-\delta}, \quad k = 1, 2, \dots$$

Hence, applying the Stolarsky principle (part (i) of Theorem 4.2), we find that for $Z_N = \{z_1, \dots, z_N\} \subset \mathbb{S}^d$,

$$\frac{2}{N^2} E_{d,\delta}(Z_N) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \rho(z_i, z_j)^\delta = I_{d,\delta}(\sigma) - D_{L^2,f}(Z_N)^2,$$

and hence

$$I_{d,\delta}(\sigma) - \frac{2}{N^2} E_{d,\delta}(Z_N) = D_{L^2,f}^2(Z_N).$$

By part (ii) of Theorem 4.2, we have

$$\inf_{Z_N} D_{L^2,f}(Z_N)^2 \leq c_d N^{-1} \max_{0 \leq \theta \leq c_d N^{-\frac{1}{d}}} \theta^\delta \leq c_d N^{-1-\frac{\delta}{d}},$$

whereas by (5.1) and part (ii) of Theorem 4.2, for any Z_N ,

$$D_{L^2,f}(Z_N)^2 \geq c \min_{1 \leq k \leq c_d N^{\frac{1}{d}}} \widehat{F}(k; \lambda) \geq c_d N^{-1-\frac{\delta}{d}}.$$

This completes the proof of the theorem for $0 < \delta < 1$.

5.2. Proof of Theorem 5.2 for $-d < \delta < 0$. For convenience, we set $\delta = -s$ with $0 < s < d$. We start with the proof of the upper estimate:

$$(5.2) \quad \frac{2}{N^2} \mathcal{E}_{d,-s}(N) \leq I_{d,\delta}(\sigma) - c_d N^{-1-\frac{\delta}{d}}.$$

Let $\varepsilon \in (0, 1)$, and set

$$F_{-s,\varepsilon}(t) = (\arccos t + \varepsilon)^{-s}, \quad t \in [-1, 1].$$

Then according to Lemma 3.2, $F_{-s,\varepsilon} \in \Phi_d$. Thus, by the Stolarsky principle (part (i) of Theorem 4.2),

$$(5.3) \quad \begin{aligned} 2N^{-2} \sum_{1 \leq i < j \leq N} (\rho(z_i, z_j) + \varepsilon)^{-s} + N^{-1} \varepsilon^{-s} \\ = D_{L^2, f_{-s,\varepsilon}}^2(Z_N) + \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y) + \varepsilon)^{-s} d\sigma(x) d\sigma(y). \end{aligned}$$

We then use (5.3) and part (ii) of Theorem 4.2 to obtain

$$\begin{aligned} \inf_{Z_N} \sum_{1 \leq i < j \leq N} (\rho(z_i, z_j) + \varepsilon)^{-s} &= \frac{1}{2} N^2 \mathcal{D}_{L^2, f_{-s,\varepsilon}, N}^2 - \frac{1}{2} N \varepsilon^{-s} \\ &\quad + \frac{1}{2} N^2 \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y) + \varepsilon)^{-s} d\sigma(x) d\sigma(y) \\ &\leq \frac{N}{2} \left(F_{-s,\varepsilon}(1) - \min_{|\theta| \leq c_d N^{-\frac{1}{d}}} F_{-s,\varepsilon}(\cos \theta) \right) \\ &\quad - \frac{N}{2} \varepsilon^{-s} + \frac{N^2}{2} I_{d,\delta}(\sigma) \\ &= -\frac{1}{2} N \min_{0 \leq t \leq c_d N^{-\frac{1}{d}}} (t + \varepsilon)^{-s} + \frac{N^2}{2} I_{d,\delta}(\sigma) \\ &= -\frac{1}{2} N (c_d N^{-\frac{1}{d}} + \varepsilon)^{-s} + \frac{N^2}{2} I_{d,\delta}(\sigma). \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ yields the upper estimate (5.2).

Next, we show the lower estimate. Let $Z_N = \{z_1, \dots, z_N\}$ be an arbitrary set of N -distinct points on $\mathbb{S}^d$. We need to prove that

$$E_{d,-s}(Z_N) = \sum_{1 \leq i < j \leq N} \rho(z_i, z_j)^{-s} \geq \frac{N^2}{2} I_{d,-s}(\sigma) - c_d N^{1+\frac{s}{d}}.$$

To this end, let k be the smallest positive integer such that $s + k + 1 > d$. For a fixed $\theta \in (0, \pi]$, define $g_\theta : [0, \pi] \rightarrow \mathbb{R}$ by

$$g_\theta(t) = (\theta + t)^{-s} + st(\theta + t)^{-s-1} + \frac{s(s+1)}{2} t^2 (\theta + t)^{-s-2} + \dots + \frac{(s)_k}{k!} t^k (\theta + t)^{-s-k},$$

where $t \geq 0$. A straightforward calculation shows that

$$g'_\theta(t) = -\frac{(s)_{k+1}}{k!} t^k (\theta + t)^{-s-k-1} \leq 0, \quad \forall t \geq 0.$$

In particular, this implies that for all $\theta, t \in (0, \pi]$,

$$(5.4) \quad 0 \leq g_\theta(0) - g_\theta(t) = \theta^{-s} - g_\theta(t) \leq \begin{cases} C\theta^{-s}, & \text{if } 0 < \theta \leq t; \\ Ct^{k+1}\theta^{-s-k-1}, & \text{if } \theta > t. \end{cases}$$

It follows that for any $\varepsilon > 0$,

$$\begin{aligned} \sum_{1 \leq i < j \leq N} \rho(z_i, z_j)^{-s} &\geq \sum_{\ell=0}^k \frac{(s)_\ell \varepsilon^\ell}{\ell!} \sum_{1 \leq i < j \leq N} (\rho(z_i, z_j) + \varepsilon)^{-s-\ell} \\ &\geq \frac{N^2}{2} \sum_{\ell=0}^k \frac{(s)_\ell \varepsilon^\ell}{\ell!} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y) + \varepsilon)^{-s-\ell} d\sigma(x) d\sigma(y) - CN\varepsilon^{-s}, \end{aligned}$$

where we have used the fact that $F_{-(s+\ell), \varepsilon} \in \Phi_d$ (by part (ii) of Lemma 3.2) and therefore the energy integral is smaller than the discrete energy: this can be deduced either from Proposition 2.1 (σ minimizes $I_{F_{-(s+\ell), \varepsilon}}$) or from the Stolarsky principle (4.3) (since $D_{L^2, f_{-(s+\ell), \varepsilon}}^2(Z_N) \geq 0$).

On the other hand, a direct computation involving (5.4) shows that

$$\begin{aligned} &\left| \sum_{\ell=0}^k \frac{(s)_\ell \varepsilon^\ell}{\ell!} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} (\rho(x, y) + \varepsilon)^{-s-\ell} d\sigma(x) d\sigma(y) - \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \rho(x, y)^{-s} d\sigma(x) d\sigma(y) \right| \\ &= c \int_0^\pi (g_\theta(0) - g_\theta(\varepsilon)) \sin^{d-1} \theta d\theta \\ &\leq c \int_0^\varepsilon \theta^{-s+d-1} d\theta + c\varepsilon^{k+1} \int_\varepsilon^\pi \theta^{-s-1-k} \sin^{d-1} \theta d\theta \leq c\varepsilon^{d-s}. \end{aligned}$$

Thus, putting the above together, we obtain

$$\sum_{1 \leq i < j \leq N} \rho(z_i, z_j)^{-s} \geq \frac{N^2}{2} I_{d, -s}(\sigma) - CN\varepsilon^{-s} - cN^2\varepsilon^{d-s}.$$

Now setting $\varepsilon = N^{-\frac{1}{d}}$, we get the desired lower estimate:

$$E_{d, -s}(Z_N) \geq \frac{1}{2} N^2 I_{d, \delta}(\sigma) - c_d N^{1+\frac{s}{d}}.$$

5.3. Proof of Theorem 5.2 in the logarithmic case $\delta = 0$. The proof is very similar to the previous case. We shall begin with the upper bound:

$$(5.5) \quad \frac{2}{N^2} \mathcal{E}_{d, 0}(N) \leq I_{d, 0}(\sigma) - c_d N^{-1} \log N.$$

Recall that for $\varepsilon \in [0, 1)$, by Lemma 3.2, $F_{0, \varepsilon}(t) = \log\left(\frac{\pi}{\arccos t + \varepsilon}\right) \in \Phi_d$. Hence, invoking Theorem 4.2),

$$\begin{aligned} &\inf_{Z_N} \frac{2}{N^2} \sum_{1 \leq i < j \leq N} \log\left(\frac{\pi}{\rho(z_i, z_j) + \varepsilon}\right) + N^{-1} \log \frac{\pi}{\varepsilon} \\ &= \mathcal{D}_{L^2, f_{-s, \varepsilon}}^2(Z) + \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \log\left(\frac{\pi}{\rho(x, y) + \varepsilon}\right) d\sigma(x) d\sigma(y) \\ &\leq N^{-1} \left(F_{0, \varepsilon}(1) - \min_{|\theta| \leq c_d N^{-\frac{1}{d}}} F_{0, \varepsilon}(\cos \theta) \right) + I_{d, 0}(\sigma) \\ &= N^{-1} \log \frac{\pi}{\varepsilon} + N^{-1} \log \left(\frac{\pi}{\varepsilon + c_d N^{-1/d}} \right) + I_{d, 0}(\sigma), \end{aligned}$$

which yields (5.5) as $\varepsilon \rightarrow 0$.

To prove the lower bound, we first observe that $\left| \log \frac{\pi}{\varepsilon + \theta} - \log \frac{\pi}{\theta} \right| \leq \frac{1}{2} \varepsilon \theta^{-1}$ for $\varepsilon, \theta > 0$. This easily implies that

$$\left| \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \left(\log \frac{\pi}{\varepsilon + \rho(x, y)} - \log \frac{\pi}{\rho(x, y)} \right) d\sigma(x) d\sigma(y) \right| \leq C\varepsilon.$$

Therefore, using the fact that σ minimizes $I_{F_0, \varepsilon}$,

$$\begin{aligned} \sum_{1 \leq i < j \leq N} \log \frac{\pi}{\rho(z_i, z_j)} &\geq \sum_{1 \leq i < j \leq N} \log \left(\frac{\pi}{\rho(z_i, z_j) + \varepsilon} \right) \\ &\geq \frac{N^2}{2} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \log \frac{\pi}{\varepsilon + \rho(x, y)} d\sigma(x) d\sigma(y) - \frac{1}{2} N \log \frac{\pi}{\varepsilon} \\ &\geq \frac{N^2}{2} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \log \frac{\pi}{\rho(x, y)} d\sigma(x) d\sigma(y) - C' N^2 \varepsilon - \frac{1}{2} N \log \frac{\pi}{\varepsilon}, \end{aligned}$$

and the proof is concluded by choosing $\varepsilon = cN^{-1} \log N$.

6. PROOF OF LEMMA 5.4

Recall that $R_n^\lambda(\cos t) = \frac{C_n^\lambda(\cos t)}{C_n^\lambda(1)}$ for $\lambda > 0$, and $R_n^\lambda(\cos t) = \cos nt$ for $\lambda = 0$. We will use the following formula ([32, p. 80-81]):

$$(6.1) \quad (R_n^\lambda(t))' = \frac{n(n+2\lambda)}{2\lambda+1} R_{n-1}^{\lambda+1}(t) = c(n, \lambda) R_{n-1}^{\lambda+1}(t).$$

For the rest of this section, we let $\eta \in C^\infty[0, \infty)$ be such that $\eta(t) = 1$ for $0 \leq t \leq \frac{\pi}{4}$ and $\eta(t) = 0$ for $t \geq \frac{\pi}{2}$.

6.1. Upper estimate. The upper estimate is a direct consequence of the following lemma:

Lemma 6.1. *Let $g \in C^\infty[0, \pi]$ and $\lambda > 0$. If $-2\lambda - 1 < \delta \leq 1$ and $\delta \neq 0$, then*

$$\left| \int_0^\pi \theta^\delta g(\theta) R_n^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \right| \leq C n^{-2\lambda-1-\delta}.$$

Proof. We first claim that it suffices to show that

$$(6.2) \quad \left| \int_0^{\frac{\pi}{2}} \theta^\delta \eta(\theta) g(\theta) R_n^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \right| \leq C n^{-2\lambda-1-\delta}.$$

Indeed, a slight modification of the proof of (6.2) below shows that for any $g_1 \in C^\infty[0, \pi]$,

$$(6.3) \quad \left| \int_0^\pi g_1(\theta) R_n^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \right| \leq C n^{-2\lambda-2}.$$

Since $\theta^\delta (1 - \eta(\theta)) g(\theta) \in C^\infty[0, \pi]$, the desired upper estimates follow directly from (6.2) and (6.3).

Next, we show (6.2) in the case when λ is a positive integer. Let $\xi_0 \in C^\infty(\mathbb{R})$ be such that $\xi_0(t) = 1$ for $|t| \leq \frac{1}{8}$ and $\xi_0(t) = 0$ for $|t| \geq \frac{1}{4}$. Let $\xi_1 = 1 - \xi_0$. Clearly,

$$\left| \int_0^{\pi/2} \theta^\delta \xi_0(\theta n) \eta(\theta) g(\theta) R_n^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \right| \leq C \int_0^{\frac{1}{4n}} \theta^{2\lambda+\delta} d\theta \leq C n^{-(2\lambda+1+\delta)}.$$

Thus, it remains to show that

$$(6.4) \quad I_n := \left| \int_0^{\pi/2} \theta^\delta \xi_1(\theta n) \eta(\theta) g(\theta) R_n^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \right| \leq C n^{-2\lambda-1-\delta}.$$

To show (6.4), set

$$(6.5) \quad D := \frac{d}{d\theta} \frac{1}{\sin \theta} = \frac{1}{\sin \theta} \frac{d}{d\theta} - \frac{\cos \theta}{\sin^2 \theta}.$$

Using (6.1) and integration by parts $\lambda + 2$ times, we then obtain

$$\begin{aligned} I_n &\leq C n^{-2\lambda} \left| \int_0^{\pi/2} D^\lambda (\xi_1(\theta n) \theta^\delta \eta(\theta) g(\theta) \sin^{2\lambda} \theta) \cos((n + \lambda)\theta) d\theta \right| \\ &\leq C n^{-2\lambda-2} \left| \int_0^{\pi/2} \left(D^\lambda (\xi_1(\theta n) \theta^\delta \eta(\theta) g(\theta) \sin^{2\lambda} \theta) \right)'' \cos((n + \lambda)\theta) d\theta \right|. \end{aligned}$$

Since η' is supported in $[\frac{\pi}{4}, \frac{\pi}{2}]$, it is easily seen that

$$\left| \left(D^\lambda (\theta^\delta \xi_1(\theta n) \eta(\theta) g(\theta) \sin^{2\lambda} \theta) \right)'' \right| \leq C \max_{0 \leq j \leq \lambda+2} n^j |\xi_1^{(j)}(n\theta)| \theta^{\delta+j-2}.$$

Since $\xi_1'(n\theta)$ is supported in $[\frac{1}{8n}, \frac{1}{4n}]$, it follows that for $1 \leq j \leq \lambda + 2$,

$$n^{-2\lambda+j-2} \int_0^{\pi/2} |\xi_1'(\theta n)| \theta^{\delta+j-2} d\theta \leq C n^{-2\lambda+j-2} n^{-\delta-j+1} = C n^{-2\lambda-1-\delta}.$$

On the other hand, since $\xi_1(n\theta)$ is supported in $[\frac{1}{8n}, \infty]$, we have

$$n^{-2\lambda-2} \int_0^{\pi/2} |\xi_1(\theta n)| \theta^{\delta-2} d\theta \leq C n^{-2-2\lambda} \int_{\frac{1}{8n}}^{\pi/2} \theta^{\delta-2} d\theta \leq C n^{-2\lambda-\delta-1}.$$

Putting these together, we deduce (6.4) and hence prove (6.2) for integer λ .

Finally, we show (6.2) for all $\lambda > 0$. We will use the following formula on ultraspherical polynomials ([32, (4.10.29), p. 99])

$$(6.6) \quad (\sin \theta)^{2\lambda} R_n^\lambda(\cos \theta) = \sum_{k=0}^{\infty} \alpha_{k,n}^{\lambda,\mu} R_{n+2k}^\mu(\cos \theta) (\sin \theta)^{2\mu},$$

where $0 < \lambda < \mu < 2\lambda + 1$ and

$$\begin{aligned} \alpha_{k,n}^{\lambda,\mu} &= \frac{\Gamma(2\lambda)\Gamma(\mu)2^{2(\mu-\lambda)}(n+2k+\mu)\Gamma(n+k+\mu)\Gamma(k+\mu-\lambda)}{\Gamma(2\mu)\Gamma(\mu-\lambda)\Gamma(\lambda)k!\Gamma(n+k+\lambda+1)} \\ (6.7) \quad &\sim (n+k)^{\mu-\lambda}(k+1)^{\mu-\lambda-1}. \end{aligned}$$

For any $\lambda > 0$, we can find an integer μ such that $\lambda < \mu < 2\lambda + 1$. Then $-1 < -\delta < 2\lambda + 1 < 2\mu + 1$. Using (6.6), (6.7) and the already proven case of (6.2), we obtain

$$\begin{aligned} \left| \int_0^{\pi/2} \theta^\delta \eta(\theta) g(\theta) R_n^\lambda(\cos \theta) (\sin \theta)^{2\lambda} d\theta \right| &= \left| \sum_{k=0}^{\infty} \alpha_{k,n}^{\lambda,\mu} \int_0^{\pi/2} \theta^\delta \eta(\theta) R_{n+2k}^\mu(\cos \theta) (\sin \theta)^{2\mu} d\theta \right| \\ &\leq C \sum_{k=0}^{\infty} (n+k)^{\mu-\lambda}(k+1)^{\mu-\lambda-1}(n+2k)^{-2\mu-1-\delta} \leq C n^{-2\lambda-1-\delta}. \end{aligned}$$

□

6.2. Lower estimate. In this subsection, we shall prove the lower estimates: for $0 < \delta < 1$,

$$(6.8) \quad \left| \int_0^\pi \theta^\delta R_n^\lambda(\cos \theta)(\sin \theta)^{2\lambda} d\theta \right| \geq C n^{-2\lambda-1-\delta}.$$

According to Lemma 3.2, it suffices to prove (6.8) as $n \rightarrow \infty$. Next, we assert that if (6.8) holds for some $\lambda = \mu > 0$, then it holds for all $\max\{\frac{\mu-1}{2}, 0\} < \lambda < \mu$. Indeed, this follows directly from (6.7) and Lemma 3.2: for $n \geq 1$,

$$\begin{aligned} \left| \int_0^\pi \theta^\delta R_n^\lambda(\cos \theta)(\sin \theta)^{2\lambda} d\theta \right| &= \sum_{k=0}^\infty \alpha_{k,n}^{\lambda,\mu} \left| \int_0^\pi \theta^\delta R_{n+2k}^\mu(\cos \theta)(\sin \theta)^{2\mu} d\theta \right| \\ &\geq c \sum_{k=0}^n (n+k)^{\mu-\lambda} (k+1)^{\mu-\lambda-1} (n+2k)^{-2\mu-1-\delta} \geq c n^{-2\lambda-1-\delta}. \end{aligned}$$

Thus, according to this last assertion, it suffices to prove (6.8) in the case when $\lambda \geq 2$ is an integer. Third, recall that $\eta \in C^\infty[0, \infty)$ is such that $\eta(t) = 1$ for $0 \leq t \leq \frac{\pi}{4}$ and $\eta(t) = 0$ for $t \geq \frac{\pi}{2}$, hence, according to (6.3),

$$\left| \int_0^\pi \theta^\delta (1 - \eta(\theta)) R_n^\lambda(\cos \theta)(\sin \theta)^{2\lambda} d\theta \right| \leq C n^{-2\lambda-2}.$$

Putting the above together, we are reduced to showing that if $\lambda \geq 2$ is an integer and $0 < \delta < 1$, then

$$\left| \int_0^\pi \theta^\delta \eta(\theta) R_n^\lambda(\cos \theta)(\sin \theta)^{2\lambda} d\theta \right| \geq C n^{-2\lambda-1-\delta}, \quad \text{as } n \rightarrow \infty.$$

For the rest of the proof, we use the notation c_n to denote a positive constant depending on n such that $c_n \sim 1$ as $n \rightarrow \infty$.

First, we observe from (6.2) for any $\beta \geq 1$, $g \in C^\infty[0, \pi]$ and $\mu \in \mathbb{N}$,

$$(6.9) \quad \left| \int_0^\pi g(\theta) \theta^\beta R_n^\mu(\cos \theta)(\sin \theta)^{2\mu} d\theta \right| \leq C_{\mu,g} n^{-2\mu-2}.$$

Thus, using (6.1), (6.5), (6.9) and integration by parts, we obtain

$$\begin{aligned} I_n^\delta &:= \int_0^\pi \theta^\delta \eta(\theta) R_n^\lambda(\cos \theta)(\sin \theta)^{2\lambda} d\theta = c_n n^{-2} \int_0^{\frac{\pi}{2}} D(\theta^\delta \eta(\theta) \sin^{2\lambda} \theta) R_{n+1}^{\lambda-1}(\cos \theta) d\theta \\ &= c_n n^{-2} \int_0^{\frac{\pi}{2}} \theta^\delta (\sin \theta)^{2\lambda-2} \eta(\theta) \left[\frac{\delta \sin \theta}{\theta} + (2\lambda-1) \cos \theta \right] R_{n+1}^{\lambda-1}(\cos \theta) d\theta + O(n^{-2\lambda-2}) \\ &= c_n n^{-2} (2\lambda-1+\delta) \int_0^{\frac{\pi}{2}} \theta^\delta (\sin \theta)^{2\lambda-2} \eta(\theta) R_{n+1}^{\lambda-1}(\cos \theta) d\theta + O(n^{-2\lambda-2}), \end{aligned}$$

where we recall that $c_n > 0$ and $c_n \sim 1$. Continuing in this way λ times, we obtain

$$I_n^\delta = c_n n^{-2\lambda} \int_0^{\frac{\pi}{2}} \theta^\delta \eta(\theta) \cos((n+\lambda)\theta) d\theta + O(n^{-2\lambda-2}),$$

where $c'_n \sim 1$. Now using integration by parts once again, we deduce

$$I_n^\delta = -c''_n n^{-2\lambda-1} \int_0^\pi \theta^{\delta-1} \eta(\theta) \sin((n+\lambda)\theta) d\theta + O(n^{-2\lambda-2}).$$

Since $\frac{\eta(\theta)-1}{\theta} = \frac{\eta(\theta)-\eta(0)}{\theta} \in C^\infty[0, \pi]$, it follows that

$$n^{-2\lambda-1} \left| \int_0^\pi \theta^{\delta-1} (\eta(\theta) - 1) \sin((n+\lambda)\theta) d\theta \right| \leq C n^{-2\lambda-2}.$$

This implies that

$$\begin{aligned} -I_n^\delta &= c_n'' n^{-2\lambda-1} \int_0^\pi \theta^{\delta-1} \sin((n+\lambda)\theta) d\theta + O(n^{-2\lambda-2}) \\ &= c_n''' n^{-2\lambda-1-\delta} \int_0^{(n+\lambda)\pi} \theta^{\delta-1} \sin \theta d\theta + O(n^{-2\lambda-2}) \\ &\geq c n^{-2\lambda-1-\delta} \int_0^{2\pi} \theta^{\delta-1} \sin \theta d\theta + O(n^{-2\lambda-2}) \geq c n^{-2\lambda-1-\delta}. \end{aligned}$$

This shows the desired lower estimate.

REFERENCES

- [1] R. Askey, G. Andrews and R. Roy, *Special Functions*. Encyclopedia of Mathematics and its Applications **71**, Cambridge University Press, Cambridge, 1999.
- [2] 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), 33–41, 1984.
- [3] A. Bonami, J. L. Clerc, Sommes de Cesàro et multiplicateurs des développements en harmonique sphériques. *Trans. Amer. Math. Soc.* **183**, 223–263, 1973
- [4] A. Bondarenko, D. Radchenko, M. Viazovska, Optimal asymptotic bounds for spherical designs. *Annals of Mathematics* **178** (2), 443–452, 2013.
- [5] D. Bilyk, F. Dai, R. Matzke, Stolarsky principle and energy optimization on the sphere, to appear in *Constr. Approx.*, 2018, available at <https://arxiv.org/abs/1611.04420>.
- [6] D. Bilyk, F. Dai, S. Steinerberger, General and refined Montgomery lemmata, preprint, 2018, available at <https://arxiv.org/abs/1801.07701>.
- [7] D. Bilyk, M. Lacey, One bit sensing, discrepancy, and Stolarsky principle. *Sbornik Math.*, **208** (6), 744–763, 2017.
- [8] G. Björck, Distributions of positive mass, which maximize a certain generalized energy integral. *Ark. Mat.* **3**, 255–269, 1956.
- [9] J. Brauchart, Invariance principles for energy functionals on the sphere. *Monatsh. Math.* **141**, 101–117, 2004.
- [10] J. Brauchart, 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.
- [11] J. Brauchart and J. Dick, A simple proof of Stolarsky’s invariance principle. *Proc. Amer. Math. Soc.* **141**, 2085–2096, 2013.
- [12] J. Brauchart, D.P. Hardin, E. B. Saff, The next-order term for optimal Riesz and logarithmic energy asymptotics on the sphere. *Contemp. Math.*, Vol. **578**, 31–61, 2012.
- [13] J. Brauchart, D. Hardin, E. Saff, Discrete energy asymptotics on a Riemannian circle. *Unif. Distrib. Theory* **7**, no. 2, 77–108, 2012.
- [14] F. Dai, and Y. Xu, *Approximation Theory and Harmonic Analysis on Spheres and Balls*. Springer Monographs in Mathematics, Springer, 2013.
- [15] L. Fejes Tóth, Über eine Punktverteilung auf der Kugel (German). *Acta Math. Acad. Sci. Hungar.* **10**, 13–19, 1959.
- [16] C.L. Frenzen and R. A. Wong, Uniform asymptotic expansion of the Jacobi polynomials with error bounds. *Canad. J. Math.* **37**, no. 5, 979–1007, 1985.
- [17] G. Gasper, Positive integrals of Bessel functions, *SIAM J. Math. Anal.* **6**, no. 5, 868–881, 1975.
- [18] Y. Katznelson, An introduction to harmonic analysis. 3rd edition. Cambridge University Press, Cambridge, 2004
- [19] A. B. J. Kuijlaars and E. B. Saff, Asymptotics for minimal discrete energy on the sphere. *Trans. Amer. Math. Soc.* **350**(2), 523–538, 1998.

- [20] H. Larcher, Solution of a gemetric problem by Fejes Tóth. *Michigan Math. J.* **9**, 45–51, 1962.
- [21] H. Montgomery, Ten Lectures at the Interface of Harmonic Analysis and Number Theory. American Mathematical Society, 1994.
- [22] H. He, K. Basu, Q. Zhao, A. Owen, Permutation p -value approximation via generalized Stolarsky invariance, to appear, available at <https://arxiv.org/abs/1603.02757>.
- [23] E. A. Rakhmanov, E. B. Saff, and Y. M. Zhou, Minimal discrete energy on the sphere. *Math. Res. Lett.* **1**(6), 647–662, 1994.
- [24] I. Schoenberg, Positive definite functions on spheres. *Duke Math. J.* **9**, 96–108, 1942.
- [25] R. Schwartz, The 5-Electron Case of Thomson’s Problem. *Experimental Math* **22**, 157–186, 2013.
- [26] M. Skriganov, Point distributions in compact metric spaces. *Mathematika* **63** (3), 1152–1171, 2017.
- [27] M. Skriganov, Point distributions in compact metric spaces, II, preprint, available at <http://www.pdmi.ras.ru/preprint/2016/16-02.html>.
- [28] M. Skriganov, Point distributions in compact metric spaces, III, 2016, preprint, available at <http://www.pdmi.ras.ru/preprint/2016/16-07.html>.
- [29] G. Sperling, Lösung einer elementargeometrischen Frage von Fejes Tóth (German). *Arch. Math.* **11**, 69–71, 1960.
- [30] S. Steinerberger, Exponential sums and Riesz energies. *J. Number Theory* **182**, 37–56, 2018.
- [31] K. B. Stolarsky. Sums of distances between points on a sphere. II. *Proc. Amer. Math. Soc.* **41**, 575–582, 1973.
- [32] G. Szegő, *Orthogonal Polynomials*. Amer. Math. Soc. Colloq. Publ. Vol.23, Providence, 4th edition, 1975.
- [33] Y. S. Tan, Energy optimization for distributions on the sphere and improvement to the Welch bound, preprint, 2016, available at <https://arxiv.org/abs/1612.06343>.
- [34] G. Wagner, On means of distances on the surface of a sphere (lower bounds). *Pacific J. Math.* **144**, no. 2, 389–398, 1990.
- [35] G. Wagner, On means of distances on the surface of a sphere. II. Upper bounds. *Pacific J. Math.* **154**, no. 2, 381–396, 1992.

SCHOOL OF MATHEMATICS, UNIVERSITY OF MINNESOTA, MINNEAPOLIS, MN 55408, USA.
E-mail address: `dbilyk@math.umn.edu`

DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES, UNIVERSITY OF ALBERTA, EDMONTON, ALBERTA T6G 2G1, CANADA.
E-mail address: `fdai@ualberta.ca`